



TECHNISCHE UNIVERSITÄT
CHEMNITZ

Clause Complexing Systems of Chinese Students' Academic Writing

Von der Philosophischen Fakultät der
Technischen Universität Chemnitz genehmigte

Dissertation

zur Erlangung des akademischen Grades

doctor philosophiae (Dr. phil.)

vorgelegt

von M.A. Ya Guo

Geburtsjahr: 1991

Gutachterin / Gutachter:

Prof. Dr. Josef Schmied

Prof. Dr. Christina Sanchez-Stockhammer

Tag der Einreichung: 20.12.2024

Tag der Verteidigung: 21.05.2025

Veröffentlichungsjahr: 2025

(URL) <https://nbn-resolving.org/urn:nbn:de:bsz:ch1-qucosa2-972566>

(DOI) <https://doi.org/10.60687/2025-0091>

This work- except the logo of TU Chemnitz - is under Creative-Commons Licence.
Attribution- 4.0 International (CC BY 4.0)
<https://creativecommons.org/licenses/by/4.0/>



For better, no better.

---- To Guo Zhongping & Xiao Ningqiong

Bibliografische Beschreibung

Guo, Ya

Clause complexing systems of Chinese students' academic writing

Technische Universität Chemnitz, Philosophische Fakultät

Dissertation (2025), 515 Seiten, 199 Abbildungen, 124 Tabellen im Hauptteil, 20 Tabellen im Anhang, 958 Zitate

Schlagwörter: Der Satzgefüge; logisch-semantische Beziehungen; Verfassen von Abschlussarbeiten; korpusbasierte Analyse; Auswahl und Wahrscheinlichkeit

Kurzfassung

Die vorliegende Studie widmet sich einer eingehenden Untersuchung von Satzgefügen (Clause Complex) in der Theorie der systemisch-funktionalen Grammatik (SFG). Dabei werden Satzverbindungen, Wissenszitate, Bedeutungsauslegungen und -vermittlungen sowie die Komplexität von Satzkombinationen beim Verfassen akademischer Texte berücksichtigt. Im Rahmen dieser Studie werden unter anderem die folgenden Unterkategorien untersucht: Expansion und Projektion (Projection), die Taxis von Satzverbindungen, der Bedeutungsausdruck im Text, die Wahl lexikalisch-grammatischer Kategorien, implizite Satzgefügen und die Komplexität von Satzgefügen innerhalb des Textes.

Ziel dieser Studie war zunächst die Untersuchung der Anordnung von Nebensätzen in einem Text durch Studierende, wobei ein besonderes Augenmerk auf Studierende gelegt wird, deren Muttersprache nicht Englisch ist und die sich bemühen, Englisch zu lernen. Ziel ist es, Ideen zu vermitteln. Die zweite Zielsetzung bestand in der Aufstellung und Bestätigung der Hypothese, dass Doktoranden im Hauptfach Englisch in China mehr und bessere Satzgefüge verwenden als Masterstudierende.

Ein weiteres Ziel dieser Forschung war die Differenzierung der Expansion von der Konjunktion und der Projektion von der Zitierung ihrer Forschungsbereiche, Positionen im entsprechenden theoretischen Rahmen, Anmerkungen und Bezeichnungen, der Betonung der Problemlösung, anwendbarer Aussichten usw. im theoretischen Teil. Ein weiterer Punkt war die Entwicklung eines Analyserahmens, der speziell auf der Grundlage des ursprünglichen SFG-Satzgefüges besser für akademische Diskurse oder das Schreiben von Abschlussarbeiten geeignet ist. Zudem sollten Kategorien im Rahmen durch Integration von Daten aus Forschungsergebnissen quantifiziert werden. Auf dieser Grundlage sollten einige neue Konzepte vorgeschlagen und dann einige neue und damit verbundene Forschungsfragen aufgezeigt werden. Im Rahmen dieser Dissertation wurde zudem die These überprüft, dass die Auswahl von Lokution (Locution) aus einem kleinen Korпустest der Doktoranden die der Masterstudenten übersteigt. Darüber hinaus wurden die folgenden Fragen erörtert: Verfügen die Studierenden über die Fähigkeit zur logischen Auswahl projizierender Verben beim Schreiben von Abschlussarbeiten? Und können diese Relationsmarker die Gedanken der Autoren, ihr Verständnis dieser Meinungen und ihre Absichten bei der Projektion und Bedeutungstransformation objektiv darstellen?

Bibliographic description

Guo, Ya

Clause complexing system of Chinese students' academic writing

Chemnitz University of Technology, Department of Philosophy

Dissertation (2025), 515 pages, 199 figures, 124 tables in main part, 20 tables in appendix, 958 citations

Keywords: The Clause Complex; Logico-semantic Relations; Thesis Writing; Corpus-based Analysis; Choice and Probability

Abstract

In light of the challenges posed by clause connection, knowledge citation, meaning construing and conveyance, and the complexity of clause combination in academic text writing, this study has been dedicated to an in-depth examination of clause complexes (CCs), with a specific emphasis on the subcategories of expansion and projection, the taxis of clause connections, the expression of meaning in the text, the choice of lexicogrammatical categories, implicit clause complexes, and the complexity of clause complexes within the text, among other aspects.

The study first intends to investigate how clauses are organized in a text by students, especially those whose mother tongue is not English and who endeavor to learn it to convey ideas. The second purpose is to confirm the hypothesis that PhD candidates of English majors in China would use more and better clause complexes than Master students whose different educational levels lead to diverse understandings and applications of grammatical structures such as clause complexes. On account of the similarity and comparability of concepts, "clause complex" in Systemic Functional Grammar and *Fuju* in traditional Chinese grammar, the denotation and connotation of the two concepts have been discussed in the theoretical part to inquire about their differences.

Another aim is to distinguish expansion from conjunction and projection from citation on their research spheres, locations in corresponding theoretical framework, annotation and denotation, emphasis on problem-solving, applicable prospects, etc., in the theoretical part. Besides, the research also inclines to deepen the theoretical foundation by following conventional definitions and the classification of conceptions in SFG categories; to develop an analyzing framework that is more suitable for academic discourse or thesis writing specifically based on the original SFG clause complex system; to quantify categories in the framework by integrating data from researching results; to put forward some new concepts on this basis and then raise some new and related research questions. Except for verifying the statement that doctoral candidates outnumber Master's students in choosing locutions obtained from a small corpus test, the questions of whether the ability to logically select projecting verbs in thesis writing has been equipped with by students or not and whether these relation markers can objectively present authors' thoughts, writers' understanding of those opinions and their intentions in projecting and meaning transforming or not have also been discussed in this dissertation.

TABLE OF CONTENTS

CHAPTER 1 INTRODUCTION	1
1.1 Research Background	1
1.1.1 Lexicogrammatical: Taxis and Expansion	3
1.1.2 Structural and Interpersonal: Projection and Projecting Verbs	5
1.1.3 Meaning Realization: The Implicit Language.....	6
1.1.4 Measuring: The Complexity of Clause Complexes	7
1.1.5 Summary	8
1.2 Purpose of the Study	9
1.3 Research Questions.....	10
1.4 Significance of the Study	11
1.5 Chapter Overview	15
CHAPTER 2 LITERATURE REVIEW.....	21
2.1 The Literature Review of Academic Writing.....	21
2.1.1 ESP/EAP/EAW	21
2.1.2 Theoretical Foundations.....	23
2.1.3 Research Methods of Academic Writing.....	24
2.1.4 Research Perspectives of Academic Writing.....	26
2.1.4.1 Discourse Analysis.....	26
2.1.4.2 Academic Literacy	36
2.1.5 Related Research of Academic Writing	41
2.1.6 English for Academic Writing Research on MA Theses versus PhD Dissertations.....	43
2.1.7 Conjunction vs. Expansion.....	43
2.1.8 Citation vs. Projection.....	45
2.1.8.1 Citation General.....	45
2.1.8.2 Citation Categories.....	46
2.1.8.3 Citation-related Research.....	49
2.1.8.4 Citation Verbs.....	50
2.1.8.5 Differences between Citation & Projection	51
2.1.9 Controversial	51
2.2 Previous Research of the Clause Complex	52
2.2.1 Research of the Clause Complex in China.....	52
2.2.1.1 On Clause Complex	52
2.2.1.2 On Projection	60
2.2.1.3 On Expansion.....	70
2.2.1.4 Related Research.....	71
2.2.2 Research of the CC in Other Countries.....	72
2.2.2.1 Theoretical Research.....	72
2.2.2.2 Discourse Analysis.....	77
2.2.2.3 Translation	87
2.2.3 Summary	88
2.3 Conclusion of Previous Research	90

CHAPTER 3 THEORY	93
3.0 Introduction.....	93
3.1 SFG: Background, Originality, and Comparison.....	93
3.1.1 SFG as a Theoretical Base	96
3.1.1.1 Choosing SFG.....	96
3.1.1.2 Socio-function for Academic Writing.....	99
3.1.1.2.1 Language as Social Semiotic	99
3.1.1.2.2 SFG as an Applicable Linguistics.....	100
3.1.1.2.3 The Context of Situation.....	100
3.1.1.3 The Contents and Main Components of SFG	101
3.1.1.3.1 Constituency	101
3.1.1.3.2 Structure, Unit, Rank	101
3.1.1.3.3 System, Delicacy.....	102
3.1.1.3.4 Function, Metafunction.....	103
3.1.1.3.5 Stratification, Realization	104
3.1.1.3.6 Instantiation.....	105
3.1.1.3.7 Text, Context, Register.....	106
3.1.1.3.8 Probabilistic, Markedness	107
3.1.1.4 Drawing for the Clause Complex	108
3.1.2 Clause & Clause Complex	108
3.1.2.1 Definitions of a <i>Clause</i>	108
3.1.2.2 Connotation of the <i>Clause Complex</i>	110
3.1.3 Expansion.....	115
3.1.3.1 Elaboration.....	116
3.1.3.1.1 Paratactic Elaboration	116
3.1.3.1.2 Hypotactic Elaboration	117
3.1.3.2 Extension.....	119
3.1.3.2.1 Paratactic Extension.....	119
3.1.3.2.2 Hypotactic Extension.....	120
3.1.3.3 Enhancement.....	122
3.1.3.3.1 Paratactic Enhancement	123
3.1.3.3.2 Hypotactic Enhancement (Finite & Non-finite)	125
3.1.4 Projection (without Embedding)	131
3.1.4.1 Quoting (“Direct Speech” - Parataxis).....	132
3.1.4.2 Reporting (“Indirect Speech” - Hypotaxis).....	133
3.1.4.3 Summary	137
3.1.5 Conclusion of SFG.....	137
3.2 Academic Writing & SFL	138
3.2.1 Combining AW with SFL.....	139
3.2.2 Seeing AW as a Register	139
3.2.2.1 The Definition of <i>Register</i>	139
3.2.2.2 The Reason for Seeing AW as a Register.....	140
3.2.2.3 The Meaning of Seeing AW as a Register	141
3.2.3 The Qualitative Analyzing Framework for the Projection CC (from AP & SFG)	143

3.3 Corpus Linguistics & SFL (Probabilistic)	143
3.3.1 A Brief History of the Intersection of CL & SFL	144
3.3.2 The Connection of CL & SFL	144
3.3.3 Related Research and Methods of CL & SFL	145
3.3.4 Related Key Concepts	146
3.3.5 Disadvantages of SFL in CL Perspective	147
3.3.6 Summary	149
3.4 Theoretical Framework for the Study	149
CHAPTER 4 DATA & METHODS	151
4.1 Data (ChAcE) Description	151
4.2 Research Methods	157
4.2.1 Qualitative & Quantitative Methods	157
4.2.1.1 Qualitative	157
4.2.1.2 Quantitative	158
4.2.1.3 Qualitative vs. Quantitative	158
4.2.2 Qualitative Methods	160
4.2.2.1 Exploratory Research (Heuristic Analysis)	161
4.2.2.2 Comparative	161
4.2.2.3 Experiential Summary	161
4.2.2.4 Grounded Theory	161
4.2.2.5 Case Study	162
4.2.2.6 Software: SysFan	162
4.2.3 Quantitative Methods	166
4.2.3.1 Descriptive Research	166
4.2.3.2 Corpus Linguistics	167
4.2.3.3 Modelling	168
4.2.3.4 JASP & Bayesian Independent T-test	168
4.2.4 Summary	174
4.3 Chosen Data: CMAC & CPhD	174
4.3.1 Corpus Design	174
4.3.2 Corpora General	175
4.3.2.1 General Information	175
4.3.2.2 Corpora Importing & Annotating	175
4.3.3 Corpora Details	175
4.3.3.1 The Number of Clauses	175
4.3.3.2 Words/Clauses	176
4.3.3.3 Clause Complexes/Sentences	178
4.3.3.4 Chosen CC/Sentences (1/2 clauses) Percentage	178
4.3.3.5 Clause Nexus/CC	181
4.3.3.6 Clause Nexuses (1-7) Information	182
4.3.3.6.1 General Distribution	182
4.3.3.6.2 Clause Nexuses on Level 1	184
4.3.3.6.3 Clause Nexuses on Level 2	186
4.3.3.6.4 Clause Nexuses on Level 3	187

4.3.3.6.5 Clause Nexuses on Level 4	188
4.3.3.7 Parataxis/Clause Nexuses Percentage	190
4.3.3.8 Hypotaxis/Clause Nexuses Percentage	192
4.3.4 Summary	192
4.4 Analysis Outline	193
CHAPTER 5 EXPANSION	195
5.1 Expansion Comparison	195
5.1.1 General Distribution of Expansion	195
5.1.1.1 Expansion/Clause Nexuses	195
5.1.2 The Distribution of Paratactic Expansion	197
5.1.2.1 Expansion Parataxis/Clause Nexuses	198
5.1.2.2 Expansion Parataxis/Parataxis	199
5.1.2.3 Expansion Parataxis/Expansion	200
5.1.3 The Distribution of Hypotactic Expansion	202
5.1.3.1 Expansion Hypotaxis/Clause Nexuses	202
5.1.3.2 Expansion Hypotaxis/Hypotaxis	203
5.1.3.3 Expansion Hypotaxis/Expansion	205
5.1.4 Expansion Composition	205
5.1.4.1 Elaboration (Descriptive Statistics; Elaboration/Expansion)	205
5.1.4.2 Extension (Descriptive Statistics; Extension/Expansion)	207
5.1.4.3 Enhancement (Descriptive Statistics; Enhancement/Expansion)	209
5.1.5 Summary of Expansion Comparison	211
5.2 Elaboration (Quantitatively Comparison)	211
5.2.1 General Distribution of Elaboration (Elaboration/Clause Nexuses)	211
5.2.2 The Distribution of Paratactic Elaboration	212
5.2.2.1 Elaboration Parataxis/Clause Nexuses	212
5.2.2.2 Elaboration Parataxis/Parataxis	214
5.2.2.3 Elaboration Parataxis/Expansion Parataxis	215
5.2.2.4 Elaboration Parataxis/Elaboration	216
5.2.3 The Distribution of Hypotactic Elaboration	218
5.2.3.1 Elaboration Hypotaxis/Clause Nexuses	218
5.2.3.2 Elaboration Hypotaxis/Hypotaxis	220
5.2.3.3 Elaboration Hypotaxis/Expansion Hypotaxis	221
5.2.3.4 Elaboration Hypotaxis/Elaboration	222
5.2.4 Summary of Elaboration Comparison	222
5.3 Elaboration (Qualitatively Analysis)	223
5.3.1 General Information of Elaboration	223
5.3.2 Exposition (Paratactic Elaboration)	223
5.3.3 Exemplification (Paratactic Elaboration)	226
5.3.4 Clarification (Paratactic Elaboration)	227
5.3.5 Hypotactic Elaboration	228
5.3.6 Summary of Elaboration Marker Choices	229
5.4 Extension (Quantitatively Comparison)	229
5.4.1 General Distribution of Extension (Extension/Clause Nexuses)	230

5.4.2 The Distribution of Paratactic Extension	231
5.4.2.1 Extension Parataxis/Clause Nexuses	231
5.4.2.2 Extension Parataxis/Parataxis	233
5.4.2.3 Extension Parataxis/Expansion Parataxis	234
5.4.2.4 Extension Parataxis/Extension.....	236
5.4.3 The Distribution of Hypotactic Extension	237
5.4.3.1 Extension Hypotaxis/Clause Nexuses.....	237
5.4.3.2 Extension Hypotaxis/Hypotaxis.....	238
5.4.3.3 Extension Hypotaxis/Expansion Hypotaxis.....	240
5.4.3.4 Extension Hypotaxis/Extension.....	241
5.4.4 Summary of Extension Comparison	242
5.5 Extension (Qualitatively Analysis)	242
5.5.1 General Information of Extension.....	242
5.5.2 Addition.....	242
5.5.2.1 Positive.....	243
5.5.2.2 Negative	245
5.5.2.3 Adversative	245
5.5.3 Variation	247
5.5.3.1 Replacive.....	247
5.5.3.2 Subtractive	249
5.5.4 Alternation.....	251
5.5.5 Summary of Extension Marker Choices	252
5.6 Enhancement (Quantitatively Comparison).....	252
5.6.1 General Distribution of Enhancement (Enhancement/Clause Nexuses).....	252
5.6.2 The Distribution of Paratactic Enhancement	253
5.6.2.1 Enhancement Parataxis/Clause Nexuses.....	254
5.6.2.2 Enhancement Parataxis/Parataxis	255
5.6.2.3 Enhancement Parataxis/Expansion Parataxis	257
5.6.2.4 Enhancement Parataxis/Enhancement	258
5.6.3 The Distribution of Hypotactic Enhancement.....	259
5.6.3.1 Enhancement Hypotaxis/Clause Nexuses.....	260
5.6.3.2 Enhancement Hypotaxis/Hypotaxis.....	261
5.6.3.3 Enhancement Hypotaxis/Expansion Hypotaxis.....	262
5.6.3.4 Enhancement Hypotaxis/Enhancement	264
5.6.4 Summary of Enhancement Comparison.....	265
5.7 Enhancement (Qualitatively Analysis)	265
5.7.1 General Information of Enhancement.....	265
5.7.2 Temporal.....	265
5.7.2.1 Same Time	265
5.7.2.2 Earlier.....	268
5.7.2.3 Later	269
5.7.3 Spatial.....	271
5.7.3.1 Same Place	271
5.7.4 Manner	272

5.7.4.1 Means.....	272
5.7.4.2 Comparison.....	273
5.7.5 Causal.....	275
5.7.5.1 Reason.....	276
5.7.5.2 Result	280
5.7.5.3 Purpose.....	281
5.7.6 Conditional.....	284
5.7.6.1 Positive.....	284
5.7.6.2 Negative	287
5.7.6.3 Concessive	289
5.7.7 Summary of Enhancement Marker Choices	292
5.8 Conclusion for Expansion Conditions	293
CHAPTER 6 PROJECTION	295
6.0 Introduction.....	295
6.1 Projection Comparison.....	295
6.1.1 General Distribution of Projection	295
6.1.1.1 Projection/Clause Nexuses.....	296
6.1.2 The Distribution of Paratactic Projection.....	297
6.1.2.1 Projection Parataxis/Clause Nexuses	298
6.1.2.2 Projection Parataxis/Parataxis.....	299
6.1.2.3 Projection Parataxis/Projection.....	301
6.1.3 The Distribution of Hypotactic Projection	302
6.1.3.1 Projection Hypotaxis/Clause Nexuses	303
6.1.3.2 Projection Hypotaxis/Hypotaxis	304
6.1.3.3 Projection Hypotaxis/Projection	306
6.1.4 Projection Composition.....	307
6.1.4.1 Locution (Descriptive Statistics; Locution/Projection)	307
6.1.4.2 Idea (Descriptive Statistics; Idea/Projection).....	309
6.1.5 Summary of Projection Comparison	310
6.2 Locution (Quantitatively Comparison)	310
6.2.1 General Distribution of Locution (Locution/Clause Nexuses)	311
6.2.2 The Distribution of Paratactic Locution.....	312
6.2.2.1 Locution Parataxis/Clause Nexuses	312
6.2.2.2 Locution Parataxis/Parataxis.....	314
6.2.2.3 Locution Parataxis/Projection Parataxis	316
6.2.2.4 Locution Parataxis/Locution.....	317
6.2.3 The Distribution of Hypotactic Locution	319
6.2.3.1 Locution Hypotaxis/Clause Nexuses	319
6.2.3.2 Locution Hypotaxis/Hypotaxis	320
6.2.3.3 Locution Hypotaxis/Projection Hypotaxis.....	322
6.2.3.4 Locution Hypotaxis/Locution	323
6.2.4 Summary of Locution Comparison	324
6.3 Locution (Qualitatively Analysis).....	324
6.3.1 General Information of Locution	324

6.3.2 Locution - Proposal	324
6.3.2.1 Objectivity (One Type)	325
6.3.2.2 Impartiality (Two Types)	325
6.3.2.3 Moderate Subjectivity (Four Types)	326
6.3.2.4 Intense Subjectivity (Six Types)	327
6.3.3 Locution - Proposition.....	329
6.3.3.1 Impartiality (Two Types)	330
6.3.3.2 Moderate Subjectivity (Three Types)	331
6.3.3.3 Intense Subjectivity (Seven Types).....	332
6.3.4 Summary of Locution Marker Choices	334
6.4 Idea (Quantitatively Comparison).....	334
6.4.1 General Distribution of Idea (Idea/Clause Nexuses)	335
6.4.2 The Distribution of Paratactic Idea	336
6.4.2.1 Idea Parataxis/Clause Nexuses	336
6.4.2.2 Idea Parataxis/Parataxis	337
6.4.2.3 Idea Parataxis/Projection Parataxis.....	338
6.4.2.4 Idea Parataxis/ Idea	339
6.4.3 The Distribution of Hypotactic Idea.....	340
6.4.3.1 Idea Hypotaxis/Clause Nexuses.....	340
6.4.3.2 Idea Hypotaxis/Hypotaxis.....	342
6.4.3.3 Idea Hypotaxis/Projection Hypotaxis	343
6.4.3.4 Idea Hypotaxis/ Idea	344
6.4.4 Summary of Idea Comparison	345
6.5 Idea (Qualitatively Analysis)	345
6.5.1 General Information of Idea	345
6.5.2 Idea - Proposal.....	345
6.5.2.1 Impartiality (Two Types)	345
6.5.2.2 Moderate Subjectivity (Three Types)	346
6.5.2.3 Intense Subjectivity (Four Types).....	347
6.5.3 Idea - Proposition	348
6.5.3.1 Impartiality (Four Types).....	349
6.5.3.2 Moderate Subjectivity (Five Types).....	350
6.5.3.3 Intense Subjectivity (Three Types)	351
6.5.4 Summary of Idea Marker Choices	352
6.6 Conclusion for Projection Conditions.....	352
CHAPTER 7 DISCUSSIONS	355
7.1 The Big System of Clause Complex in Academic Writing.....	355
7.1.1 System Comparison	355
7.1.1.1 Category Distribution of CC System in Master's Theses	355
7.1.1.2 Category Distribution of CC System in Doctoral Dissertations	356
7.1.1.3 Category Distribution of CC System in the Corpus.....	357
7.1.1.4 Comparison with Wu's Outcome	358
7.1.1.5 Category Range of CC System in the Corpus.....	359
7.1.2 Grammatical Comparison	360

7.1.2.1 BF Value Comparison	360
7.1.2.2 Comparison within Range, Mean, and BF Values	363
7.1.3 Lexical Comparison	367
7.1.3.1 Annotating Markers in Elaboration.....	367
7.1.3.2 Annotating Markers in Extension	368
7.1.3.3 Annotating Markers in Enhancement.....	368
7.1.3.4 Annotating Markers in Locution.....	371
7.1.3.5 Annotating Markers in Idea	372
7.1.4 Realization Analysis.....	373
7.1.5 Summary of CC System Comparison	373
7.2 Four Kinds within AW & Error Analysis.....	374
7.2.1 Four Kinds of Additions within Academic Writing	374
7.2.1.1 Unrelated.....	374
7.2.1.2 Unequal.....	374
7.2.1.3 Wrongly.....	375
7.2.1.4 Implicit.....	376
7.2.2 The Location of Four Additions	376
7.2.3 The Location of Four Layers in any Category of CC	377
7.2.4 Annotation for Wrong CCs.....	377
7.2.4.1 Adding.....	377
7.2.4.2 Missing.....	378
7.2.4.3 Substituting	380
7.2.4.4 Inconsistent	380
7.2.4.5 Not-splitting	381
7.2.4.6 Punctuation Mistakes	381
7.2.4.7 Wrong Usage.....	381
7.2.5 Summary of Four Additions and Error Analysis.....	382
7.3 Implicit & Explicit Clause Complex.....	382
7.4 Common Recursion Patterns and Clause Order.....	382
7.4.1 Common Patterns of Clause Complex (Nesting)	382
7.4.2 Clause Order in a Clause Nexus.....	382
7.5 The Complexity of the CC or the <i>Clause Complexical Complexity</i>	384
7.5.1 Origins.....	384
7.5.2 Theoretical Background for the Complexity of the CC	386
7.5.3 Definition of Clause Complexical Complexity.....	386
7.5.4 Three Formulas with Comparison.....	387
7.5.5 Formula Application.....	389
7.5.6 Significance & Limitation.....	390
7.6 Conclusion for Discussions.....	390
CHAPTER 8 CONCLUSION	393
8.1 Conclusion	393
8.2 Contributions of the Study	399
8.3 Implications (for Researchers, Translators, and Teachers)	402
8.4 Limitations of the Study.....	404

8.5 Suggestions (for Theories, Translation, and Teaching).....	406
REFERENCES	409
APPENDIX.....	445

ABSTRACT

In light of the challenges posed by clause connection, knowledge citation, meaning construing and conveyance, and the complexity of clause combination in academic text writing, this study has been dedicated to an in-depth examination of clause complexes (CCs), with a specific emphasis on the subcategories of expansion and projection, the taxis of clause connections, the expression of meaning in the text, the choice of lexicogrammatical categories, implicit clause complexes, and the complexity of clause complexes within the text, among other aspects.

The study first intends to investigate how clauses are organized in a text by students, especially those whose mother tongue is not English and who endeavor to learn it to convey ideas. The second purpose is to confirm the hypothesis that PhD candidates of English majors in China would use more and better clause complexes than Master students whose different educational levels lead to diverse understandings and applications of grammatical structures such as clause complexes. On account of the similarity and comparability of concepts, “clause complex” in Systemic Functional Grammar and *Fuju* in traditional Chinese grammar, the denotation and connotation of the two concepts have been discussed in the theoretical part to inquire about their differences.

Another aim is to distinguish expansion from conjunction and projection from citation on their research spheres, locations in corresponding theoretical framework, annotation and denotation, emphasis on problem-solving, applicable prospects, etc., in the theoretical part. Besides, the research also inclines to deepen the theoretical foundation by following conventional definitions and the classification of conceptions in SFG categories; to develop an analyzing framework that is more suitable for academic discourse or thesis writing specifically based on the original SFG clause complex system; to quantify categories in the framework by integrating data from researching results; to put forward some new concepts on this basis and then raise some new and related research questions. Except for verifying the statement that doctoral candidates outnumber Master’s students in choosing locations obtained from a small corpus test, the questions of whether the ability to logically select projecting verbs in thesis writing has been equipped with by students or not and whether these relation markers can objectively present authors’ thoughts, writers’ understanding of those opinions and their intentions in projecting and meaning transforming or not have also been discussed in this dissertation.

The particular research issues can be broken down into five distinct categories: theory (the construction of a framework for text analysis of clause complexes, discussions of the scope of research), methods (the corpus design, category quantification, software selection), practice (the way in which the lexicogrammatical continuum realizes logico-semantic relations in texts; the similarities and differences between master’s theses and doctoral dissertations in the selected corpus in some basic grammatical categories; the choice of the intersection of taxis and logico-semantic relations in the two groups of corpora to ascertain whether doctoral students’ thesis texts are more complex than those of master’s students; any preferences for logico-semantic relationship markers in theses; the recursive patterns of clause complexes, etc.), analysis (the interpretation of the results of the selected texts, expectation conformity, and the identification of any sources of data deviation), and application (the construction of the mathematical models for text complexity analysis; the use in a variety of fields including writing, teaching, and translation; the research methods improvement used for clause complexes).

The findings of the literature review (LR) indicated that academic writing (AW) primarily drew attention to some basic academic writing concepts, annotations, and denotations (concepts of EAP&ESP&EAW; research theories of discourse analysis, academic literacies, cognitive perspective, etc.; research methods of corpus linguistics (CL) from quantitative and qualitative perspectives; research directions in discourse analysis, academic literacies, cognition, identity construction, etc.), writing-related knowledge (writing purposes and motives, writing modes, writing analysis framework, and writing process), specific research on English thesis writing of Master and PhD students, conjunction and citation which are analogical to expansion and projection concepts in the clause complex system, and also studies on academic writing from systemic functional linguistics or grammar perspective.

There have been few studies on clause complexes, whether in China or abroad, in the past, let alone its integration with corpus linguistics and academic writing theory. After teasing out research on clause complex from its combination with academic writing or corpus linguistics, the intersection of academic writing and corpus linguistics directions, a multiple connection of research aspects on the clause complex, corpus linguistics and academic discourse writing, and other related works published both in domestic and abroad in the past thirty years, it was discovered that most of these studies would adopt qualitative rather than quantitative methods, and conduct discourse analysis instead of theoretical exploration; application of theories were easier for publishment while ontological research was hard for

researchers; themes on projection were quite common while on expansion were seldom; journal articles were constantly published though the contents and scope of which were limited because of length, and theses or dissertations were insufficient as to the number and thus deserve improvement for research depth though their length were considerable; a plenty of works were written in Chinese but only a handful were in English.

Based on the phenomena mentioned above, this research has established the necessity of taking corpus linguistics as a means of processing data and regarding academic text writing as a genre to design a comparative research program on the concept of clause complexes in Master's theses and doctoral dissertations. Before taking on the research, a series of theoretical concepts were also discussed, such as clearly defining the clause complex and its categories, limiting the scope of the clause complex for this study, distinguishing the connotation and denotation of expansion with conjunction, pointing out the difference within projection and citation, precisely stating the particularity of drawing on and integrating important concepts in the Appraisal Theory, etc. To reflect the ability to implement theories, a systematic and comprehensive framework for thoroughly describing and analyzing the clause complex has been established by combining SFG (supplies detailed and in-depth description and functional analysis for grammatical categories), Corpus Linguistics (provides theoretical supports for annotation, statistics and the analysis of a large amount of e-texts), and academic writing (directly affects the choice of clause complex from theoretical perspectives as a special genre), in which the qualitative research on Taxis in CC was based on the description of SFG, that of Expansion was built on basic ideas of CC and AW, and the qualitative investigation of Projection came from the CC framework and part of the Appraisal Theory.

The research object was the corpus of master's and doctoral theses written by English major students from Chinese universities. The researcher and colleagues from the Chemnitz University of Technology jointly established the corpus ChAcE (the Chinese Academic English). It contains 303 master's theses (CMAC, about 5.1 million words) and 112 doctoral dissertations (CPhD, approximately 5.3 million words). In this study, however, only a subset of the ChAcE corpus was selected for analysis, specifically 80 master's theses (1.28 million words, 28,350 clause complexes) and 32 doctoral dissertations (1.44 million words, 28,257 clause complexes) to examine clause complexes. By annotating nearly 50,000 clause complexes of their relations and then recording results in the software *SysFan*, the pioneering activity of manually annotating a large-scale corpus was realized, corresponding connecting words of CCs with markedness were taken notes, and grammatical categories in the whole system were counted. Other software, such as the auxiliary use of *AntConc*, reflected the consistency of results from the vocabulary indexing and the annotation, and the adoption of the statistical software *JASP* played a decisive role in both the Independent Samples T-test and the Bayesian factors numerical comparison of the grammatical-lexical selection results by Master's and doctoral students. The research also introduced a novel analytical method, namely an upgraded mixed research method, which entails the integration of qualitative (specifically exploratory research, comparative research, experiential summary, grounded theory, case study, etc.) and quantitative methods, or a combination of these two, following the requirements of diverse research subjects. The method was then applied in a unique analytical direction (a comparison of the clause complex structures chosen in Master's and doctoral theses from the field of AW), to be detailed, descriptive and qualitative methods (case study and grounded theory, in particular) were primarily used for annotation, and example analysis; quantitative and comparative methods were applied in data calculation; experimental summary and modeling analysis were utilized for system construction; and the exploratory method was adopted for outcome discussing, yielding a series of comparable results.

In terms of quantitative research, unlike previous studies that compared data by means or sums of consequences, a series of quantitative indexes (a measurement system of 73 indexes including numbers and proportion) within the clause complex system has been modeled, and plenty of annotated data were compared from statistical significance in the study.

The selected master's and doctoral theses exhibit certain discrepancies in fundamental aspects:

The number of words in each sentence (clause complexes) in each text in the master's group is around 18 - 31, with a mean number of 23.10, and in the doctoral group is about 18 - 33, with a mean number of 25.68.

The proportion of clause complexes/sentences in each text in the master's group is around 39% - 80%, with a mean proportion of 57.06%, and in the doctoral group is about 42% - 67%, with a mean proportion of 56.05%. Conversely, the proportion of clause simplexes/sentences in each text in the master's group is around 20% - 61%, with a mean proportion of 42.94%, and in the doctoral group is about 33% - 58%, with a mean proportion of 43.95%.

The proportion of clause nexuses per clause complexes in each text in the master's group is around 1.25 - 2.02 with a mean number of 1.56, while in the doctoral group is about 1.32 - 1.76 with a

mean number of 1.52.

For the recursion proportion of clause nexuses on each level in each text: on L1, the proportion in the master's group is around 64.88% while in the doctoral group is about 66.17%; on L2, the proportion in the master's group is around 25.83% while in the doctoral group is about 25.34%; on L3, the proportion in the master's group is around 7.33% while in the doctoral group is about 6.64%; on L4, the proportion in the master's group is around 1.55% while in the doctoral group is about 1.52%.

The proportion of parataxis per clause nexuses in each text in the master's group is around 24.39%-55.72% with a mean proportion of 39%, while in the doctoral group is about 18.12% - 46.20% with a mean proportion of 34.89%. Correspondingly, the proportion of hypotaxis per clause nexuses in each text in the master's group is around 44.28% - 75.61% with a mean proportion of 61%, while in the doctoral group is about 53.8% - 81.88% with a mean proportion of 65.11%.

The practice, hence, confirmed the feasibility of conducting quantitative analysis within grammatical categories, especially that of SFG, developed Halliday's probability theory on grammatical categories, and in the meantime, made the grammatical quantification from relations within and between categories extensively and profoundly.

By showing the proportion of each category within the clause complex system in Master's and doctoral theses from the corpus, the selection tendency of expansion or projection components intersected with taxis in the writing genre of academic papers can be analyzed and interpreted. The comparison results between the two groups thoroughly reflected the essential diversities in texts chosen by master's and doctoral students. The combining results of the two groups objectively revealed the fundamental reasons for diversely choosing each grammatical category, that is, from the characteristics of language expression itself on the one hand and attributed to writers' preferences and personal habits in writing papers on the other hand. Moreover, the proportion of each category in the clause complex system provided a dynamic perspective that some categories must be employed in academic writing with considerable frequency, some do not need to be chosen at all, and this kind of absence would not affect the integrality of meaning in the discourse and the accuracy of meaning expression.

In general, with regard to the choice of subcategories within the clause complex system as a whole, the distinction between the master's group and the doctoral one has been that the latter was more likely to exhibit a preference for hypotaxis, enhancement in expansion, hypotactic locution in hypotaxis, hypotactic extension in hypotaxis, hypotactic enhancement in hypotaxis, hypotactic enhancement in enhancement in texts writing; while the former had a higher probability in indexes such as idea in projection, extension in expansion, paratactic enhancement in parataxis, hypotactic idea in hypotaxis, hypotactic elaboration in hypotaxis, paratactic extension in extension. The corpus of written texts demonstrated a prevalence of hypotaxis, with approximately twice the frequency of parataxis. Expansion was approximately four times as prevalent as projection, with approximately 70% of projection being accounted for by ideas. Enhancement accounted for about 44% of expansions, extension accounted for approximately 41%, and elaboration accounted for only around 15%. In all paratactic relations, paratactic extension accounted for around 80%, with paratactic enhancement accounting for approximately 10%. In hypotactic relations, hypotactic enhancement accounted for approximately 51%, with hypotactic ideas accounting for approximately 21% and hypotactic elaboration accounting for nearly 15%. It is noteworthy that over 98% of ideas are hypotactic in nature. Furthermore, approximately 90% of enhancement was hypotactic enhancement, 84% of locution was hypotactic locution, and 80% of elaboration was hypotactic elaboration. In contrast, 90% of extensions were the paratactic extensions. These probabilities are contingent upon the intrinsic characteristics of the language itself.

Besides, the choice of subcategories within the system influences the choice of subcategories at the upper level, which is ultimately reflected in the distinction in lexical choice. It has been demonstrated that the positive addition relation marked by *and* was the most significant and challenging for students to master. As anticipated, the comparative analysis of the choices of paratactic enhancement and hypotactic enhancement between master's and doctoral students suggested that doctoral students were more inclined to the latter. This preference was observed regardless of the parallel measuring indexes, whether it was the average proportion of hypotactic enhancement in all hypotactic expansions or all enhancement structures (See Section 5.6). The qualitative analysis of enhancement was divided into five main functions as follows: temporal (same time, earlier, or later), spatial (same place), manner (means or comparison), causal (from reason, result, and purpose), and conditional (positive, negative, or concessive). The latter two functions were more significant when students employed causal relations to provide explanations for principles, indicate the influence that one matter has on another, and present researchers or their intentions to readers; and used conditional relations to identify sufficient or necessary conditions that were needed to fulfill something, highlight something that could not be realized despite conditions being met, or state the endeavor for something that has been confirmed as unachievable.

Whether PhD dissertations are more complex than Master's theses in choosing logico-semantic relations depends on which categories are regarded as markers for the complexity of clause complexes. Students did exhibit proclivity when choosing specific words and phrases that mark logico-semantic relations within clause complexes in thesis writing. The analysis demonstrated that there was a plethora of words and phrases that could be chosen to realize the same function. Moreover, the same words and phrases could be deployed to achieve subtly different functions in disparate contexts. It is important to note that certain words or phrases may serve multiple functions, and thus, the specific function they perform in a given text must be considered in its context. Generally, the choice of expansion lexis was primarily concerned with achieving discourse functions, connecting clauses, and establishing connections between preceding and subsequent elements in a clause complex. Conversely, the choice of projection lexis was predominantly focused on fulfilling interpersonal functions, demonstrating the interaction between individuals, which could also be observed in the writer's affirmation or negation of his/her perspectives, the expression of positive or negative attitudes, the indication of support or opposition, and the anticipation of readers' expectations.

Moreover, a qualitative comparison of lexical choices in annotated texts showed the reasons for preferentially choosing diverse grammatical functions by Master's and doctoral students, perspectives of making diverse words or phrases realized for the same grammatical function, and abilities to implement various expressing ways through one kind of grammatical functions corresponded to the same lexis in fulfilling meaning transmission; pointed out the fundamental reason for the difference in grammatical choices, i.e., the diversity in lexical expression; demonstrated the necessity of the existence of a delicate rank system; and explained the point that grammar and lexis are on a continuum and locate at two ends of the cline instead of being a superordinate-subordinate relation.

There are notable similarities and differences between the master's and doctoral students on recursion patterns, for example, both groups have a high probability of $1+2$, $\alpha x\beta$, $\alpha'\beta$, $x\beta\alpha$, $\alpha=\beta$, $\alpha'\beta(1+2)$, $x\beta\alpha(1+2)$, $\alpha'\beta(\alpha x\beta)$, $1+2(\alpha x\beta)$, $\alpha=\beta(1+2)$, and so forth. Furthermore, various patterns manifest at all levels, with the identification of repeating patterns becoming increasingly challenging at L3 and above levels in recursion.

Theoretical innovation of the research can be reflected in the proposal, characterization, and classification of the concept of *implicit clause complex*, the distinction between the implicit and explicit clause complexes, the methodological exploration of transforming implicit clause complexes into explicit ones, and the statistics and description of clause complex recursion patterns at each level. Being different from the practice of previous scholars that applied "syntactic complexity" in measuring structures within sentences to compare the complexity of texts, this study has focused on the clause relations and functions that primary annotating markers play in clause connection, proposed the concept of *clause complex complexity*, attempted to construct three calculation formulas based on category probabilities, and compared the sophistication of texts by calculating the complexity of clause complexes. In fact, the method of building the clause complex framework and then calculating categories in the framework for comparison adopted in this study is applicable to any subsystem of the systemic functional grammar, that is, bringing ways of annotating and quantifying categories in subsystem to solve problems such as: analyzing meaning expression in texts and calculating the complexity of the system structure from a functional perspective.

Admittedly, this study has some limitations, such as the condition of embedded and factual clause complexes that have not been regarded as research objects. Overall, the outcome of the Master's and PhD students' theses comparison can be inspiring and instructive for thesis writing, text reading, writing education, Chinese-English translation, machine translation, etc., error analysis can be enlightening for preventing more mistakes, the grammatical-lexical analyzing practices can be referential for conducting discourse analysis, the complexity of clause complex can be creative for further texts calculation and comparison and the clause combining modes from functional perspective can be innovatory for computational linguistics studies.

Key Words: The Clause Complex; Logico-semantic Relations; Thesis Writing; Corpus-based Analysis; Choice and Probability

ZUSAMMENFASSUNG

Die vorliegende Studie widmet sich einer eingehenden Untersuchung von Satzgefügen (Clause Complex) in der Theorie der systemisch-funktionalen Grammatik (SFG). Dabei werden Satzverbindungen, Wissenszitate, Bedeutungsauslegungen und -vermittlungen sowie die Komplexität von Satzkombinationen beim Verfassen akademischer Texte berücksichtigt. Im Rahmen dieser Studie werden unter anderem die folgenden Unterkategorien untersucht: Expansion¹ und Projektion (Projection)², die Taxis von Satzverbindungen, der Bedeutungsausdruck im Text, die Wahl lexikalisch-grammatischer Kategorien, implizite Satzgefügen und die Komplexität von Satzgefügen innerhalb des Textes.

Ziel dieser Studie war zunächst die Untersuchung der Anordnung von Nebensätzen in einem Text durch Studierende, wobei ein besonderes Augenmerk auf Studierende gelegt wird, deren Muttersprache nicht Englisch ist und die sich bemühen, Englisch zu lernen. Ziel ist es, Ideen zu vermitteln. Die zweite Zielsetzung bestand in der Aufstellung und Bestätigung der Hypothese, dass Doktoranden im Hauptfach Englisch in China mehr und bessere Satzgefüge verwenden als Masterstudierende. Dies lässt sich darauf zurückführen, dass die unterschiedlichen Bildungsniveaus der beiden Gruppen ein unterschiedliches Verständnis und eine unterschiedliche Anwendung grammatischer Strukturen, insbesondere von Satzgefügen, bedingen. Aufgrund der Ähnlichkeit und Vergleichbarkeit der Konzepte „Satzgefüge“ in der systemisch-funktionalen Grammatik und *Fuju* in der traditionellen chinesischen Grammatik wurden im theoretischen Teil die Denotation und Konnotation dieser beiden Konzepte diskutiert, um eine Differenzierung der beiden Begriffe vorzunehmen.

Ein weiteres Ziel dieser Forschung war die Differenzierung der Expansion von der Konjunktion und der Projektion von der Zitierung ihrer Forschungsbereiche, Positionen im entsprechenden theoretischen Rahmen, Anmerkungen und Bezeichnungen, der Betonung der Problemlösung, anwendbarer Aussichten usw. im theoretischen Teil. Darüber hinaus zielte die Forschung darauf ab, die theoretischen Grundlagen durch Befolgung konventioneller Definitionen und der Klassifizierung von Konzepten in SFG-Kategorien zu vertiefen. Ein weiterer Punkt war die Entwicklung eines Analyserahmens, der speziell auf der Grundlage des ursprünglichen SFG-Satzgefüges besser für akademische Diskurse oder das Schreiben von Abschlussarbeiten geeignet ist. Zudem sollten Kategorien im Rahmen durch Integration von Daten aus Forschungsergebnissen quantifiziert werden. Auf dieser Grundlage sollten einige neue Konzepte vorgeschlagen und dann einige neue und damit verbundene Forschungsfragen aufgezeigt werden. Im Rahmen dieser Dissertation wurde zudem die These überprüft, dass die Auswahl von Lokution (Locution)³ aus einem kleinen Korpus der Doktoranden die der Masterstudenten übersteigt. Darüber hinaus wurden die folgenden Fragen erörtert: Verfügen die Studierenden über die Fähigkeit zur logischen Auswahl projizierender Verben beim Schreiben von Abschlussarbeiten? Und können diese Relationsmarker die Gedanken der Autoren, ihr Verständnis dieser Meinungen und ihre Absichten bei der Projektion und Bedeutungstransformation objektiv darstellen?

Die jeweiligen Forschungsfragen lassen sich in fünf verschiedene Kategorien unterteilen: Theorie (Aufbau eines Rahmens für die Textanalyse von Satzgefügen, Diskussionen des Forschungsumfangs), Methoden (Korpusdesign, Kategorienquantifizierung, Softwareauswahl), Praxis (die Art und Weise, wie das lexikalisch-grammatische Kontinuum logisch-semantische Beziehungen in Texten realisiert; die Ähnlichkeiten und Unterschiede zwischen Masterarbeiten und Doktorarbeiten im ausgewählten Korpus in einigen grundlegenden grammatikalischen Kategorien; die Wahl der Schnittmenge von Taxis und logisch-semantischen Beziehungen in den beiden Korpusgruppen, um festzustellen, ob die Texte der Doktorarbeiten komplexer sind als die der Masterstudenten; etwaige Präferenzen für logisch-semantische Beziehungsmarker in Arbeiten; die rekursiven Muster von Satzgefügen usw.), Analyse (dazu gehören die Interpretation der Ergebnisse der ausgewählten Texte, die Erwartungskonformität und die Ermittlung etwaiger Quellen von Datenabweichungen) und Anwendung (Aufbau der mathematischen Modelle für die Textkomplexitätsanalyse; die Verwendung in zahlreichen Bereichen, darunter Schreiben, Lehren und Übersetzen; die Verbesserung der Forschungsmethoden für Satzgefügen).

Die Ergebnisse der Literaturrecherche verdeutlichten, dass im Rahmen des wissenschaftlichen Schreibens insbesondere einige grundlegende Konzepte, Anmerkungen und Denotationen des wissenschaftlichen Schreibens im Vordergrund standen (Konzepte von EAP&ESP&EAW;

¹ Eine Form der adverbialen Erweiterung.

² Konstituentensatz, im Unterschied zum Adverbialsatz.

³ Projizierte Rede.

Forschungstheorien der Diskursanalyse, wissenschaftliche Lese- und Schreibfähigkeiten, kognitive Perspektive usw.; Forschungsmethoden der Korpuslinguistik aus quantitativer und qualitativer Perspektive; Forschungsrichtungen in der Diskursanalyse, wissenschaftliche Lese- und Schreibfähigkeiten, Kognition, Identitätskonstruktion usw.), schreibbezogenes Wissen (Schreibzwecke und -motive, Schreibmodi, Rahmen der Schreibanalyse und Schreibprozess), spezifische Forschungen zum Verfassen von Abschlussarbeiten in Englisch von Master- und PhD-Studenten, Konjunktionen und Zitaten, die den Konzepten von Expansion und Projektion im System des Satzgefüges analog sind, sowie Studien zum wissenschaftlichen Schreiben aus der Perspektive der systemisch-funktionalen Linguistik oder Grammatik.

In der Vergangenheit gab es weder in China noch im Ausland Studien zu Satzgefügen, ganz zu schweigen von ihrer Integration mit der Korpuslinguistik und der Theorie des akademischen Schreibens. Nachdem man die Forschung zu Satzgefügen aus ihrer Kombination mit dem akademischen Schreiben oder der Korpuslinguistik, der Schnittstelle zwischen akademischem Schreiben und Korpuslinguistik, einer Vielzahl von Forschungsaspekten zu Satzgefügen, Korpuslinguistik und akademischem Diskursschreiben sowie anderen verwandten Arbeiten, die in den letzten dreißig Jahren im In- und Ausland veröffentlicht wurden, herausgearbeitet hatte, stellte man fest, dass die meisten dieser Studien eher qualitative als quantitative Methoden verwendeten; Diskursanalysen statt theoretischer Untersuchungen durchführten; die Anwendung von Theorien für die Veröffentlichung einfacher war, während ontologische Forschung für Forscher schwierig war; Projektionsthemen waren recht häufig, während Themen zur Expansion selten waren; ständig wurden Zeitschriftenartikel veröffentlicht, deren Inhalt und Umfang jedoch aufgrund der Länge begrenzt waren, und Abschlussarbeiten oder Dissertationen waren zahlenmäßig unzureichend und verdienten trotz ihrer beträchtlichen Länge eine Verbesserung hinsichtlich der Forschungstiefe; viele Arbeiten wurden auf Chinesisch verfasst, aber nur eine Handvoll englischer Literatur.

Die zuvor genannten Phänomene haben aufgezeigt, dass die Korpuslinguistik als Mittel der Datenverarbeitung genutzt werden sollte. Zudem sollte das Verfassen akademischer Texte als eigenes Genre betrachtet werden, um ein vergleichendes Forschungsprogramm zum Konzept von Satzgefügen in Master- und Doktorarbeiten zu entwerfen. Vor Beginn der Forschung wurden auch eine Reihe theoretischer Konzepte geklärt, beispielsweise die klare Definition des Satzgefüges und seiner Kategorien, die Begrenzung des Umfangs des Satzgefüges für diese Studie, die Unterscheidung der Konnotation und Denotation von Expansion mit Konjunktion, das Aufzeigen des Unterschieds zwischen Projektion und Zitat sowie die genaue Angabe der Besonderheit der Verwendung und Integration wichtiger Konzepte in der Appraisal Theory⁴. Um die Fähigkeit zur Umsetzung von Theorien widerzuspiegeln, wurde ein systematischer und umfassender Rahmen für die gründliche Beschreibung und Analyse des Satzgefüges durch die Kombination von SFG (liefert detaillierte und gründliche Beschreibungen und Funktionsanalysen für grammatische Kategorien), Corpuslinguistik (bietet theoretische Unterstützung für Annotationen, Statistiken und die Analyse großer Mengen von E-Texten) und akademischem Schreiben (beeinflusst direkt die Wahl des Satzgefüges aus theoretischen Perspektiven als spezielles Genre) erstellt, wobei die qualitative Forschung zu Taxis in CC auf der Beschreibung von SFG basierte, die von Expansion auf den Grundideen von CC und AW aufbaute und die qualitative Untersuchung von Projektion aus dem CC-Rahmen und einem Teil des Appraisal Theories.

Im Rahmen der vorliegenden Untersuchung wurde das Korpus der Master- und Doktorarbeiten von Anglistik-Studierenden chinesischer Universitäten als Forschungsobjekt herangezogen. Dieses Korpus mit dem Namen ChAcE (Chinese Academic English) wurde von der Forscherin in Zusammenarbeit mit Kollegen der TU Chemnitz erstellt. Das Korpus umfasst 303 Masterarbeiten (CMAC, rund 5,1 Millionen Wörter) und 112 Doktorarbeiten (CPhD, rund 5,3 Millionen Wörter). In der vorliegenden Studie wurde lediglich eine Teilmenge des ChAcE-Korpus zur Analyse ausgewählt, nämlich 80 Masterarbeiten (1,28 Millionen Wörter, 28.350 Satzgefügen) und 32 Doktorarbeiten (1,44 Millionen Wörter, 28.257 Satzgefügen), um Satzgefüge zu untersuchen. Durch die Annotation von fast 50.000 Satzgefügen ihrer Beziehungen und die anschließende Aufzeichnung der Ergebnisse in der Software *SysFan* wurde die Pioniertätigkeit der manuellen Annotation eines umfangreichen Korpus realisiert, entsprechende Verbindungswörter von Satzgefügen mit Markierungen wurden notiert und grammatische Kategorien im gesamten System wurden gezählt. Andere Software, wie die Hilfsverwendung von *AntConc*, spiegelte die Konsistenz der Ergebnisse aus der Vokabelindizierung und der Annotation wider, und die Einführung der Statistiksoftware *JASP* spielte sowohl beim T-Test für unabhängige Stichproben als auch beim numerischen Vergleich der grammatikalisch-lexikalischen

⁴ Eine Theorie sprachlicher Wertungen im Diskurs.

Auswahlergebnisse durch Master- und Doktoranden mit Bayesschen Faktoren eine entscheidende Rolle. Die Forschung führte auch eine neuartige Analysemethode ein, nämlich eine verbesserte gemischte Forschungsmethode, die die Integration qualitativer (insbesondere explorativer Forschung, vergleichender Forschung, Erfahrungszusammenfassung, Grounded Theory, Fallstudie usw.) und quantitativer Methoden oder einer Kombination dieser beiden gemäß den Anforderungen verschiedener Forschungsthemen beinhaltet. Die Methode wurde dann in einer einzigartigen analytischen Richtung angewendet (ein Vergleich der in Master- und Doktorarbeiten aus dem Bereich AW gewählten Satzgefügestrukturen), um detailliert zu sein. Für die Annotation und Beispielanalyse wurden hauptsächlich beschreibende und qualitative Methoden (insbesondere Fallstudien und Grounded Theory) verwendet. Für die Datenberechnung wurden quantitative und vergleichende Methoden angewendet. Für die Systemkonstruktion wurden experimentelle Zusammenfassungen und Modellanalysen verwendet. Für die Ergebnisdiskussion wurde die explorative Methode übernommen, was zu einer Reihe vergleichbarer Ergebnisse führte.

Im Hinblick auf die quantitative Forschung wurde im Gegensatz zu früheren Studien, die Daten einfach anhand von Mittelwerten oder Summen von Konsequenzen verglichen, eine Reihe von quantitativen Indizes (ein Messsystem mit 73 Indizes, darunter Zahlen und Proportionen) innerhalb des Satzgefüge Systems modelliert und anschließend zahlreiche annotierte Daten hinsichtlich ihrer statistischen Signifikanz verglichen.

Die ausgewählten Master- und Doktorarbeiten weisen interessante Unterschiede in grundlegenden Aspekten auf:

Die Anzahl der Wörter in jedem Satz (Satzgefügen) in jedem Text beträgt in der Mastergruppe etwa 18 - 31 mit einem Mittelwert von 23,10 und in der Doktorgruppe etwa 18 - 33 mit einem Mittelwert von 25,68.

Der Anteil der Satzgefüge/Sätze in jedem Text beträgt in der Mastergruppe etwa 39-80% bei einem durchschnittlichen Anteil von 57,06%, in der Doktorandengruppe etwa 42-67% bei einem durchschnittlichen Anteil von 56,05%. Umgekehrt beträgt der Anteil der Satzsimplexe (Clause Simplexes)/Sätze in jedem Text in der Mastergruppe etwa 20-61% bei einem durchschnittlichen Anteil von 42,94%, in der Doktorandengruppe etwa 33-58% bei einem durchschnittlichen Anteil von 43,95%.

Der Anteil der Satzverknüpfung (Clause Nexus) pro Satzgefüge in jedem Text beträgt in der Mastergruppe etwa 1,25-2,02 bei einem durchschnittlichen Anteil von 1,56, während er in der Doktorandengruppe etwa 1,32-1,76 bei einem durchschnittlichen Anteil von 1,52 beträgt.

Für den Rekursionsanteil der Satzverknüpfungen auf jeder Ebene in jedem Text gilt: Auf L1 beträgt der Anteil in der Mastergruppe etwa 64,88 %, während er in der Doktorandengruppe etwa 66,17% beträgt; auf L2 beträgt der Anteil in der Mastergruppe etwa 25,83%, während er in der Doktorandengruppe etwa 25,34% beträgt; auf L3 beträgt der Anteil in der Mastergruppe etwa 7,33%, während er in der Doktorandengruppe etwa 6,64% beträgt; auf L4 beträgt der Anteil in der Mastergruppe etwa 1,55%, während er in der Doktorandengruppe etwa 1,52% beträgt.

Der Anteil der Parataxen pro Satzverknüpfungen (Clause Nexuses) in jedem Text beträgt in der Mastergruppe etwa 24,39% - 55,72% bei einem durchschnittlichen Anteil von 39%, während er in der Doktorandengruppe etwa 18,12% - 46,20% bei einem durchschnittlichen Anteil von 34,89% beträgt. Entsprechend beträgt der Anteil der Hypotaxen pro Satzverknüpfungen in jedem Text in der Mastergruppe etwa 44,28% - 75,61% bei einem durchschnittlichen Anteil von 61%, während er in der Doktorandengruppe etwa 53,8% - 81,88% bei einem durchschnittlichen Anteil von 65,11% beträgt.

Die Praxis bestätigte somit die Durchführbarkeit der Durchführung quantitativer Analysen innerhalb grammatischer Kategorien, insbesondere der SFG, entwickelte Hallidays Wahrscheinlichkeitstheorie zu grammatischen Kategorien und führte in der Zwischenzeit eine umfassende und gründliche grammatische Quantifizierung aus Beziehungen innerhalb und zwischen Kategorien durch.

Die Darstellung des Anteils jeder Kategorie im Satzgefüge System in Master- und Doktorarbeiten aus dem Korpus ermöglicht eine Analyse und Interpretation der Auswahltendenz von Expansions- oder Projektionskomponenten, die sich mit Taxis in der Schreibart akademischer Arbeiten überschneiden. Die Vergleichsergebnisse innerhalb der beiden Gruppen spiegelten die wesentlichen Unterschiede in den Texten von Master- und Doktoranden genau wider, und die kombinierten Ergebnisse der beiden Gruppen enthüllten objektiv die grundlegenden Gründe für die unterschiedliche Auswahl jeder grammatikalischen Kategorie. Dabei zeigten sich einerseits die Merkmale des Sprachausdrucks selbst und andererseits die Vorlieben und persönlichen Gewohnheiten des Autors beim Schreiben von Arbeiten als ausschlaggebend. Darüber hinaus bot der Anteil jeder Kategorie im Satzgefüge System eine dynamische Perspektive, dass einige Kategorien in akademischen Arbeiten mit beträchtlicher Häufigkeit verwendet werden müssen, während andere überhaupt nicht ausgewählt werden müssen, und diese Art der Abwesenheit würde die Integrität der Bedeutung im Diskurs und die Genauigkeit des

Bedeutungsausdrucks nicht beeinträchtigen.

Im Allgemeinen bestand der Unterschied zwischen der Master- und der Doktorandengruppe bei der Wahl der Unterkategorien innerhalb des gesamten Satzgefüge Systems darin, dass letztere eher eine Präferenz für Hypotaxis, Enhancement⁵ in der Expansion, hypotaktische Lokution in Hypotaxis, hypotaktische Extension⁶ in Hypotaxis, hypotaktische Enhancement in Hypotaxis, hypotaktische Enhancement in Enhancement beim Schreiben von Texten zeigten, während erstere eine höhere Wahrscheinlichkeit für Indizes wie Idea⁷ in Projektion, Extension in Expansion, parataktische Enhancement in Parataxis, hypotaktische Idea in Hypotaxis, hypotaktische Elaboration⁸ in Hypotaxis und parataktische Extension in Extension hatten. Der Korpus der geschriebenen Texte wies eine Prävalenz von Hypotaxis auf, wobei Parataxis etwa doppelt so häufig war. Expansion war etwa viermal so häufig wie Projektion, wobei etwa 70% der Projektionen auf Ideas zurückzuführen waren. Enhancement machte etwa 44% der Expansionen aus, Extension etwa 41% und Elaboration nur etwa 15%. In allen parataktischen Beziehungen machte die parataktische Extension etwa 80% aus, während die parataktische Enhancement etwa 10% ausmachte. In hypotaktischen Beziehungen machte die hypotaktische Enhancement etwa 51% aus, wobei hypotaktische Ideas etwa 21% und hypotaktische Elaboration fast 15% ausmachten. Es ist bemerkenswert, dass über 98% der Ideas hypotaktischer Natur sind. Das Weiteren lässt sich feststellen, dass etwa 90% der Enhancement hypotaktische Enhancement, 84% der Lokution hypotaktische Lokution und 80% der Elaboration hypotaktische Elaboration sind. Im Gegensatz dazu handelte es sich bei 90% der Extension um parataktische Extension. Diese Wahrscheinlichkeiten hängen von den intrinsischen Eigenschaften der Sprache selbst ab.

Das Weiteren determiniert die Selektion der Subkategorien innerhalb des Systems die Selektion der Subkategorien auf der höheren Ebene, was sich letztlich in der Differenzierung der lexikalischen Präferenzen widerspiegelt. Es konnte nachgewiesen werden, dass die positive Additionsrelation, gekennzeichnet durch die Konjunktion *und*, für die Studierenden die größte Bedeutung hatte und zugleich die größte Herausforderung darstellte. Wie prognostiziert, zeigte die vergleichende Analyse der Wahl der parataktischen und hypotaktischen Enhancement zwischen Master- und Doktoranden, dass Doktoranden eine höhere Präferenz für die letztere aufwiesen. Diese Präferenz manifestierte sich unabhängig von den parallelen Messindizes, sei es der durchschnittliche Anteil der hypotaktischen Enhancement in allen hypotaktischen Expansionen oder allen Enhancement Strukturen (vgl. Abschnitt 5.6). Die qualitative Analyse der Enhancement wurde in fünf Hauptfunktionen unterteilt: zeitlich (gleiche Zeit, früher oder später), räumlich (gleicher Ort), Art und Weise (Mittel oder Vergleich), kausal (aus Grund, Ergebnis und Zweck) und bedingt (positiv, negativ oder konzessiv). Die beiden letzteren Funktionen waren bedeutender, wenn Studenten kausale Beziehungen verwendeten, um Erklärungen für Prinzipien zu liefern, den Einfluss einer Sache auf eine andere anzuzeigen und den Lesern Forscher oder ihre Absichten vorzustellen; und sie verwendeten Bedingungsrelationen, um hinreichende oder notwendige Bedingungen zu identifizieren, die für die Erfüllung von etwas erforderlich sind, um etwas hervorzuheben, das trotz erfüllter Bedingungen nicht realisiert werden konnte oder um das Bestreben nach etwas auszudrücken, von dem bestätigt wurde, dass es nicht erreichbar ist.

Die These, dass Dissertationen bei der Auswahl logisch-semantischer Beziehungen komplexer sind als Masterarbeiten, ist davon abhängig, welche Kategorien als Indikatoren für die Komplexität von Satzgefüge herangezogen werden. Studenten zeigten eine Neigung bei der Auswahl bestimmter Wörter und Ausdrücke, die logisch-semantische Beziehungen innerhalb von Satzgefügen beim Schreiben von Abschlussarbeiten markieren. Die Analyse ergab, dass eine Vielzahl von Wörtern und Ausdrücken zur Verfügung stand, die dieselbe Funktion erfüllen konnten. Darüber hinaus konnten dieselben Wörter und Ausdrücke eingesetzt werden, um in unterschiedlichen Kontexten leicht unterschiedliche Funktionen zu erreichen. Es ist wichtig zu beachten, dass bestimmte Wörter oder Ausdrücke mehrere Funktionen erfüllen können und daher die spezifische Funktion, die sie in einem bestimmten Text erfüllen, in seinem Kontext berücksichtigt werden muss. Im Allgemeinen ging es bei der Wahl der Expansionslexik in erster Linie darum, Diskursfunktionen zu erreichen, Sätze zu verbinden und Verbindungen zwischen vorhergehenden und nachfolgenden Elementen in einem Satzgefüge herzustellen. Im Gegensatz dazu war die Wahl der Projektionslexika vor allem auf die Erfüllung zwischenmenschlicher Funktionen ausgerichtet und zeigte die Interaktion zwischen Individuen. Diese Beobachtungen lassen sich auch auf die Bestätigung oder Verneinung der eigenen Ansichten durch den

⁵ Eine Form der adverbialen Erweiterung.

⁶ Eine Form der adverbialen Erweiterung.

⁷ Projizierte Idee.

⁸ Eine Form der adverbialen Erweiterung.

Autor, den Ausdruck positiver oder negativer Einstellungen, die Anzeige von Unterstützung oder Opposition sowie die Vorwegnahme der Erwartungen der Leserschaft übertragen.

Darüber hinaus zeigte ein qualitativer Vergleich der lexikalischen Auswahl bei der Annotation von Texten die Gründe dafür, warum Master- und Doktoranden bevorzugt unterschiedliche grammatische Funktionen auswählen, die Perspektiven für die Verwendung unterschiedlicher Wörter oder Ausdrücke für dieselbe grammatische Funktion und die Fähigkeit, verschiedene Ausdrucksweisen durch eine Art grammatischer Funktionen zu implementieren, die bei der Bedeutungsübertragung derselben Lexik entsprechen. Er wies auf den grundlegenden Grund für die Unterschiede bei der grammatikalischen Auswahl hin, d. h. die Vielfalt im lexikalischen Ausdruck. Er demonstrierte die Notwendigkeit der Existenz eines feinen Rangsystems und erläuterte den Punkt, dass Grammatik und Lexik auf einem Kontinuum liegen und an den beiden Enden des Kontinuums liegen, statt in einer Übergeordnete-Untergeordnete-Beziehung zu stehen.

Es lassen sich bemerkenswerte Ähnlichkeiten und Unterschiede zwischen den Master- und Doktoranden hinsichtlich der Rekursionsmuster feststellen. Diesbezüglich sei auf die hohe Wahrscheinlichkeit für Rekursionsmuster wie $1+2$, $\alpha\beta$, $\alpha'\beta$, $x\beta\alpha$, $\alpha=\beta$, $\alpha'\beta(1+2)$, $x\beta\alpha(1+2)$, $\alpha'\beta(\alpha\beta)$, $1+2(\alpha\beta)$, $\alpha=\beta(1+2)$, usw. verwiesen. Das Weiteren lassen sich zahlreiche unterschiedliche Muster beobachten. Mit dem Erreichen der Rekursionsebene L3 und höher wird die Identifizierung sich wiederholender Muster zunehmend schwieriger.

Die theoretische Innovation der Forschung manifestiert sich in dem Vorschlag, das Konzept des *impliziten Satzgefüges* zu charakterisieren und zu klassifizieren, implizite und explizite Satzgefüge voneinander zu unterscheiden, die Umwandlung impliziter Satzgefüge in explizite Satzgefüge methodisch zu untersuchen sowie Statistiken und Beschreibungen von Rekursionsmustern von Satzgefüge auf jeder Ebene zu erstellen. Im Gegensatz zur Vorgehensweise früherer Wissenschaftler, die den Begriff der „syntaktischen Komplexität“ zur Messung von Strukturen innerhalb von Sätzen anwandten, um die Komplexität von Texten zu vergleichen, konzentrierte sich diese Studie auf die Satzbeziehungen und Funktionen, die primäre Anmerkungsmarker bei der Satzverbindung spielen. Es wurde das Konzept der *Komplexität von Satzgefüge* vorgeschlagen, drei Berechnungsformeln auf Grundlage von Kategoriewahrscheinlichkeiten konstruiert und die Komplexität von Texten durch Berechnung der Komplexität von Satzgefügen verglichen. Die in dieser Studie verwendete Methode zum Aufbau des Klauselkomplexrahmens und zum anschließenden Berechnen von Kategorien im Vergleichsrahmen ist auf jedes Subsystem der systemisch-funktionalen Grammatik anwendbar. Dadurch ergeben sich Möglichkeiten zum Annotieren und Quantifizieren von Kategorien im Subsystem, um Probleme zu lösen, wie beispielsweise: Die vorliegende Studie befasst sich mit der Analyse des Bedeutungsausdrucks in Texten sowie der Berechnung der Komplexität der Systemstruktur aus einer funktionalen Perspektive.

Zugegebenermaßen hat diese Studie auch einige Einschränkungen, z. B. wurden eingebettete und faktische Satzgefüge nicht als Forschungsobjekte betrachtet. Insgesamt kann das Ergebnis des Vergleichs der Abschlussarbeiten von Master- und Doktoranden inspirierend und lehrreich für das Verfassen von Abschlussarbeiten, das Lesen von Texten, die Ausbildung im Bereich des Schreibens, die Übersetzung von Chinesisch nach Englisch, die maschinelle Übersetzung usw. sein, die Fehleranalyse kann aufschlussreich sein, um weitere Fehler zu vermeiden, die grammatikalisch-lexikalischen Analysepraktiken können als Referenz für die Durchführung von Diskursanalysen dienen, die Komplexität von Satzgefügen kann für die Berechnung und den Vergleich weiterer Texte herangezogen werden und die Kombination von Satzmodi aus funktionaler Perspektive kann für Studien zur Computerlinguistik innovativ sein.

Schlüsselwörter: Der Satzgefüge; logisch-semantische Beziehungen; Verfassen von Abschlussarbeiten; korpusbasierte Analyse; Auswahl und Wahrscheinlichkeit

ACKNOWLEDGMENT

The time spent at TU Chemnitz passes rapidly, akin to the fleeting shadow of a white pony across a crevice. I initially made the acquaintance of Professor Josef Schmied at Sun Yat-sen University, where he was engaged in pedagogical activities related to the Academic English Course for the Exchanging Students Program. I was fortunate to be selected as one of four students from my exchange program to attend TU Chemnitz in 2014. Subsequently, approximately six months after I had obtained my Master's degree, Prof. Schmied accepted me as his doctoral candidate, thereby commencing my PhD studies. However, the path to obtaining a degree was fraught with challenges. In the initial year, I was overwhelmed by the vast array of linguistics literature, which proved a significant challenge in navigating new ideas and approaches. It was Prof. Schmied who encouraged me, provided me with a structured academic framework, offered unwavering support, and facilitated my understanding of the role of a researcher. I have consistently recalled his counsel: "Keep thinking, keep writing." With his professional dedication and kindness, I could pursue my studies in Germany and obtain my doctoral degree. I am therefore extremely grateful to Prof. Dr. Josef Schmied.

Secondly, I am immensely grateful to TU Chemnitz for offering the InProTUC and DAAD-program STIBET Doktoranden Scholarship on an annual evaluation for doctoral candidates. I was fortunate to attend academic conferences in Shanghai, Xi'an, and Nanjing in China, where I had the opportunity to present my research. Additionally, I conducted a research visit at SYSU in Guangzhou, China, and completed my dissertation without significant financial constraints. I am grateful to Ms. Antje Pfeifer and Ms. Susan Kuchs, the coordinators of these scholarships, as well as my supervisor, for making these opportunities possible. As my research is focused on the English academic writing of Chinese students, these invaluable academic experiences in my home country have provided me with further inspiration and evidence for my ongoing studies. Without financial support for travel, accommodation, conference fees, and other related expenses, the finish of my dissertation would have been unfeasible.

As the adage states, optimal health is the foundation of revolutionary endeavors, as it is only when our bodies are in an optimal condition that we can achieve our goals. In the second year, I developed a benign tumor of considerable size, approximately the dimensions of a human fist. I subsequently underwent surgical intervention at Klinikum Chemnitz. Following a recuperation period of roughly six weeks, I could resume my research activities. During my recovery, I realized the importance of happiness and good health. I am profoundly grateful to the attending doctor, Lutz Kaltoven, and the associate doctors, Vincent Prochnow and Anja Gebbert, for performing the successful operation. I am also indebted to the careful nursing staff, including the assistant doctors Anja Leonhardt and Maria Kaiser, who provided fluent German-English translation, and other nurses who assisted me during my recovery. I am, therefore, profoundly grateful to them.

I would also like to express my gratitude to my colleagues and friends, including Xinlei Zhang and her husband Liang Chen, Wenyi Zhang and her family, Jessica Dheskali, Vincenzo Dheskali, Jacinta Edusei, and others, who have provided me with valuable insights and guidance on my research, as well as demonstrated their commitment to my well-being, particularly Xinlei Zhang. She frequently prepares Chinese cuisine, including dumplings, pancakes, sweet-and-sour pork, and other dishes, then takes them to me, demonstrating an understanding of my longing for home and the inconvenience of cooking. When I was indisposed, she even drove me to the hospital with decisive action, provided me with stewed beef and devoted caretakers, assisted my mother with visa documents wholeheartedly, and accompanied me uncomplainingly. I consider myself extremely fortunate to have her as a friend.

In the third year of my studies, I had the opportunity to meet my future husband, Fan XU. As a "pistachio," he consistently elicits laughter with his witty remarks and actions. Being a partner, he demonstrates remarkable patience and concern, providing care and support. As a chef, he dedicates considerable time and effort to preparing exquisite cuisines, showcasing remarkable skills. As a student,

his dedication and seriousness about studies motivate me to pursue my academic endeavors with greater devotion. As a brother, he joins me in various activities, including swimming, fitness, opera, and Christmas markets, demonstrating enthusiasm and companionship. And as a friend, he often sends gifts, showcasing his thoughtfulness and effort. In light of these observations, I extend my sincerest gratitude to him for his unwavering support and constructive feedback, particularly regarding my self-perceived deficiencies in confidence and ability. The completion of the dissertation would not have been possible without the contribution of Fan and his family, whose input should, therefore, be acknowledged as half of the dissertation.

I am indebted to Prof. M.A.K. Halliday and Prof. Christian Matthiessen for their invaluable contributions to the field of linguistics, which have significantly enhanced my understanding of language and the world. Halliday's perspective on "natural language as a dynamic open system" (Halliday, 1987) and Matthiessen's approach to conducting research between instantiation and probabilities (Matthiessen, 2002) provided me with insights and a theoretical foundation for this dissertation. I am also indebted to Professor Erich Steiner for offering me the opportunity to present a research report at Saarland University, and helping me with the Zusammenfassung in this thesis. I have derived considerable benefit from our discussions thus far. Additionally, other scholars in the field of SFG, including Robin Fawcett, Lise Fontaine, Geoff Thompson, Susan Hunston, Jonathan Webster, Jim Martin, John Bateman, Marina Bondi, Akila Sellami-Baklouti, Monika Bednarek, etc.; C. S. Butler in FG; Douglas Biber, John Sinclair, Tony McEnery, Susan Conrad, Jan Svartvik, etc. in corpus linguistics; John Swales, Ken Hyland, Lachlan Mackenzie and Ian Bruce in academic writing, Talmy Givón and Susan Hunston for grammar, and Zhiwei Feng in computational linguistics, etc. have provided valuable insights and guidance that have influenced my work. I find great joy in immersing myself in their works, akin to the exhilarating experience of uncovering precious gems from the depths of the sea.

For the entire defense process, I would like to thank Professor Klaus Stolz (In fact, he played a crucial role in my entire doctoral and research career and provided me with unlimited help), Professor Christina Sanchez-Stockhammer, and Jun.-Prof. Jennifer Schluer for their strong support. They raised many constructive questions about the thesis, which are worth pondering. I am also very grateful to my colleagues and friends in the English Department who attended my defense. I would also like to thank Qi Cai for providing me with accommodation for my defense preparation, and especially thank Fan Xu and Weiwei Chen for giving me enough strength and courage to face a series of difficult things that happened before and during the defense.

These acknowledgments would be incomplete without expressing gratitude to my esteemed parents, Zhongping Guo and Ningqiong Xiao, who have provided unwavering guidance and support, both emotionally and financially, throughout my academic journey. It is difficult to envisage the trajectory of this dissertation without the influence of my parents, whose wisdom, foresight, and kindness have been instrumental in my development. I consider myself extremely fortunate to be their child.

I would be remiss if I did not express gratitude to Prof. Ding Jianxin. He encouraged me to contribute articles on the current state of SFG research in Europe and consistently support throughout my academic career despite no longer serving as my master's supervisor.

Furthermore, I would like to show respect to my relatives for their financial assistance and my friends for their blessings. Again, I would be remiss if I did not acknowledge those individuals who have assisted me in the past but are not included in this list.

Attaining the degree is merely the initial stage of my academic career. Given the length of the journey ahead, I will pursue a wide range of avenues to achieve my goal!

LIST OF FIGURES

Figure 1 The “Create A Research Space (CARS) model” put forward by Swales (1990)	30
Figure 2 The structural model of academic writing by Murray & Moore (2006, adapted from Moore, 2003)	42
Figure 3 The organization of different kinds of realization statements in SFG.....	105
Figure 4 The intersection of the cline of instantiation and stratification	106
Figure 5 The system network of the clause complex in English (ibid, p. 44)	109
Figure 6 The general clause complex system from SFG (Halliday & Matthiessen, 2014, p. 438)	112
Figure 7 The analyzing framework of clause complexes in corpora for this thesis (from the SFG perspective).....	137
Figure 8 Register variates along the hierarchy of stratification within semantic variation and the context of use (functional variation) crossed by the cline of instantiation between the meaning potential and the instance pole.....	140
Figure 9 The framework for analyzing situational characteristics of registers and genres (Biber & Conrad, 2009, p. 40)	142
Figure 11 The theoretical framework for analyzing the clause complex system in this thesis under relevant theories of SFL, AW, and CL	149
Figure 12 The homepage of the Chinese version website CNKI.....	151
Figure 13 The page that shows search results with presumptive keywords or phrases.....	151
Figure 14 A dissertation example that can be downloaded.....	152
Figure 15 The increase of data annotation labor intensity and the location of the software SysFan for text analysis along with stratification and rank.....	163
Figure 16 The text import interface of SysFan with related information filling in columns.....	164
Figure 17 The chunk interface of texts into clause complexes in SysFan	164
Figure 18 The clause complex annotation interface in SysFan	165
Figure 19 The interface of clause complex annotation notes in SysFan.....	165
Figure 20 The interface of analyzing the matrix of clause complex categories in one text in SysFan.....	166
Figure 21 The plots choosing interface of descriptive statistics in JASP	168
Figure 22 The testing parameters of descriptive statistics in JASP in this project	169
Figure 23 The interface of the Independent Samples T-test in JASP	169
Figure 24 The Vovk-Sellke maximum p-ratio calculation application	171
Figure 25 The interpretation of the Bayes factor (Jeffreys, 1961; Kass & Raftery, 1995)	172
Figure 26 The interface of the Bayesian Independent Samples T-test in JASP	172
Figure 27 The inferential plot interface of prior and posterior in JASP (words per clauses of Masters (H0) and PhD (H+) groups as an example)	173
Figure 28 The inferential plot interface of the Bayes factor robustness check in JASP (clause nexus percentage on Level 1 of Masters (H0) and PhD (H+) groups as an example).....	173
Figure 29 The inferential plot interface of the sequential analysis in JASP (clause nexus percentage on Level 1 of Masters (H0) and PhD (H+) groups as an example)	173
Figure 30 The distribution boxplot of the number of clauses in Master and PhD sub-corpora.....	176
Figure 31 The distribution boxplot of the number of words in clauses in Master and PhD sub-corpora	176
Figure 32 The inferential plot of the number of words in clauses in Master and PhD sub-corpora.....	177
Figure 33 The boxplots of the number of clause complexes chosen for annotation in Master and PhD sub-corpora	178
Figure 34 The boxplots of the percentage of clause complexes in chosen sentences in Master and PhD sub-corpora	179
Figure 35 The inferential plots comparing the clause complex per chosen sentences percentage in Master and PhD sub-corpora with the null hypothesis group $D > \text{group M}$	180
Figure 36 The inferential plots comparing the clause complex per chosen sentence percentage in Master and PhD sub-corpora with the null hypothesis group $D < \text{group M}$	180
Figure 37 The inferential plots comparing the clause complex per chosen sentence percentage in Master and PhD sub-corpora with the null hypothesis group $D = \text{group M}$	180
Figure 38 The boxplots of clause nexuses per clause complex in Master and PhD sub-corpora	181
Figure 39 The inferential plots of the clause nexuses per clause complex in Master and PhD sub-corpora	182
Figure 40 The clause complex recursion annotation interface of E.g., 4.3.1 in SysFan.....	182
Figure 41 The boxplots of the clause nexuses in Master and PhD sub-corpora	183

Figure 42 The boxplots of clause nexuses on levels 1-3 in Master and PhD sub-corpora	184
Figure 43 The distribution plots of clause nexuses percentage on Level 1 in Master and PhD sub-corpora	184
Figure 44 The Bayesian Independent samples t-test inferential plots of the clause nexus percentage on Level 1 in Master and PhD sub-corpora	185
Figure 45 The distribution plots of clause nexuses percentage on Level 2 in Master and PhD sub-corpora	186
Figure 46 The Bayesian Independent samples t-test inferential plots of the clause nexus percentage on Level 2 in Master and PhD sub-corpora	187
Figure 47 The distribution plots of clause nexuses percentage on Level 3 in Master and PhD sub-corpora	187
Figure 48 The Bayesian Independent samples t-test inferential plots of the clause nexus percentage on Level 3 in Master and PhD sub-corpora	188
Figure 49 The distribution plots of clause nexuses percentage on Level 4 in Master and PhD sub-corpora	188
Figure 50 The inferential plots (prior and posterior) of the clause nexus percentage on Level 4 in Master and PhD sub-corpora	189
Figure 51 The boxplots of the amount of parataxis in Master and PhD sub-corpora.....	190
Figure 52 The boxplots comparing the paratactic clause complex percentage in Master and PhD sub-corpora	191
Figure 53 The inferential plots (prior and posterior) of the paratactic clause complex percentage in Master and PhD sub-corpora	191
Figure 54 The integration of SFG, Corpus Linguistics, and Academic Writing theories and methods for conducting this research.....	193
Figure 55 The boxplots of expansion distribution in Master and PhD sub-corpora.....	195
Figure 56 The boxplots of expansion per clause nexuses proportion in Master and PhD sub-corpora	196
Figure 57 The inferential plots comparing the expansion per clause nexuses proportion in Master and PhD sub-corpora with different alternative hypotheses.....	197
Figure 58 The boxplots of the number of paratactic expansions in Master and PhD sub-corpora.....	197
Figure 59 The boxplots comparing the proportion of paratactic expansion per clause nexuses in Master and PhD sub-corpora.....	198
Figure 60 The inferential plots comparing the paratactic expansion per clause nexus proportion in Master and PhD sub-corpora (with the original hypothesis group $D < \text{group M}$).....	199
Figure 61 The boxplots comparing the paratactic expansion in parataxis in Master and PhD sub-corpora	199
Figure 62 The inferential plots comparing the paratactic expansion per parataxis in Master and PhD sub-corpora (with different original hypotheses).....	200
Figure 63 The boxplots comparing the proportion of paratactic expansion in expansion in Master and PhD sub-corpora	201
Figure 64 The inferential plots comparing the proportion of paratactic expansion in expansions in Master and PhD sub-corpora (with the original hypothesis group $D < \text{group M}$).....	201
Figure 65 The boxplots comparing the number of hypotactic expansions in Master and PhD sub-corpora	202
Figure 66 The boxplots comparing the proportion of hypotactic expansion in clause nexuses in Master and PhD sub-corpora.....	202
Figure 67 The inferential plots comparing the proportion of hypotactic expansion in clause nexuses in Master and PhD sub-corpora (with the original hypothesis group $D > \text{group M}$).....	203
Figure 68 The boxplots comparing the proportion of hypotactic expansion in hypotaxis in Master and PhD sub-corpora	204
Figure 69 The inferential plots comparing the proportion of hypotactic expansion in hypotaxis in Master and PhD sub-corpora with diverse original hypotheses.....	205
Figure 70 The boxplots of the number of elaborations in Master and PhD sub-corpora.....	205
Figure 71 The boxplots comparing the proportion of elaboration per expansion in Master and PhD sub-corpora	206
Figure 72 The inferential plots comparing the proportion of elaboration in expansion in Master and PhD sub-corpora (with diverse original hypotheses).....	207
Figure 73 The boxplots of the number of extensions in Master and PhD sub-corpora	207
Figure 74 The boxplots comparing the proportion of extension in expansion in Master and PhD sub-corpora	208
Figure 75 The inferential plots comparing the proportion of extension in expansion in Master and PhD	

sub-corpora (with the original hypothesis group $D < \text{group M}$).....	209
Figure 76 The boxplots of the number of enhancements in Master and PhD sub-corpora	209
Figure 77 The boxplots comparing the proportion of enhancement in expansion in Master and PhD sub-corpora	210
Figure 78 The inferential plots comparing the proportion of enhancement in expansions in Master and PhD sub-corpora (with the original hypothesis group $D > \text{group M}$).....	210
Figure 79 The inferential plots comparing the proportion of elaboration of clause nexuses in Master and PhD sub-corpora (with the original hypothesis group $D < \text{group M}$).....	212
Figure 80 The boxplots of the number of paratactic elaborations in Master and PhD sub-corpora	212
Figure 81 The boxplots comparing the proportion of paratactic elaboration in clause nexuses in Master and PhD sub-corpora.....	213
Figure 82 The inferential plots comparing the proportion of paratactic elaboration in clause nexuses in Master and PhD sub-corpora (with the original hypothesis group $D < \text{group M}$).....	213
Figure 83 The boxplots comparing the paratactic elaboration in parataxis in Master and PhD sub-corpora	214
Figure 84 The inferential plots comparing the paratactic elaboration in parataxis in Master and PhD sub-corpora (with diverse original hypotheses).....	215
Figure 85 The boxplots comparing the proportion of paratactic elaboration in paratactic expansion in Master and PhD sub-corpora	215
Figure 86 The inferential plots comparing the proportion of paratactic elaboration in paratactic expansion in Master and PhD sub-corpora (with diverse original hypotheses).....	216
Figure 87 The boxplots comparing the proportion of paratactic elaboration in elaboration in Master and PhD sub-corpora	217
Figure 88 The inferential plots comparing the proportion of paratactic elaboration in elaborations in Master and PhD sub-corpora	217
Figure 89 The boxplots comparing the number of hypotactic elaborations in Master and PhD sub-corpora	218
Figure 90 The boxplots comparing the proportion of hypotactic elaboration in clause nexuses in Master and PhD sub-corpora.....	218
Figure 91 The inferential plots comparing the proportion of hypotactic elaboration in clause nexuses in Master and PhD sub-corpora (with diverse original hypotheses).....	219
Figure 92 The boxplots comparing the proportion of hypotactic elaboration in hypotaxis in Master and PhD sub-corpora	220
Figure 93 The inferential plots comparing the proportion of hypotactic elaboration in hypotaxis in Master and PhD sub-corpora	220
Figure 94 The inferential plots comparing the proportion of hypotactic elaboration in hypotactic expansion in Master and PhD sub-corpora (with the original hypothesis group $D < \text{group M}$)	221
Figure 95 The inferential plots comparing the proportion of hypotactic elaboration in elaborations in Master and PhD sub-corpora (with diverse original hypotheses).....	222
Figure 96 The boxplots comparing the proportion of extension in clause nexuses in Master and PhD sub-corpora	230
Figure 97 The inferential plots comparing the proportion of extension in clause nexuses in Master and PhD sub-corpora (with the original hypothesis group $D < \text{group M}$).....	231
Figure 98 The boxplots of the number of paratactic extensions in Master and PhD sub-corpora.....	231
Figure 99 The boxplots comparing the proportion of paratactic extension in clause nexuses in Master and PhD sub-corpora.....	232
Figure 100 The inferential plots comparing the proportion of paratactic extension in clause nexuses in Master and PhD sub-corpora (with the original hypothesis group $D < \text{group M}$).....	232
Figure 101 The boxplots comparing the paratactic extension in parataxis in Master and PhD sub-corpora	233
Figure 102 The inferential plots comparing the paratactic extension in parataxis in Master and PhD sub-corpora (with diverse original hypotheses).....	234
Figure 103 The boxplots comparing the proportion of paratactic extension in paratactic expansions in Master and PhD sub-corpora	234
Figure 104 The inferential plots comparing the proportion of paratactic extension in paratactic expansions in Master and PhD sub-corpora.....	235
Figure 105 The boxplots comparing the proportion of paratactic extension in extensions in Master and PhD sub-corpora	236
Figure 106 The inferential plots comparing the proportion of paratactic extension in extensions in Master and PhD sub-corpora (with the original hypothesis group $D < \text{group M}$).....	236

Figure 107 The boxplots comparing the number of hypotactic extensions in Master and PhD sub-corpora	237
Figure 108 The boxplots comparing the proportion of hypotactic extension in clause nexuses in Master and PhD sub-corpora.....	237
Figure 109 The inferential plots comparing the proportion of hypotactic extension in clause nexuses in Master and PhD sub-corpora (with the original hypothesis group $D > \text{group M}$).....	238
Figure 110 The boxplots comparing the proportion of hypotactic extension in hypotaxis in Master and PhD sub-corpora	239
Figure 111 The inferential plots comparing the proportion of hypotactic extension in hypotaxis in Master and PhD sub-corpora (with the original hypothesis group $D > \text{group M}$).....	239
Figure 112 The boxplots comparing the proportion of hypotactic extension in hypotactic expansion in Master and PhD sub-corpora	240
Figure 113 The inferential plots comparing the proportion of hypotactic extension in hypotactic expansion in Master and PhD sub-corpora (with the original hypothesis group $D > \text{group M}$)	241
Figure 114 The boxplots comparing the proportion of hypotactic extension in extensions in Master and PhD sub-corpora	241
Figure 115 The inferential plots comparing the proportion of hypotactic extension in extensions in Master and PhD sub-corpora	242
Figure 116 The boxplots comparing the proportion of enhancement in clause nexuses in Master and PhD sub-corpora	252
Figure 117 The inferential plots comparing the proportion of extension in clause nexuses in Master and PhD sub-corpora (with the original hypothesis group $D > \text{group M}$).....	253
Figure 118 The boxplots of the number of paratactic enhancements in Master and PhD sub-corpora	254
Figure 119 The boxplots comparing the proportion of paratactic enhancement in clause nexuses in Master and PhD sub-corpora	254
Figure 120 The inferential plots comparing the proportion of paratactic enhancement in clause nexuses in Master and PhD sub-corpora (with the original hypothesis group $D < \text{group M}$).....	255
Figure 121 The boxplots comparing the paratactic enhancement in parataxis in Master and PhD sub-corpora	255
Figure 122 The inferential plots comparing the paratactic enhancement in parataxis in Master and PhD sub-corpora (with diverse original hypotheses).....	256
Figure 123 The boxplots comparing the proportion of paratactic enhancement in paratactic expansions in Master and PhD sub-corpora	257
Figure 124 The inferential plots comparing the proportion of paratactic enhancement in paratactic expansions in Master and PhD sub-corpora.....	258
Figure 125 The boxplots comparing the proportion of paratactic enhancement in enhancements in Master and PhD sub-corpora	258
Figure 126 The inferential plots comparing the proportion of paratactic enhancement in enhancements in Master and PhD sub-corpora (with the original hypothesis group $D < \text{group M}$).....	259
Figure 127 The boxplots comparing the number of hypotactic enhancements in Master and PhD sub-corpora	260
Figure 128 The boxplots comparing the proportion of hypotactic enhancement in clause nexuses in Master and PhD sub-corpora	260
Figure 129 The inferential plots comparing the proportion of hypotactic enhancement in clause nexuses in Master and PhD sub-corpora (with the original hypothesis group $D > \text{group M}$).....	261
Figure 130 The boxplots comparing the proportion of hypotactic enhancement in hypotaxis in Master and PhD sub-corpora.....	261
Figure 131 The inferential plots comparing the proportion of hypotactic enhancement in hypotaxis in Master and PhD sub-corpora (with the original hypothesis group $D > \text{group M}$).....	262
Figure 132 The boxplots comparing the proportion of hypotactic enhancement in hypotactic expansion in Master and PhD sub-corpora	263
Figure 133 The inferential plots comparing the proportion of hypotactic enhancement in hypotactic expansion in Master and PhD sub-corpora (with the original hypothesis group $D > \text{group M}$)	263
Figure 134 The boxplots comparing the proportion of hypotactic enhancement in enhancements in Master and PhD sub-corpora	264
Figure 135 The inferential plots comparing the proportion of hypotactic enhancement in enhancements in Master and PhD sub-corpora	264
Figure 136 The boxplots of projection distribution in Master and PhD sub-corpora.....	295
Figure 137 The boxplots of expansion per clause nexuses proportion in Master and PhD sub-corpora	296

Figure 138 The inferential plots comparing the projection per clause nexuses proportion in Master and PhD sub-corpora with different alternative hypotheses.....	297
Figure 139 The boxplots of the number of paratactic projections in Master and PhD sub-corpora.....	298
Figure 140 The boxplots comparing the proportion of paratactic projection per clause nexuses in Master and PhD sub-corpora	298
Figure 141 The inferential plots comparing the paratactic projection per clause nexuses proportion in Master and PhD sub-corpora	299
Figure 142 The boxplots comparing the paratactic projection in parataxis in Master and PhD sub-corpora	300
Figure 143 The inferential plots comparing the paratactic projection per parataxis in Master and PhD sub-corpora (with different alternative hypotheses)	301
Figure 144 The boxplots comparing the proportion of paratactic projection in projections in Master and PhD sub-corpora	301
Figure 145 The inferential plots comparing the proportion of paratactic projection in projections in Master and PhD sub-corpora (with different alternative hypotheses)	302
Figure 146 The boxplots comparing the number of hypotactic projections in Master and PhD sub-corpora	303
Figure 147 The boxplots comparing the proportion of hypotactic projection in clause nexuses in Master and PhD sub-corpora.....	303
Figure 148 The inferential plots comparing the proportion of hypotactic projection in clause nexuses in Master and PhD sub-corpora (with different alternative hypotheses)	304
Figure 149 The boxplots comparing the proportion of hypotactic projection in hypotaxis in Master and PhD sub-corpora	305
Figure 150 The inferential plots comparing the proportion of hypotactic projection in hypotaxis in Master and PhD sub-corpora with diverse alternative hypotheses	306
Figure 151 The inferential plots comparing the proportion of hypotactic projection in projections in Master and PhD sub-corpora with diverse alternative hypotheses	307
Figure 152 The boxplots of the number of locutions in Master and PhD sub-corpora	308
Figure 153 The boxplots comparing the proportion of locution per projection in Master and PhD sub-corpora	308
Figure 154 The inferential plots comparing the proportion of locution in projections in Master and PhD sub-corpora	309
Figure 155 The boxplots of the number of ideas in Master and PhD sub-corpora.....	310
Figure 156 The boxplots comparing the proportion of ideas in projections in Master and PhD sub-corpora	310
Figure 157 The boxplots comparing the proportion of locution in clause nexuses in Master and PhD sub-corpora	311
Figure 158 The inferential plot comparing the proportion of locution in all clause nexuses in Master and PhD sub-corpora (with the alternative hypothesis group D < group M)	312
Figure 159 The boxplots of the number of paratactic locutions in Master and PhD sub-corpora.....	312
Figure 160 The boxplots comparing the proportion of paratactic locution in clause nexuses in Master and PhD sub-corpora.....	313
Figure 161 The inferential plots comparing the proportion of paratactic locution in clause nexuses in Master and PhD sub-corpora (with diverse alternative hypotheses)	314
Figure 162 The distribution plots comparing the paratactic locution in parataxis in Master and PhD sub-corpora	314
Figure 163 The boxplots comparing the paratactic locution in parataxis in Master and PhD sub-corpora	314
Figure 164 The inferential plots comparing the paratactic locution in parataxis in Master and PhD sub-corpora (with diverse alternative hypotheses)	315
Figure 165 The boxplots comparing the proportion of paratactic locution in paratactic projection in Master and PhD sub-corpora	316
Figure 166 The inferential plots comparing the proportion of paratactic locution in paratactic projections in Master and PhD sub-corpora.....	317
Figure 167 The boxplots comparing the proportion of paratactic locution in locutions in Master and PhD sub-corpora	317
Figure 168 The inferential plots comparing the proportion of paratactic locution in locutions in Master and PhD sub-corpora.....	318
Figure 169 The boxplots comparing the number of hypotactic locutions in Master and PhD sub-corpora	319

Figure 170 The boxplots comparing the proportion of hypotactic locution in clause nexuses in Master and PhD sub-corpora.....	319
Figure 171 The inferential plot comparing the proportion of hypotactic locution in clause nexuses in Master and PhD sub-corpora (with the alternative hypothesis group D < group M)	320
Figure 172 The boxplots comparing the proportion of hypotactic locution in hypotaxis in Master and PhD sub-corpora	321
Figure 173 The inferential plot comparing the proportion of hypotactic locution in hypotaxis in Master and PhD sub-corpora (with the alternative hypothesis group D < group M)	321
Figure 174 The boxplots comparing the proportion of hypotactic locution in hypotactic projections in Master and PhD sub-corpora	322
Figure 175 The inferential plot comparing the proportion of hypotactic locution in hypotactic projections in Master and PhD sub-corpora (with the alternative hypothesis group D < group M)	323
Figure 176 The inferential plots comparing the proportion of hypotactic locution in locutions in Master and PhD sub-corpora (with diverse alternative hypotheses).....	323
Figure 177 The boxplots comparing the proportion of ideas in clause nexuses in Master and PhD sub-corpora	335
Figure 178 The inferential plots comparing the proportion of ideas in all clause nexuses in Master and PhD sub-corpora (with diverse alternative hypotheses)	336
Figure 179 The boxplots of the number of paratactic ideas in Master and PhD sub-corpora	336
Figure 180 The inferential plots comparing the proportion of paratactic ideas in clause nexuses in Master and PhD sub-corpora (with diverse alternative hypotheses)	337
Figure 181 The inferential plots comparing the proportion of paratactic ideas in parataxis in Master and PhD sub-corpora (with diverse alternative hypotheses)	338
Figure 182 The inferential plots comparing the proportion of paratactic ideas in paratactic projections in Master and PhD sub-corpora (with diverse alternative hypotheses)	338
Figure 183 The inferential plots comparing the proportion of paratactic idea in ideas in Master and PhD sub-corpora (with diverse alternative hypotheses)	339
Figure 184 The boxplots comparing the number of hypotactic ideas in Master and PhD sub-corpora	340
Figure 185 The boxplots comparing the proportion of hypotactic ideas in clause nexuses in Master and PhD sub-corpora	340
Figure 186 The inferential plot comparing the proportion of hypotactic ideas in clause nexuses in Master and PhD sub-corpora (with diverse alternative hypotheses)	341
Figure 187 The boxplots comparing the proportion of hypotactic ideas in hypotaxis in Master and PhD sub-corpora	342
Figure 188 The inferential plots comparing the proportion of hypotactic ideas in hypotaxis in Master and PhD sub-corpora (with diverse alternative hypotheses).....	343
Figure 189 The inferential plots comparing the proportion of hypotactic ideas in hypotactic projections in Master and PhD sub-corpora (with diverse alternative hypotheses)	344
Figure 190 The inferential plots comparing the proportion of hypotactic ideas in ideas in Master and PhD sub-corpora (with diverse alternative hypotheses)	344
Figure 191 The number with corresponding frequency for taxis and logico-semantic types in the clause complex system (Master sub-corpus)	355
Figure 192 The number with corresponding frequency for taxis and logico-semantic types in the clause complex system (PhD sub-corpus)	356
Figure 193 The number with corresponding frequency for taxis and logico-semantic types in the clause complex system in Master and PhD sub-corpora	357
Figure 194 The number with corresponding frequency for taxis and logico-semantic types in the clause complex system (From Wu, 2000, p. 255).....	358
Figure 195 The range of frequencies for taxis and logico-semantic types in the clause complex system (from Master and PhD sub-corpora).....	360
Figure 196 The location of four kinds of additions with their general proportion	377
Figure 197 The location of four layers in any logico-semantic category with their general proportion	377
Figure 200 Formula 1 for measuring the clause complex complexity.....	387
Figure 201 Formula 2 for measuring the clause complex complexity.....	388
Figure 202 Formula 3 for measuring the clause complex complexity.....	388

LIST OF TABLES

Table 1 Modified CARS model for Ph.D. thesis Introduction.....	31
Table 2 The comparison of Dudley-Evans and Swales' moves in the discussion section in articles research (italics indicate difference)	32
Table 3 The location of clause complex and sentence in the stratification of ranking units (From Halliday & Matthiessen, 2014, p. 436)	112
Table 4 The properties of parataxis and hypotaxis intersected with logico-semantic relations (ibid, p. 453)	114
Table 5 The hierarchical presentation of a clause complex (with the clause ID CMAC15_05_590).....	115
Table 6 The categories of the extension relation in diverse clause connecting structures with relative markers (Halliday, 2014, p. 471)	122
Table 7 The categories of the enhancement relation in diverse clauses connecting structures with relative markers (Halliday, 2014, p. 477-8)	130
Table 8 Projecting types of the speech function, level, and mode intersection with examples and annotations	131
Table 9 Main verbs for introducing paratactic quoting clauses in the verbal process according to the speech function	132
Table 10 Main verbs for introducing hypotactic reporting clauses in the verbal process according to the speech function	135
Table 11 Main verbs for introducing hypotactic reporting clauses in the mental process according to the speech function (adapted from Halliday, 2014, p. 517).....	136
Table 12 The downloading numbers of CMAC corpus from each disciplinary section and year	152
Table 13 The proportion of disciplinary sections of CMAC corpus from the least to the most	153
Table 14 The percentage of university categories and corresponding thesis numbers of CMAC corpus	153
Table 15 The percentage of university and thesis numbers of CMAC corpus from each district	154
Table 16 The downloading numbers of CPhD corpus from each disciplinary section and year	154
Table 17 The proportion of disciplinary sections of CPhD corpus from the least to the most	155
Table 18 The number distribution of dissertations from the CPhD corpus in each University	155
Table 19 The (average) words number and their percentage in CMAC and PhD corpus by year.....	156
Table 20 The (average) words number and their percentage in CMAC and PhD corpus by section ...	156
Table 21 The detailed mixed methods research designs in this project	160
Table 22 The Independent Samples T-test (Mann-Whitney U test) for the number of words in clauses in Master and PhD sub-corpora	177
Table 23 The Bayesian Mann-Whitney U test for the number of words in clauses in Master and PhD sub-corpora	177
Table 24 The Mann-Whitney U test for comparing the clause complex per chosen sentences percentage in Master and PhD sub-corpora	179
Table 25 The Bayesian Mann-Whitney U test for comparing the clause complex per chosen sentences percentage in Master and PhD sub-corpora	179
Table 26 The Mann-Whitney U test for clause nexuses per clause complex in Master and PhD sub-corpora	181
Table 27 The Bayesian Mann-Whitney U test for the clause nexuses per clause complex in Master and PhD sub-corpora	182
Table 28 The Welch's test of the clause nexus percentage on Level 1 in Master and PhD sub-corpora	185
Table 29 The Bayesian independent t-test of the clause nexus percentage on Level 1 in Master and PhD sub-corpora	185
Table 30 The Welch's test of the clause nexus percentage on Level 2 in Master and PhD sub-corpora	186
Table 31 The Bayesian independent t-test of the clause nexus percentage on Level 2 in Master and PhD sub-corpora	186
Table 32 The Welch's test of the clause nexus percentage on Level 3 in Master and PhD sub-corpora	188
Table 33 The Welch's t-test comparing the clause nexus percentage on Level 4 in Master and PhD sub-corpora	189
Table 34 The Bayesian Mann-Whitney U Test for comparing the clause nexus percentage on Level 4 in Master and PhD sub-corpora	189
Table 35 The number of texts on each paratactic clause complex percentage stage in Master and PhD	

sub-corpora	190
Table 36 The Welch's t-test of the paratactic clause complex percentage in Master and PhD sub-corpora	191
Table 37 The general information of the chosen Master and PhD sub-corpora for the research.....	193
Table 38 The specific methods with software for diverse research activities based on corresponding theories in this research.....	194
Table 39 The Mann-Whitney U test for comparing the expansion per clause nexuses proportion in Master and PhD sub-corpora	196
Table 40 The Bayesian Mann-Whitney U test for comparing the expansion per clause nexuses proportion in Master and PhD sub-corpora	196
Table 41 The Welch's t-test for comparing the paratactic expansion per clause nexuses proportion in Master and PhD sub-corpora	198
Table 42 The Mann-Whitney U test for comparing the paratactic expansion in parataxis in Master and PhD sub-corpora (with the null hypothesis group D=group M).....	199
Table 43 The Welch's t-test for comparing the proportion of paratactic expansion in expansions in Master and PhD sub-corpora	201
Table 44 The independent samples t-test for comparing the proportion of hypotactic expansion in clause nexuses in Master and PhD sub-corpora.....	203
Table 45 The Mann-Whitney U test for comparing the proportion of hypotactic expansion in hypotaxis in Master and PhD sub-corpora	204
Table 46 The Welch's t-test for comparing the proportion of elaboration in expansion in Master and PhD sub-corpora	206
Table 47 The independent samples t-test for comparing the proportion of extension in expansion in Master and PhD sub-corpora	208
Table 48 The Mann-Whitney U test for comparing the proportion of enhancement in expansion in Master and PhD sub-corpora	210
Table 49 The Welch's t-test for comparing the proportion of elaboration in clause nexuses in Master and PhD sub-corpora.....	211
Table 50 The Mann-Whitney U test for comparing the proportion of paratactic elaboration in clause nexuses in Master and PhD sub-corpora.....	213
Table 51 The Mann-Whitney U test for comparing the paratactic elaboration in parataxis in Master and PhD sub-corpora	214
Table 52 The Mann-Whitney U test of comparing the proportion of paratactic elaboration in paratactic expansion in Master and PhD sub-corpora.....	216
Table 53 The Mann-Whitney U test for comparing the proportion of paratactic elaboration in elaborations in Master and PhD sub-corpora.....	217
Table 54 The Mann-Whitney U test for comparing the proportion of hypotactic elaboration in clause nexuses in Master and PhD sub-corpora.....	219
Table 55 The Welch's t-test for comparing the proportion of hypotactic elaboration in hypotaxis in Master and PhD sub-corpora	220
Table 56 The Welch's t-test for comparing the proportion of hypotactic elaboration in hypotactic expansion in Master and PhD sub-corpora	221
Table 57 The number of each lexis for paratactic exposition clause complexes applied in Master and PhD sub-corpora	223
Table 58 The number of each lexis for paratactic exemplification clause complexes applied in Master and PhD sub-corpora.....	226
Table 59 The number of hypotactic elaborations with markers annotated in Master and PhD sub-corpora	228
Table 60 The Welch's t-test for comparing the proportion of extension in clause nexuses in Master and PhD sub-corpora	230
Table 61 The Welch's t-test for comparing the proportion of paratactic extension in clause nexuses in Master and PhD sub-corpora	232
Table 62 The Mann-Whitney U test for comparing the paratactic extension in parataxis in Master and PhD sub-corpora	233
Table 63 The Mann-Whitney U test for comparing the proportion of paratactic extension in paratactic expansions in Master and PhD sub-corpora.....	235
Table 64 The Mann-Whitney U test for comparing the proportion of paratactic extension in extensions in Master and PhD sub-corpora	236
Table 65 The Mann-Whitney U test for comparing the proportion of hypotactic extension in clause nexuses in Master and PhD sub-corpora.....	238

Table 66 The Mann-Whitney U test for comparing the proportion of hypotactic extension in hypotaxis in Master and PhD sub-corpora	239
Table 67 The Mann-Whitney U test for comparing the proportion of hypotactic extension in hypotactic expansion in Master and PhD sub-corpora	240
Table 68 The number of each lexis for positive addition clause complexes applied in Master and PhD sub-corpora	243
Table 69 The number of each lexis for adversative addition clause complexes applied in Master and PhD sub-corpora	245
Table 70 The number of each lexis for variation: replacive clause complexes applied in Master and PhD sub-corpora	248
Table 71 The number of each lexis for variation: subtractive clause complexes applied in Master and PhD sub-corpora	250
Table 72 The Welch's t-test for comparing the proportion of enhancement in clause nexuses in Master and PhD sub-corpora.....	253
Table 73 The Mann-Whitney test for comparing the proportion of paratactic enhancement in clause nexuses in Master and PhD sub-corpora.....	254
Table 74 The Mann-Whitney U test for comparing the paratactic enhancement in parataxis in Master and PhD sub-corpora.....	256
Table 75 The Mann-Whitney U test for comparing the proportion of paratactic enhancement in paratactic expansions in Master and PhD sub-corpora.....	257
Table 76 The Mann-Whitney U test for comparing the proportion of paratactic enhancement in enhancements in Master and PhD sub-corpora	259
Table 77 The Mann-Whitney U test for comparing the proportion of hypotactic enhancement in clause nexuses in Master and PhD sub-corpora.....	260
Table 78 The Mann-Whitney U test for comparing the proportion of hypotactic enhancement in hypotaxis in Master and PhD sub-corpora.....	262
Table 79 The Welch's t-test for comparing the proportion of hypotactic enhancement in hypotactic expansion in Master and PhD sub-corpora.....	263
Table 80 The number of each lexis for temporal: same time clause complexes applied in Master and PhD sub-corpora	266
Table 81 The number of each lexis for temporal: earlier clause complexes applied in Master and PhD sub-corpora	268
Table 82 The number of each lexis for temporal: later clause complexes applied in Master and PhD sub-corpora	269
Table 83 The number of each lexis for spatial: same place clause complexes applied in Master and PhD sub-corpora	271
Table 84 The number of each lexis for manner: means clause complexes applied in Master and PhD sub-corpora	272
Table 85 The number of each lexis for manner: comparison clause complexes applied in Master and PhD sub-corpora	274
Table 86 The number of each lexis for cause: reason clause complexes applied in Master and PhD sub-corpora	276
Table 87 The number of each lexis for cause: result clause complexes applied in Master and PhD sub-corpora	280
Table 88 The number of each lexis for cause: purpose clause complexes applied in Master and PhD sub-corpora	282
Table 89 The number of each lexis for condition: positive clause complexes applied in Master and PhD sub-corpora	284
Table 90 The number of each lexis for condition: negative clause complexes applied in Master and PhD sub-corpora	288
Table 91 The number of each lexis for condition: concessive clause complexes applied in Master and PhD sub-corpora	289
Table 92 The Mann-Whitney U test for comparing the projection per clause nexuses proportion in Master and PhD sub-corpora	296
Table 93 The Bayesian Mann-Whitney U test for comparing the projection per clause nexuses proportion in Master and PhD sub-corpora	297
Table 94 The Mann-Whitney U test for comparing the paratactic projection per clause nexuses proportion in Master and PhD sub-corpora	299
Table 95 The Mann-Whitney U test for comparing the paratactic projection in parataxis in Master and PhD sub-corpora (with the null hypothesis group D < group M).....	300

Table 96 The Mann-Whitney U test for comparing the proportion of paratactic projection in projections in Master and PhD sub-corpora	302
Table 97 The independent samples t-test for comparing the proportion of hypotactic projection in clause nexuses in Master and PhD sub-corpora.....	304
Table 98 The independent samples t-tests for comparing the proportion of hypotactic projection in hypotaxis in Master and PhD sub-corpora.....	305
Table 99 The Mann-Whitney U test for comparing the proportion of hypotactic projection in projections in Master and PhD sub-corpora.....	306
Table 100 The independent samples t-tests for comparing the proportion of locution in projections in Master and PhD sub-corpora	308
Table 101 The Mann-Whitney U test for comparing the proportion of locution in clause nexuses in Master and PhD sub-corpora	311
Table 102 The Mann-Whitney U test for comparing the proportion of paratactic locution in clause nexuses in Master and PhD sub-corpora.....	313
Table 103 The Mann-Whitney U test for comparing the paratactic locution in parataxis in Master and PhD sub-corpora	315
Table 104 The Mann-Whitney U test for comparing the proportion of paratactic locution in paratactic projections in Master and PhD sub-corpora.....	316
Table 105 The Mann-Whitney U test for comparing the proportion of paratactic locution in locutions in Master and PhD sub-corpora	318
Table 106 The independent samples t-test for comparing the proportion of hypotactic locution in clause nexuses in Master and PhD sub-corpora.....	320
Table 107 The Mann-Whitney U test for comparing the proportion of hypotactic locution in hypotaxis in Master and PhD sub-corpora	321
Table 108 The Mann-Whitney U test for comparing the proportion of hypotactic locution in hypotactic projections in Master and PhD sub-corpora.....	322
Table 109 The Mann-Whitney U test for comparing the proportion of ideas in clause nexuses in Master and PhD sub-corpora.....	335
Table 110 The independent samples t-test for comparing the proportion of hypotactic idea in clause nexuses in Master and PhD sub-corpora.....	341
Table 111 The Mann-Whitney U test for comparing the proportion of hypotactic idea in hypotaxis in Master and PhD sub-corpora	342
Table 112 Summarized categories of grammatical indexes for comparing clause complexes in Master and PhD sub-corpora (according to Bayes Factor values).....	361
Table 113 The list of comparing results of clause complex indexes in Master and PhD sub-corpora (with BF values and key categories).....	363
Table 114 The list of comparing results of clause complex indexes in Master and PhD sub-corpora (with ranges and means)	365
Table 115 Main markers for annotating the elaboration relations in Master and PhD sub-corpora	367
Table 116 Main markers for annotating the extension relations in Master and PhD sub-corpora.....	368
Table 117 Main markers for annotating the enhancement relations in Master and PhD sub-corpora..	370
Table 118 Main markers for annotating the locution relations in Master and PhD sub-corpora.....	371
Table 119 Main markers for annotating the idea relations in Master and PhD sub-corpora	372
Table 131 The clause order in a clause nexus with markers in the clause complex categories.....	384
Table 132 The number of clause complexes and clause nexuses on each level (CMAC10SE_18)	388
Table 133 The clause complex complexity value by Formula 1 (Master sub-corpus)	389
Table 134 The clause complex complexity value by Formula 1 (PhD sub-corpus)	389
Table 135 The number of texts within a complexity value scope in Master and PhD sub-corpora	389

LIST OF SYMBOLS

	clause divider
	clause complex divider
^	clause conjunction
$\alpha, \beta, \gamma \dots$	hypotactic related clauses
1,2,3...	paratactic related clauses
[[]]	downranked clause
[]	downranked phrase/group
<<>>	enclosed clause
< >	enclosed phrase/group
'	idea
“	locution
.	proposition
!	proposal
+	extension
=	elaboration
x	enhancement
[] in section 4.2.1	a core mixed methods design
→	sequence that the latter follows the former
←	comes from
()	the latter is embedded within a larger intervention design
H ₁	alternative hypothesis
H ₀	original hypothesis

LIST OF ABBREVIATIONS

A	antecedent
adj.	adjective
ADS	American diplomatic speeches
AFL	academic formulas list
AI	artificial intelligence
AL	academic literacy
ALs	applied linguistics
ANOVA	analysis of variance
AP	appraisal theory
AVL	academic vocabulary list
AW	academic writing
BNC	the British national corpus
C	clause
C/S	clause per sentence
C/T	clause per T-unit
CARS	create a research space model
CC	clause complex
CDA	critical discourse analysis
ChAcE	the Chinese academic English corpus
CL	corpus linguistics
CMAC	the Chinese master's students' corpus
CMCR	characteristic matrix of clause relations
CN	complex nominal
CN/C	complex nominal per clause
CN/T	complex nominal per T-unit
CNKI	Chinese national knowledge infrastructure
CNN	cable news network
CNs	Clause nexuses
CNSs	Chinese non-native speakers
COCA	the corpus of contemporary American English
COD	correlation of discourses
COHA	the corpus of historic American English
CP	coordinate phrase
CP/C	coordinate phrase per clause
CP/T	coordinate phrase per T-unit
CPhD	the Chinese PhD corpus
CRFs	conditional random fields
CT	complex T-unit
CT/T	complex T-unit ratio
DA	discourse analysis
DC	dependent clause
DC/C	dependent clause per clause
DC/T	dependent clause per T-unit
DUJOPAS	the Dutse journal of pure and applied science
EAL	English as an additional language
EAP	English for academic purposes
EAW	English for academic writing
EFL	English as a foreign language
EGP	English for general purposes
ELT	English language teaching

EM	exact match rate
EOP	English for occupational purposes
EPP	English for professional purposes
ESL	English as a second language
ESP	English for specific purposes
EVP	English for vocational purposes
FG	functional grammar
HRAs	the humanities research articles
HSS	humanities and social sciences
hypo	hypotactic relations or hypotaxis
IC	identity construction
IELTS	the international English language testing system
IMRD	introduction-method-results-discussion method
IPVC	infinitival perception verb complements
IT	information technology
LFG	lexical functional grammar
LOB	The Lancaster-Oslo/Bergen corpus
LOCNESS	the Louvain corpus of native English essays
LR	literature review
LSR	logico-semantic relation
MA	master of art
MCMC estimates	the Markov chain Monte Carlo (MCMC) simulations
MIARN	marine incident and accident reports news
MLC	mean length of clause
MLS	mean length of sentence
MLT	mean length of T-unit
MRC	machine reading comprehension
MS	master of science
NES	native English speaking
NNES	non-native English-speaking
NNS	non-native speaker
NS	native speaker
NT clause	the naming-telling clause
para	paratactic relations or parataxis
POS	parts-of-speech
QUAL	qualitative
QUAN	quantitative
RA	research articles
RAA	research article abstracts
RC	relative clause
RRG	role and reference grammar
RST	rhetorical structure theory
S	sentence
SE Difference	the standard error for the difference between two means
SFG	systemic functional grammar
SFL	systemic functional linguistics
SIA	<i>and</i> as the initial position in sentences
SIB	<i>but</i> as the initial position in sentences
SPSS	statistic package for social science
SRAs	the sciences research articles
SSCI	social science citation index
ST	source text
T	T-unit (terminable unit)

T/S	T-unit per sentence
TC	translation corpus
TCI	technology choice index
TEM-4	test for English majors-band 4
TESL	teaching English as second language
TL	target language
TT	target text
V/T	verb phrase per T-unit
VOA	voice of America
VP	verb phrase
W	word count

Chapter 1 Introduction

The principles of choosing research objects in this thesis: “1. The systems to be counted had to be very general ones, not those of a more “delicate” kind; they should be systems that apply to a large number of instances. 2. The features to be counted must, obviously, be systemic. 3. The systems should be ones that are highly loaded semantically: that do a large amount of work in the grammar. 4. The systems should be such that we could formulate and test the hypothesis already outlined. 5. And finally, the system must be recognizable: that is, it must be such that instances of each term could be clearly identified in the corpus.” (p. 144-6)

---- Halliday, M. A. K. (1993/2005b)

1.1 Research Background

When writing English academic texts, writers typically encounter a series of challenges, including those related to vocabulary/lexical or chunk choice; grammatical, syntactical structure and sentence connection; textual structure, content organization, and discourse genre; literature review (LR) and format of reference; academic style and corresponding tone; lack of critical thinking and the inability to express meaning, not knowing how to construe meaning and interpretation; communication difficulties caused by cultural differences; the complexity of the text, etc.

In particular, within academic writing (AW), the necessity to adopt precise and professional vocabulary in conveying intricate concepts, particularly in the case of proper nouns, technical terminology, and academic vocabulary, can give rise to a sense of inadequacy among authors about their vocabulary. This can result in inaccurate or vague expressions, which ultimately confuses readers. Nevertheless, this issue can be addressed by combining reading extensively, utilizing vocabulary learning tools (such as academic vocabulary lists), and engaging in vocabulary exercises. Furthermore, consulting professional dictionaries and making use of online translation tools can facilitate comprehension and application of new vocabulary.

Some chunks, for example, fixed or semi-fixed lexical combinations such as phrases (e.g., *in conclusion*, *on the other hand*, *as a result*), collocations (e.g., *conduct research*, *make a decision*, *provide evidence*), common expressions (e.g., *the aim of this study*, *it is widely accepted that*, *according to recent studies*), and sentence patterns (e.g., *It is important to note that...*, *The results indicate that...*, *This study aims to...*), and so on, are not yet fully mastered by writers, resulting in somewhat stiff and disfluent expressions. Furthermore, these chunks are frequently associated with particular contexts, which may lead to inappropriate use or misunderstanding of meaning in writing. However, the accumulation and mastery of these chunks can also be achieved through the reading of literature in related research fields, the imitation of the writing style of excellent academic articles, the use of learning software or chunk dictionaries, and other methods in systematically developing learning and improving memory techniques to enhance language fluency and accuracy.

Problems of grammar or syntactic structures in academic writing usually stem from the transfer of mother tongue or native language of the second language writers, which can lead to confusion regarding the appropriate use of tenses, inconsistency between the subject and predicate, improper use of clauses (C), fragmented sentence structure, inconsistent grammatical choices in parallel structure, excessive use of the passive voice, and the wrong selection of articles, numerals, pronouns, conjunctions, etc. Despite implementing various strategies, including extensive grammar learning, exercise practice, teacher guidance, the utilization of grammar tools, and other measures that can improve part of the problems, the vast array of grammatical structures continues to result in a considerable error rate in specific writing categories.

Take the connection between sentences for example, as it involves a series of grammatical and lexical features such as conjunctions or phrases *first*, *second*, *next*, *finally* for time order; *because*, *therefore*, *thus*, *consequently* for causal relationships; *however*, *on the other hand*, *in contrast*, *nevertheless* for contrast; *in addition*, *furthermore*, *moreover*, *also* for complementary information; *in conclusion*, *to sum up*, *in summary* for summaries, etc.; keywords or synonyms for repetition; personal pronouns like *he*, *she*, *it*, *they*; demonstrative pronouns like *this*, *that*, *these*, *those*; relative pronouns such as *who*, *which*, *that*; transitional sentences; parallel structures in the enumeration or comparison functions, and even more, has been widely discussed in recent years. However, as the focus of the connection problem is the relationship between sentences, educators will deliberately foster students' capacity to choose connectors between sentences in teaching, and writers also pay particular attention to

this aspect due to the evaluation criteria in composing texts. Consequently, the selection of connectors between sentences would not be prioritized in this research.

The third aspect concerns the issue of structure and organization. Given the rigorous structural requirements typically imposed on academic texts, including an introduction, literature review, methodology, results, discussion, summary, and other sections can present a significant challenge to writers lacking awareness of these requirements, writers who are absent of such awareness may struggle to organize and connect these elements appropriately and effectively, and reasonably layout the text, resulting in a lack of clarity in the research logic because of structural disorder or chaos and raising questions about the reliability of the results presented by readers. However, an in-depth familiarity with the existing literature in the professional field, coupled with an understanding of its structure and organizational nuances; and the use of mind maps, outlines, templates, and other methods in writing can all assist in planning and organizing the content of the article, thereby improving the structural layout; furthermore, the reference of genre-related research papers is an invaluable aid for writers seeking to enhance their theoretical understanding, the issue of structural organization therefore, is not the primary focus of this project.

As an integral component of academic writing, references must be presented following the prescribed format. In the condition that the author is not conversant with the format, there is a high probability that the cited material will be perceived as an anomalous or erroneous reference. Consequently, the format of citations has consistently been a topic of considerable debate within academic writing. In addition to the APA, MLA, Chicago, and other citation format-related guides and books on citation formats, literature management software such as EndNote and Zotero are also frequently employed to manage and generate literature citations within the text. But more often, what bothers the author is not only the format problem but also the substance of their references, namely the content they cite to substantiate their arguments; how they cite the findings or theories of predecessors to advance their novel ideas; the choice of verbs to introduce the cited material; and the selection of verbs and related modality, tense, tone, and so forth to express their stance on the cited content, i.e., whether they support for, neutralize in or oppose to what they cited, etc. These issues will be addressed subsequently in the thesis.

Given that the intended audience of academic texts is primarily researchers in related fields and that the content is the research topic itself, the author is required to maintain a formal and objective tone in terms of writing style. This necessitates the avoidance of the first person and informal or colloquial expressions. Such issues can also be addressed through extensive reading and emulation of academic articles. However, the development of a writing style is not a process that can be completed in a short period. It is a complex and advanced concern that most novice writers will require significant guidance to navigate effectively. It is therefore not included in this study, despite its evident importance.

For academic writers, a greater challenge may be related to critical thinking because academic writing requires not only the description and summary of the research results and theories of predecessors but also the analysis and criticism of these aforementioned elements, to demonstrate their knowledge of related research, highlight the shortcomings of previous research, and provide rationales and justifications for their research projects. If the writer is overly inclined to accept previous knowledge without subjecting it to critical scrutiny and exploration, proposing novel ideas will be very challenging, relevant evidence will be difficult to utilize to support the writer's arguments, and the advancement of knowledge will not be reflected. Nevertheless, despite the efforts of many writers to cultivate their critical thinking and apply it in practice, they often remain uncertain about how to begin composing their theses. Despite a critical attitude towards previous research, some writers are reluctant to present it in their theses due to concerns about professional reputation on the one hand, and some indeed lack clarity regarding the expression of critical thinking on the other hand. As these attitudes and meanings are conveyed through lexis and grammar, this will be addressed in the present study.

The difficulties encountered by those writing in a second language are inextricably linked to the challenges posed by cultural differences. This is because the norms and styles of academic writing in different cultural contexts are not uniform. In some cultures, the direct expression of personal opinions is encouraged, whereas in some, a more implicit and euphemistic approach is favored. Similarly, some cultures value independent thinking and critical analysis, whereas some require writers to demonstrate high respect for authority. The structure of a text may also be influenced by the cultural norms of the writer as in some cultures, the text structure is linear, with each paragraph having a clear topic sentence and supporting sentences, the overall structure is, therefore, rigorous. While in some cultures, the text structure is spiral, with a high degree of paragraph freedom. In this case, the position of the topic sentence is not fixed, and the central argument is gradually presented during the analysis process. Some cultures advocate concise and clear sentences, while some prefer beautiful and complex language

expressions and gorgeous words. When authors from non-English-speaking countries are required to produce academic texts in English, it is therefore essential that they first gain an understanding of the norms and requirements belonging to written communication in English-speaking countries. This will enable them to achieve the desired level of publication. However, the concept of culture is a broad system. While the discussion of linguistic characteristics or category differences is inextricably linked to the influence of culture, it is imperative to ensure that these differences are not attributed to cultural factors. Furthermore, culture should not be regarded as the primary focus, and the discussion of the language should not detract from the linguistic category itself. Accordingly, the influence of cultural variations on academic writing is only touched upon in this thesis but is not the primary focus of discussion.

Upon closer examination of the text, it becomes evident that there is a discrepancy in the selection of specific lexical and grammatical features among writers with varying educational backgrounds. Those with higher levels of education tend to employ specific linguistic categories in an attempt to imbue their texts with a greater degree of complexity and sophistication. This leads to whether complex texts are inherently superior to simpler ones. Is this interpretation of complexity generally accepted within the academic community? How might this complexity be defined, and which features might be used to measure it? It would be interesting to ascertain whether there is a correlation between the complexity of these text features and the writer's educational level. It would also be beneficial to ascertain whether it is necessary to implement teaching measures to enhance the measurement indicators of specific features. What impact will an enhancement in the measurement criteria for particular lexicogrammatical elements have on the outcomes of other features? One might posit that the complexities of choosing linguistic features are determined by the intrinsic characteristics of the language itself. An alternative hypothesis is that the complexities are determined by the users themselves in accordance with their usage environments. The present study addresses some of the aforementioned questions, despite some of which have been addressed by other researchers. Nevertheless, some questions remain unresolved, either temporarily or permanently.

In light of the recurring issues encountered by writers in the composition of academic texts, this thesis will concentrate on clauses rather than sentence linkage issues in grammatical aspect, literature citation challenges as to structure, difficulties in construing the intended meaning during the writing process, and obstacles in measuring text complexity in the following part.

1.1.1 Lexicogrammatical: Taxis and Expansion

In traditional grammar, the relationship between clauses is typically characterized as either coordination or subordination, in which coordination refers to “a construction consisting of two or more members which are equivalent as to grammatical function and bound together at the same level of structural hierarchy by means of a linking device” (Dik, 1968, p. 25). Quirk et al. (1985, p. 918) distinguished two types of coordination, i.e., syndetic coordination and asyndetic coordination. The former exhibits an explicit conjunctive expression “with a coordinator,” thereby representing the unmarked or “more usual form” of coordination. In contrast, the latter lacks an explicit conjunctive expression, thus representing the “stylistically” marked or unusual form of coordination and being “used for dramatic intensification, or to suggest an open-ended list.” While coordination is typically conveyed through a coordinating conjunction, one type elucidates the equivalence between the meanings expressed in the clauses, while the other specifies whether those meanings also represent a cause-effect relationship, a time sequence, a contrast, or a condition, among other possibilities. In the absence of a conjunction, the equal relation of the two clauses being paralleled is evident. However, the nature of the relationship between them must be judged by the meaning of these clauses themselves and the context in which they occur. (cf. Li, 1991, p. 14)

The term *subordination* is used to describe clauses “that are embedded as a constituent of a matrix sentence” (Holler, 2008, p. 188). This involves two clauses that are not equivalent, “with one functioning as a grammatical constituent of the other,” limiting to “those constructions where one clause is embedded within another as a grammatical constituent.” (Li, 1991, p. 17) Normally in traditional grammar, it encompasses a range of clauses, including adverbial, subject, object, relative, and appositive clauses.

The relationship between coordination and subordination has traditionally been conceptualized as a continuum rather than as a dichotomy. This view is supported by a number of studies (e.g., O'Dowd, 1992; Kortmann, 1996; Johannessen, 1998; Givón, 2001; Cristofaro, 2003; Haspelmath, 2004; Verstraete, 2007), which posit that there are “a series of intermediate stages” between typical coordination and typical subordination. These include coordination with a subordination-inclined structure led by the word *and* and pseudo-coordination connected by subordinate conjunctions. The

same condition goes for the two kinds of subordination, i.e., “it is also a continuum from the typical embedding (e.g., restrictive relative clauses) to the typical dependency (e.g., adverbial clauses).” (He, 2019, p. 4)

Nevertheless, the distinction between coordination and subordination within clause connection is more accurately defined from the structural and formal perspectives. This is determined by the use of conjunctions, as Blühdorn (2008, p. 68) states, “subordinators and coordinators establish connections on sentence level.” When the conjunction in a sentence is not evident, it is challenging to ascertain whether the clause structure is that of coordination or subordination. Some conjunctions have multiple meanings when conveying meaning, and it is difficult to determine the connection relationship they represent based solely on their form when connecting clauses. The same structure, due to its inherent changes in a continuum, can realize different and more complex functions when situated in contexts, which cannot be reflected by traditional structural division, however. In the field of systemic functional grammar, the relationship between clauses is divided into two categories: parataxis and hypotaxis. Additionally, the functional perspective is subdivided into two subcategories: expansion and projection. A comparison between the two reveals that, in traditional grammar, coordination corresponds to paratactic extension and paratactic enhancement in systemic functional grammar. In contrast, subordination or clause combining is the antithesis of coordination, corresponding instead to hypotactic extension and hypotactic enhancement. Apposition is seen to be paralleled with paratactic elaboration, while the non-restrictive relative clause (RC) is regarded as a resemblance of hypotactic elaboration in SFG. Quoting is associated with paratactic locution and paratactic idea, whereas reported speech, which is adversative in linguistic features, is pertinent to hypotactic locution and hypotactic idea. The initial three categories are situated at the intersection of taxis (parataxis/hypotaxis) and expansion, while the subsequent two are located at the intersection of taxis and projection.

Consequently, an examination of the choice of coordination/subordination in the text reveals that apposition, non-restrictive relative clauses, quoting, and reported speech are not considered. Furthermore, a considerable proportion of the latter also demonstrates the connection relationship between clauses. Concurrently, the structures that correspond to extension and enhancement in coordination and subordination are not clearly distinguished in terms of function. Furthermore, subordination encompasses a substantial proportion of structures, including noun clauses, subject clauses, and adjectival clauses, which exist as a constituent of another clause rather than forming a relationship with another clause in sentences. Such structures are regarded as embedded clauses in SFG. This indicates that a clause should undergo “a double transitivity analysis” whereby one of the two clauses is a constituent of a grammatical constituent in another clause rather than “a double syntactic analysis” whereby “one of the two clauses is a constituent of another clause.” (He, 2019, p. 4) In traditional grammar, “adverbial clauses, subject clauses, and complement clauses are all subject to double syntactic and double transitivity analyses” (ibid). Consequently, they can all be treated as either clause complexes (CCs) or simple clauses with embedded ones. The investigation of subordination necessarily encompasses that of embedded structures. In SFG, however, hypotaxis is treated as a distinct area of inquiry, separate from that of embedded structures. It is, therefore, imperative that the former be the primary focus of research into the investigation of the relationship between clauses.

Similarly, the linguistic features that realize the logico-semantic relationship of expansion appear to overlap with conjunction and preposition. On the one hand, the distinction between lexis and grammar renders the latter unsuitable for text analysis, particularly when the same word or phrase connects disparate structures in diverse contexts to realize corresponding functions. This is exemplified by *and*, *while*, *if*, *as*, and so forth, as it is challenging to ascertain their role in connecting clauses solely based on their form. On the other hand, several distinguished words exhibit disparate structural forms when connecting clauses, yet they serve the same function. To illustrate, when the preposition *without* connects hypotactic non-finite clauses and the conjunction *but* connects paratactic clauses, it can perform enhancement: condition: concessive function, as well as addition: adversative function. Both require an examination of the relationship between the clauses in the text from the perspective of meaning realization rather than a mere analysis of the structural form. Furthermore, as previously stated, the absence of a conjunction between clauses renders it challenging to ascertain the nature of the relationship expressed between clauses through conventional classification methods. In conclusion, by considering the applicability of the structural distinction of taxis and the functional distinction of expansion/projection, this study will examine the connection relationship between clauses and the reasons and probabilities of the choice of linking words or phrases when realizing the relationship from the perspective of systemic functional grammar, rather than the traditional grammatical category classification method.

1.1.2 Structural and Interpersonal: Projection and Projecting Verbs

It is standard practice in academic writing to include citations of relevant literature. Such references may be found in the sections of academic papers that include background introduction, literature review, theoretical presentation, method clarification, discussion, conclusion, and so on, especially in the literature review and theoretical sections. Writers will typically cite the views or results of previous researchers to support their arguments and provide compelling evidence for them. They will also demonstrate their understanding of related research, the research background, and context and then put forward new viewpoints or findings. Furthermore, they can justify their arguments and present the novelty of their position (Gilbert, 1977; Dubois, 1988). They can also critically analyze and integrate diverse sources of information to develop new ideas and prepare for the breadth and depth of their research, taking into account the research conducted by predecessors. Additionally, they can provide readers with a more comprehensive understanding of relevant research, enhance the transparency and credibility of their research to avoid plagiarism, “display an allegiance to a particular community or orientation, create a rhetorical gap for his or her research, and establish a credible writer ethos” by “acknowledging a debt of precedent” (Hyland, 2000, p. 20), situate themselves within the academic community, engage with the research outputs of others with either confirmation or criticism, contribute to the development of subsequent discussions, promote the advancement of related research, and so forth.

Previous discussions on citation can be formally divided into direct citation (which involves the verbatim quotation of a paragraph or sentence from the original text and its integration into one’s own text, marked with double quotation marks to enhance the credibility and authority of the argument) and indirect citation (which involves the integration of the quoted content into other language structures, whereby the content of the original text is repeated in the writer’s words while maintaining the main meaning of the original text without using the specific wording of the original text; being employed to simplify the content, integrate information, and maintain the coherence of the thesis) or “integral citations (are those where the name of the cited author occurs in the citing reference) and non-integral citation (make reference to the author in parentheses or by superscript numbers)” (Swales, 1990, p. 148), as what Weissberg & Buker (1990, p. 386) distinguish between them: “the former is placed within the sentence and plays an explicit role within the syntax of the sentence, while the latter is typically placed outside the sentence in brackets), and postulate that integral citations put focus on the cited author(s) while non-integral citations make the information contained in the sentence prominent”; contently related to the manner in which source material is employed in the writer’s argument (Dubois, 1988; Thompson, 1996b) such as choices largely concerning “the extent to which the report duplicates the original language event, the options for using short direct quotes (up to six or eight words), extensive use of original wording set as indented blocks, the summary from a single source, or a generalization, where the material is ascribed to two or more authors” (Hyland, 2000, p. 25 - 26), etc.; motives for citing, such as highlighting important concepts, narrowing the gap with the original authors, reviewing and evaluating literature, (Case & Higgins, 2000) citing theories and demonstrating methods, constructing interpersonal networks within the academic community, (Vinkler, 1987) persuading readers, and questioning other scholars, (Brooks, 1986) etc.; the rhetorical functions like “attribution, exemplification, further reference, statement of use, application, evaluation, establishing links between sources, and comparison of one’s own work with that of other authors (Hyland & Shaw, 2016, p. 386), etc.

It should be noted, however, that the term *citation* encompasses not only different forms of projection between clauses but also those between words/phrases and clauses, the integration of words or phrases or clauses into sentences, the citation of entire sentences or paragraphs in texts, and even that of examples or formulas from others research. Because the combination of clauses in terms of form and function can be classified as paratactic or hypotactic quoting and reported speech, this study has focused on projection as the research object rather than citation. In the process of citing or projecting content or information, the selection of the appropriate verb is of great importance. Despite the existence of numerous classification systems for citation verbs (e.g., Bloch, 2010; Francis et al., 1996; Hyland, 1999; Manan & Noor, 2014; Thomas & Hawes, 1994; Thompson, 2002; Thompson & Ye, 1991; Yang, 2013), among which Thompson & Ye (1991) is widely accepted as they categorize the processes performed by verbs in text into research acts, mental acts and textual acts (verbs performing textual acts are similar to locutionary verbs in SFG, and those performing mental acts are parallel to mental verbs), the majority of the classifications from the perspective of interpersonal and rhetorical functions conflate form with function of related citation verbs. The citation structures involved extend beyond the relationship between clauses; they also encompass the mixed relationship between phrases, words, and phrases and clauses. Consequently, the absence of a clear distinction between citation verbs and reporting verbs results in a discrepancy between the conventional understanding of reporting verbs and those being

interpreted within the framework of systemic functional grammar. In addition to focusing on reporting verbs, this study also discusses the choice and meaning of quoting verbs and the realization of projecting verbs on the projected content and its impact on the realization of the writer's interpersonal function.

1.1.3 Meaning Realization: The Implicit Language

The expression and realization of meaning entail the utilization of tacit knowledge (Polanyi, 1958), which is inherently challenging to delineate and convey through linguistic or written means. Its acquisition is contingent upon prolonged individual practice and the accumulation of experience, relying on subjective sentiments, intuition, and insight. Such knowledge is frequently context-specific, making it challenging to comprehend and apply it generally without recourse to the specific context. As tacit knowledge is inherently difficult to express clearly and concisely, its transmission typically necessitates observation, imitation, participation, and actual operation. Furthermore, tacit knowledge is highly individualized, with each person's tacit knowledge reflecting their unique background, experience, and mode of understanding. Though it is often present in the minds of individuals, it is not a static entity as it will continue to develop and evolve due to the accumulation of personal experience and the influence of external factors. Consequently, the influence of implicit knowledge renders the expression of certain meanings challenging in textual form. This is exemplified by the aforementioned comprehension of the interconnectivity between clauses: the ability to discern the optimal connecting method for conveying the intended meaning, the suitability of a paratactic or hypotactic structural relationship between clauses, and the capacity to express the meaning scope of such connections through a single word in facing of kinds of lexical choices, etc., all represent significant challenges. Similarly, when projecting the views, methods, and results of other scholars, which content should be selected for projection, whether to present them directly or indirectly, whether to present the other party's locutions or ideas, whether to transform the other people's views into propositions or proposals for projection, how to choose the projecting verbs and their related tenses/modals/voices/tones to express the writer's attitude towards the presented content, and what attitude should be held towards the projected information, etc. all derived from personal experience, requiring the writer to undergo and encapsulate the experience through extensive writing practice. However, when the subject, content, or genre of the writing tasks changes, some knowledge that has been acquired may become inapplicable and require modification and adaptation.

The difficulty in solving the problems mentioned above is not solely attributable to their tacit nature as being the knowledge itself. From the perspective of their being lexis and grammar, some expressions are absent from the process of conveying meaning. This can be described as *implicit meaning* in the former case while *implicit grammar* in the latter.

Whorf once put forth the concept of *cryptotype* or *covert categories* of a language in his paper *Grammatical Categories* in 1937 (Lee, 1996, p. 160), which describes "semantic or syntactic features" that lack "morphological implementation" but are indispensable for the organization and interpretation of a phrase. "A covert category may also be termed a cryptotype, a name which calls attention to the rather hidden, cryptic nature of such word-groups, especially when they are not strongly contrasted in idea, nor marked by frequently occurring reactances such as pronouns." (Whorf, 1937, p. 4) In contrast to "the phenotype or overt category," which are categories that are "overtly marked as such," the cryptotype is not readily apparent. (Lee, 1996, p. 168) Just as Halliday (1999, p. 569) described: "Whorf (1956) distinguished between overt and covert categories and pointed out that covert categories were often also 'cryptotypes' - categories whose meanings were complex and difficult to access. Many aspects of clause grammar, and of the grammar of clause complexes, are essentially cryptotypic." This concept is predicated on the assumption that all grammatical categories must be marked in some way within the language to contribute to the overall meaning. However, Whorf drew attention to the fact that not all linguistic categories are overtly marked. Some categories are only overtly marked in specific instances, while they are covertly marked in the majority of or even all other cases. Whorf himself employed the term *cryptotype* to differentiate it from the *covert category*, characterizing it as "a special, highly concealed subdivision of covertness, amounting sometimes to a second degree of covertness." (see Lee, 1996, p. 164 - 171) The distinction between "phenotypes and cryptotypes" indicates that "it is generally easier for a person interested in becoming more metalinguistically aware to become conscious of phenotypes rather than cryptotypes" whereas "cryptotypes may be difficult for the linguist to define because the semantic ramifications of items in the category may be so diffuse that choosing a word or phrase to characterize what is communicated when they manifest in speaking, writing, or thinking is problematic." (ibid, p. 173) But this does not mean that cryptotypes are impossible to be aware of.

In addition to the tacit knowledge involved in the realization of meaning, this study primarily focuses on certain meanings that can be expressed but are obscured in language due to the constraints of lexis and grammar (personally or linguistic in nature) rather than being explicitly conveyed. Accordingly, the former is designated as *implicit language*, and the latter is classified as *explicit language*. A detailed examination of the specific content will be undertaken in Sections 7.2 and 7.3. The distinction between explicit and implicit language allows for the interpretation of meanings that would otherwise be hidden in the grammar. This enables language users to make more informed choices regarding the lexis and grammar they apply in realizing meaning in the future.

1.1.4 Measuring: The Complexity of Clause Complexes

The conventional linguistic approach to assessing the intricacy of sentence construction in text is the concept of *syntactic complexity* (also designated as *syntactic maturity* or *linguistic complexity*). This concept relates to the spectrum of linguistic forms that emerge during language production and the extent of complexity exhibited by such forms. In light of the assumption that language development encompasses, among other processes, the expansion of “L2 learners’ syntactic repertoire” and their capacity to utilize this repertoire effectively in a range of contexts, the construct of “syntactic complexity” is a pivotal element in second language research. (cf. Ortega, 2003, p. 492)

The focus is typically on the diversity and complexity of the production units or grammatical structures in the investigation (Foster & Skehan, 1996; Ortega, 2003; Wolfe-Quintero et al., 1998). Thus, a multitude of disparate global measures have been put forth to characterize it, to quantify one or more of the following in diverse ways: length of production units such as clauses, sentences, and T-units, etc., the extent of embedding or subordination of clauses, degree of coordination, “range of surface syntactic structures, and degree of sophistication of particular syntactic structures (Ortega, 2003).” (Lu, 2010, p. 474-5)

In the field of syntactic complexity measurement, a *clause* is defined as a structure comprising a subject and a finite verb (Hunt, 1965; Polio, 1997). This encompasses a range of clause types, including independent clauses, adjective clauses, adverbial clauses, and nominal clauses. In contrast, a *dependent clause* is defined as a finite adjectival, adverbial, or nominal clause (Cooper, 1976; Kameen, 1979). The key concept *T-unit* (terminable unit) is defined as “one main clause plus any subordinate clause or nonclausal structure that is attached to or embedded in it” (Hunt, 1970, p. 4), and a *complex T-unit* is the one that contains a dependent clause (Casa-nave, 1994). (cf. Lu, 2010, p. 481-2)

While syntactic complexity measures have been employed in L2 writing research for a variety of purposes, including evaluating the impact of pedagogical interventions on the acquisition of grammar, writing proficiency, or both; “investigating task-related variation in L2 writing”; and assessing differences in L2 texts produced by learners at various proficiency levels and over time (see Polio, 2001), etc., (cf. Ortega, 2003, p. 492-3) the concept of *syntactic complexity* is inherently, fundamentally defective. On the one hand, the concept of *T-unit* has been the subject of considerable debate, with differing connotations and denotations (will not be discussed here in detail), which has led to the development of numerous studies and algorithms based on it is questionable as a viable approach. On the other hand, traditional grammatical divisions have shaped the focus of research, with an emphasis on structural divisions within and around a sentence and between sentences. Consequently, the complexity of the concept is largely attributed to its structural aspects, with other functions, including the discourse function, receiving less attention. Furthermore, the separation of grammar and lexical semantics has the effect of generalizing the realization of originally distinct meanings while simultaneously obscuring a considerable amount of lexical semantic information. In addition, general classification discussions and quantitative statistics are unable to demonstrate the sentence or clause structure created by users for different purposes. This is because a single linguistic device in the same structural position may result in entirely distinct interpersonal effects due to the varying purposes of the speaker. It can, therefore, be seen that the relationship between clauses is not simply dependent or independent, nor is the traditional division of clauses, such as adverbial, predicative, complement, object, or attributive (which examine the relationship between lexis and clauses) that is relevant here. It is also the case that there is no measurement of its complexity in the study of syntactic complexity!

In addition to analyzing the relationships between clause components or clauses, the complexity of the clause should also take into account the recursion of clauses, as proposed by Martin (1988) regarding the interconnectivity of clauses: a “synoptic” approach that treats lexicogrammatical constructions as “planned” and a “dynamic” approach that focuses on their “linear emergence and development.” The synoptic view is appropriate for the analysis of “experiential patterns,” such as those observed in the simplex clause, the dynamic view is, however, more suitable when considering “logical recursive structures” whereby “the construction can be expanded continuously in the running text.” From the

perspective of systemic functional grammar, *complexity* is a property inherent to “organized entities,” whether at the level of “organisms or systems.” (Givón, 2009, p. 3) “Individual entities” that lack “internal organization” are defined as maximally “simple.” In an organized system, the simple entities may be related to “the system as a whole, to its sub-parts, or to each other.” “At the most abstract level, a system may be described as a network of nodes and connections, where the nodes stand for either the simplest entities or more abstract, higher-level sub-parts of the system, and the connections stand for the nodes’ relations within the system.” (ibid) This is what Simon’s (1962) “seminal work on the architecture of complexity” has revealed: “increased complexity is, at the most general level, ... an increase in the number of hierarchic levels within a system.” (Givón, 2009, p. 4) But the recursion of clauses is different from *recursivity*, which is defined as “embedding a clause inside another clause” by Chomsky (1957, 1965). Therefore, the study of the combination of clauses in this thesis is what Givón (2009, p. 4) described as concatenated (conjoined, chained) clauses. Concerning the methodology employed in the development or organization of the system structure, this study adopts a kind of “expansion analysis,” i.e., “start from an undifferentiated holistic signal and then reanalyze it as a composite of morphemes (a word), or of words (a clause)” instead of the “combination (synthesis) analysis” by “starting from the small units (morphemes, words) and then building up the word or clause combinatorially.” (ibid, p. 8)

After a thorough examination of the functional division of the system of complex clauses and the choices of its sub-categories in the text, i.e., “the overall probabilities of linguistic systems (in a simultaneous system, the selection of one feature is independent of another, i.e., the overall probabilities are unaffected)” and “the conditional probabilities (the selection of one feature is dependent on the other, and the first constitutes the environment for the second)” (Wu, 2000, p. 261), this study will propose a concept for calculating the complexity of clause complexes based on the system probabilities. More details will be presented in the discussion Section 7.5.

1.1.5 Summary

In conclusion, this study is concerned with the examination of clause complexes, with a particular focus on the subcategories of expansion and projection, the taxis of clause connections, the expression of meaning in the text, the choice of lexicogrammatical categories, implicit clause complexes, and the complexity of clause complexes within the text, etc. The texts in the study come from a self-constructed corpus because “corpora can inform understanding of academic discourse, from lexical, phraseological, grammatical, and genre perspectives.” (Hyland & Shaw, 2016, p. 206) The information on the aforementioned language features can certainly be extracted from the self-built corpus, and the extensive quantity of text in the corpus can provide a viable foundation for the research. The genre of the text for research is English academic writing, with their writers being students of the English major in Chinese universities and their native language being Chinese. On the one hand, the researcher has the same identity and thus possesses a natural advantage in collecting relevant corpora, which can ensure the authenticity and accuracy of the corpus and provide a solid foundation for the feasibility and reliability of this research. On the other, as the texts of English majors in university students’ writing are predominantly composed in English, the challenge of devising and structuring writing tasks is circumvented. Furthermore, given that students in Chinese universities have generally been studying English for over a decade, their texts possess intrinsic value as reference materials and can also illuminate the challenges associated with learning English as a second language. The text for examination is a thesis or a dissertation for getting a degree, which is distinguished from other forms of university student writing by its length, writing difficulties, and the depth and breadth of knowledge and language reflected. To ascertain whether the choice of grammar and lexis improves with an elevated level of education, the graduation theses of master’s and doctoral students are selected for comparative analysis. Given that the majority of EAP (English for academic purposes) corpus-related activity incorporates “a comparative element,” for instance, the differentiation of “texts belonging to disparate disciplinary domains (Hyland, 2008) or cultural contexts (Gardezi & Nesi, 2009),” the observation of advancements “in texts produced across stages of study (Issitt, 2011),” and so forth, (Hyland & Shaw, 2016, p. 206) “Hyland (2004a), for example, outlines a methodology whereby sample texts are manually coded to “identify metadiscursive items,” which are then searched for in a corpus of master’s theses and doctoral dissertations. This research does not seek to contrast findings with those from Hyland’s research described above. Instead, it adopts a text complexity perspective and focuses on graduate theses produced by students at the master’s and doctoral levels. Although the theses are not from the same group of students, i.e., a comparative study of a group of students’ master’s theses and doctoral dissertations (as master’s theses from the same group of doctoral students are not necessarily searchable on the Internet) is conducted, by increasing the number of texts at each

educational stage, this research endeavors to avoid the bias of research results caused by differences in individual selection tendencies. The particular research questions will be enumerated in Section 1.3.

1.2 Purpose of the Study

This study firstly intends to investigate how clauses are organized in a text by students, especially those whose mother tongue is not English and who endeavor to learn it in conveying ideas, apply it in academic activities, transmit knowledge through it, and often suffer from it because of insufficient preparation for or consciousness about it. By demonstrating and explaining how clauses are connected into a clause complex, students can understand how knowledge of clause complexes enables them to comprehend structures and threads within textual media more effectively. This is because clause complexes have the same importance as coherence in texts, with the former concerning the connection of clauses and the latter concerning the connection and intersection of clause complexes. Suppose students can identify, isolate, and even comprehend them. In that case, they will undoubtedly master them, thereby experiencing fewer or no issues or difficulties in conveying their meanings through colloquial or textual expressions. Once the obstacles to reading and writing clauses and clause complexes are removed, the psychological obstacles that impede their language abilities can also be deleted, eliminating obstacles to their advancement in the English language.

The second objective is to propose the hypothesis that doctoral candidates in English studies in China would utilize a greater number and more sophisticated clause complexes than master's students, whose differing educational levels entail diverse understandings and applications of grammatical structures, such as clause complexes. If this hypothesis can be substantiated, it would be prudent to consider whether Master students should prioritize enhancing their abilities and consciousness in expressing thought. Additionally, it would be valuable to ascertain whether students should devote attention to clause complexes. It would be advantageous to examine whether students' capacity to articulate themselves would be augmented by their acquisition of clause-complex structures. To prove this hypothesis, a series of research designs, text collections, corpus building, data selection, theoretical framework confirmation, corpus annotation, statistical index construction, data calculation and comparison, result analysis, and a list of explanations must be developed. These will enable us a thorough examination of the differences between the two educational levels in question. On this basis, can students be advised to bring in a new perspective for grammar learning and application; teachers be suggested to see sentences in a unique angle in teaching and researching; translators be enlightened in finding out structural, cognitive ontological diversities of these two languages in C-E/E-C translating; and can intelligent machines be exercised of recognizing, construing, comprehending and creating clause complexes in machine translation.

On account of the similarity and comparability of concepts, *clause complex* in Systemic Functional Grammar and *Fuji* in traditional Chinese grammar, the denotation and connotation of these two concepts will be discussed in this research in the theoretical part to inquire how much different they are. If similarity outweighs diversities, the question of whether SFG's perspective of seeing grammatical structures is more beneficial for Chinese students in understanding and learning English just like Chinese compared with other traditional grammar is then worthy of discussion. The comparison would depend on grammar classifications of clauses connection relationships in Mandarin sentences and English clause complexes that the former includes six categories with 42 subcategories in Lv Shuxiang's *Essentials of Chinese Grammar* (《中国文法要略》) (which emphasizes more on traditional Chinese language in addition to the modern one and not all subcategories are divided according to clauses divisions); the latter mentions categories and relationships from *A Comprehensive Grammar of the English Language* (Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J., 1985) and then Halliday's system of clauses relations. The goal of this lies in the advocacy of learning English grammar (for Chinese students) from the perspective of Halliday's Systemic Functional Grammar, which might be more suitable, at least from the aspect of clause relations, for English learners with Mandarin being their mother tongue because of the history and development of SFG, this ontological similarity, distinct and intelligible analyzing framework, delicate ranks with the continuum of the cline of the instantiation, and also its applicability especially for discourse analysis.

Another aim of this research is to distinguish expansion from conjunction and projection from citation on their research spheres, locations in corresponding theoretical framework, annotation, and denotation, emphasis on problem-solving, applicable prospect, etc., in the theoretical part that though these two groups of concepts have intersection on their research objects (grammatical structures such as connecting words/phrases, citation verbs, clauses connection with unequal relations, meaning addition, etc.) and the former two (expansion & projection) are not acquainted by scholars outside of the SFG field, the misunderstanding or mixture of these concepts indeed bring confusion and questions for not only

researchers but also teachers and students. It is therefore imperative to differentiate between the concepts in order to establish a foundation for further research. Conjunction and citation are more akin to structural, formal, or surface analysis in that they do not prioritize the meanings that writers intend to convey or the functions that clauses perform in the process of communication or during social activities; nor do they focus on the logico-semantic relationships within clauses, their status and even recursion of their relations. The conjunction can also be the connector of two sentences in traditional grammar, which is not the sphere of clause complexes that only focuses on relations within two clauses or embedded clauses but on the coherence in texture or discourse analysis. In addition, citation covers non-integral citations (Source, Identification, Reference, etc.), integral citations (naming), and non-citation, according to Thompson & Tribble (2001), paying close attention to information originality and confirming the effectiveness and credibility of authors' references, which is beyond the scope of projection.

Besides, the research also inclines to deepen the theoretical foundation by following conventional definitions and the classification of conceptions in SFG categories; to develop an analyzing framework that is more suitable for academic discourse or thesis writing specifically based on the original SFG clause complex system; to quantify categories in the framework by integrating data from researching results; to put forward some new concepts on this basis and then raise some new and related research questions; and also to attract some attention from other scholars in studying corpus linguistics, academic writing, genre analysis, etc., and even computational linguistics. SFG is proved in this project to be suitable for not only popular discourse analysis but also genres like academic discourse, especially theses and dissertation writing; for not only text studies but also clauses and logico-semantic relations; for not only comparing studies within Chinese and Foreign languages but also level comparison within mono-language; for not only synchronic but also diachronic research. This research cannot be able to expand the applicable fields of SFG but at least expand that of clause complexes in a sense. It would be proved from the literature review section that most researchers apply SFG in their theses qualitatively, leaving only a few to be quantitative, and they are short of enough data to support their viewpoints. Consequently, the research will circumvent this deficiency in identifying regularities by employing a substantial corpus of data and subsequently conducting an objective qualitative analysis of this database. The process mode of the data, i.e., the application of annotation and calculation software, will enable the expansion of research methods for clause complex, as will be shown in the following sections. In general, a broad research range and more possibilities of SFG implication are to be confirmed by this conduction, with an additional intention to design and construct a quantified model of clause complex in text analysis for further comparisons and discussions.

In the soft fields, convincing readers that an argument is both novel and sound often depends on the use of reporting structures not only to build a shared theoretical basis for one's arguments but also to establish a common perspective on the reliability of the claims one reports. The application of a greater number of projections by doctoral candidates in constructing their arguments is indicative of a desire to present a reasonable and powerful case. This is accompanied by an eagerness to establish relationships within the academic community and a wish to gain knowledge or viewpoints confirmed by scholars in the world. If more locution structure appears in Ph.D. texts, then the hypothesis that Master students may use more locution than PhD candidates is at least not so convincing, especially when the locution is regarded as a less complex structure than the idea for students to organize. Under this condition, more locutions should be taken as the better way for objectively transmitting ideas or sayings of original authors as any transformation or paraphrase can change the primary meanings of these utterances and result in writers' or students' extending or minimizing viewpoints and explanations. Hence, except for verifying the statement that doctoral candidates outnumber Master's students in choosing locutions obtained from a small corpus test, whether students can logically select projecting verbs in thesis writing or not and whether these relation markers can objectively present authors' thoughts, writers' understanding of those opinions and their intentions in projecting and meaning transforming or not, will also be discussed in this dissertation.

1.3 Research Questions

Purpose-matched research questions of this project come from five main parts: theoretical, methodological, practical, analytical, and applicable, in which the practical aspect holds the core position, and the former two entail problems that need to be answered and then served for the practical conduction.

In detail, the dissertation intends to answer two questions in theoretical:

RQ1T1. How can the clause complex be better applied in text analysis in specific genres, such as academic thesis writing?

And RQ1T2. How different are research scopes of the clause complex when compared with concepts such as *conjunction*, *coherence*, *citation*, etc., in traditional grammar (sense)?

Methodological questions will include:

RQ2M1. How many texts are enough for the corpus comparison analysis, especially for the annotation of clause complexes, as an oversized corpus may consume manual energy?

RQ2M2. Can clause complex categories be quantified? If yes, then how (in what ways can they be calculated in texts, and how can they be compared)?

And RQ2M3. What kinds of research methods software can be used for promoting clause complex analysis?

In practice, the dissertation mainly attempts to solve the problems below:

RQ3P1. How are logico-semantic relations realized by the lexicogrammatical continuum in detailed texts?

RQ3P2. Are chosen texts of Master theses and PhD dissertations different or the same on some basic conditions? Like:

the number of words in each sentence (clause complexes) in each text on different educational levels;

the proportion of clause complexes/sentences and clause simplexes/sentences in each text on the same educational level;

the proportion of clause nexuses (CNs) per clause complexes in each text on different educational levels;

the proportion of clause nexuses on each recursion level in each text on different educational levels;

the proportion of hypotaxis and parataxis per clause nexuses in each text on different educational levels; etc.

RQ3P3. To be more specific, do Master theses and PhD dissertations have variance in each grammatical category, and where do these diversities reflect? Statistical tallies would be the number of instances and the proportion of categories in each text within, around, and outside of the clause complex system: expansion, hypotactic expansion, paratactic expansion, addition, hypotactic addition, paratactic addition, extension, hypotactic extension, paratactic extension, enhancement, hypotactic enhancement, paratactic enhancement, projection, hypotactic projection, paratactic projection, locution, hypotactic locution, paratactic locution, idea, hypotactic idea, paratactic idea, etc. Can comparison results of these categories prove that PhD dissertations are more complex than MA (Master of Arts) theses in choosing logico-semantic relations?

RQ3P4. Do these theses have a preference in choosing words/phrases that mark logico-semantic relations in clause complexes? What meanings do these markers convey, and what functions do they perform in academic discourse?

And RQ3P5. In considering the recursion of clause complexes, are there unique patterns in thesis writing, and how can differences or similarities be contrasted between educational levels?

In analyzing the research results obtained above, a list of questions still deserved to be answered:

RQ4N1. How do we explain research results from the corpus? And whether the chosen texts are defective or not.

RQ4N2. Do these results meet expectations before we conduct the research? If the answer is not, then where does deviation come from?

Being applicable,

RQ5A1. Is it possible to construct mathematical models for text analysis of the clause complexes' complexity?

RQ5A2. How can outcomes of this research be useful for fields like academic writing, classroom teaching (on clause relations), translation (on essential cultural comparison), and even computational linguistics in automatically analyzing meaning-based logico-semantic relations?

And RQ5A3. How can research methods be improved for future clause complex studies?

All these questions will be answered interwoven through the dissertation.

1.4 Significance of the Study

This project conducted quantitative and qualitative research on the concept of clause complex in systemic functional linguistics based on the comparison of a large number of academic texts, hence undoubtedly has theoretical and practical significance research:

Theoretically, a series of queries are raised firstly by examining how tacit knowledge is expressed outside of the clause complex system: at the semantic level, the connotation and denotation of words can be expanded, but at the grammatical level, especially when it comes to logical relationships within clauses, some implicit relationships do not have corresponding connecting words, while some words can even express a variety of logical relationships, hence bring troubles to students' understanding and expression or communication of the implicit knowledge. Through an in-depth discussion of logical relations, this study presents the transmission of tacit knowledge at the grammatical level, pointing out the methods of digging out deeper functions of some specific grammatical-lexical structures in the face

of numerous logical relations and fundamentally embodying the importance of meaning in construing relationships.

Another question concerns whether syntactic complexity can indeed calculate and compare various aspects of the complexity of clauses. Given the existing syntactic complexity theories, regardless of the internal structure of the clause, scholars usually examine the length of the sentence, the usage of certain structures within the sentence, T-unit-related conditions, etc. But the term “T-unit” has not been clearly defined, the calculation of complexity is limited to the frequency of use, and the calculating indexes are simply divided based on the structure, the complexity of texts hence cannot be comprehensively or deeply measured by these syntactic complexity measuring indexes. Having noticed these insufficiencies, the clause complexity measurement of texts from the perspective of clause complex in SFG is explored in this research instead. Considering that the division of categories of the clause complex is based on functions that language plays in meaning expression, the attempt can be called an innovation.

Inspecting from the systemic functional linguistics theory internally, three principal approaches to the analysis of text model are roughly taken: 1. exploring the patterns (nets and bonds) of lexical organization in a text from bottom to up (Sinclair, 1970, 1991, 2004; Hoey, 1991, 2001, 2005); 2. hierarchically displaying the text structure from top to down or a procedure called Rhetorical Structure Theory (Mann & Thompson, 1988, 1992; Matthiessen & Bateman, 1992; Webster, 2002, 2014); and 3. examining text composition that lies beyond the clause (Halliday, 1978a; cf. Hoey, 1983; Hunston & Francis, 2000; Hunston & Thompson, 2000), i.e., a lexicogrammatical paradigm or the known Cohesion Theory (Halliday & Hasan, 1976, 1985), in which the third one was developed by Hasan (1984) and Martin (1992a) as a dynamic model of cohesion analysis that describes aspects of “text as process” (studies in Lemke, 1991; Hoey, 1991; Parsons, 1991; Martin, 2001; Fries, 2002; a collection on Theme in Ghadessy, 1995; Forey & Thompson, 2008; Berry, 2013a & b; Matthiessen, 2013; Thompson, 2013; Hasan, 2015 & in press; cf. Stoddard, 1991; Goutsos, 1997; Esser, 2009), the first perspective also treats “texts as process” and considers the role of time (see also Gregory, 2002 on Phasal Model of text construction; Cloran, Stuart-Smith & Young, 2007 for introducing Phrasal Model of text; Cloran, 2010 for rhetorical unit analysis of text from the perspective of historical, biographical and social time-space configuration) while the second focuses more on space and analyses the “text more as a product” that scholars in this direction treat texts statically (see also Cloran, Stuart-Smith & Young, 2007 for introducing Rhetorical Unit Model). (cf. Peng, 2016) By consulting the text analysis paradigm and seeing the text as a product, this research inspects the selection of the internal structure of the clause complex system from top to down, summarizes the clause complex recursion paradigm from bottom to up, and statically explores the mutual effects of grammatical-lexical choices and the meaning transmitting of clause complexes, is thus of great theoretical value.

There have been few studies on clause complexes, whether in China or abroad, in the past, let alone its integration with corpus linguistics and academic writing theory. After teasing out research on clause complex from its combination with academic writing or corpus linguistics, the intersection of academic writing and corpus linguistics directions, a multiple connection of research aspects on the clause complex, corpus linguistics and academic discourse writing, and other related works published both in domestic and abroad in the past thirty years, it was discovered that most of these studies would adopt qualitative rather than quantitative methods; conduct discourse analysis instead of theoretical exploration; application of theories are easier for publishment while ontological research is hard for researchers; themes on projection are quite common while on expansion are seldom; journal articles are constantly published though the contents and scope of which are limited because of length, and theses or dissertations are insufficient as to the number and deserve improvement for research depth though their length are considerable; a plenty of works are written in Chinese but only a handful of English literature. Based on the phenomena mentioned above, this research hereby establishes the necessity of taking corpus linguistics as a means of processing data and regarding academic text writing as a genre to design a comparative research program on the concept of clause complexes in Master theses and doctoral dissertations. Before taking on the research, a series of theoretical concepts were also clarified, such as clearly defining the clause complex and its categories, limiting the scope of the clause complex for this study, distinguishing the connotation and denotation of expansion with conjunction, pointing out the difference within projection and citation, precisely stating the particularity of drawing on and integrating important concepts in the Appraisal Theory, etc.

To reflect the ability of implementing theories, a systematic and comprehensive framework for thoroughly describing and analyzing the clause complex would be established by combining SFG (supplies detailed and in-depth description and functional analysis for grammatical categories), Corpus Linguistics (provides theoretical supports for annotation, statistics and the analysis of large amounts of e-texts), and academic writing (directly affects the choice of clause complex from theoretical perspectives as a special genre), in which the qualitative research on Taxis in CC is based on the

description of SFG, that of Expansion is built on basic ideas of CC and AW, and the qualitative investigation of Projection comes from the CC framework and part of the Appraisal Theory, as expansion is regarded as more textual-cline on the continuum of three meta-functions in realizing clause complex relations while projection more interpersonal-cline though both of them are seen as the highly embodiment of logico-semantic functions in the conceptual meta-function. The combination of three principal theories and the perspective of the functional continuum has the potential to facilitate the development of a new research field in AW. This would entail a re-examination of the exploration of clause relations from a functional perspective, as well as the provision of a novel research direction for the resolution of specific problems in AW, which involves the seeking of inspiration from other aspects of SFG to address measurement issues such as syntactic complexity in domains that are not adequately addressed by existing theoretical frameworks. It also aims to advance the research scope and analytical framework of AW by, for instance, developing comparative paradigms for structures that transcend the structural divisions typically employed in traditional grammar.

The research object is the corpus of master's and doctoral theses written by English major students from Chinese universities. The researcher and colleagues from Chemnitz University of Technology jointly established this corpus, called ChAcE (Chinese Academic English), which contains 303 master's theses (CMAC, about 5.1 million words) and 112 doctoral dissertations (CPhD, approximately 5.3 million words). This corpus is available for interested researchers to conduct in-depth studies. In the present study, however, only a portion of the ChAcE corpus is utilized. By manually annotating nearly 50,000 clause complexes and recording the results in the software SysFan, a pioneering activity was realized: the annotation of a large-scale corpus. Other software, such as the auxiliary use of AntConc, reflects the consistency of results from the vocabulary indexing and the annotation. The adoption of the statistical software JASP plays a decisive role in the Bayesian factors numerical comparison of the grammatical-lexical selection results by Master's and doctoral students. The research also introduces a novel analytical method, namely an enhanced mixed research method, which employs a combination of qualitative and quantitative methods, or a single qualitative or quantitative method, according to the specific requirements of the research topic. This method is then applied in a unique analytical direction, in this case, a comparison of the clause complex structures chosen in Master and doctoral theses from the field of AW. The approach has yielded a series of comparable results. It is evident that corpus linguistics theory and its practical methods have had a significant impact on this research.

In terms of quantitative research, unlike previous studies that simply compared data by means or sums of consequences, a series of quantitative indexes (a measurement system of 73 indexes including numbers and proportion) within the clause complex system is modeled, and then plenty of annotated data are compared from statistical significance in the study. The practice, hence, confirms the feasibility of conducting quantitative analysis within grammatical categories, especially that of SFG, develops Halliday's probability theory on grammatical categories, and at the same time, makes the grammatical quantification from relations within and between categories extensively and profoundly. Besides, a qualitative comparison of lexical choices in annotating texts shows the reasons for preferentially selecting diverse grammatical functions by Master's and doctoral students, perspectives of making diverse words or phrases realized for the same grammatical function, and abilities to implement various expressing ways through one kind of grammatical functions corresponded to the same lexis in fulfilling meaning transmission; points out the fundamental reason of the difference in grammatical choices, i.e., the diversity in lexical expression; demonstrates the necessity of the existence of a delicate rank system; and explains the point that grammar and lexis are on a continuum and locate at two ends of the cline instead of being a superordinate-subordinate relation.

By showing the proportion of each category within the clause complex system in Master and doctoral theses from the corpus, the selection tendency of expansion or projection components intersected with taxis in the writing genre of academic papers can be analyzed and interpreted. The comparing results within the two groups thoroughly reflect the essential diversities in texts by master's and doctoral students, and the combining results of the two groups objectively reveal the fundamental reasons for diversely choosing each grammatical category. On the one hand, this can be attributed to the characteristics of language expression itself. On the other hand, it is influenced by the preferences and habits of individual writers when writing academic papers. After comparing the data results on grammatical categories from this research with Wu's, the consistency of opting tendency for each category in the paratactic or hypotactic system can be concluded, which is undoubtedly a meaningful direction for new research. Moreover, the proportion of each category in the clause complex system provides a dynamic perspective that some categories must be employed in academic writing with considerable frequency while some do not need to be chosen at all, and this kind of absence would not affect the integrality of meaning in the discourse and the accuracy of meaning expression.

Theoretical innovation can also be reflected in the proposal, characterization, and classification of

the concept *implicit clause complex*, the distinction between the implicit and explicit clause complexes, the methodological exploration of transforming implicit clause complexes into explicit ones, and the statistics and description of clause complex recursion patterns at each level. Being different from the practice of previous scholars that apply “syntactic complexity” in measuring structures within sentences to compare the complexity of texts, this study focuses on the clause relations and functions that primary annotating markers play in clause connection, proposes the concept *clause complex complexity*, attempts to construct three calculation formulas based on category probabilities, and compares the sophistication of texts by calculating the complexity of clause complexes. While the third formula is too complex and challenging for inexperienced researchers to operate, the entire computational framework offers valuable insights for comparing the structure of clause complexes. In fact, the method of constructing the clause complex framework and then calculating categories in the framework for comparison adopted in this study is applicable to any subsystem of the systemic functional grammar, that is, bringing ways of annotating and quantifying categories in subsystem to solve problems such as: analyzing meaning expression in texts and calculating the complexity of the system structure from a functional perspective.

Based on this research, the specific usage of some rigid indexes such as the proportion of hypotactic and paratactic subcategories, the proportion of implicit and explicit clause complexes, the proportion of each category within the system, recursion patterns, etc. at each clause complex level can be contrasted and interpreted in depth; the research of which complexity measurement method is closer to the ideal state of comparatively measuring text structures, and other research on the fitting, construction, and optimization of the measurement model of clause complexes complexity in discourse can be carried out; the impact of embedded structures in clause complexes on measuring indexes referred to above cannot be ignored; the comparison of selecting each category of the system in any functioning chunk in a text or in texts from diverse genres, in theses written by students from diverse educational levels, from different external conditions for example: countries, ethnic backgrounds, regions, social identities, genders, etc., are all worthy of exploration and discussion.

Therefore, this study has theoretical innovation and development while integrating systemic functional linguistics, corpus linguistics academic writing in a foreign language, applied linguistics (ALs), and other theories and applying them to solve practical problems, and has a certain enlightenment on related theories.

Practically, since the research results come from students’ written texts, they can first be applied to text writing, especially theses for getting a degree. Therefore, while explaining and studying the combination of clauses, students should also pay attention to the decisive role of grammatical structure and keywords or phrases in the expression of meaning that can help them make appropriate language choices for the message they intend to deliver in writing practice, to improve their writing and communicating abilities, promote their transmitting efficiency of thoughts and display their thinking processes in the text. For readers, the author’s thorough understanding and appropriate realization of the relation between clauses can greatly simplify readers’ comprehension procedures and eliminate their reading obstacles as much as possible on the one hand; and the reader’s solid grasp of clause relations can strengthen their capacity of meaning construal and thus avoid misunderstandings on the other hand. In regards to teaching, teachers can raise awareness of the importance of clause complexes in students’ reading and writing and selectively teach students with diverse cognitive abilities at different educational levels in terms of content. In operations, teachers can focus on language functions rather than structures to interpret texts from multitudinous perspectives on the one side and carry out more in-depth and extensive research based on this study in combination with daily teaching practice on the other. In research for discourse analysis, the idea of joining a mixed scientific research method with a grammatical-lexical analyzing process for academic discourse demonstrated in this study is also applicable to other discourses that distinct results may emerge for contrast. Back to the language, the particular comparative analysis of English clause complexes and Chinese complex sentence structures enables teachers to explain certain difficult points from a brand-new perspective in translation teaching and can also help students realize the differentiation in clause combining relations between English and Chinese languages, to more accurately convey authors’ ideas in translating English texts, especially literary texts, into Chinese, refraining from possible errors to truly reach the point where the form and meaning remain intact; and to appropriately add keywords or phrases to connect clauses in denoting their relations instead of being limited to common expressions in translating Chinese texts into English. One phenomenon deserves attention that the multi-functioning clause relations of one word such as “and” can cause troubles in English-to-Chinese translation and can also make Chinese-to-English translation tedious, the distinction of such kind in the research hence eliminates translators’ concerns about those troubles and increases their delicacy and richness of translation language. In the field of machine translation, the forming processes of clause complexes from lexis to grammar presented in the paper

prompt computer software to learn and attempt to comprehend, to reach the aim of automatically annotating clause complexes in texts; probabilities of each grammatical category and lexical choice obtained from the calculation results can provide a reference for the software, so that it can make a selection by examining the probability of lexis and grammars in corresponding semantic environment in case of being ambiguous for clause relations in translation, thereby to improve the efficiency and accuracy of automatic translation. Finally, the endeavor to connect systemic functional grammar with computational linguistics in this study is inspiring.

1.5 Chapter Overview

The whole dissertation is composed of eight chapters. Except for the first and the last one, each chapter follows a similar framework with an introduction part as a beginning and a summary as an ending.

Chapter One presents an introduction to some basic issues about the current study: the research background of where this research comes from and why this research should be conducted, including 1. difficulties that English learners are facing in writing their theses, especially on clause combinations of the English language when they transfer from a Master student to a PhD candidate, 2. the importance of focusing on some grammatical categories such as taxis instead of coordination or subordination, expansion rather than conjunction, 3. the structural diversities of projection from the traditionally called citation, 4. the meaning realization of implicit or explicit knowledge of language and the mention of the implicit language, and 5. the complexity of language grammar and lexical choices, and how they should be measured; the purpose of the present investigation from the clause organization and meaning transferring, language development, concepts clarification, SFG adaptability, and theoretical application deepening; detailed research questions to be answered from theoretical, methodological, practical, analytical, and applicable aspects; theoretical significances on research necessity excavating, research direction confirming, theoretical analyzing framework constructing, a mixed research method adopting, corpus building, texts annotating, the feasibility of quantitative and qualitative analysis verifying, theory denotation developing, and new research directions discovering; practical significances that the outcome of Master and PhD students' theses comparison can be inspiring and instructive for thesis writing, texts reading, writing education, Chinese-English translation, machine translation, etc., error analysis can be enlightening for preventing more mistakes, the grammatical-lexical analyzing practices can be referential for conducting discourse analysis, the complexity of clause complex can be creative for further texts calculation and comparison and the clause combining modes from functional perspective can be innovatory for computational linguistics studies; and the overall organization of the dissertation that is described in this section.

Chapter Two then devotes to the literature review from both academic writing and the clause complex part that the former draws attention to some basic academic writing concepts, annotations and denotations (concepts of EAP&ESP&EAW; research theories of discourse analysis, academic literacies, cognitive perspective, etc.; research methods of corpus linguistics from quantitative and qualitative perspective; research directions in discourse analysis, academic literacies, cognition, identity construction, etc.), writing-related knowledge (writing purposes and motives, writing modes, writing analysis framework, and writing process), specific researches on English theses writing of Master and PhD students, conjunction and citation which are analogical to expansion and projection concepts in the clause complex system, and also studies on academic writing from systemic functional linguistics or grammar perspective; and the latter demonstrates more than one hundred works published in Journals, books and opened online dissertation and thesis platform from both China and other countries (because half of the studies are conducted in China and in Chinese language, and also the theme of this project is on Chinese students' conditions and background) with themes of clause complex, projection, expansion intersected with theoretical studies, discourse analysis, English-Chinese language comparison, stylistics, translation and related, in order to demonstrate researchers' mutual views as well as conflicts in previous works, find gaps from methodological, theoretical and directional aspects within past research, state the location of this study in related areas, declare the significance of this study and to denote probable creativities that will be presented in following works.

Chapter Three begins by laying out the theoretical dimensions of this project and looks at how three directions intersect with each other and mutually influence the design, performance, and analysis of the research: one is systemic functional grammar (SFG), which is the core theory for the whole research, including its basic knowledge, definition of the term clause and clause complex, and contents of expansion and projection; another is about academic writing (AW), which would be roughly depicted of its way to write a thesis and relevant teaching methods; and the third part is corpus linguistics (CL) with the probabilistic theory in SFG. In detail, the SFG section contains the originality that explains the reason for being chosen as a theoretical basis, its differences from other functional or related grammars, the

socio-function of AW, contents that diversify eight groups of keywords, and the focus of this research: clause complex; gives definitions and shows the location of clause and clause complex in and around SFG theoretical frame, underlines three meta-functions of a clause; and fully describes the framework of the clause complex that Taxis is divided into Hypotaxis and Parataxis, Expansion is composed of Elaboration (exposition, exemplification, and clarification), Extension (addition, variation, and alternation) and Enhancement (from temporal, spatial, manner, causal, and conditional), and Projection is usually distinguished as quoting or reporting, with examples of their intersections from both Master and PhD students' theses. The AW section sees AW as a register and addresses the following questions: how can SFG be applied to AW and subsequently influence AW theories and methods; furthermore, how AW practical problems can benefit SFG linguists and thus extend the SFG theoretical development? Followed by CL description with the definition of corpus and corpus linguistics, classifications of corpus, choosing principles of corpora, historical development of CL, CL methodology, tools and software for CL annotation, advantages and disadvantages of their usage, the distinction between corpus-based and corpus-driven approach; and then the narrow sense of CL under the influence of SFG from its generality, history, relationship, classification or research methods, key words especially probability and instantiation, and the angel that CL was regarded as a perspective for SFG studies in this project. At the end, a theoretical framework is outlined, a conclusion will be drawn from a schematic diagram with three circles that qualitative analysis of register analysis in SFG is emphasized from SFG & AW research overlap, quantitative analysis of grammatical ones is executed from SFG & CL crossing, lexical analysis from quantitative methods is conducted within CL & AW research overlap, and a whole clause complex system for this project is put forward by scoping the SFG&AW&CL intersection. Thus, a theoretical foundation is established for analyzing practical problems encountered in thesis writing for MA and PhD students.

In Chapter Four, the corpus chosen for analysis in this study and the corresponding analytical methods will be described in detail. The Chinese Academic English (ChAcE) Corpus was first built and introduced of its data collection conditions, data conversion, and cleaning processes, and its related information, then the data, which includes 80 MA theses (1.28 million words, 28,350 clause complexes) and 32 PhD dissertations (1.44 million words, 28257 clause complexes) for clause complexes analysis in this research are selected from the ChAcE corpus. Before dealing with the chosen corpus, the advantages and disadvantages of qualitative versus quantitative methods are listed with a mixed design being set up for the research, and then some fixed methods are opted according to requirement: descriptive and qualitative methods (case study and grounded theory especially) are primarily used for annotation, and examples analysis; quantitative and comparative methods are applied in data calculating; experimental summary and modeling analysis are utilized for constructing a system; and exploratory method is adopted for outcome discussing. As to tools and software, *SysFan* is for manually annotating data, recording clause complexes and their connecting words with markedness, and counting grammatical categories in the whole system; *AntConc* is useful for calculating keywords in each clause complex and excluding unnecessary ones; *JASP* is beneficial for statistical analysis from both Independent Samples T-test and Bayesian Independent Samples T-test with figures and diagrams; and *Matlab* is inclined for modeling the clause complex complexity in the last but two chapters. After the chosen corpus is departed as CMAC and CPhD groups, depicted of its general information, and presented of its importing and annotating procedures, several indexes such as the number of clauses in each text, the average number of words in each clause, of clause complexes per sentences, of clause nexuses per clause complexes, of expansion and projection structures in each text, the number and the proportion of clause nexuses on each recursion level (from 1 to 7), the average proportion of parataxis or hypotaxis in clause nexuses in each text, etc., are counted, calculated and compared within Master and PhD levels. In Section 4.4, the theoretical framework presented in Chapter Three is synthesized with the relevant analytical methods mentioned above to outline the data analysis.

Chapter Five and Chapter Six are the most important parts of this dissertation for their data comparison and explanations. Chapter Five starts with general conditions of expansion and then focuses on the composition of expansion, i.e., elaboration, extension, and enhancement that quantitative (including descriptive statistics, independent samples t-test, and the Bayesian independent samples t-test of every index) and qualitative comparisons within MA theses and PhD dissertations are orderly conducted. In detail, the expansion part in Section 5.1 encompasses the average proportion of expansion in all clause nexuses, indexes to be compared about paratactic (the average proportion of paratactic expansion in all clause nexuses, paratactic structures, and expansions) and hypotactic expansion (the average proportion of hypotactic expansion in all clause nexuses, hypotactic structures, and expansions), and the average proportion of each composition in expansion structures in two groups. The quantitative analysis of elaboration will involve the average proportion of elaboration in all clause nexuses and conditions of paratactic (the average proportion of paratactic elaboration in all clause nexuses, all

paratactic structures, paratactic expansions, and all elaborations) or hypotactic elaboration (the average proportion of hypotactic elaboration in all clause nexuses, all hypotactic structures, hypotactic expansions, and all elaborations) in two groups. Qualitative analysis of elaboration in Section 5.3 refers to how elaboration structures, from the classification of exposition, exemplification, clarification, and hypotactic elaboration, are fulfilled through the realization of corresponding markers or lexis by both Master's and doctoral students in their writing with fine distinctions. Though the number of these markers chosen by students are listed and compared, the status of the overuse of which lexis causes a higher proportion of related grammatical structure is still not transparent until all data of lexicogrammatical comparisons are comprehensively considered. The quantitative analysis of extension is analogous to that of elaboration, though the numerator of comparing indexes will be changed into paratactic or hypotactic extension, and the denominator of the last group is an extension but not elaboration. Then, a qualitative discussion of extension covers additional (positive, negative, and adversative), variational (replacive and subtractive), and alternative clause relations, in which the positive addition relation marked by *and* is the most important and difficult one for students to master. An evident comparison result would be received as imagined from paratactic enhancement versus hypotactic enhancement choices by Master and doctoral students that the latter is more preferred to doctoral candidates after the quantitative analysis of enhancement in Section 5.6 is parallelly completed no matter the measuring index is the average proportion of hypotactic enhancement in all hypotactic expansions or all enhancement structures. The qualitative analysis of enhancement occupies certain space as it is divided into five main functions: temporal (same time, earlier, or later), spatial (same place), manner (means or comparison), causal (from reason, result, and purpose) and conditional (positive, negative or concessive), the latter two of which are of particular significance when students rely on causal relations to give explanations for principles, state the influence that one matter has on another, and present researchers or their intentions to readers; on conditional relations to manifest sufficient or necessary conditions that are needed to fulfill something, point out something that cannot be realized despite that conditions are all equipped with, or state the endeavor for something though they have been confirmed unachievable. This chapter will not present a detailed analysis of the subtle differences between the two groups. Instead, it will focus on how similar and different the two groups are in their use of lexicogrammatical structures to convey meaning and express their viewpoints.

The same data calculating and comparison methods are adopted in Chapter Six just as those in Chapter Five, only with categories' changes from hypotactic and paratactic projection to Quoting Locution and Reporting Locution, to Quoting Idea and Reporting Idea; from Locution and Idea to propositions and proposals; from grammatical choices to lexical (projection keywords) instances, for horizontally and vertically digging out how different Masters and Doctoral students are in referencing and citing other scholars' words directly and indirectly. To be specific, Section 6.1 constitutes indexes for comparison within Master theses and doctoral dissertations: the average proportion of projection in all clause nexuses; paratactic projections in clause nexuses, paratactic structures, and all projections; hypotactic projection in all clause nexuses, hypotactic structures, and all projections; and the composition of projection, i.e., the number and average proportion of locution and idea in all projecting structures. For the quantitative analysis of the locution part, besides hypothesis tests within two groups on the index, the average proportion of locution in all clause nexuses, paratactic locutions are compared from denominators being clause nexuses, paratactic structures, paratactic projections, and locutions to calculate the average proportions in each group, and hypotactic locutions are compared when denominators change to clause nexuses, hypotactic structures, hypotactic projections and again locutions. The qualitative analysis for locution pays attention to projecting verbs and divides them into verbs that lead to proposals or propositions according to the speech functions that they perform in a clause complex. By presenting each verb chosen in theses with their frequencies, the condition of which verbs are usually seen in each group, which markers are both common in two groups with high frequencies, which are common but have varieties of their preferences in two groups, and which verbs are unique or peculiar in each group will be plain that the variety may come from the number of texts on the one hand and the direction difference within literary studies or linguistics research on the other hand, and not all projection structures appear in the literature review part in theses, which deserves further investigation. By combining the SFG theoretical framework with Appraisal Theory in analyzing examples, the interpersonal metafunction of projections in thesis writing will undoubtedly be revealed. For the quantitative analysis of the idea part, except for turning the numerator into the paratactic or hypotactic idea and changing the denominator from locution to idea for hypothesis tests within two groups, the method and principle for index comparison remain unchanged. Though the number of paratactic ideas in most theses is zero, which would contaminate calculating statistics, the Bayesian independent samples t-test is also activated for a precise result and as a supplement to means comparison. The qualitative analysis of *idea* will emphasize some verbs that lead to propositions, such as: *show*, *find*, *indicate*, *believe*,

mean, see, think, know, suggest, etc., to uncover how projected ideas are originally, objectively, faithfully, or in reverse, personally, subjectively, forcefully organized in theses writing and then transmitted to readers to establish a certain kind of relationship with them. The summary for Chapter Six would stress some research points with their results and, at the same time, mention likely research directions.

For discussions in Chapter Seven, the general conditions of data outcome covering 1. the grammatical categories distribution of the clause complex system, 2. range, mean and Bayes Factor values of 73 measuring indexes of clause complex grammatical comparison within MA and PhD groups, and 3. lexical choices that mark clause complex relations will firstly be demonstrated that by comparing category distributions within Master and doctoral theses, and within the whole corpus and Wu's outcome, more diversities than similarities can be discovered, and some interesting phenomena can be dug out; the range of proportions of each grammatical category in the system in the whole corpus will indicate some necessary categories in thesis writing genre while some are optional; by judging the Bayes Factor values of all indexes within two groups, the preference for hypotactic versus paratactic structures or enhancements versus elaborations can be obvious; the inconsistency of range and mean of these indexes suggest a multivariate distributed data in both groups, and the comparison of mean values and BF values in each index results to a controversial outcome that needs further exploration; annotation markers in five categories, i.e., elaboration, extension, enhancement, locution and idea will be particularly compared with their numbers in sub-categories, to answer questions raised in Section 1.3 that how the lexical choices influence the selection of relating grammatical relations to realize functions that clause complexes perform in texts, how different are lexicogrammatical choices within Master and doctoral students, and how to explain these diversities, etc., to provide students with possibilities for improving their grammatical complexity if more hypotactic or complex structures indeed mean development, and to confirm the viewpoint of regarding lexis and grammar as two ends of a cline instead of hyponymy at last. Four kinds of additions that are concluded within academic writing texts and the annotation for wrong clause complexes will then be presented in Section 7.2, in which the former denotes unrelated, unequal, wrong, and implicit relations in each grammatical category in the clause complex system that are put forward from the addition relation as an investigation of peculiarities analysis except for depicting the location of these four additions as well as the location of four layers in any category of the system; and the latter means some kinds of clause complexes with wrong relations or markers in writing, the way to identify and then recompose them, to be specific, adding, missing, substituting, inconsistent, not-splitting, punctuation mistakes, and wrong usages. Section 7.3 is concerned more with the differentiation of implicit clause complexes from the explicit ones, in which the theoretical foundation for putting forward the concept, including division that originates to clause relations, tacit knowledge, and choice or the probability theory are provided; before getting right to the point, two groups of close terms are defined and irrespective of due to their inappropriateness, i.e., "overt" - "covert" and "communicative" - "tacit," the counterpart *implicit* versus *explicit* is then be confirmed; followed by the average proportion of implicit clause complexes in all CCs in each text within MA and PhD groups; characteristics such as recessivity, integrality, stability, occasionality, situationality, culturality, relativity, individuality, and exclusivity of the implicit clause complex; the classification; some related structures that are hard to identify: unique finite clause complexes, clauses connected by *and*, non-finite ones especially when the implicit process is played in dependent clauses, one finite clause plus one non-finite, implicit projection, etc.; and steps of transforming implicit clause complexes into explicit ones if possible. The discussion of implicit and explicit clause complexes can be treated as a theoretical expansion to SFG, which will enhance the practical raising of students' consciousness regarding the organization of clauses in academic writing. Furthermore, the generalization and summarization of common patterns for the recursion of clause complexes represent another innovation. This will be demonstrated in Section 7.4, where it will be shown that, on each recursion level, similarities can be more significant than diversities for the choices made by master's and doctoral students. Furthermore, the clause order of any grammatical category in a clause nexus with corresponding logico-semantic markers is presented in order to provide a comprehensive deconstruction of the clause complex. In Section 7.5, the concept of *the complexity of the clause complex* or *the clause complex complexity* will be proposed as an alternative perspective. This perspective treats the connection of clauses from a functional or dynamic angle, rather than the traditional structural or static angle, which is commonly referred to as "syntactic complexity." After the origins and theoretical background are introduced, the term *the clause complex complexity* is provided with a complete definition, and three formulas synthesizing categories with proportions are modeled based on the definition. In applying the first formula to test the complexity of texts from MA and doctoral groups, the results may appear reversed, indicating the need for further investigation and improvement of the formula. Though the two other formulas are more delicate on account of the proportion of every sub-category on each level, problems are still obvious when these proportions are calculated according to manual recording. Nevertheless, the attempt to depict the complexity of each text

through the use of clause complexes is a valuable avenue for further exploration.

The dissertation will be concluded by Chapter Eight with answers to all questions presented in Section 1.3 and summaries of both qualitative and quantitative findings from the data (especially from Chapters 5 to 7), contributions that have been made as compared to studies presented in the Literature Review part, implications that this study can bring to other scholars from both theoretical and practical perspectives, limitations of theories and methods that are adopted in the research and data that are chosen for comparison because of irresistible factors, and suggestions that are provided for future research in theoretical development as well as application scopes such as writing, teaching, reading, translating, machine translation, computational analyzing, data annotation and processing with software, etc.

Chapter 2 Literature Review

This chapter is a literature review of mainly two parts: one is English academic writing, and the other is clause complex. The aim is to set the background for further discussion that theories, methods, and analyzing strategies in academic writing as a genre or a register in previous studies provide practical perspectives for clause complex research, and the unique organizational and functional angle of the latter in investigating the English language promotes the development of texts analysis from writing materials, of students' reading and writing abilities as well as of teachers' pedagogical implication.

2.1 The Literature Review of Academic Writing

This section revolves around a list of research for academic writing that, after explaining some key terms such as ESP, EAP, and EAW, related theoretical foundations from the discourse analysis (DA), academic literacy (AL), cognitive including identity construction (IC) perspective are presented, followed up by the mention of some research methods such as ethnography and case study from the qualitative aspect and corpus, especially from the quantitative aspect. In corresponding to Hyland's (2009a) framework of academic discourse analysis, i.e., from textual (in academic words or lexis, discourse, register, genre, corpus, multimodal, etc. tendency), contextual (in the sociology of science, ethnography, sociohistorical and cultural fields), and critical (CDA specifically) division, past works are generally reviewed to perform a connecting link function for the preceding and the following section. However, AL-related research is separated from the critical perspective as a parallel section to DA, in which the cognitive angle, including the identity construction, the publication of theses, and the teaching of EAW-relevant works, are displayed. Moreover, writing purposes and motives, writing modes, and the process of writing are also necessary in reference. Then comes the most correlative research to compare the English Master's theses and doctoral dissertations from the academic writing field to discover their strengths and weaknesses and find the gap between this research and previous ones. Sections 2.1.7 & 2.1.8 aim to explain expansion instead of conjunction and projection rather than citation as the object in this research because of the view of the language they perform in communication or meaning conveying on the one hand and the research scope of these concepts on the other. The ending part endeavors to bring the scope of discussion back to clause complex in the next section by raising some controversial points of the mentioned works above.

2.1.1 ESP/EAP/EAW

The relationship between ESP/EAP/EAW is hierarchical, though ESP and EAP are widely accepted by scholars as an abbreviation for its denoting concept.

ESP

As the earliest monograph discussing English for Specific Purposes with corresponding teaching methods, *English for Specific Purposes: A Learning-Centered Approach* (Hutchinson & Waters 1987) thoroughly demonstrated the concept of ESP, its course design, application including syllables, materials evaluation, materials design, methodology and evaluation, and role of teachers. For them, ESP should better be seen as an approach rather than a product that "ESP is not a particular kind of language or methodology, nor does it consist of a particular type of teaching material." (ibid, p. 19) Rather, "it is an approach to language learning" that is based on learners' need, or "an approach to language teaching in which all decisions as to content and method are based on the learner's reason for learning." (ibid) Therefore, Johns et al. (2011, p. 1) considered that "studying language, discourses, and contexts of use" in addition to "students' needs" from general senses and the application of all these research results "to the pedagogical practices" is what ESP differs from other "branches of applied linguistics and language teaching." Or the "heavy demands on its practitioners" in "collecting empirical needs-assessment data," "creating or adapting materials" for specific identified needs, and "coping with often unfamiliar subject matter and even language use" make ESP pedagogy balanced within "specific needs-based" and "theory-driven" perspectives compared to other approaches. (Belcher, 2004, p. 166) No matter from the sociodiscourse, sociocultural, or sociopolitical approach, being antagonistic to EGP (English for general purposes), ESP can be divided into either EAP (English for academic purposes) or EOP (English for occupational purposes) with related EVP/EPP (English for vocational/professional purposes).

EAP

One significant work of EAP is *English for Academic Purposes: A Guide and Resource Book for*

Teachers, in which its definitions and scope, the definition and methods of needs analysis, EAP syllables and course design, the evaluation of students and courses, methodologies and materials, study skills and practice of English for general academic purposes, etc. are systematically introduced, and some research of English for specific academic purposes with their applications are also demonstrated. Originating from the broader aspect of ESP, i.e., a “theoretically and pedagogically” combined concept, EAP emphasizes “the specific communicative needs and practices of particular groups in academic contexts” for language research and instruction by absorbing the understanding from “the cognitive, social and linguistic demands of specific academic disciplines.” (Hyland & Hamp-Lyons, 2002, p. 2) It is generally defined as “the teaching of English with the specific aim of helping learners to study, conduct research or teach in that language.” (Flowerdew & Peacock, 2001, p. 8) Having the same characterization of paying attention to “strong interdisciplinary research as a means of illuminating the constraints of social contexts on language use and the ways learners can gain control over these,” it also commits to “linguistic analysis, contextual relevance and the classroom replication of community-specific communicative events” as well as refraining from some ESP disadvantages such as “a tendency to work for rather than with subject specialists, a vulnerability to claims that it ignores students’ cultures, and a reluctance to critically engage with the values of institutional goals and practices.” (Hyland & Hamp-Lyons, 2002, p. 2-3) Or in Tribble’s (2009, p. 401) words, “rooted in ground-breaking work by Halliday, McIntosh & Strevens (1964) in register analysis and later work by Halliday & Hasan (1985) on genre,” EAP demands for thoroughly “describing both the communicative context and the linguistic behaviour arising from this context as the starting points for any pedagogic solutions that are developed to meet learners’ needs.” In detail, beside including study skills that represent for “abilities, techniques and strategies” applied in “reading, writing or listening for study purposes” (Richards, Platt & Platt, 1992), EAP also concerns about additional features which “can be summarized as a general academic English register, incorporating a formal, academic style, with proficiency in the language use.” (Jordan, 1997, p. 5)

Normally, EAP can be divided into two parts that “may be either ‘common-core’ or ‘subject-specific’” in which the former “shades off into ‘language study skills’ (Coffey, 1984, p. 4)” and the latter “comes to categories for special situations or setting, special communicative needs or functions, special subject matter, and special skills or tasks” (see Robinson, 1980), or in Jordan’s (1997, p. 5) words, “Subject-specific English is the language needed for a particular academic subject ... together with its disciplinary culture. It includes the language structure, vocabulary, the particular skills needed for the subject, and the appropriate academic conventions.” Therefore, EAP is depicted by Blue (1988, p. 96) as English for General Academic Purposes (EGAP) and English for Specific Academic Purposes (ESAP).

Recently, the research trends of EAP have been transferred from corpus-based analysis to multi-dimensional and multi-perspective genre analysis and interpretation; from traditional academic genres such as journal articles to relatively non-mainstream, (semi) closed genres and to those triggered by new media and new technologies; from texts to individuals such as students and scholars in academic English communication; from pragmatism to critical construction orientation, etc. The EAP writing research has gradually shifted from the superficial description of discourse to the macro-structure of discourse and the in-depth explanation of communicative intention regarding genres; from the combination with other specific disciplines to the division of sub-areas in directions; from the research on individual writers to the research institution of social groups that the individual works as to writer’s perspective; from higher education and the university to K-12 education for education levels (Lea & Street, 2006, p. 368) with a linkage to the Content Areas (Wilcox & Jeffery, 2014, p. 168-176), etc.

Though Swales (1990, p. 2) put forward the concept of *remediation* for EAP teaching to “understand the forces which variously shape the language of the academy, and which stands against the view that our teaching of academic English is at bottom nothing more than remedial,” the contemporary methods refer to teaching English “in the academy of all age and proficiency levels” by absorbing in interdisciplinary advances on “research methods, theories and practices” for the sake of deeply understanding “the structures and meanings of academic texts,” “the demands on communicative behaviors” in particular academic contexts, and “the pedagogic practices” by which these behaviors can be developed. (Hyland & Hamp-Lyons, 2002, p. 3) And from students’ views, the setting for EAP may involve the accommodation to “a new academic system” with conventions from a diverse “cultural environment” and the observation of “the nature of the relationships between academic staff and students, and among students themselves” in which “attitudes and expectations” are conveyed via language. (see Jordan, 1997, p. 6)

EAW

Being one main branch of EAP, EAW (English for Academic Writing) emphasizes cultivating students’ writing abilities in capturing academic trends, expressing academic thoughts, and participating in international academic exchanges, therefore possesses characteristics of strong explanatory and argumentative tendency in language expression, explicit disciplinary directions in content selection, and

professional standardization in texts formats. The main concepts of EAW, including *English* (as a second language) *Academic Literacy*, *speech genre*, *discourse community*, users of academic English, *authors* vs. *writers*, etc., can all be referred to the journal *English for Specific Purposes* and *Journal of Second Language Writing*. Take the concept of *genre*, for example, Ramanathan & Kaplan (2000, p. 180-2) pointed out that genre, on the one hand, has conventional stability and on the other, evolves “and changes to meet the growing and changing socio-cognitive needs of discourse communities; and develops to meet the needs of changing technology; to adapt to changes in ideology and worldviews in discourse communities;” and “changes as individual writers take liberties with textual conventions.” For the concept of *discourse community*, Swales (1990, p. 24-7) put forward six significant features that are “necessary and sufficient” for the identification of a group of individuals as a discourse community:

1. a discourse community has a broadly agreed-upon set of common public goals;
2. a discourse community has mechanisms for intercommunication among its members;
3. a discourse community uses its participatory mechanisms primarily to provide information and feedback;
4. a discourse community utilizes and hence possesses one or more genres in the communicative furtherance of its aims;
5. in addition to owning genres, a discourse community has acquired some specific lexis;
6. a discourse community has a threshold level of members with suitable degrees of relevant content and discursual expertise.

2.1.2 Theoretical Foundations

For the research and teaching of *academic literacy*, Coffin & Donohue (2012, p. 64) connected and compared two perspectives that are considered as overwhelming “over the last decade” and are closely related to “both first and second language contexts”: one is called *Academic Literacy* (AL) by emphasizing more on “practices in context,” and the other is *Systemic Functional Linguistics* (SFL) in focusing on “texts in context.” In this section, the theoretical foundation for academic writing research or theories that can be applied in researching and teaching related materials are classified into three directions: discourse analysis that follows the framework of Hyland (2009a), including SFL, academic literacy, as well as cognitive theory with identity construction encompassed.

Discourse Analysis (including SFL)

The analyzing approaches to academic discourse are generally categorized into three aspects by Hyland (2009a, p. 20), i.e., the textual, the contextual, and the critical, in which the textual is more about “language choices, meanings and patterns in texts including those based on genre, corpora, and multimodal analyses,” the contextual concerns about broader “situational aspects, such as the sociology of science, ethnography, and sociohistorical perspectives,” and the critical mainly refers to critical discourse analysis developed by Fairclough (1989, 1992, 1995, 2003) and Wodak (1989). Though academic literacy is seen as a critical perspective, its cognitive attribute mixed with the social angle in linguistic practices and its status in academic research with quick development makes it worthwhile for a separate discussion.

Academic writing research under the SFL framework mainly includes the structural potential of genre and register, the interpersonal or ideational metafunction, the grammatical metaphor, etc., in academic discourse. Being a proven, systematic, and applicable linguistic theory, SFL can be a panacea to provide theoretical support for language analysis from the textual, contextual, and critical perspectives mentioned above.

Academic Literacies

Academic literacy is based on the concept of *literacy*, which is hard to define but is provided with eight characteristics by Barton (1994, p. 34-35)⁹, where literacy is a “social activity” that relates to “people’s literacy practices” in particular “literacy events” and hence has various kinds according to diverse “domains of life” (cultural or historical periods); is based on a “symbolic system used for communication” and “representing the world” as well as relating to other information exchanging systems; guides our actions through “awareness, attitudes and values” that we have in social settings; “has a history” as well as “social history” that would change through time and then affects individual’s practices in literacy events. In other words, “Literacy is primarily something people do; it is an activity, located in the space between thought and text. ... Like all human activity, literacy is essentially social, and it is located in the interaction between people.” (Barton & Hamilton, 1998, p. 3)

Instead of taking the “study skills” approach to see “literacy as a set of atomized skills,” focusing on

⁹ For a thorough version of characteristics of literacy with explanations, see p. 34-35.

the “attempts to ‘fix’ problems with student learning” on their own “surface features, grammar, and spelling,” and the “academic socialization perspective” to ask students to notice the cultural aspect and focus on their “orientation to learning and interpretation of learning tasks,” the Academic Literacy approach still takes literacies as “social practices.” By absorbing the social and ideological orientation of the “New Literacy Studies,” it holds epistemological and identifying attitudes toward “student writing and learning” instead of seeing them as skills or from a socialized view and the “institutions” where “academic practices” are conducted encompass and are influenced by “discourse and power.” (Lea & Street, 1998, p. 158-9) The same viewpoint can be proved by Jones et al. (1999, p. xvi) when higher education and corresponding writing practices are required not for “skills and effectiveness” but for “epistemology” in making clear the essence of “knowledge” and the “authority over it,” for “identity” in finding out “the relation between forms of writing and the constitution of self and agency,” and for “power” in understanding “how partial and ideological positions and claims are presented as neutral and as given” in “academic activities” such as “requirement” and then “feedback and assessment” processes from teachers. Hence, the curriculum for academic literacy is also asked to be designed with kinds of “communicative practices” by taking “genres, fields and disciplines” into account. (Lea & Street, 1998, p. 159) And the teaching and learning methods are conceptualized and affected as the literacy demands “cultural and contextual aspects of reading and writing” that language in AL, like that in CDA, is seen as discourse practices applied in contexts instead of “a set of discrete skills.” (Hyland, 2009a, p. 42)

For students, one main characteristic of AL practices is its requirement for switching within settings with appropriate linguistic practices in each setting and the ability to convey and understand social meanings and find identities in that setting. (see Lea & Street, 1998, p. 159) It is no wonder that “technical aspects of grammar and organization” are not regarded as difficulties in academic writing for students, but their techniques of using different strands of learning and “previous experiences.” “Entering the academy” indicates an appropriate “cultural shift” for being identified as one of the members in relevant communities; and “academic success” signifies representing oneself as the discipline required by “adopting the values, beliefs and identities embodied in academic discourses.” (Hyland, 2009a, p. 43) Therefore, the AL approach coincides with viewpoints from CDA, SFL, and cultural anthropology perspectives concerning the processes of meaning-making and meaning-related problems instead of skills or deficits in student writing.

Cognitive Theory (Identity Construction)

Cognitive theory is a comprehensive concept with numerous viewing angles and tremendous application fields. Though scholars from academic writing research would take cognition for granted and ignore the influence that it has on students, on teaching, on the process of academic writing, on the relationship between teachers and students in writing classroom, on the motives of thesis writing, etc., problems related to it remain and questions urgently to be answered can be various: How do learners recognize the importance, difficulty, and the complexity of academic writing? Is there any deviation in understanding these factors between teachers and students? What strategies do learners use to process English academic writing? What factors may influence the acquisition process of academic writing literacy? What are the individual differences? What kind of changes and effects do individual learners have in the process of constructing academic texts? What are the interactive features between learners and academic discourse communities or institutions? Hence, correlated aspects are usually of identity construction through which students earn their fame such as getting a degree by thesis writing; publication where deficiency of language, structure, style, concept cognition, etc. can be informed of and improved, and the contents are viewed as breakthrough or creative; teaching course design where teachers can be aware of the unbalanced relationship with students and learners are retrospectively able to give feedback for; writing motives with which students gradually transfer their attitudes from passive to active even though sometimes they are urged to write; writing processes of how to organize the structure, what to choose for meaning expression and what can be regarded as reasoning via genre or register changes, etc. The effects of cognition or cognitive psychology on academic writing research, therefore, primarily lie in social, developmental, educational, and personal perspectives.

2.1.3 Research Methods of Academic Writing

Research methods for academic writing research can be qualitative, such as ethnography, case study, interviews, questionnaires, etc., and quantitative primarily by corpus analysis and statistics calculating. This means that methods for linguistics or applied linguistics can normally be applied in AW research, and often, a mingle of them would be taken for most scholars in a project. For example, by adopting research methods such as “case study, ethnographic techniques, rhetorical and discourse analyses of changing features in large corpuses and in the texts of individual writers” (p. ix), Berkenkotter & Huckin (1995) “illustrate the social dynamic nature of genres,” discuss “the textual dynamics of scientific

publication,” investigate “the relationship between disciplinary formation and the features of evolving genres in disciplinary communities,” and demonstrate “a socialization study of a doctoral student’s acquisition of the registers and genre conventions” (p. x-xi). For them, genres are essentially dynamic rhetorical structures that are influenced by the context in which they are used, and an approach for “the textual character of disciplinary communication is to examine both the situated actions of writers, and the communicative systems” that these “disciplinary actors” take part in. (p. ix)

Qualitative: Ethnography

Originating from anthropology for describing how people behave in social groups, the ethnography approach focuses more on “people’s behavior in naturally occurring, ongoing settings” and endeavors to provide a “cultural interpretation” with corresponding patterns for the behavior. To conduct “an interpretive-explanatory account of what people do in a setting, the outcome of their interactions, and the way they understand what they are doing (the meaning interactions have for them)” besides holding “a holistic view” that the description and explanation of any cultural or behavioral strand should be considered of its systems, and following theory-driven and comparative principles, ethnographers would normally take out “systematic, intensive, detailed observation of that behavior - examining how behavior and interaction are socially organized - and the social rules, interactional expectations, and cultural values underlying behavior.” (Watson-Gegeo, 1988, p. 576-7)

The ethnographic method used by Mur Dueñas (2012) enables the study of the English-language paper publishing processes in an international environment from “Spanish researchers in the field of Finance” and results to findings that the key to their success of publication lies in persistence and being good at communicating with reviewers by taking various strategies. Curry & Lillis (2010) adopted “a longitudinal text-ethnographic” methods and through drawing out “network histories” with core foregrounding dimensions: “local and transnational, formal and informal, strong and weak, durable and temporary,” they discovered that “strong, local, durable cooperation network” can help scholars participate in “transnational” academic networks, thus providing strong support to scholars Publishing research results in both English and native language. By auto-ethnographically describing the process of writing his doctoral thesis, Weatherall (2018) presented deep thought on “conventions and procedures of doctoral writing,” answered the question such as “how could doctoral writing be done differently to expand the boundaries of thought in management,” and thus gives supports for doctoral students to “reconsider the conventional approach” during writing processes. Usually, the ethnographical method is widely applied in AL research, which will be demonstrated later.

Qualitative: Case study

Normally, a case study is “an empirical inquiry” in investigating “a contemporary phenomenon in depth within its real-life context” that its result can either be data “converging in a triangulating fashion” from “variables of interest” or be a theory or a method “to guide data collection and analysis” from previous “theoretical propositions.” (Yin, 2009, p. 18) Stake (1995, p. 3-4) characterized three main types of case study: intrinsic (study a unique phenomenon), instrumental (study a particular case for a broader appreciation of an issue or a phenomenon), and collective (simultaneously or sequentially study multiple cases for a broader appreciation of a particular issue).

The case study by Flowerdew (1999b) was not conducted on an individual but a group, i.e., “Hong Kong Chinese academics who have Cantonese as their first language” (non-native speakers of English) that in their processes of publishing in “international refereed journals,” a large-scale survey concerns on their previous exposure to English, their attitudes to publishing, their problems in the process, strategies to success and any change that would influence their publishing language. From the same case in English publication and by taking “in-depth interviews,” Flowerdew (1999a) discovered a list of problems that may be confronted in the process: expression inefficiency, time-consuming, sparse vocabulary, inappropriate amount of force in “making claims for their research,” L1 language transfer, difficulties in writing qualitative articles, style simplicity, and problems for introductions and discussions writing. Though the case is a group of scholars, the research results can be generalized for non-native English speakers. In the “socio-politically-oriented case study” of a doctoral student (a nonnative speaker of English) from the physics field in publishing papers, Li (2006) pins on the notion of “legitimate peripheral participation” and takes a “social constructionist perspective” on research writing that the “power-infused relationships” between the case and the institutional context indeed are factors of influencing his publication processes, including the publication being a graduation requirement from the school, the supervisors possessing “expertise and authority,” as well as editors from target journals for publication. Given the challenges, “affordances and constraints” of two English as an Additional Language (EAL) doctoral candidates in writing their PhD theses in “social and academic contexts” and the opportunities they possess, online conversation with them and a teacher of “EAL writing and doctoral education researcher” is conducted to investigate the changes of identity, voice, and agency as well as their roles in writing processes. The research (Magalhães et al., 2018) suggests that candidates’

confidence is increased when they have supportive interactions in social networking and institutional opportunities. Then their agentive behavior increases, new chances of expressing their voices are quested, resource constraints are therefore conquered, and their identities are accepted as scholars. But as two cases are “relatively privileged EAL writers” of proficiency, “experienced language teachers and staff members at their respective institutions,” unbalanced institutional or academic resources and opportunities may be supplemented with inadequacy of English language ability for other EAL doctoral candidates. In the “exploratory” case study of the composing process from a master’s student in writing a 4000-word essay by using in-depth logs, questionnaires, and interviews, the difference in allocating time to composing behaviors is obvious from those under the writing task of immediately responding to a prompt, and the “cognitive resources used in an electronic environment” may have differed from those using pen and paper by L2 university students. (Stapleton, 2010)

Quantitative: Corpus Analysis

Being diverse from intuitive research methods, the biggest advantage of corpus research is that through retrieval and statistics analysis of large samples of language usage in daily life, information about the usage frequency and the distribution of lexical and grammatical features can be provided, leading to more objective and scientific results, and more persuasive conclusions. In addition, as multiple types of corpora can represent different languages, language varieties, or genres, learners’ acquisition characteristics through comparative analysis of these materials can be further explored.

In conducting the quantitative diachronically investigation of critical (Cr) or non-critical (NCr) references in English medical discourse over 185 years (from 1810 to 1995) by building a corpus of 90 medical articles from 34 diverse journals, Salager-Meyer (1999) discovered that Cr references were significantly more frequent in the year 1810-1929, but NCr references outnumbered Cr ones from the year 1930 to 1995. This means that 19th and early 20th century medical papers took a more critical view, and “the evolution of the tone of voice of Cr references” over the 185 years signifies a transfer from “an author-centered and privately-based medicine (an “involved production context”) to a fact-invoking, professionalized and highly competitive scientific community (an “informational production context”).”

However, corpus research, especially data-driven analyses emphasizing description, makes it hard to provide a systematic theoretical explanation for researching phenomena. That is why, in retrospection of corpus studies on ESP and the New Rhetoric approaches to genre analysis, Flowerdew (2015) recognized the major criticisms that corpus analyses apply more “bottom-up rather than top-down methodologies” and their ignorance of socio-cultural context leads to “a decontextualized corpus data.” Therefore, he put forward an “integration of both corpus-based and genre-based approaches to text analysis in EAP/ESP.” In recent years, the research for academic writing using the corpus methodology has mushroomed. It is not the method itself that counts, but research objects such as grammatical or lexical features from texts, feedback or viewpoints from questionnaires and interviews, etc., and subjects, i.e., students, teachers, or scholars who using these academic strategies in writing and who are interviewed for either publishing or teaching that are of necessity for corpus analysis.

2.1.4 Research Perspectives of Academic Writing

Angels for conducting academic writing research can be related to lexical, grammatical including hedges, fossilization, redundancy, etc. problems, and textual analysis; structural framework, moves and steps, or citation patterns of texts; genre analysis; corpus analysis; critical discourse analysis of texts written by multi-lingual students; writing system and skills; cognitive or psychological studies; teaching methods of course designs; specific terms/words/phrases or EFL writing; processes of academic writing, etc., but in this section, they will be generalized into two directions: discourse analysis which mainly pays attention to texts or contexts of the producing process, and academic literacy which concerns more about writers or authors, scholars and teachers of their cognition on writing, publication problems, methods of teaching, writing processes and strategies applied in writing.

2.1.4.1 Discourse Analysis

The literature review of academic writing from the discourse analysis perspective mainly focuses on the textual, contextual, and critical aspects described in Hyland (2009a, p. 20), though sub-categories do not strictly follow the division, such as the corpus or multimodal angles.

Textual

The textual discourse analysis for academic writing usually pays attention to aspects such as grammatical-lexical choices, the discourse itself, seeing the discourse as a register or a genre, compiling the discourse as a corpus and investigating grammatical-lexical patterns or genre features via the databases, adopting a multimodal angel to discuss the relationship within texts and graphics, etc. In

which genre and register are most welcomed, followed by grammatical-lexical features and patterns. The corpus aspect, in fact, runs through others as its property of a large number of texts, leaving multimodal a peripheral and unusual stance.

Grammatical-Lexical

Grammatical-lexical research refers to features that are paid attention to in academic writing materials, such as syntax or sentence-related characteristics, lexis like nouns, citation verbs, deixis, etc., multi-word expressions like clusters, chunk or bundles, concordance, formulaic sequence, etc.

A detailed quantitative analysis of 1,083 L1 and L2 academic texts by Hinkel (2003b) can indicate that advanced nonnative-English-speaking students in U.S. universities would frequently use “simple syntactic and lexical constructions,” such as be-copula as the prevalent verb associated with high employment of predicative adjectives and then vague nouns. Public, private, and expecting/tentative verbs are significantly higher than those found in basic texts by native English speakers (newly admitted first-year NS students) on comparing their median frequency rates. However, the choice of academic and formal it-cleft constructions is quite low in L2 texts.

After Aktas & Cortes (2008) studied *shell nouns* of their frequency, corresponding lexicogrammatical patterns, and associated functions in articles and MA/MS or PhD theses by combining corpus-based and qualitative methodology, then discovered that shell nouns are more frequently used by international graduate students (from one university) than published authors (from the Published Research Articles Corpus), and the functional patterns of characterization, temporary concept-formation, and linking functions in these two groups vary.

Other research on lexis can be *reporting verbs* in MA theses (Lou, 2011), “adverbials of result” in learners’ corpus and native speakers (Lou, 2017), “deictic, modifying, and intensifying adverbials” in academic essays written by first-year NS and academically-advanced NNS students (Hinkel, 2003a), *directives* in a 2.5-million-word corpus organized by published articles, textbook, L2 student essays that the choice highly depends on purposes and perceptions of rhetorical context (Hyland, 2002c), disciplinary preferences and discourse functions of *self-mentions* between the Chinese PhD students’ texts and the international journal articles (Wang & Lv, 2017), *hedges, boosters, attitude markers, self-mention*, etc., in interactive metadiscourse from L2 postgraduate dissertations that the choice of these features enables students to make connections with readers, anticipate readers’ reflection, and project themselves into texts as credible academics and writers (Hyland, 2010).

As to Lexical Bundles (Biber et al., 1999) or Clusters (Scott, 1996), Hyland (2008a) advocated that the master of multi-word expressions, i.e., clusters, chunks, or bundles greatly embody the fluency of “linguistic production,” more randomly application of extended collocations than expectation can enable us a better meaning expression in particular contexts and understanding or awareness of coherence in texts, a 3.5 million words corpus of research articles, doctoral and Master’s theses across four contrasting disciplines is thus built to probe into “disciplinary variations” of frequencies and preference of choosing forms, structures, and functions of 3-, 4- and 5-word bundles. By clarifying diverse functions of these bundles from “research-oriented” (location, procedure, quantification, description, topic), “text-oriented” (transition signals, resultative signals, structuring signals, framing signals), and “participant-oriented” (stance features, engagement features) aspects, the analysis suggests that bundles are not only of significance in organizing academic discourse, but also provides evidence in classifying written texts into disciplines. Hyland (2008b) also compared “structural differences in clusters across genres” and “functional diversities in genre clusters” and proved that writers from diverse disciplines or directions have numerous choices in conveying their arguments, building up their credibility, and convincing readers of their academic abilities. Further research can be variation of bundles in various genres such as in university teaching and textbooks (Biber, 2006; Biber, Conrad & Cortes, 2004), in postgraduate learner writing and articles from the BNC (Scott & Tribble, 2006), in dissertation abstracts by Chinese and L1 English doctoral students (Lu & Deng, 2019), etc.

One of the important purposes of researching academic vocabulary is to acquire vocabulary expressions used in academic contexts that are different from general English, and because Hyland (2008a) asks for a pedagogical focus on bundles and a building of a single core vocabulary for academic study, some scholars tend to create representative academic vocabulary lists from academic text corpora. Based on empirical research, Simpson-Vlach & Ellis (2010) built up an Academic Formulas List (AFL) for both spoken and written language with pedagogical implications. Containing high frequency of formulaic sequences in academic spoken and written corpora, the list is classified into three categories and described from the pragmatic functions of the discourse: referential expressions (specification of attributes, identification, and focus, contrast and comparison, deictics and locatives, vagueness marker), stance expressions (hedges, epistemic stance, (obligation and directive, expressions of ability and possibility, evaluation, intention/volition, and prediction), and discourse organizing function (metadiscourse and textual reference, topic introduction, and focus, topic elaboration, discourse marker).

Based on the COCA (Corpus of Contemporary American English Academic, Davies 2012), extracted 120 million words for an academic sub-database by using detailed word selection criteria such as ratio, range, dispersion, and discipline measure, a new Academic Vocabulary List (AVL) was created by Gardner & Davies (2014), then the robust frequency and dispersion statistics were used to identify the list, which is available for download from two versions: the lemma and the word-family.

Discourse

Research on discourse mainly refers to the academic discourse built from existing corpora or self-built corpora organized by theses/articles/dissertations or just texts written by subjects such as students, scholars, teachers, etc. Perspectives on academic discourse can be genre analysis, SFL, New Rhetorical studies, etc., or a combinational choice from existing theories. Excepting aspects such as hedges and boosters, chunks or bundles mentioned above, analyses can also focus on indirect expressions, writing results from structure, metaphor on grammar, etc.

What deserves mentioning here is the concept *metadiscourse* and its definition, revolving around proposition (originating from SFL, refers to ideational metadiscourse) and non-proposition (refers to interpersonal and textual metafunctions) in language structure, which has been controversial all along. For Hyland (2005a, p. 37), metadiscourse denotes “self-reflective expressions used to negotiate interactional meanings in a text, assisting the writer (or speaker) to express a viewpoint and engage with readers as members of a particular community.” Hence, it is “a functionally oriented perspective, which sees writers as conducting interaction with their readers” (Hyland & Tse, 2004) with three main principles: 1. it is diverse from “prepositional aspects of discourse”; 2. it refers to “writer-reader interactions in texts”; and 3. it indicates relations that are only “internal to the discourse” (Hyland, 2005a, p. 38) and is classified into two categories (see also p. 48-53); interactive (including transition markers, frame markers, endophoric markers, evidentials, code glosses) and interactional (encompassing hedges, boosters, attitude markers, self-mention, engagement markers). For taking this definition, Hyland (2004a) probed into how second language writers (Hong Kong students) apply metadiscourse resources in theses writing (a corpus of 240 doctoral and master’s dissertations in total four million words), proposed a model of “metadiscourse as the interpersonal resources” to properly “present propositional materials” through the change of disciplines and genre contexts, and revealed the function of metadiscourse in discovering “rhetorical and social diversities” in different disciplinary communities.

However, for Halliday & Matthiessen (2014), if the discourse is endowed with ideational meanings, would not be seen as a metadiscourse. In contrast, the discourse that plays the role of organizing structures and guiding readers in expression can then be regarded as a metadiscourse. Flowerdew (2015, p. 18) disapproved of this view by taking the claim that “while this distinction is helpful in highlighting two dimensions of meaning within metadiscourse, it nevertheless disguises the holistic nature of meaning and the fact that any stretch of text will carry with it simultaneously all three levels” (including ideational, textual and interpersonal metafunctions). Based on this dispute, by investigating a corpus of research articles from two disciplines, Linguistics and Computer Science, Triki (2018) critically explored how the discourse function of *definition* is performed in academic writing. Results suggest that definitional chunks can not only be seen as non-propositional structures in “explaining the meaning of utterances and ideas” but also as propositional ones in “construing meaning” and “building argumentation,” therefore have dual functions of both ideational and metadiscoursal, though the extent may change via “functional reasons” and “disciplinary practices.”

Following research for *metadiscourse* can be the diverse distribution of metadiscourse markers across moves in the discussion section from research articles in the field of applied linguistics by Liu (2018) and the construction of persuasive arguments in essays writing by the application of metadiscourse (interpersonal especially) patterns from first-year university students in Ho (2018) that the suggestion for early implementation of metadiscourse in pedagogical course design in creating convincing arguments is put forward.

Register

Being defined as “a variety associated with a particular situation of use (including particular communicative purposes)” (Biber & Conrad, 2009, p. 6), the concept *register* can often be mixed and then compared with the concept “genre” and “style” that all three are basic functional variations of the language, in which “the situational characteristics of registers are more basic than the linguistic features” in realizing the principle that the application of the language is always in specific situation or context. “Genre” is realized through “register,” the latter of which, however, depends largely on the “style of language.” Discourse is no wonder the product of both “register” and “genre”. In making clear of the denotation and connotation of these three perspectives in discourse analysis, four major characteristics are put forward for comparison: “(1) the *texts* considered for the analysis, (2) the linguistic characteristics considered for the analysis, (3) the distribution of those linguistic characteristics, and (4) the interpretation of linguistic differences.” (ibid, p. 16) Normally, *register* and *style* focus on any

lexicogrammatical feature in a sample of text excerpts to investigate the frequent and persuasive linguistic distribution according to the variety of texts though the interpretation of linguistic features emphasizes more on important communicative functions in register analysis while on aesthetic values in style analysis. Genre analysis, however, differs from these two in all four aspects as it focuses on specialized expression, rhetorical organization, or formatting of complete texts to discover patterns that once occurred in the text or a fixed place, and the interpretation of these features is conventional instead of functional. That is why, in distinguishing register variation from genre variation, Biber & Conrad (ibid, p. 23) assumed that the former looks for persuasive patterns from functionally varied linguistic features across situations but the latter for “the conventional ways in which complete texts of different types are structured.”

Specifically, the description of a register correlates to three aspects: “the situational context, the linguistic features, and the functional relationships between the first two components” (ibid, p. 6), which means that register analysis has to “note the situational characteristics that distinguishes it from other registers, describe the typical (persuasive) linguistic features” (ibid, p. 7) as well as clarifying the relationship within them in functional terms. Besides, registers can be “identified on different levels of specificity” with its analysis depending on representative samples of texts. (See ibid, p. 9-10 for key points of *register*)

For studies in the register of academic writing, the situational characteristics can be constant in a sense, with analyses of linguistic features being various on lexis, grammatical aspects, or pragmatical and contextual variation, resulting in numerous interpretations of the relationship between them functionally. Only to name a few: reporting verbs in medical journal articles (Thomas & Hawes, 1994), personal pronouns in scientific journal articles (Kuo, 1999), collocation in research article introductions (Gledhill, 2000), self-mention in research articles (Hyland, 2001), lexical bundles in disciplinary writing (Cortes, 2004), etc. on account of the choice of specific lexis; imperatives in scholarly writing (Swales et al., 1998), the expression of directives (Hyland, 2002c), extraposed constructions in university student writing (Hewings & Hewings, 2002), etc. as to grammatical structures; and modality in medical research writing Vilha's (1999), question in academic writing (Hyland, 2002), questions and modality in written dating advertisement (Marley, 2002), hedges in medical English written discourse (Salager, 1994), hedges in scientific research articles (Hyland, 1996, 1998; Grabe & Kaplan, 1997; Salager-Meyer & Defives, 1998), nouns or noun patterns Charles (2003, 2007) as well as reporting clauses (Charles, 2006b) in constructing stance in theses from contrasting disciplines, evaluation (in abstracts: Hyland & Tse, 2005; in linguistic book reviews: Römer, 2005; in the art-historical research articles: Tucker, 2003; and in research articles abstracts from the narrative and hard sciences disciplines: Stotesbury, 2003), etc. for contextual variation.

Genre

Though *genre* has been regarded as “a particular type of discourse classification, a classification based in rhetorical practice and consequently open rather than closed and organized around situated actions” by Miller (1984, p. 155), as an abbreviation for the elaborate phrase “genre-specific semantic potential” that signifies “language doing the job appropriate to that class of social happenings when the contextual configuration is a class of situation type” by Halliday & Hasan (1985, p. 108), and later on, as “regularities of staged (because it usually takes more than one step for participants to achieve their goals), goal oriented (because they have evolved to get things done) social processes (because members of a culture interact with each other to achieve them)” by Martin, et al. (1987, p. 117), it is Swales' (1990, p. 58) definition that has been widely accepted for scholars on academic discourse: “a genre comprises a class of communicative events” of which members have the same kind of “communicative purposes” that are confirmed by “the expert members of the parent discourse community and thereby constitute the rationale for the genre.” *Communicative purpose* is a “privileged criterion and one that operates to keep the scope of a genre” within “narrowly focused comparable rhetorical action,” and the rationale forms the “schematic structure of the genre” as well as affecting and limiting the “choice of content and style” of texts. Hence, the genre offers a kind of direction of action to both doers and acceptors and provides people with ways of conducting actions by choosing recognizable languages in interaction.

As has been mentioned earlier, that genre has conventional stability and, at the same time, evolvement and changes (Ramanathan & Kaplan, 2000, p. 180-2; see also Berkenkotter & Huckin, 1995, p. 6), users can make an innovation of the genre, and variations are thus created. The relation within genres is a kind of “intertextuality” that they may “interact with, draw on and respond to others” (Bakhtin, 1986; see Hyland, 2009a, p. 27) by either “manifest intertextuality” (“specific other texts are overtly drawn upon within a text” by quoting, paraphrasing, citing, etc.) or “constitutive intertextuality” / “interdiscursivity” (“a discourse type constituted through a combination of elements of orders of discourse”). (Fairclough, 1992, p. 117-8)

The discordance of the concept *genre* because of “difference understandings of language and its

relationship to social contexts” results in various perspectives for genre analysis under the influence of theories such as SFL, ESP, New Rhetoric, etc. (Hyland, 2009a, p. 26) For these scholars, genre analysis can be conducted from the investigation of either action of individuals who produced specific texts in contexts, or the dispersion of diverse linguistic features in complementarily organizing into texts from changing situations, just as what Bhatia (2022, p. 22) advocated, to “study situated linguistic behavior in institutionalized academic or professional settings.”

Research on academic writing discourse, especially research articles, however, takes a more global perspective in investigating the overall organization of texts to make clear “how one episode of a narrative develops from another and how paragraphs and chapters are built into cohesive wholes” Hutchins (1977) because of constraints from the scientific community for knowledge presenting purposes within researchers. This kind of “macro-structure” (van Dijk, 1972, 1977) analysis has introduced to some famous models: 1. Dogma - Dissonance - Crisis - Search - New Model format (Kinneavy, 1971) from a Rhetorical tradition; 2. processes of current approach - demonstration of inadequacies - statement of the problem - statement of hypotheses & testing of hypotheses - solution: “proof” of hypothesis by Hutchins (1977) for scientific papers based on Kinneavy & Brémond’s cycles (1970); 3. rhetorical divisions of experimental-research papers from Introduction, Procedure (Materials, Methods, Results), and Discussion by Hill et al. (1982, p. 335); 4. macro-structure in medical research report writing: Introduction-Method-Result-Discussion (IMRD) format by Bruce (1983); and 5. a problem-solution structure in Stanley (1984).

However, detailed works for describing the macro nature of texts for novice learners or researchers come from Swales (1981, 1990) and Hopkins & Dudley-Evans (1988), in particular moves or steps from article sections such as the Introduction and the Discussion part. For Dudley-Evans (1986, p. 135), six moves are necessary for the Introduction section in dissertation writing:

Move 1: Introducing the field

Move 2: Introducing the general topic (within the field)

Move 3: Introducing the particular topic (within the general topic)

Move 4: Defining the scope of the particular topic by:

(i) introducing research parameters

(ii) summarizing previous research

Move 5: Preparing for the present research by

(i) indicating a gap in previous research

(ii) indicating a possible extension of previous research

Move 6: Introducing the present research by

(i) stating the aim of the research, or

(ii) describing briefly the work carried out

(iii) justifying the research

Swales (1990, p. 141) provided a more useful model, i.e., the “Create A Research Space (CARS) model” for article introductions from an ecological analogy:

Move 1 Establishing a territory

Step 1 Claiming centrality

and/or

Step 2 Making topic generalization(s)

and/or

Step 3 Reviewing items of previous research

Move 2 Establishing a niche

Step 1A Counter-claiming

or

Step 1B Indicating a gap

or

Step 1C Question-raising

or

Step 1D Continuing a tradition

Move 3 Occupying the niche

Step 1A Outlining purposes

or

Step 1B Announcing present research

Step 2 Announcing principal findings

Step 3 Indicating RA structure

Declining
rhetorical-
effort

Weakening
knowledge
claims

Increasing
explicitness

Figure 1 The “Create A Research Space (CARS) model” put forward by Swales (1990)

Later on, Bunton (2002, p. 74) offered a modification of the former two models, especially the CARS model (see Table 1 Table 1), for writing the introduction section based on a corpus of 45 PhD dissertations from various disciplines written by students in the University of Hong Kong. In establishing a territory, *defining terms* is a must step for engineering, arts, and science disciplines; in establishing a niche, the step *indicating a problem or need*, though is added, resembles that of the 1A, i.e., *indicating a gap in research*; for the *occupying the niche* move, steps such as *work carried out*, *method*, *materials or subjects*, and *findings or results*, though also are offered in the model and appeared in some dissertations, in fact, are not conventional contents for an introduction section, which means that they should rather be arranged in *method* or *discussion* or even the *conclusion* part, the whole model hence is controversial and requires for contemplation for application although a corpus study seems to be persuasive.

Often Present	Occasionally present
Move 1: Establishing a Territory	
STEPS	
1: Claiming centrality	
2: Making topic generalizations and giving background information	Research parameters
3: <i>Defining terms</i> (Eg, A, So)	
4: Reviewing previous research	
Move 2: Establishing a Niche	
STEPS	
1A: Indicating a gap in research	
1B: <i>Indicating a problem or need</i>	
1C: Question-raising (So, A)	Counter-claiming
1D: Continuing a tradition (M, So)	
Move 3: Announcing the Present Research (Occupying the Niche)	
STEPS	
1: Purposes, <i>aims</i> , or <i>objects</i>	<i>Chapter structure</i>
2: Work carried out (Eg, Si)	<i>Research questions/Hypotheses</i>
3: <i>Method</i>	<i>Theoretical position</i> (So)
4: <i>Materials or Subjects</i>	<i>Defining terms</i>
5: Findings or Results	Parameters of research
6: <i>Product of research</i> (Eg) / <i>Model proposed</i> (So)	
7: Significance / Justification	<i>Application of product</i> (Eg)
8: Thesis structure	<i>Evaluation of product</i> (Eg)
NB The moves in this model may not occur in a single progression, but may well be cyclical. For example, the sequence of moves may be: 1-2, 1-2, 1-2-3	
A=Arts, So=Social Sciences, Eg=Engineering, Si=Science, Si=Science, M=Medicine.	
Newly identified steps are in <i>italics</i>	

Table 1 Modified CARS model for Ph.D. thesis Introduction

For the Discussion section, Dudley-Evans (1995, p. 298-299) once proposed nine moves, while Swales (1990, p. 172-173) adopted and simplified it into an 8-move model according to the frequency of occurrence. (See Table 2) To compare Swales from Dudley-Evans, Moves 2 and 3 in the latter are merged into one, i.e., Move 2 in the former, indicating that findings are sure to be presented in stating the results in Swales'. The move *deduction and hypothesis* in the former means to *make a claim about the generalizability of some or all of the reported results* (ibid, p. 173), which includes Move 7 *claim* of Dudley-Evans', while the move *exemplification* in the former (to list examples that are used to support for an explanation) seems to be combined with the *explanation* move in the latter; and Move 8 *limitation* in Dudley-Evans has not been considered about by Swales.

Moves in the Discussion Section	Dudley-Evans (1995)	Swales (1990)
Move 1	Information Move	Background Information
Move 2	Statement of Result	Statement of results
Move 3	<i>Finding</i>	-
Move 4	(Un)expected Outcome	(Un) expected outcome
Move 5	Reference to Previous Research	Reference to previous research for comparison or support
Move 6	Explanation	Explanation
Move 7	<i>Claim</i>	<i>Exemplification</i>
Move 8	<i>Limitation</i>	<i>Deduction and Hypothesis</i>
Move 9	Recommendation	Recommendation

Table 2 The comparison of Dudley-Evans and Swales' moves in the discussion section in articles research (italics indicate difference)

The moves and steps in the introduction and discussion sections by Swales, including Method and Result, are what is usually called the IMRD (Introduction-Method-Results-Discussion) structure for thesis writing, which has been repeatedly applied in academic discourse research from diverse disciplines to explore linguistic features within these rhetorical moves and genres, and the linguistic mechanisms in knowledge construction. For example, Kanoksilapatham (2005) adopted a two-level rhetorical structure, i.e., moves and steps in researching the rhetorical organization of biochemistry research articles, and found that this kind of structure can have "15 distinct moves: three moves for the Introduction section, four for the Methods section, four for the Results section, and four for the Discussion section." In the same genre, Kanoksilapatham (2007) demonstrated the typical discourse organization of moves and steps in each article section from a multi-dimensional perspective. In the genre personal statement for a doctoral application, Chiu (2016) investigated rhetorical patterns of structure of texts written by 21 PhD students at one UK- and US-based university in the context of PhD admissions to compare their perspectives through patterns choosing. The result suggests that rhetorical moves and the construction of a writer's identity through the discourse highly correspond to their awareness of positioning, sensitivity to target readers, and the context for personal statement writing.

More studies based on the IMRD structure focus on rhetorical moves in particular sections from research articles such as the acknowledgment, abstract, introduction, literature reviews, discussion or conclusion section, etc. Regarding the acknowledgment section, except for offering opportunities for writers in gratitude conveying, the scholarly identity of these writers by "displaying their immersion in scholarly networks, their active disciplinary membership, and their observance of the valued academic ideals of modesty, gratitude and appropriate self-effacement" can also be constructed. Hence, in the corpus of 20 MA and 20 PhD dissertations in six disciplines written by students from five Hong Kong universities, Hyland (2004) discussed the move structure of this section and concluded that three main moves are performed in the genre, i.e., the reflecting move, the thanking move (including presenting participants, thanking for academic assistance, for resources and Moral support), and the announcing move (including accepting responsibility and dedicating the thesis). In genre and move analysis of abstracts, Sukan & Mohammadzadeh (2022) built a corpus of MA and PhD abstracts written by Turkish students at Cyprus International University from 2009 to 2019 and then randomly chose 50 abstracts from the ELT department. By taking Hyland's (2000) five-move model for the abstract section, i.e., introduction, purpose, method, product, and conclusion, the research found that 40 did not follow the model, thus bringing constraints for ELT students in meaning conveying and academic communication realization. Similar research can be referred to Sun (2010) in MA theses, Li (2017) in doctoral dissertations from English and Chinese students, Samraj (2005) for disciplinary comparison, etc.

Besides Bunton's (2002) generic moves in his doctoral dissertation for the introduction section mentioned above, Samraj (2002) compared the move structure in two related fields, Wildlife Behavior and Conservation Biology, and discovered that introductions in the latter journal have "a greater promotional function" than the former by using steps like "centrality claims," and justify the concluded research as from real world instead of the epistemic world. At last, a revised CARS model for accounting corresponding features was put forward. Later, Ozturk (2007) probed into variabilities of move structures in 20 articles from two subdisciplines of applied linguistics, second language acquisition, and second language writing research of their introduction section based on Swales's CARS model. Results indicate that the main structure of the former can be traditionally "M1: establishing a territory - M2: establishing a niche - M3: occupying the niche" while that of the latter is M1-M2-M1-M3 or simply M1-M3 type,

explaining the diversities of a “established” field (the former) from an “emerging” field (the latter). Lim (2012) transformed his research field into high-impact management research papers for discussing how the “establish research niches” move is performed by experienced writers in the introductory section. By employing a wide spread of “persuasive communicative resources and linguistic mechanisms” in the “indicating a gap” and “adding to what is known” moves, the research uncovers how the importance of past research is downgraded while the creativity of writers’ works is foregrounded, at the same time presents methods of solving the problems of distinguishing diverse but correlated rhetorical steps under specific circumstances. Another research based on Bunton’s CARS model comes from Kwan (2006) when generic moves in the Literature Review (LR) section from 20 doctoral theses produced by native English students from applied linguistics discipline was examined to distinguish it from the introduction section that though the Move is Introduction-Body-Conclusion in general, the recursive pattern of Move 1-2 recurs the most often while the conclusion move occurs the least frequently. The writer concluded that the literature review section should not be seen as the same as the introduction, although the two are usually mixed up in thesis writing and therefore difficult to distinguish.

The very early research of moves or communicative categories in the discussion section is Holmes (1997) when the attempt to discover disciplinary differences between social and natural ones was conducted on 30 social science research articles (10 each from disciplines of history, political science, and sociology). Though the result suggested that the history discipline is the most distinctive and has the least likeness to those from the natural sciences in all three, the condition of natural sciences has not been demonstrated. Further advocacy for the division of such a subgenre and the development of materials designation for academic reading and writing are raised by the author. Another research is to probe into the challenges in EL1 and EL2 doctoral thesis writing in the discussion section through a digital survey performed on 295 doctoral candidates from their “self-defined learning needs.” (Shen et al., 2019) After questions about factors such as the English language background, discipline, writing stage, research methods, etc., students’ awareness of challenges and genre conventions because of these factors as well as their “high expectations for the specificity and depth of support” are presented.

For the generic structure of the conclusion section, from 45 PhD theses in science and technology disciplines as well as humanities and social sciences (HSS), indexes such as the status and titles of concluding chapters, length, and references, section headings, moves and steps in two disciplinary divisions from thesis-oriented perspective, etc. are comparatively discussed by Bunton (2005). The research discovered that the conclusion section usually plays functions such as “restate purpose, consolidate research space with a varied array of steps, recommend future research and cover practical applications, implications or recommendations, etc.” in emphasizing the content of the thesis though a few of them would focus more on the field in taking a “problem-solution” text structure. Typical research for the results section (based on Brett, 1994 for corpus selection) comes from Lim (2010) when a mixed-method genre-based perspective is adopted to quantitatively and qualitatively probe into disciplinary and methodological diversities on choosing certain comments (particularly comparison of present and past research findings) and various categories of these comments according to linguistic mechanisms in 30 writing research reports written by university students (15 each in educational and applied linguistics disciplines). The research does suggest disciplinary diversities and methodological similarities. Among four commentary steps: explaining the finding/s, evaluating the findings, comparing findings with literature, and making recommendations for future research, the former three are common in the applied linguistics discipline. More than 80% of the comments are found in ALs and 80% of which include at least one kind; while comments are rarely discovered in educational discipline, though reasonable explanations have not been provided by the writer.

Hence, the importance of genre analysis for academic discourse is self-evident. Though Swales (2004) considered genre research as discourse-oriented that the analysis should not follow fixed templates but comprehend constantly changing essential characteristics of the genre in regarding relevant analysis as a heuristic tool, Bhatia (2008, p. 176) advocated a series of other perspectives and dimensions such as the ethnographic perspective to conduct research on “‘critical sites of engagement’ through ‘lived narratives’ (see Scollon, 1998) of expert and experienced professionals” and the socio-cognitive perspective to interpret “behavior of specialist and non-specialist audiences” in critically understanding how a specific group of generic resources is creatively applied in changing “the socially shared generic norms” for the achievement of “private” regarded corporate inclinations.

Corpus

Being the study of language used in real life, corpus analysis is normally chosen for explaining the usage of language in particular domains and genres as it focuses on large databases of electronically annotated texts rather than only a small number of them and describes usual but not grammatically potentially characters. In describing how the genre is conventionally constructed by its users, the frequency of specific words or strings, as well as grammatical patterns that are frequently applied in this genre or

sub-set of language, can be of significance, and qualitative analyses of the correlating methods of features in collocational patterns by concordance data being adopted as a supplementary, patterns in the surrounding co-text for providing information to the use of target words can be uncovered. (see Hyland, 2009a, p. 28-9) Except for O’Keeffe et al. (2007, p. 200-3) in researching and discovering pedagogical implications of chunks, idioms, grammatical and lexical patterns, clauses, response tokens, relational language, etc. in written academic discourse from the corpus perspective, Thurston & Candlin (1997) presented rhetorical functions of regularly used words in academic writing to teach college-level writing skills to students through concordancing, and Starfield (2004, p. 154) clarified how second language PhD students “strategically engaging with the resources of authoritative English” by taking the method of concordancing to develop their writing skills and empower themselves.

Multimodal

The multimodal perspective requires researchers to notice the particular ways that diverse modes provide for in configuring the world, and the accompanying changes “in authority, in forms of meaning and in forms of human interaction with the social and natural world” (Kress, 2003; Kress & Van Leeuwen, 2002), indicating that multimodal analyses ask for the description of “potentials and limitations” in meaning-making via diverse modes. For Kress (2003, p. 1), “The two modes of writing and of image are each governed by distinct logics, and have distinctly different affordances.” The former is temporally governed in logic as it is organized by a sequence of elements arranged in time like a speech; while the latter is spatially governed that it is structured by simultaneously arranged visual or describing elements in a place, the meaning in writing is thus sequentially conveyed while in the picture is positionally expressed. Therefore, graphics or pictures in this perspective are usually more focused on by researchers as they not only illustrate or supply information but also bear “the informative and persuasive weight of an argument” normally. (see Hyland, 2009a, p. 30 - 31)

In academic writing discourse, the “multimodal” seems to be contradictory as the mode is fixed in text writing, while pictures, tables, graphs, etc., are also applied in genres such as researching articles, Master’s theses, and doctoral dissertations, essays, etc., the arrangements of how these materials are organized and what kind of function each structural type has can be discussed.

Contextual

When a discourse is seen as language in action in a particular context, factors from social, cultural, and historical background for producing the discourse are useful for consultation and comprehension.

Sociology of science

According to Hyland (2009a, p. 32-33), the “strong social constructionist position” and research methods such as “ethnography, participant observation, conversation analysis,” etc., have led to sociologists’ consistent emphasis on the “context” in creating knowledge that “the social structure of science” constructs our comprehension of regarding “scientific writing as a social act” being accepted by “the academic community” in which scientists share the same knowledge, “methods and rules, and define what a valid research really is” (Lakatos, 1978). Taking this perspective means that the research conducted by researchers must have a connection to previous works and even contribute to the whole deemed knowledge, the process of which is often decided by the scientific community that a claim or a viewpoint would not be called knowledge until it is “adequately tested” and agreed by most of the scientists in the field (who have been specially trained and have authority upon the verdict). Hence, Polanyi (1964, p. 13) considered that the action of research would rather be taken as a quest for agreement than a search for truth in a sense. Studies of academic discourse in this perspective can uncover influences of institutional and social contexts on academic writing materials, demonstrate the process of writing articles in describing the reformulation of research activity from a suitable but controversial “disciplinary ideology,” and suggest scientific writers’ purposes of constructing a text with knowledge claims being unquestionably “related to observed data,” to present “a disinterested, inductive, democratic and goal-directed activity” while at the same time “concealing contingent factors, downplaying the role of social allegiance, self-interest, power and editorial bias.”

Ethnography

Originating in anthropology and sociology, ethnography provides a participant or insider with an “oriented description of individuals’ practices or behaviors” by collecting “naturally occurring data under normal conditions from numerous sources” along a period. In seeing the language application from a broader scope, i.e., “the physical and experiential contexts,” researchers of this perspective are required to “locate acts of communication in the behavior of groups,” adopt a variety of “watching and asking” methods (including “participant and non-participant observation, in-depth interviews, surveys, focus group discussions, diaries and biographical histories,” etc.) and data materials, conduct research for quite a long time, reconsider the results by participant confirmation and “member checking,” and present an informative as well as fresh interpretation on locally contextualized interaction with individuals. (see Hyland, 2009a, p. 36) Hence, Prior’s (1998) work demonstrated the mutual influences of academic

practices, peers, mentors, students' personal experiences and changing goals, etc., on students' writing activities and their processes of becoming academic writers, writing is proven to be the product of mutual decision by other people and stuff, classroom tasks and speech genres, discourses and disciplinary practices, awareness, and cognition, etc.

Sociohistorical

As for constructionists, the truth of stuff is limited by a specific group of people at a fixed time, the research for the historical circumstance that produces academic discourse and keeps "relatively stable in particular time" is also affected. Numerous studies have revealed that "writing conventions" have been intentionally developed via time and social situational changes, such as the de-emphasis of methods while tremendous "expansion of theoretical discussions" (Atkinson, 1999; Bazerman, 1988); variabilities because of "the standardization of experimental procedures" and the greater necessity of contextualizing "work in discipline-recognized problems"; increasing citations of theory and references due to the common theory's "strong force in structuring articles and binding articles to each other" (Bazerman, 1988, p. 157); more attentions paid to visuals and graphics with its incremental integration into arguments, vastly multiple authorship of writers in journal articles, simpler syntax while overuse of nominalization, more distinct division of sections in typography and stylistics; the development of "titles, abstracts, sub-heads and graphics" to "foreground novelty and significance" in satisfying "the scanning reading patterns of information-saturated readers" efficiently looking for correlation and creativity (Berkenkotter & Huckin, 1995), etc. These studies from the sociohistorical perspective suggest that methods taken for presenting arguments, listing procedures, citing pieces of literature, probing into data results, engaging with readers, etc., are only persuasive or negotiating strategies in writing when social environments are considered; demonstrate the significance to us in considering about cultural practices from wider social contexts; and greatly contribute to our comprehension of academic discourse as well as its research. (see Hyland, 2009a, p. 34 - 35)

Cultural

The discussion about cultural influences on academic writing discourse for English as a second or foreign language students concentrates on "Contrastive Rhetoric" when Kaplan observed ESL students' paragraph writing and provocatively asked for attention to cultural diversities in their writing by putting forward a model of concerns in contrastive rhetoric (Kaplan, 2005, p. 384). Defined as "an area of research in second language acquisition that identifies problems in composition encountered by second language writers, and by referring to the rhetorical strategies of the first language, attempts to explain them" (Connor, 1996, p. 5), and is taken as an interchangeable term with "cross-cultural or inter-cultural rhetoric" by Enkvist to denote the "study of patterns of text and discourse in different languages that vary in structural and in cultural background" (Enkvist, 1997, p. 188), Connor (2011, p. 1) would prefer to the term "Intercultural Rhetoric" in which "intercultural" denotes to "collaborative interaction between and among cultures and individuals" as well as those within cultures, and the whole term refers to "the study of written discourse between and among individuals with different cultural backgrounds" with "discourse" means "language use beyond the sentence as well as social and ideological assumptions" (ibid, p. 2) during the activity of communication. The term "rhetoric" then resembles Aristotle's trichotomy of "rhetoric" as "invention, style, and arrangement," with three proving directions: "logos (the use of reasoning), ethos (the credibility of the speaker/writer), and pathos (the use of emotional appeals)" (ibid, p. 5), and its definition should better be based on "situation" and on the "concern about how rhetorical situation guides production" (Sullivan & Porter, 1997, p. 25). For scholars approving of this approach, the research of writing should not be limited to lexicogrammatical features or patterns in texts on the surface but take into consideration the social contexts and practices that the text resides in; the potential culture for macroscopically analyzing discourse should include not only small or disciplinary cultures but also as complex as the large or national/ethnic cultures; and the "dynamic, interactive patterns of communication" among interlocutors as well as their "negotiation and accommodation" (Connor, 2011, p. 2), which result in convergences of cultural diversities, should all be noticed. In solving rhetorical structural problems in academic writing for English as a second language students, besides investigating the effects of first language, culture, and education on the creation of texts, the cross-cultural research, i.e., the comparison of one stuff in two cultures and studies of interaction and negotiation by writers from various linguistic, cultural, educational and social backgrounds are also welcomed by scholars. Although be divided into subfields such as comparative discourse linguistics, writing as cultural activity studies, classroom-based writing research, and special genre comparative research, etc., the cornerstone that rhetorical preference would be based on the diversities of language and cultures has never been denied, and the preference normally does not judge by goods or bads, all of which undoubtedly and unexpectedly have effects on English as second language writing.

In comparing rhetorical differences within texts written by academics from diverse cultural backgrounds, Mauranen (1993) had once explored metatext use in papers from economics journals

written by Finnish and Anglo-American academics and found that Anglo-American writers were more reader-oriented in guiding them, presented a more positive tendency for politeness and quite explicit textual rhetoric in revealing writers themselves; while Finnish writers demonstrated a more negative tendency of politeness and implicit rhetoric in writing. Though both rhetorical strategies are legitimate, reasonable, polite in the corresponding culture, and can be persuasive in guiding readers, they would be unintentionally inefficient when the target readers share diverse attitudes and assumptions from the writer and a certain discontinuing and illogical impression from readers with different cultural background then be caused. Hence, research from the contrastive rhetorical perspective can help ESL/EFL writers be aware of cultural differentiations in case of negative transfer of rhetoric from their mother tongue in second language writing.

Critical (CDA)

In taking “critical” as an attitude or a way of considering about and dealing with discourse, critical discourse analysis speculates “language as a form of social practice” (Fairclough, 1989, p. 20) and intends to uncover or reveal those fundamental but have long been regarded as “common, acceptable and natural features of discourse” and then overlooked, or in Fairclough’s (1992, p. 135) words, the aims of CDA are to “systematically explore often opaque relationships of causality and determination between (a) discursive practices, events and texts, and (b) wider social and cultural structures, relations and processes; to investigate how such practices, events and texts arise out of and are ideologically shaped by relations of power and struggles over power; and to explore how the opacity of these relationships between discourse and society is itself a factor securing power and Hegemony.” Therefore, CDA connects language with the activities happening in its context, investigates how factors such as “social relations, identity, knowledge, power,” etc. are constructed in texts in diverse situations as well as constraining the discourse, and discusses the further construction on “social and political realities” from the discourse. Belonging to “neither a qualitative nor quantitative research method,” CDA is rather a “questioning” to the main assumptions of all these methods. And the biggest difference from other discourse analysis perspectives lies in its great emphasis on power relations and “ideologies in maintaining these relations” in social context though all of them devote to contextually situated texts. (see Hyland, 2009a, p. 38-39)

Normally, particular features of discourse such as vocabulary, transitivity, nominalization and passivization, mood and modality, theme, text structure, intertextuality and interdiscursivity (using generic conventions, registers, or styles from other texts), recontextualization (the colonization of elements from discourses in one context on another), etc. are what CDA pay attention to in practice. (ibid: 40) Besides largely conducting research in mass media and public discourses for topics like class, genre, racism, power, etc., research in academic discourse has sprung up (Rogers & Schaenen, 2014) such as writing tasks assigned by teachers in the classroom when students are reluctant to finish them that the autonomy of writing from students are restrained by teachers’ requirement through performing their power. It is from teachers in this situation that claims are deemed as knowledge, texts are judged by teachers as good or bad, textual practices are performed, and social relations are gradually legitimated.

Drawing on SFL, CDA also focuses on choices of specific linguistic features in discourse, though explanations from situations under unequal power are often provided for these choices. (Fairclough, 1992, 2003; Wodak, 1989) They mutually look upon “language as a social construct” to discuss the reciprocal relation within language and society, then search for mutual influence of “particular discursive events” and its contexts by holding “a dialectical view of language” (Young & Harrison, 2004, p. 1) and follow trends of “cultural and historical aspects” on meaning conveying, (Graham, 2004, p. 56/64) but CDA seldom locates power relations in realizing certain linguistic features or patterns in text analysis and fails to consult participants’ opinions in particular situations in discourse analyzing and interpreting. Hence, many studies tend to float on the surface in revealing power relations, just as Slembrouck (2001, p. 42) called, to “appeal to explanatory social theory” rather than provide an “absolute” or a “pure” explanation (Blommaert, 2005, p. 32) of ideology for texts.

2.1.4.2 Academic Literacy

Mainly adopts an emic perspective and uses research methods such as ethnography, the academic literacy theoretical paradigm focuses on language users to examine how they participate in the practical activities of academic English, including learners’ different feelings, expectations, and understandings of writing tasks; how they learn to think, behave, judge and communicate in academic writing contexts; and how new methods of teaching writing can be designed for those of necessity. (Coffin & Donohue, 2012) Primarily, three paths are included for AL research in academic writing: cognitive with identity construction for writers, publishing problems writers may encounter, and pedagogical methods or designs for teaching students academic writing-related knowledge and literacies.

Cognitive

The cognitive path for academic writing research can cover diachronically case studies of the writing

processes, especially tracking characteristics of the development from individual students, exploration of writing processes of doctoral students from the social cognitive aspect, queries of academic writing strategies of learners by using think-aloud and texts analysis research methods, etc. The first of which was presented by Cheng (2006, 2008) when the development processes of being aware of the academic criticism and generic structures as well as rhetorical literacy in texts from second language learners in academic writing are revealed by tracking the academic reading and writing processes of a Chinese master's student. Research on the cognitive characteristics of writing, in addition, demonstrated the cognition of the importance, difficulties, and complexity of academic writing from both teachers and students and their attitudes toward plagiarism and academic citations in texts.

As language can be used to express psychological activities, in which hedging, being a rhetorical device, can be applied in "projecting honesty, modesty and proper caution in self-reports" as well as "diplomatically creating research space in areas heavily populated by other researchers," (Swales, 1990, p. 175) it is a phenomenon not only interactively resulted from authors, readers, and discourse, but also produced by language forms in reflecting authors' psychological orientation as a cognitive modality. Later on, Ding (2002) considered that academic writing of articles is heavily influenced by writers' subjectivity, and the application of hedging strategy in texts enabled better communicative achievement. Another linguistic feature that can signify psychological activities from writers is projection in transferring or repeating locution or viewpoints from other people or one's own. Hence, Zeng (2007) probed into the clause structure and semantic features of projection in detail from interpersonal, ideational, and textual metafunctions based on grammatical metaphor in SFL. Being the product of psychological cognition, Tang & Fan (2008) have analyzed the impact of thought from the mother tongue on English academic writing papers at three levels: vocabulary, sentence structure, and discourse. They considered that interaction was significant in achieving the ultimate goal of the interpersonal meaning through academic writing, i.e., to gain academic cognition from other scholars, the thought from students' mother tongue may have a negative transfer on their expression, then may pull them far away from this goal. As the psychological health of doctoral students is of great significance during their process of writing, data from semi-structured interviews with 13 EAL doctoral students in an Australian university suggest that individual (or one-on-one) consultations from advisors who can identify their writing problems and providing personalized advice indeed have pedagogical and psychological impacts on them as well as their writing processes. (Ma, 2017) But this service may not sweep the world, leaving the consultation more valuable for doctoral experience.

Identity Construction

For academic writing of diverse genres, no matter theses, dissertations, journal articles, argumentative essays, or even classroom tasks, a key point that is relevant to students' cognitive literacy is their identity construction through the process of writing that for those whose mother tongue is English, the necessity of mastering corresponding academic rhetorical structures in genres or disciplines for their entering into the academic community and constructing their identity as scholars in their fields in the future is self-evident, and for ESL/EFL students, besides seeking for the identity as academic researchers, they still need to have acquaintance with thinking modes or even ideologies of English as well as abandoning their own, and endeavor to avoid language transfer or negative impacts of thinking methods, values, conventions on their second language expression from the mother tongue. Generally, students are not aware of the impacts that this concept has on their writing, nor do they realize the difficulties from deeper social, cultural, historical, or ideological aspects that they would face in identity construction. Just as Cadman (1997) indicated, different epistemologies in which international postgraduates have been trained and on which their identities as learners are based directly bring challenges to them in using English for argument texts writing and then for identity transformation from students to scholars.

The discursal construction of identity in academic writing was intensively discussed by Ivanic (1998, p. 24), in which three ways of focusing on the identity of a writer in text writing, i.e., the writer's "autobiographical self," the "discursal self" which the writer constructs in the act of writing, and the "self as author" (referring to a writer's relative authoritativeness) was presented, and then the discursal identity was primarily investigated through research methods such as text analysis, case study, interviews, observations, etc. For the writer, "these three 'selves' are all socially constructed and socially constructing" because "they are shaped and shape the more abstract 'possibilities for self-hood' that located in the writer's "socio-cultural context," and "writing is an act of identity in which people align themselves with socio-culturally shaped possibilities for self-hood," performing themselves in "reproducing or challenging dominant practices and discourses," in which "values, beliefs and interests" are demonstrated. (ibid, p. 32) Hence, whether the texts written by students meet traditional requirements or standards or not is not decided by their literacy but by identity. As writing is highly influenced by social-cultural aspects in expressing the power relationship between writers and readers that the established first language identity, value system, and thinking modes cannot completely be consistent

with the rhetorical requirements from the second language, students need to constantly adjust the relationship of their concepts and the mainstream disciplinary concepts during the process of writing, making academic writing a dynamic identity constructing activity. This means that the direct effects of bilingual, multicultural identity, as well as ideological diversities on L2 writers, prompt them to construct a new identity through writing, in which various languages, activities, ways of thinking, values, beliefs, etc., are integrated, kinds of symbols, tools, and objects are jointly applied in achieving special social identity to gain recognition from target language readers. From the experiences of two EAL doctoral candidates, de Magalhães et al. (2018) found that self-decisions in chapter writing, interactions with peers and supervisors, network with scholars beyond institutions and methods searching to overcome resource constraints can boost their confidence, increase their agentive behavior, encourage them to express voices, create their networking opportunities and then support their development of identity. Though interviewers' (supervisors') "language background, disciplinary specialties, personal histories, and preferences shaped impressions of voice" and bring challenges to "define and locate voice" in EAL PhD candidates' texts as well as to discover progress in their abilities to present "a convincing authorial voice," Morton & Storch (2019) did provide a new angle, i.e., from the readers' perceptions in identifying the importance of researching and assessing voices in students' writing.

In self-presentation and specific personas projection in academic writing, one significant rhetorical strategy that enables writers to demonstrate diverse discourse functions in texts and construct a convincing argument in persuading readers of the validity and creativity of their claims, as well as their abilities is first person pronouns that through an examination of essays from 27 first-year undergraduates at the National University of Singapore, Tang & John (1999) claimed that it is "not a homogeneous entity" but represents six different identities, i.e., *I* "as the representative, as the guide through the essay, as the architect of the essay, as the recounter of the research process, as the opinion-holder, and as the originator." The awareness to empower students so they are no longer manipulated by conventions is advocated hereby. The underuse of authorial reference has also been presented in Hyland (2002b) when the choice of personal pronouns in 64 Hong Kong undergraduate theses is compared with that in other research articles, and interviews with students and supervisors are conducted to explore the concept of identity. Although writers' credibility can be improved by projecting their identity with individual authority, and confidence can be gained by evaluating and presenting attitudes towards other scholars' ideas, these students still avoid using personal pronouns in emphasizing contributions from themselves. Yang (2015) held the same opinion that Chinese students usually have a relatively "implicit" identity with authority deficiency when the feature of self-referencing from MA theses in the linguistics field is compared to that of North American students.

As to the first-person plural pronoun *we*, the research of Spanish EFL Engineering students' 55 multi-authored reports from Luzón (2009) suggested that the failure to make clear of using pronouns as well as conventional corresponding phraseological patterns in constructing authorial identities may be due to their unawareness in academic writing genres; and the comparison of English academic journal articles written by Chinese science and engineering doctoral students and international scholars in position expressing from Wang & Jiang (2019) revealed the extensive and overuse of "*we*" by the former group in avoiding self-presentation.

By exploring rhetorical positioning in the academic writing of Italian English, Carla (2011) probed into the choice and function of the first-person pronoun *I*, *we*, *me*, *us*, and respective determiners, claiming that the difficulty for them comes from the effects of their genre and cultural conventions. Later, Wang & Xu (2019) confirmed the influences of social-cultural aspects such as social expectations, teaching orientations, institutional regulation, etc. and individual cognitive aspects like identity positioning, learning experience, learning beliefs, etc., on the choice of first-person pronouns in constructing academic identity through a longitudinal research of Chinese English learners at the undergraduate, master and doctoral stages by the interlanguage comparative analysis, and the relatively obscure identity constructed by these students.

Unlike works before, Martínez (2005) explored the use of first-person pronouns in diverse sections of research articles from the field of biology. By comparing those from native English-speaking (NES) writers and research article manuscripts from non-native English-speaking (NNES) writers, he found that the preference by the former group in the Results section indicates their responsibility for methodological forces on results while the under-use, overuse, and phraseological problems by the latter group suggest students' unawareness of the rhetorical strategy.

Except for first-person pronouns, lexical bundles and citations are also of great significance in identity construction: the underuse of constructions with life subject and content words of Chinese students' English academic writing compared to those from the journal article corpora by Xu (2011) predicts the ability discrepancy of handling the relationship within confidence and cautious in identity expressing; a comparative analysis of using citations in English and Chinese scholars by Ju (2016)

indicates a more confident and belonging image of the former while a more objective and alienated one of the latter in writing; and a mixed approach of textual analysis and interviews of citation practices from Chinese EFL learners' MA thesis writing by Liu & Deng (2019) suggests the function of linguistic, academic, disciplinary and novice identity building in choosing citations that suitable number and relevant references represent "knowledgeable and credible" selves while ineffective ones signify their inability of constructing an authorial identity and leave an uncredible impression on readers.

In academic writing, writers sometimes assume the existence of readers in creating texts that the way they express ideas and make claims reflect cautious negotiation with peers or fellow scholars, while readers would "recognize and expect" or "draw on assumptions" about what the writer intends to convey. This kind of interaction concerns readers' engagement when "writers relate to their readers concerning the positions advanced in the text." As a rhetorical mode, it is "an alignment dimension where writers acknowledge and connect to others, recognizing the presence of their readers, pulling them along with their argument, focusing their attention, acknowledging their uncertainties, including them as discourse participants, and guiding them to interpretations." (Hyland, 2005b, p. 176) Hence, academic writing involves the process of persuasion that writers' endeavor to convince readers "of the knowledge claim at the heart of a research article or dissertation; of an evaluation of others' work in a book review, or of one's understanding and intellectual autonomy in an undergraduate essay," (ibid, p. 177) making the process also an identity constructing one. Beside discussing about "engagement" from five main categories: reader pronouns, personal asides, appeals to shared knowledge, directives, and questions in disciplinary corpus with contrast, Hyland & Jiang (2016) also confirmed the continuous changes of academic discourse conventions in choosing engagement resources by drawing on a 2.2 million words corpus collected from top five journals in four major disciplines at three time periods. Reasons are provided for "substantial drops in asides, knowledge reference and reader mention" in applied linguistics and sociology while "a huge rise in directives" in electrical engineering. Based on the same model, Jiang (2018) concentrated on engagement features in the genre of research reports for PhD candidature confirmation and compared results with those from journal articles. The lower occurrence of especially questions and imperatives in the former group suggested an undermining of confidence and abilities to persuade readers to engage when the research genre is endowed with assessment purposes, and the difficulty in building an inclusive relationship between readers and writers because of the complexity of engagement rhetorical features.

The influence of disciplines on constructing identity in academic writing resides in the unique discourse standards and conventions in each discipline that the field, mode, and tenor would all constrain the choice of discourse contents and expression methods from members of the academic discourse community. (Hyland, 2012) Except for three main paths for academic identity research: discourse analysis, critical discourse analysis, and narrative analysis, the combination of corpus which provides researchers with frequency tables, vocabulary indexes, and keywords can reveal disciplinary tendencies of ontological characteristics, ideologies, knowledge structures, rhetorical paradigms, etc. in identity construction. For legal research articles for example, from 18 English and Chinese articles exploring IC features based on engagement system in Appraisal Theory by Jiang & Yang (2018), the common characteristics, i.e., "multiplicity, imbalance and context-dependence" of these resources help construct four kinds of identities for writers: "researcher, arguer, evaluator and information provider," but socio-cultural diversities in "personal values, interpersonal relationships, risk awareness and cultural ideology" cause to English writers' "direct and open" attitudes with the inclination to creating a "flexible dialogue space with readers" while Chinese writers' "prudent and conservative" performance in leaving "relatively less room" for readers.

Because "successful academic writing depends on writers' projections of a shared professional context as they seek to embed their writing in a particular social world which they reflect and conjure up through approved discourses" (Hyland, 2005b, p. 178), when relating identity construction of academic writing to international publication, Xu (2017) advocated the transfer from introducing and verifying knowledge of Western academic research to co-constructing a critical environment in the field or discipline; from monolingual writing to bilingual writing and publishing for capacities; and from critical review to collaborative peer review when articles are judged for publication. For personal statements in applying for graduate study, Chiu (2016) discovered that the sense of the writer's positioning, the sensitivity to the target audience, and the context for writing this genre mutually influence their choice of rhetorical patterns of structure in presenting and positioning themselves to construct an applicant's identity. And for the role of writing groups in assisting students' identity construction, Mochizuki & Starfield (2021) considered that writing conferences can "facilitate their inter-psychological and intra-psychological processes" to make them have an acquaintance of "the power and social structure, value systems and ideologies" in scholarly writing, revising through the exchange of oral feedback provides them "multi-voiced activity" in developing their voice construction.

Publishing

Another aspect of academic literacy of writing comes from the publishing problems ESL/EFL writers or authors encounter in especially international major journal articles, which are usually relevant to their political, economic, psychological, linguistic, cognitive, etc. background and problems can be what kind of difficulties and complexities do they face in international publishing, what are the main constraints for their publishing, i.e., from editors or their textual deficiency, what processes would they go through in successfully constructing their publishing manuscripts, what problems of academic writing literacy do these authors have, can they be aware of the real reason for acceptance or rejection from the editorial board, what strategies can they take to solve corresponding problems, etc.

Targeting discursive and non-discursive problems that researchers in periphery countries (against center countries) face, Salager-Meyer (2008) presented a proposal to help these scientists be fully integrated into the worldwide network and facilitate their contribution to the “promotion of scientific multilingualism.” Then, in revealing how “the overall anglicization of academic publishing” is negotiated with local scholars, “in-depth semi-structured interviews” were conducted with 20 academics across some humanities and social sciences disciplines at a top-ranking research-based university in China by Flowerdew & Li (2009) to investigate “how a potentially competing relationship between English and Chinese” might be demonstrated in China for scholars’ publishing language. Based on the extent and methods that English is adopted and attitudes toward the official requirement of English publishing from these participants, the study pointed out the dominant position of Chinese in publication as well as the increasing status of English in their international journal publishing. In facing the disadvantage that EAL (English as an additional language) writers have in international publishing, the viewpoint of the mutual endeavor from EAL writers and English L1 writers in improving the understanding and bearing of the awkward situation that EAL writers experienced based on Goffman’s rationale does not work in some situations, Flowerdew (2008) therefore, asked for the intelligibility instead of the conformity to “standard English” be regarded as a dominant criterion in judging the acceptability of articles written by EAL authors, and a building of “large corpora of EAL-authored disciplinary texts” in assisting the identification of the acceptability on intelligibility. For the pressure of doctoral students in publishing during their candidature, Lei & Hu (2019) discovered that some mediating resources such as “cultural artifacts and social others in facilitating publishing efforts as well as student’s socialization into the academic culture and community” would be taken in relieving tension after a multiple-case study of six nursing doctoral students and one supervisor from a main research-inclined university in China was conducted. Wilkins et al. (2021) concentrated on experienced doctoral educators of their attitudes and practices toward publishing problems of doctoral students in the fields of business and management. Through an online questionnaire survey of 54 professors or scholars worldwide, the authors found an agreement from these participants on the necessity for doctoral students to publish before graduation and ended with recommendations for students, institutions, and supervisors/advisors to mutually take responsibility to endow students with the publishing abilities to become successful academic scholars in the future. The choice of which language to publish by ESL/EFL scholars has also been discussed by Duszak & Lewkowicz (2008) when attitudes toward and experiences of publishing from Polish academics in fields of “applied linguistics and foreign language studies, psychology and medicine” were referred to. Though advantages and difficulties of choosing any language can both exist, which made the decision a complex question and the tension between two languages would last for a long time as Polish is straightforward for local academics while English facilitates them to access the international research community, the refusal of English publishing may suggest a resistance for its language dominance. In revealing how some “off network scholars” find their methods of publishing in “mainstream Anglophone journals” by continuously negotiating with journal gatekeeping reviewers as well as keeping others (both off-networked and networked) from success, the compilation and analysis of “a sample of submission history documents” given to applied linguistics journal by Belcher (2007) suggested that authorial persistence or willingness in continuous “revising and resubmitting” when confronting with “extensive critical commentary from reviewers” can promote the publishing efficiency.

Pedagogical

The third aspect of academic literacy of writing is pedagogical studies to explore and then find more effective teaching models of academic writing suitable for students and build up a tight connection among theories, practices, and research. Problems urgently seeking answers are whether a new pedagogic method of teaching strategy is feasible and effective; what kinds of theories and methods are adopted in writing course design; how can the teaching project of EAP academic writing be comprehensively evaluated; how can teaching materials and measures including corpora be developed, applied and evaluated; whether supervisors are endowed with expertise in guiding students’ theses writing, etc., hence research in this direction is mainly threefold: teaching strategy for academic writing,

the development and application of teaching materials and methods, and the curriculum/course or project design with its evaluation for teaching writing.

In which studies from the teaching strategy are mostly conducted, including writing elements from the EAP courses, discussions on teaching strategies for writing with their practicing effects, and strategies taken during the writing process or in each structural section, etc. For example, in helping NNS graduate students successfully enter their discourse communities, “disciplinary portfolios” are advocated by Hirvela (1997) to collect corresponding articles, put them in the portfolio, read all of them, search for writing processes and research results in each article, and then take notes of the retrospection on linguistics features of this discipline. A similar strategy was performed by Charles (2012) when the discipline-specific corpora-building method was “trialed in multidisciplinary classes” of graduates after a six-weekly 2-hour sessions experiment. The initial and final questionnaires from 50 participants indicate that these students can build a corpus including 10-15 research articles, and most of them are confirmative of the effective effects of corpora building on their thesis writing. Regarding the development and application of teaching materials and methods, besides corpora building and the utilization of corpora software in aiding academic writing teaching, other software such as the anti-plagiarism services from Turnitin, can significantly decrease the plagiarism rates when it is introduced to students in raising their awareness of the importance of originality in writing. (Stapleton, 2012) In the ESP course designed for EFL students in intensively teaching their thesis/dissertation writing, based on the target learning needs and the interactive needs analysis of participants, and by discussing the diversity within text-based and total discourse-based courses, Richards (1988) constructed a one-month syllable and through the evaluation from students on their pinning on organization and outlining, revising, prerequisite organizing and editing skills, advocated the necessity of equipping students with rhetorical structures in specific genres in the course design. Another program for ESL graduate students referring to writing thesis proposals is conducted by Paltridge (1997) when the detailed process is offered in the course, such as writing preparation, structure establishing, details inclusion, purpose clarification, relevant background literature reviews, thesis expectation explanation, etc. Also aimed at ESL students of their difficulties in presenting acceptable standard dissertations, a program based on the diagnostic assessment profile was described and evaluated by Allison et al. (1998) in the English Centre at the University of Hong Kong. Through interviews with graduate supervisors and a survey of graduate students, shortcomings of the students writing, including untransparent research purposes, failure to substantiate arguments with evidence, strong or overgeneralized claims, the inability to organize paragraphs and texts, typographic or grammatical problems, etc. were demonstrated.

In general, academic literacy sees the context of culture as literacy practices and the context of situation as literacy events to focus on writers and writing in context, i.e., what socially situated writers do by the main method of ethnography, while SFL sees the context of culture as genres and the context of situation as registers (including field, tenor and mode) to focus on written texts, i.e., how linguistic resources construe meanings in context mainly by linguistic analysis from the experiential, interpersonal and textual systems. (See Gardner, 2012, p. 54- 55)

2.1.5 Related Research of Academic Writing

Some peripheral research of academic writing can include purposes and motives, modes or models, and processes in which strategies are used, especially media chosen for writing.

Writing Purposes and Motives

Based on the iterative dynamics’ theory, the engagement theory, and empirical data, Murray & Moore (2006, p. 179) put forward a structural model for academic writing to explain the motives and purposes during this process, which includes mainly nine factors: writing triggers or blockers, writing facilitators, writing prohibitors, engaging in efforts to write, environmental moderators, individual moderators, achieving effective writing outcomes, intrinsic rewards, and extrinsic rewards. As they emphasized the function of dynamics, the process of writing starts from triggers or blockers, then influences facilitators or prohibitors respectively, and these two factors mutually work on “engaging in efforts to write,” which functions on “achieving effective writing outcomes” together with environmental and individual moderators, then intrinsic and extrinsic rewards of writing are gained. (see Figure 2) Hence, the model reflects social-psychological processes during academic writing. And except for moderators and writing outcomes, all others belong to the dynamics category. The system is also a recycle but not a recursive one when triggers or blockers are regarded as the starter in influencing efforts to write and then on outcomes with kinds of moderators, at last, bring to intrinsic or extrinsic rewards, which become motives or stimulates for further writing.

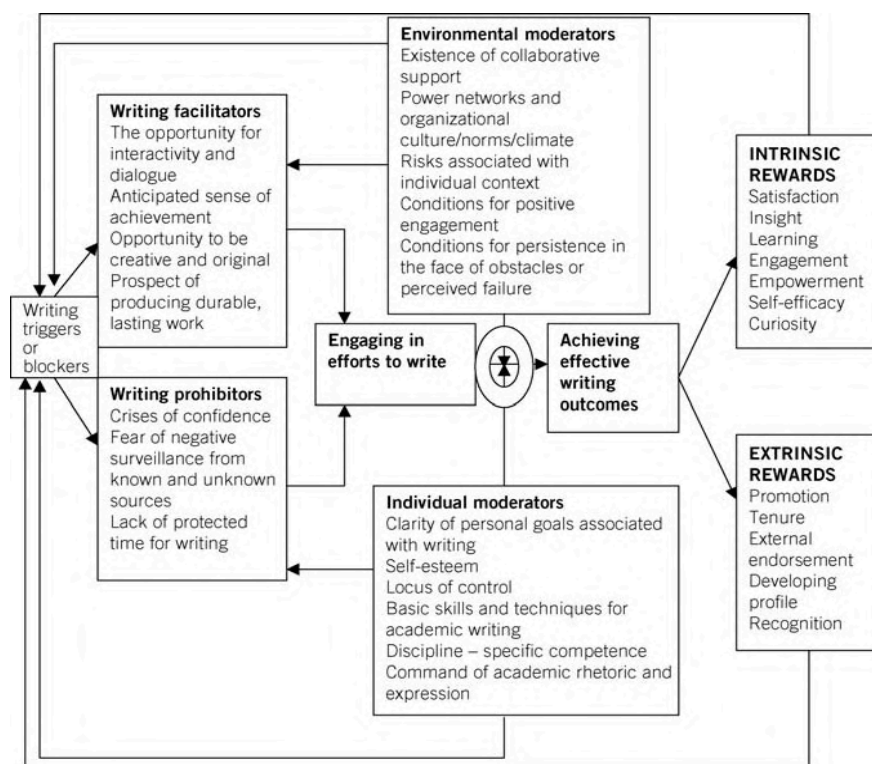


Figure 2 The structural model of academic writing by Murray & Moore (2006, adapted from Moore, 2003)

As to the purpose of writing and publishing in English for ESL scholars, based on email interviews of twelve Asia-based scholars, Lee (2014) categorized the reasons into three aspects: the academic/research community-related issues that research results can be shared and communication can be maintained with other researchers, the membership of the academic community can be gained, and the benefit of the language education institutions can be increased; job-related for job security, career advancement, and research funding, or because of job duty and peer pressures; and personal satisfaction that self-actualization can be gained through conducting challenging tasks, ideas can be developed through writing, can be rewarded after publication, or inner/intrinsic motivation/drive and love for writing.

Writing Modes

Based on previous cognitive writing models and written language-producing models, Paltridge et al. (2009) integrated multiple perspectives such as cognitive, social, and textual, and proposed a new second language academic writing model, which mainly held the opinions that academic writers, like general writers, have to experience different internal cognitive sub-processes before producing texts; and academic writing is not a behavior that occurs in a vacuum, but that would be affected by many external factors of the writer; both internal and external factors function together in enabling writers create a text. Hence, the model not only presents a multi-angel exploration of second language academic writing but also confirms the components of academic writing and the content of each part, as well as highlights the role of external knowledge resources.

Processes of Academic Writing

Research on the process of academic writing mainly focuses on writing strategies, their correlated influencing factors as well as the application of these strategies in sub-processes, including traditional language learning strategy systems or strategies specifically for academic writing, and discussions on overall or comprehensive writing strategy or a simple strategy with its functions. From the exploration of composing strategies, i.e., the “mental representations of the intended audience” and “the rhetorical purposes for writing” by four English-major students who also attended a postgraduate teacher education program through interviews and think-aloud videotapes, Wong (2005) discovered an obvious diversity of the extent that strategies such as cognitive, metacognitive and affective, etc. are made use of and of purposes in different junctures by writers adopting the similar strategies in the composing process. By taking the same research methods on the summary writing processes of six first-year MBA students, Yang & Shi (2003) noticed strategies such as “verbalizing what is being written, planning content, referring to pieces of literature, reading written materials, reviewing and modifying one’s writing, and commenting on source texts” in making them feel confident or floundering during the writing process,

the influence of their previous writing expertise on learning writing, the unawareness of the complexity of assignment and teachers' unclear and implicit expectations on them, and the necessity of mastering these composing strategies in disciplinary writing. The detailed composing processes of reading-to-write tasks and traditional impromptu writing-only tasks compared by Lia (2008) from ten non-native English-speaking students in a US university indicated a more interactive process of the former while more initial planning of the latter, hence explaining the participants' preference for the reading-to-write tasks in course design.

The choice of different media for completing written tasks by writers has also been focused on by some scholars, such as Lee (2002), who compared the composing processes of written texts in fixed time in paper-and-pencil mode and the computer mode though slight differences in essay scores have been discovered because of a small number of testing participants. The case study of an L2 university student finishing a 4000-word essay by Paul (2010) also signified the diversity of the cognitive processes in an electronic environment from those chosen in a pen-and-paper writing environment though the qualitative and temporal result may differ when more cases are taken into account.

2.1.6 English for Academic Writing Research on MA Theses versus PhD Dissertations

The comparative research of texts written by master's students and doctoral candidates normally concentrates on citation types based on Coffin's (2009) categorization of citations and the engagement sub-system in Appraisal Theory (Martin & White, 2005). Song & Wang (2020) probed into the choice of citations in Masters' and doctoral theses (ten for each) written by Applied Linguistics students around the year 2015 and discovered that both groups have a preference for author-cited integral with authors being the subject rather than the proposition being cited. Without slight differences in non-integral citations chosen by doctoral students, the two groups show little diversities in texts-cited integral or author stance types. On the one hand, standards and conventions of citing are emphasized by teachers in case of plagiarism, on the other hand, the number of cases is not sufficient to provide enough evidence for comparison. By choosing 30 theses in each group (Chinese English-major Master and doctoral students) and based on an expansion of Xu's (2016) research framework, Li & Zhang (2021) deemed that doctoral candidates are more skilled in evaluating and applying previous theories in their research as well as confirming their identity by directly supporting for other scholars' viewpoints than Master students as two functions of citations, i.e., "application" and "agreement" from the interpersonal dimension are significantly diversified in two groups after quantitative comparison.

2.1.7 Conjunction vs. Expansion

Conjunctive elements, also called as "conjuncts" (Quirk & Greenbaum, 1973; Greenbaum & Quirk, 1990), "connectives" (e.g. Crewe et al., 1985), "discourse markers" (Fraser, 1999), "logical connectors" (e.g., Celce-Murcia & Larsen-Freeman, 1983) or "logical devices" (Byrne, 1982), have been defined by Halliday & Hasan (1976a, p. 226) as "cohesive not in themselves but indirect, by virtue of their specific meanings; they are not primarily devices for reaching out into the preceding (or following) text, but they express certain meanings which presuppose the presence of other components in the discourse." In addition, "with conjunction, we move into ... a specification of the way in which what is to follow is systematically connected to what has gone before". (ibid, p. 227) In fact, as scholars adopt different names for denoting to devices that can combine clauses or sentences to "presuppose the presence of other components" by meanings, definitions of these terms usually have variants, so does the classification and corresponding research for them. Early and classic research can be dated back to Milton & Tsang (1993) when the usage of 25 single-word logical connectors by comparing an NNS corpus (a learners' corpus collected from 2000 files written by 800 first-year Chinese undergraduates in learning English) with an NS corpus (a combination of the American Brown corpus, the LOB corpus, and HKUST corpus) through concordancing is investigated from four main types: additive, adversative, causal and sequential. They found that Chinese students at Hong Kong University tend to overuse a ratio of logical connectors while having a misuse of *moreover* and *therefore*. The same classification and term were taken by Wu (2012), though the theoretical framework is based on functional linguistics, Halliday & Hasan's (1989) discussions on cohesion and coherence specifically. From the comparison of a native corpus (18 articles and the LOsESS corpus) with 350 argumentative essays written by Chinese freshmen of English majors, the latter group is deemed to apply more logical connectors while insufficient types in writing, in which the "additive" type is most preferred and some connectors such as *and*, *but*, *or*, *for example*, *at least*, *in fact*, *in addition*, *for instance*, *however*, and *because* are overused, leaving complex ones rarely being chosen. In Wu's opinion, the transfer of the mother tongue (Chinese) is the main reason for misuse, not to mention the influence of their Chinese thinking modes on choosing specific words or phrases, and

students have little awareness of the stylistic characteristics of the argumentative register or genre. However, how SFL theories can fully explain the misuse has not been presented, and the confusion of sentence connectors with clause connectors for quantitative analysis seems to lower the reliability of the research. Though exploring the usage of adversative connectors by Iranian EFL students, Housseinpur & Pour (2022) accepted Sorensen's (1997) classification as "subordinating conjunctions, coordinating conjunctions, prepositions, and transitions and conjunctive adverbs" to compare those applied in a learner corpus compiled of 188 essays and the Louvain Corpus of Native English Essays (LOCNESS). Excepting using *but* the most in two groups, for Iranian students, the overuse of *while*, the underuse of *despite*, *yet*, *instead*, propositions, conjunctive adverbs and transitions, and the misuse of *whereas* and *though* in the initial positions by learners are typical characteristics.

Focusing on additive, causal, and adversative connectives, a list of experiments was conducted by Murray (1997) to explore the psychological mechanism underlying the differential contributions made by three kinds of connectives to integrative processes. Although adversative connectives were found to cause longer time of disruption by participants in reading, how the influence occurs and which exact word/phrase results in a longer reaction time have not been answered yet. In combining Murray's with Fahnestock's (1983) taxonomy, Pretorius (2006) chose three types of logical connectives: illustrative, causal, and adversative in probing into their influences on students in reading expository texts. After two reading comprehension texts over three weeks from 82 medical and health science students, research results strongly indicate the importance of understanding causal and adversative logical relations in corresponding materials. However, as detailed words or phrases for testing have not been listed by the author, the reliability of the tests was doubtful. Logical connectives have also been discussed from the misuse, overuse, and positioning (initial or non-initial) in ESL students' writing (Tang & Ng, 1995), the (in)appropriate and (in)correct choice in unplanned argumentative essays by ELT students from Turkish (Karahan, 2015), the underuse of adversatives by Chinese students compared with journal article scholars because of weak genre awareness (Wang, 2014), etc.

Research has also been concerned about linking adverbials when Gao (2016) comparatively investigated the differences and similarities in their application in English academic writing by English native and Chinese non-native speakers (CNSs) based on Liu's (2008) taxonomy of LAs as additive, adversative, causal and sequential. Statistics from research articles in the fields of Physics, Computer Science, Linguistics, and Management suggest that the writing of CNSs has a relatively underuse of additive and adversative linking adverbials, and soft non-science disciplines prefer to choose these devices over hard science disciplines generally. However, the concordancing method of not taking clause relations apart from sentence relations makes the result doubtful, and the slight difference of choice between CNSs and native speakers as to density is where the peculiarity lies.

Another related term is "sentence connectors," such as in Carrió-Pastor's (2013) research on the variation in Native (NES) and non-native English speakers (Spanish writers of English specifically, NNES). Based on the classification from Quirk et al. (1985, p. 634), i.e., listing, summative, appositional, resultive, inferential, contrastive, and transitional, the purposes of identifying categories of sentence connectors, relative use across different sections in academic papers (twenty for each group in the field of engineering), and possible causes to various production were reached that contrasting and listing previous knowledge are considered to be important in the introduction section by both groups, more summative connectors are chosen by NNES while all other categories are more preferred by NES. Hence, texts written by native English speakers are "writer-oriented" in guiding readers by emphasizing more connectors than Spanish writers. However, the choice of detailed devices in each section has not been listed by the author, and the explanation for research results is not enough.

The status of conjunction *and* and *but* is reinforced by Bell (2007) when the initial position in sentences is speculated, i.e., SIA and SIB. From one million words of data in 11 journals representing science, social science, and humanities, Biber et al.'s (1999) statement that "while coordinator *and* is more frequent in academic prose than *but*, SIA is much less frequent than SIB" has been confirmed. For scientific writing, the occurrence of SIA and SIB is relatively low, while for social science and humanities, it is much higher. And SIB is the second most popular contrastive connective after *however*. As based on Halliday & Hasan (1976a) and Schiffrin (1986, 1987, 2006), SIA in academic writing can be chosen for performing three primary purposes: to indicate the last item in forming the text as lists; to develop arguments by topic organization; and to transform from previous discourse to authors' elaborative, contrastive and parenthetical comments, the research proved that the listing function is the most important one in academic writing. For SIB, the main purpose is to develop an argument "through the wide variety of cancellation that can be signaled in the current message" (Bell, 2007, p. 198). However, how each purpose can be realized under what kind of condition has not been elaborated by the author.

Regarding conjunctions, Clachar (2003) chose the paratactic conjunction *because* and *but* to

explore whether Creole English learners vary from native English speakers in application. Depending on Schleppegrell's (1992, 1996) examination of *because* derived from corpus-based studies as non-subordinate *because*, *because* and main proposition, and segment-linking *because* and Edmondson's (1981) and Altenberg's (1986) classification of *but* as interactional countering *but*, concession *but*, *but* after disarmers, and topic resumption *but*, the choice of them in essays written by students in a community college in South Florida was investigated. From 50 hours of tape-recorded conversations and 48 essays, it is discovered that creole-English-speaking learners draw more on paratactic conjunction function because of the constant directional shifting within the creole and standard English and due to the similarity of creole and standard English vocabulary while diverse grammar systems. Other conjunctions have not been inspected, and the classification is not uniform or unanimous in a system, leaving explanations of conjunctions within sentences unconvincing. The mixture of spoken and written materials for calculating in a small corpus also needs improvement. However, the advocate that Creole-English speakers be taken apart from native English speakers for researching EFL/ESL studies is worthy of consideration. Also, on conjunctions, Mohamed (2016) absorbed the theory of Halliday & Hasan (1976b) and followed Muftah's (2014) classification of their appropriate usage into additive, adversative, causal, and temporal. To explore the frequency of each category used by Malaysian ESL students, to recognize the most popular type, and to check whether there is a problematic one, 50 argumentative essays on a particular topic written by 50 Universiti Teknologi Mara undergraduates as well as corresponding semi-structured interviews were referred to. From the discussion of misuse, unnecessary addition, omission, and repetition of conjunctions, additive (*and*) was proved to be the most frequent semantic category, while the most inappropriate type turned out to be causal (*because* and *so*) due to the lack of introducing from textbooks, unawareness of teachers' information from students and their characteristic of forgettable. Another research can be the misuse of coordinating conjunctions *and*, *but* and *or* in Chinese ESL learners' writings of English under the theory of Error Analysis as from the comparison of a learner corpus with 21 theses of Chinese graduate students of English major and 22 theses from Proquest-PQDT, "and" and "but" are confirmed by Chen (2023) to be overused while "or" is slightly underused by Chinese learners; relative phrases such as "*and so, and then, more and more, and so on, home and abroad (at home and abroad), not only... but also, last but not least, he or she, more or less*" are more preferred by Chinese students while "*all but, whether or not, positive or negative*" are underused; the frequency of SIA and SIB is six times and three times more than native speakers correspondingly with a strong tendency of overusing *and then* in the initial position in sentences from the Chinese corpus. Strange are the mingle of sentence combinations and phrases or words connection concerning conjunctions, and the viewpoint of taking overuse as misuse by the writer.

Therefore, no matter whether the research object is conjuncts, connectives, discourse markers, logical connectors, logical devices, or conjunctions, as the term is changeable, the definition, connotation, or denotation of it and its classification are all controversial. And, by structurally investigating relationships within sentences, phrases, and words together by concordancing method for example in corpus analysis, the result would be contradictory, and the comparison of these devices within native and non-native English speakers is often superficial due to the generosity of the concept. Functions that each device (conjunctive word or phrase) performs in realizing meaning transmitting and, at the same time, structure combining are rarely the focus of these scholars, not to mention a uniform system for analysis. The clause or sentence relation relating to structural status as subordination or coordination remains peripheral, leaving definite or non-definite clauses, as well as embedded clauses, in a different category. The temporal relation leading by *WH-* or not is not included, and so are the relative clauses for elaborating initiated by other *WH-* words.

2.1.8 Citation vs. Projection

Before comparing the citation and projection for leading to the following section, a literature review for the research of citation will be presented, including the basic information, the categories of citation (including the division of subject/object, form which related to the integration of cited information, the rhetorical function that citation performs in demonstrating source knowledge, the stance/attitude that the writer takes in citing author's viewpoints or the author takes), related citation research (the disciplinary diversity; writings by L1 students compared to L1 professionals, L2 compared to L1 writers from not only novice but also professional ones; genre difference such as citation choices in essays, task essays, master's theses, PhD dissertations, research articles, etc.; choice from different sections such as in introduction, methodology, discussion and conclusion parts), and various classification of citation verbs.

2.1.8.1 Citation General

Citation in literature usually obviously marks the use of resources with sources, which is "visible

citation,” as Perorari (2006) called it, and described as a rhetorical feature that is “central to the social context of persuasion” (Hyland, 1999, p. 342). Besides gaining ideas about the topic, shaping opinion and using source texts to support opinion; gaining language support from source texts and using them as a model for organization (Plakans & Gebril, 2012) or supporting claims and presenting justifications (Swales, 1990, 2014; Thompson, 2002; Thompson & Tribble, 2001), the citation of information can also help establish one’s study on a solid basis (Hyland, 2009b; Petrić & Harwood, 2013; Samraj, 2013), making academic writing where citation is inevitable, a unique genre except for its high degree of formality. Citations are normally employed by scholars “to position their research in the right context; to show the significance and relevance of their research; to demonstrate their competence in their research field; to exhibit the adequacy and accuracy of methodology and procedures; to prove the legitimacy of their claims; to establish their argument and justify their findings, to persuade readers to accept their new claims as scientific facts, to contribute to new knowledge based on prior knowledge within the field, (Mansourizadeh & Ahmad, 2011) and finally, to claim membership of their disciplinary community” (Hewings, etc., 2010) through theses or papers writing. Research for citation mainly depends on Dialogism, SFL, Intertextuality, Appraisal Theory (Ma & Qin, 2016) etc., and varies significantly (Dontcheva-Navratilova, 2016) according to genres, culture, experience, and disciplines (Lillis et al., 2010) and is therefore challenging for novice writers, particularly EFL writers (Hyland, 2006; Jomaa, 2017; Jomaa & Bidin, 2017) in texts writing.

2.1.8.2 Citation Categories

In categorizing citation, researchers take diverse perspectives from aspects such as the subject or object of citing practices, the form whether the information of author(s) is integrated or not, the form of citation expressions or citation verbs, the rhetorical functions that citation performs in knowledge construction and information transmission, the stance or attitude from authors or writers in choosing citation toward cited sources, etc. For the subject of citation, it can be either the writer or the cited writer (author), and the object has been seen as past research or the current research being cited. (Zhang & Chen, 2020; see also Zhao & Zhang, 2022)

Form-integration

The categorization of form, though seems to be in chaos or unsystematically, usually contributes to three directions: one is from the integration put forward by Swales (1986) when the non-integral citation means that the researcher, the cited texts or the cited research information are deemphasized with the citation being outside of a sentence and the integral citation indicates that researcher, cited texts and information are paid attention to with citation being encompassed in a sentence, which later on is also called as the “information prominent” or the “author prominent” by Weissberg & Buker (1990). Related research can be found in Zhao & Zhong (2016), Zhao & Zhang (2022), Zhang & Chen (2020), Charles (2005), Willis et al. (2010), Hewings et al. (2010) and Hemaïs (2001) by following the same type from Swales (1990, p. 141). The division of integral citation into verb controlling in highlighting the authority of cited authors and naming in focusing on the status of authors by Thompson & Tribble (2001) has been adopted by some scholars such as Ma & Qin (2016); Xu (2012; 2016); Kuo et al., (2011). It is discovered that these two types are roughly equally chosen by professional researchers, while verb controlling is more outstanding than naming by novice writers. (Mansourizadeh & Ahmad, 2011)

Another direction is the focus on integrating cited information by Dubois (1988, p. 184), which is classified into direct quotation, paraphrase, summary, and generalization. As the name of the researcher or the author can be integrated as subject, agent, adjunct, noun phrase (Swales, 1990), and others in sentences, taking the tense (present, past, present perfect) of citation verbs into account, Pickard (1995) demonstrated the lexical and grammatical choices of quotations and citations by academic writers. This also leads to Peng’s (2019) theoretical framework in including direct quotation types (quotation fragment, brief quotation, extended quotation, number of words quoted, mean length of DQ) in investigating variations of authorial voice in citations in literature review thesis’s part by doctoral candidates. The combination of the cited author being “agent” and “adjunct” into “non-subject” with two other integral types (Hyland, 1999, p. 347) is absorbed by Okamura (2008) in comparing L1 and L2 professional writing of scientific texts, and the advocate of “block quotation (direct quotes of 40 or more words)” rather than paraphrase in citing information from Hyland (1999) is adopted by Ge & Wang (2023) for authorial construction research and by Lee et al. (2018) for paper research.

The third direction should rather be considered as the deepening of Swales (1990, p. 141) by Thompson (2000, 2001) in referencing to rhetorical functions of citation when the non-integral type performs the source (to attribute research findings, information, or ideas to an author in demonstrating acquired knowledge of a specific field), identification (to identify the actor or the agent in the cited sentence), reference (to demonstrate sources with introduction), and origin (to signify the “originator of a concept or product” (Thompson & Tribble, 2001, p. 95) functions and integral type includes the

verb-controlling (citation including a lexical verb being either active or passive), naming (citation being used as a noun phrase or part of a noun phrase) and non-citation (the name of the cited author without any information of publication year) subsections. This frame is widely accepted by scholars such as Petrić (2007), Mansourizadeh & Ahmad (2011), Thompson & Tribble (2001), Thompson (2002), Esfandiari & Sales (2024), etc. By removing non-integral origin and integral non-citation from Thompson (2002, 2005), Samraj (2013) compared citation forms and functions in discussion sections within Master's theses and research articles and found that non-integral citations are preferred by professional writers than novice writers. But by adding non-integral examples (the cited work illustrating those being stated in the sentence) to Thompson's, Master's theses in high grades are compared with those in low grades (Petrić, 2007, p. 240); and in taking direct quotations together with Thompson's (2002) non-integral citations, corresponding features in the Literature Review part in PhD theses by Spanish and English writers are probed into. (Soler-Monreal & Gil-Salom, 2011)

The practice of citation by students is sometimes perceived as a form of plagiarism. This is because, if a student does not provide a reference for information that they have cited, it is unclear where the information came from. In addition to direct quotations and references to the author or the source text, the cited sources can be classified as follows: (a) information that has been copied exactly, (b) information that has been slightly modified by adding or deleting words or using synonyms for content words, and (c) information that has been closely paraphrased by reorganizing sentence structures or replacing the wording from the original text with others (Shi, 2004, p. 178). The research results indicate that, particularly among L2 students, the a and b approaches are more preferred than the c and direct quotation approaches, with the former being the more frequently used. This is a relatively new finding. For the textual integration of sources, Coffin (2009) once mentioned a trichotomy, i.e., insertion, assimilation, insertion plus assimilation.

Although the categorization of citation types is various, being either complementary or contradictory, can mainly include the distribution of integral (author as a subject, agent, adjunct, in NP and others) versus parenthetical citations, the choice of reporting verbs, the effect of citing systems, the occurrence of selected features such as the choice of a citee's first name and last name, the employment of direct quotation from sources, the incorporation of source names within nominal groups, etc. (Swales, 2014)

Function

The rhetorical functions that citation performs in knowledge transforming in text writing have been discussed in Odell et al. (1983) and distinguished as "signposting, supporting, credit, position, engaging, building, typing, advertising, suturing, competence, topical." (see also Harwood, 2009) But it was Thompson (2005) who emphasized on four kinds, i.e., "attribution" that the source of the proposition or information cited in the sentence was indebted to, "reference" that a text for further information was listed, "origin" that where the theory, technique or product comes from were indicated, and "identification" that specific studies or texts in the sentence were alluded to. (see also Thompson, 2000; Thompson & Tribble, 2001) Later on, Petrić (2007, p. 243-247) extended the rhetorical functions of citations to eight main categories: attribution, exemplification, further reference, statement of use (to state source(s) and purpose of source(s) used in the paper), application (to make connections between source(s) and writer's work for her/his purposes), evaluation, establishing links between sources, comparison with sources and others (instances with an unclear relationship between citing sentence and source). (see Lee et al., 2018) or as described by Kuo et al. (2011): citation can provide for texts "a term or a definition of a term," the "background information for a research area/topic," views or findings, support for research result, explanations, examples, comparison, and "a method/approach/instrument/model/theory." The analogous categories such as "attribution," "support," "reference," "links establishing," "identification," and "comparison" of his/her findings with other sources from both similarities and dissimilarities are stated by Mansourizadeh & Ahmad (2011), though the "support" function indicates supports for the topic of the study, the procedures and materials, the writers' claims or argument, the justification for findings, etc., and the "establishing links between sources" function includes sources with "similar research findings," "similar focus" and "similar argument." For Samraj (2013), the research results and evaluation are excessively focused on as rhetorical functions of citation can not only compare, interpret, or explain results, evaluate previous studies and the whole field, but can also provide research or applied recommendations for readers as well as the background information. However, the "establishing links" and the "evaluation" functions seem to be covered by others as these two can both be realized when the citation is used to perform other functions in text writing. Hence Xu (2011, 2016) only discussed "attribution, exemplification, further reference, accept and application, linking/comparison, and summary functions" (see also Ma & Qin, 2016; Ge & Wang, 2023) while attributing "links establishing" and "support" to interpersonal functions. Another taxonomy is "position, defining, supporting, application, topic relevance, disagreement,

acknowledgment and agreement” that these functions would be diversified according to the disciplines and writing tasks of students. (Petric & Harwood, 2013; see also Li & Zhang, 2021) Still Coffin (2009) advocates the distinguish of “comparing and contrasting theoretical views” from “comparing and contrasting theoretical views with writer’s position” as the former is limited in scholars’ viewpoints instead of the writer’s, and that of “using authorities to reinforce own position” from “using one authority to reinforce another” when the writer’s stance or purposes are considered into, besides paying attention to normal functions such as “providing a rational or niche for thesis, showing the origins of a theory, adapting ideas to new contexts, building on definitions, strengthening position by referencing evidence.”

Although Agbaglo & Bonsu (2022) adopt the taxonomy of “background, claim asserted or denied, example, definition and hortatory,” in which the last one has never been mentioned, the typology of these functions can either be “essential,” such as attributing ideas/activities/concepts or methods/approaches/procedure to an author; stating the choice of works, concepts and theory approaches; establishing links between sources; comparing no matter the writer’s or other scholars’ findings with other sources, etc. or be “peripheral” like exemplification, further reference, explicit evaluation, etc. that has widely accepted by researchers.

Stance/Attitude

As citing and referring to other scholars’ research involves the persuasiveness and the evaluation of academic viewpoints for agreement or disagreement (Maroko, 2013; Nguyen & Pramoolsook, 2015; Shoohtari & Jalilifar, 2010), the relationships within the writer and cited literature or texts, among the writer, author, and readers are hereby taken into account that the writer, on the one hand, needs to show the attitude toward previous research to become part of the research community, and on the other intends to evaluate those works through citation in attempt to prove the creativity, reliability or worth of his/her research with gaining more supports from readers. While professional writers have good acquaintance with stance expressing, novice writers generally manifest difficulties, especially for those whose mother tongue is not English. Just as Hyland (2005b) or Gray & Biber (2012) have noticed, novice writers usually are stanceless in academic writing, making the research for stance or attitude conveying through citation important and necessary. Buckingham & Neville (1997) had once constructed a framework for stance research of citation in combining the concept of “colloquy,” the negotiated or challenged “controversy,” engagement theory, “author acts” vs. “writer acts” (Thompson & Ye, 1991), denotational or evaluational, and integral vs. non-integral (Swales, 1990, p. 148) division together that the first condition can be the writer totally accepts the knowledge with no controversy or any reader engagement in not recognizing the colloquy (research community context) through non-integral structure with no citing verb phrase; the second is that the writer has potential controversy to knowledge as it is negotiable, in recognizing the colloquy with possible reader engagement through non-integral, author citing verb phrases (either denotational or evaluational); the third is that the writer has actual controversy toward negotiated knowledge in engaging the colloquy with likely reader engagement through integral, author citing verb phrases (either denotational or evaluational); and the last one is the writer creates knowledge and then initiates controversy in affecting the colloquy with actual reader engagement through integral, writer citing verb phrases. But this complex integration of theories from a broad academic context seems to be not widely accepted by most scholars. Another model concerns the propositional responsibility and corresponding textual voice that when the writer takes the responsibility for the cited proposition, the voice is writer-monologic as the author is absent hereby; when the author takes the responsibility, the voice is author-monologic with the writer being deferred; and when the writer and the author both take responsibilities, the writer may be dominant with the author being subordinate or the other way around, in addition to the condition that they have a corresponding textual voice in meaning conveying. (Groom, 2000, p. 22; see also Thompson, 2002) But it is Coffin’s (2009) framework which is workable and warmly welcomed by related researchers by combining the Appraisal Theory (Martin & White, 2005; White, 2003) with the interpersonal dimension from the systemic functional linguistics: for the engagement of writers in attributing to sources, the writer’s stance can be “acknowledge,” “distance,” “endorse,” or “contest,” and the textual integration of sources relates to either insertion or assimilation or the mixture of these two strategies; (see also Hu & Wang, 2014; Lee et al., 2018) for the nature of source devised from van Leeuwen (1996), one aspect refers to “personalization” that the human dimension of the source may be foregrounded as “human” (a particular human, a group of humans, an institutional source) or be hidden as “abstract human” and “non-human,” the other aspect indicates the “identification” of sources whether they are attributed to a “named” author or an “unnamed,” produced by an “individual” or a “collective” group, and have high or low status in academia. (see also Wang & Hu, 2022) A similar framework is on the author’s stance (neutral, positive, or negative) and the writer’s (acceptance, neutrality, or rejection) toward cited sources through “factive,” “non-factive” and “counter-factive” options of citation features in Soler-Monreal & Gil-Salom (2011) and on general interpersonal dimension of citation from acceptance, disagreement or no clear signal (positive, neutral, tentative, critical) in Li &

Zhang (2021). The linguistic features for writers to take stance in citing sources usually include the phraseological patterns of reporting evidential being realized by lexicogrammatical terms such as author+ date; verb + *that* structure, be *verbed* structure, it is verbed structure, *as* structure or realized by non-verbal terms such as noun patterns or adjuncts; information sources as human (specific, non-specific), non-human or concealed; reporting verbs (e.g., Bloch, 2010; Hyland, 2002a; Thompson & Ye, 1991), and nouns like fact and finding, idea, argument, etc. (Yang, 2013) Still some reporting structures (Charles, 2006a, b; Jalilifar, 2012; Kwon et al., 2018) or tenses of verbs (Davidse & Vandelanotte, 2011; Hawes & Thomas, 1997) can signify writers' stance in writing.

2.1.8.3 Citation-related Research

Related research for citation involving choices because of disciplinary varieties; in L1 writing texts, the educational level difference within novice writers and professional writers, the cultural diversities within L1 and L2 writers; in different genres such as task essays (Sun, 2016), course papers (Xu, 2012), essays by undergraduate students (Schembri, 2009; Ma & Qin, 2016; Santos, 2018), MA theses (see Ma & Qin, 2014; Petrić, 2007; Agbaglo & Bonsu, 2022; Loan & Pramoolsook, 2015/2016; Charles, 2006a; Samraj, 2013), PhD dissertations (Jomaa & Bidin, 2016/2019a; Dong, 1996; Thompson, 2000/2002/2005; Thompson & Tribble, 2001; Jalilifar & Dabbi, 2012; Ge & Wang, 2023), the comparison within MA and PhD dissertations (Song & Wang, 2020; Li & Zhang, 2021), within MA theses and professional writings (Samraj, 2013), among these three levels (Xu, 2016), and research articles (Moravcsik & Murugesan, 1975; Thomas & Hawes, 1994; Kuo et al., 2011; Yang, 2013; Hu & Wang, 2014; Dontcheva-Navratilova, 2016; Agbaglo, 2017; Hyland & Jiang, 2017); in sections from theses written by scholars or students.

Disciplinary

Cross-disciplinary research mainly focus on the density of citation-related linguistic features compared in various disciplines, such as sub-types of integral and non-integral citations in agricultural botany and agricultural economics (Thompson & Tribble, 2001); two corpora of theses by native speakers in politics/international relations and in materials science (Charles, 2005); in politics versus material sciences (Charles, 2006b); among biology, civil engineering, education, and linguistics (Pecorari, 2006); among biology, philosophy and linguistics (Samraj, 2008); or in biology, chemistry and physics written by writers in L1 and L2 contexts (Okamura, 2008); in computer science and sociology (Harwood, 2009); variation in terms of citation integration, presentation, frequency, and reporting verb usage in hard disciplines (biology, electric engineer, technology engineer, physics) versus soft disciplines (philosophy, sociology, applied linguistics, marketing) (Hyland, 2009b); in applied linguistics versus medicine (Hu & Wang, 2014); in the Literature review part of PhD theses by EFL Arab postgraduates from IT (information technology) and applied linguistics (Jomaa & Bidin, 2016); in applied linguistics and electrical engineering compared to sociology and biology (Hyland & Jiang, 2017); among accounting, medicine, applied linguistics and engineering (Uba, 2020); applied linguistics versus nursing (Barghamadi, 2021); in linguistics, sociology, biology, and IT (Zhao & Zhang, 2022); in research articles sampled from 12 journals published in Chinese/English from applied linguistics or medical sciences (Wang & Hu, 2022); in economics versus Industrial and Manufacturing Engineering (IME) (Esfandiari & Sales, 2024), etc.

L1, L2 or professional

For novice writers, the practice of citation is not as familiar as professionals or experts in related aspects mentioned above, the research for comparing texts within two groups is hence increasing, such as comparing research papers written by English as a second language novice writers with those by experts in University from Malaysia (Mansourizadeh & Ahmad, 2011), students' writing with professionals (Zhao & Zhong, 2016), focusing on professional citation features or functions to provide an empirical foundation for language teaching (Xu, 2012; Zhang & Chen, 2020), etc. But the practice is also influenced by social or cultural factors, as language transfer from the mother tongue of L2 writers would cause differentiations in writing when their texts are matched to those by native English writers from not only novice but also professional levels. Such as the citation strategies realized in English-medium journal articles published in national contexts by Hungary, Slovakia, Spain, and Portugal authors are compared to those from international journals by them as well as international journals by UK and US scholars (Willis et al., 2010; Hewings et al., 2010), in research articles published in Chinese or in English from applied linguistics or general medicine discipline (Hu & Wang, 2014), in native and non-native English speakers' writing materials (Esfandiari & Sales, 2024; Campbell, 1990; Borg, 2000; Yeh, 2009/2010; Rabab'ah & Al-Marshadi, 2013; Rowley- Jolivet & Carter-Thomas, 2014), in Chinese and English academic discourse (Bloch & Chi, 1995), in English and Spanish (Soler-Monreal & Gil-Salom, 2011), etc.

Sectional

Citation in texts can be realized in diverse rhetorical sections such as Introduction (Posteguillo, 1999;

Gledhill, 2000), Method (Lim, 2006), Results (Brett, 1994), Discussion and Conclusion (Yang & Allison, 2003), results and closing sections (Kwan & Chan, 2014), in different sections for comparison as introduction, LR, methods, context, analysis, conclusion (Petrić, 2007); introduction, experimental, results & discussion, conclusion (Mansourizadeh & Ahmad, 2011); Introduction, Methods, Results, Discussion (Dontcheva-Navratilova, 2016), etc. In addition, citation density varies in transmitting knowledge of concepts/definitions, terminology, theoretical model, viewpoints/explanations, research topics, research methodology, and results/findings (Xu, 2012/2016; see also Zhao & Zhong, 2016; Ge & Wang, 2023; Li & Zhang, 2021) or just without the viewpoints/explanation sections (Zhao & Zhang, 2022; Zhang & Chen, 2020).

2.1.8.4 Citation Verbs

One aspect that closely relates to projection and cannot be ignored is citation verbs or “reporting verbs” adopted by most scholars that Chubin & Morita (1975) modified Moravcsik & Murugesan’s (1975) typology and put forward an analyzing frame from “affirmative essential” (being basic or subsidiary), “affirmative supplementary” (the cited sources are additional information or perfunctory) and “negational” (being partial or total) angles; (see Swales, 1986; Meyer, 1997) later on classified citation verbs from eight main types: “(a) mental verbs (i.e., *realize, recognize*); (b) speech-act verbs (i.e., *argue, assumption, explain, describe, recommend*); (c) given an object verbs (i.e., *adjudge, consider, diagnose, identify, interpret, judge, regard as*); (d) logic related verbs (i.e., *conclude, infer, deduce*); (e) source related (i.e., *learn, gather*); (f) emphasis of the elusiveness of object (i.e., *detect, trace*); (g) novelty related (i.e., *discover, reveal, invent*); and (h) on certainty or precision (i.e., *determine, ascertain*)” according to function as most verbs for him, have some commonalities in academic discourse “(a) the verb involves the human participant as a scholar; (b) the verb flows into a described event; (c) the object of the verb is knowledge of the object studied; and (d) the verb describes the cognitive achievement, or knowing, as the result of some intentional action.” It was Onwuegbuzie & Slate (2010) who even mentioned 15 categories for these verbs: (a) evidence-based/data driven verbs (e.g., verbs that acquire data or evidence); (b) explicit verbs (e.g., verbs that directly state); (c) implicit verbs (e.g., verbs that imply); (d) inclusive verbs (e.g., verbs that encompass more than one element); (e) procedural verbs (e.g., verbs that specify the procedure used); (f) interpretation verbs (e.g., verbs that specify the form of inferences made); (g) proposition verbs (e.g., verbs that suggest); (g) visual verbs (e.g., verbs that display); (h) comparison verbs (e.g., verbs that link two or more elements); (i) verification verbs (e.g., verbs that verify or confirm); (j) creation verbs (e.g., verbs that originate); (k) cognitive process verbs (e.g., verbs that refer to thoughts); (l) perception verbs (e.g., verbs that refer to observation); (m) direct object verbs (e.g., verbs that refer to an object or act); and (n) reference verbs (e.g., verbs that refer to another element or act). (see Msuya, 2019) But these taxonomies mix textual functions, rhetorical functions, and the writer’s cognitive functions, and I have not noticed whether the verb comes from writers or cited authors or scholars. Hence, Thompson & Ye (1991) classified reporting verbs into “textual, mental and research” types from author acts and “comparing and theorizing” from writer acts in terms of the denotation of these verbs, and into “positive, negative and neutral” from author’s stance and “factive, non-factive and counter-factive” on writer’s stance toward cited information in terms of the evaluation perspective. (see also Thomas & Hawes, 1994) According to the status of the cited proposition, the writer’s interpretation is referred to four kinds: “author’s discourse interpretation, author’s behaviour interpretation, status interpretation (indicating the functional status within the writer’s framework of the reported information/opinion) and non-interpretation (presenting the report as objective)”. (Thompson & Ye, 1991, p. 373) And the layers of reporting act are arranged in a reverse chronological way in presenting how citation performance is fulfilled: “1 writer writes; 2 writer evaluates; 3 writer reads; 4 author writes; 5 author thinks; 6 author researches.” (ibid, p. 378) Based on this, Francis et al. (1996) emphasized three main types for citation verbs, i.e., show and find verbs, think verbs and argue verbs. Following the same “denotation - evaluation” division of reporting verbs, the former was considered to include “research acts for describing (findings, procedures), cognition acts for representing mental processes and discourse acts for depicting communicative or argumentative expressions” and the latter encompass “factive (writer acceptance), non-factive with no clear signal (author positive, author neutral, author tentative, author critical) and counter-factive (writer disagreement).” (Hyland, 1999, p. 350; see also Yang, 2013) But the division is canceled when research acts of reporting verbs chosen by writers are discussed from findings (factive, counter factive, non-factive) and procedures; cognitive acts from positive, critical, tentative, neutral; and discourse acts from assurance (factive, non-factive), doubt (tentative, critical) and counters. (Hyland, 2002a, p. 119; see also Peng, 2019; Agbaglo & Bonsu, 2022; Nguyen & Pramoolsook, 2015) Genre or disciplinary-related research can be reporting verbs in Master’s theses (Lou, 2011; Manan & Noor, 2014), undergraduate essays from Chinese and UK (Lou, 2013), in applied linguistics journal articles (Un-udom & Un-udom,

2020), in medical journal articles (Thomas & Hawes, 1994), etc.; and form-related research comes from tenses of reporting verbs in chemical engineering technical articles (Oster, 1981), voice, tense and sentence functions of reporting verbs in introductory of PhD theses in agricultural biology and biochemistry (Shaw, 1992), choice of voice and tense in two disciplines by following Thompson & Ye's (1991) taxonomy (Thompson, 2001), type of verbs in introductions by Thai students (Jogthong, 2001), etc.

2.1.8.5 Differences between Citation & Projection

Except for traditional reporting and quoting research (Jomaa & Bidin, 2019b) and on process verbs (Jomaa & Bidin, 2019a) from the SFL perspective in which processes include “material, mental (cognition, affection, perception), verbal, behavioral, relational (attributive, identifying) - (intensive, circumstantial, possessive, causative) and existential” in looking at citations, though the writer considered the structure as citation, the research scope of citation and that of projection has almost no intersection. In terms of the form, the research object is different as citation includes integral and non-integral forms in providing it with a broader scope and making it belong to the intertextuality field that the cited information may be longer than one sentence or larger than a page, while most of the forms with quoting or non-quoting features are not included in the projection research, and the projection, in a sense if it has to be defined, then can be paralleled with the integral citation. In terms of the function, the perspective for classification tends to be a rhetorical one in investigating the function that citation performs in diverse theses sections, in treating a proposition or the content of the cited information, and in demonstrating the relationship among author, writer, and readers. While the classification of projection realizes the function of signifying the relationship between a clause and another, within a clause and a projecting verb, and a verb and cited propositions and proposals, which would not be referred to in citation research. A part of the projection is paralleled to the verb-dominating integral citation that intertextual strategies, rhetorical and interpersonal functions, or those concerned about stance and evaluation of the latter can be borrowed for the research of the former, although lexicogrammatical realization in instantiations with choices instead of citation behavior is what the projection research more focuses on. The comparing methods of citation choices because of different cultural backgrounds (L1 versus L2 writers), disciplines, genre sections, educational levels within novice and professional writers, etc., are to be absorbed in projection research. As to verbs, those in verb-dominating structures in the citation that resemble the projection are frequently paid attention to. Although the classification for citing verbs is diversified, in which “textual verbs” resemble “locution” and “mental verbs” denote “idea” in projection, most of them take an interpersonal or rhetorical perspective except for structural taxonomy. The general, complicated types and the undistinguished citation verbs from reporting verbs result in totally not the same concepts in two directions. In addition, the unincluded structures such as embedded clause complexes, facts, or acts in projection, except the non-finite or finite features, make the projection different from the citation as well as their research.

2.1.9 Controversial

Through explaining key terms ESP, EAP, and EAW, and presenting theoretical foundations from the discourse analysis, academic literacy, and cognitive including identity construction perspective in academic writing, whether the studies, especially the academic literacy and cognitive aspect on clause complexes, are common or not deserves inspection in the next section. Qualitative methods such as ethnography and case studies in analyzing academic discourse are referential in clause-complex related studies. Though various corpora are selected for quantitative or empirical comparison, the small size and insufficient data analysis would usually lead to some unpersuasive research. As to discourse analysis seldom research is conducted from the logico-semantic relation in the textual sphere, not even to mention clause complex related from the contextual and critical aspects. The review of AL, including IC studies, reveals that the discussion about reporting verbs for inspecting lexical diversities in evaluation or stance presentation is not enough, and the undistinguishable reporting verbs from citation verbs make some research results doubtful. Although master's theses and doctoral dissertations are regularly regarded as the object for academic writing research, most of which are conducted from genre analysis in taking up move comparison in diverse sections rather than from register field. The comparison of linguistic features within master's theses and doctoral dissertations for development inspection deserves concern. Given the scope of conjunction or conjunctive, part of which has an intersection with that of expansion, and the left part concerned not only with clause combining or connecting relation but also about sentence binding or connection, the “expansion” described in the next section is thus preferred in this thesis.

As the scope of the citation is also far-ranging, while that of the projection focuses more on the relationship between reporting verbs and reported clauses; and as citation verbs have not been departed

from reporting verbs in academic writing research, the function that each reporting verb performs in diverse projecting environment would change according to writers' different aims, while citation verbs related studies have not mentioned this point, the projection frame from SFL seems to be more practical.

2.2 Previous Research of the Clause Complex

This section will concentrate on clause complexes to synthesize previous research literature from three aspects: their theoretical development, the application of theory in practice, and the detailed examination of subsystems in texts. Two main parts will be presented in this section: the first part will examine the development of clause complex-related research in China, and the second part will focus on the development of clause complex-related research in other countries and regions outside of China. The reason for this dichotomy from the perspective of countries can be attributed to:

1. The specific development of subsystems in systemic functional grammar has its roots in Chinese grammar because Halliday once studied Chinese grammar in China and subsequently adopted the ideas and system classification of Chinese grammar when teaching English as a second language. For further details, please refer to Hu Zhuanglin's analysis of Chinese clause complexes, the concepts of Chinese complex sentences (*Fuju*) put forth by other scholars, and the comparative study of the concepts of Chinese and English complex sentences (*Fuju*) and clause complexes.

2. Beginning with a comparison of the concepts of clause complexes and complex sentences, Chinese scholars have devoted greater attention to the particularities of this phenomenon and have conducted a range of theoretical and applied studies on it. In other words, the quantity and scope of research in China far exceed those in foreign countries, although the concept was initially proposed by Halliday.

3. Given that the focus of this study is on written texts produced by Chinese students in English, it is inevitable that the results will be influenced by the transfer of the Chinese language. Accordingly, the research findings of Chinese scholars on this subject are initially examined.

4. The researcher is Chinese, thus facilitating the retrieval of research findings from Chinese scholars in the compilation of materials. Moreover, research utilizing Chinese as the written language or even the Chinese language itself as the subject matter is more accessible, comprehensible, and amenable to textual organization than being conducted by scholars from other regions.

2.2.1 Research of the Clause Complex in China

The author has collected 91 papers that employ the Chinese language to examine English clause complexes or utilize Chinese clause complexes as the subject of investigation. Of these, 39 papers address the overarching framework of clause complexes, 45 papers focus on projection and its constituent subsystems, six papers examine expansion-related systems, and one paper is dedicated to embedded clause complexes. While this study does not examine embedding in great detail, it is still a topic of significant importance.

2.2.1.1 On Clause Complex

The research on clause complex systems can be broadly classified into three categories: theoretical discussions, discourse analysis, and its application in text translation.

Theoretical

Theoretical research on clause complexes in China encompasses a range of linguistic phenomena, including the study of the relationship between clauses, the relationship between the main and the subordinate, and the stylistic function of the language. Furthermore, it encompasses the comparative analysis of "clause complexes" and "complex sentences" (*Fuju*) and the comparative examination of the paratactic and hypotactic structures of English and Chinese clause complexes. Additionally, it involves the investigation of issues relating to clause complexes and their translation, logico-semantic relations based on corpus data, and the quantitative analysis of clause relations in discourse.

Chen, A. (1997) was the first researcher who paid attention to the clause complex in China and studied the relationship between it and style. Just as Halliday (1985a, p. 198) thought, parataxis is symmetrical and transitive, while hypotaxis is non-symmetrical and non-transitive. She found stylistic differences between *King Edward VIII 1936* and *Functions of American English* by Jones C. V. Baeyer that parataxis represents loose meaning, and hypotaxis conveys tense and preciseness. Then, four kinds of stylistic discourse, i.e., law, technical, toast, and prose, were chosen and analyzed from their mode, field, and tenor, respectively, finally concluded that the more formal the textual style is, the more complicated their clause complexes are, with a higher ratio of hypotaxis and embedding for exposition

and modification; and vice versa for extension and direct speech projection. Though reasons for their differences can be penetrated, the article did open a new road for stylistic analysis by interpreting the logical relations between clauses.

Xiao, J. (2000) argued some details on Halliday's (1994a) Hypotaxis and held the opinions that: 1. Hypotaxis is different from subordination in traditional grammar; 2. Hypotaxis should be seen as a category in embedding syntactically; 3. Clause complex could be "above" and also the same level as Clause (Berry, 1975, p. 108); 4. The logico-semantic relationship could be applied only to paratactic clauses but not hypotaxis. In addition, dominate clauses could be independent only when dependent clauses were non-obligatory adverbial clauses or non-finite attributive clauses. (Huang & Xiao, 1996) The location of dominant clauses and dependent clauses could not be changed except that dependent clauses played adverbial or complement roles. (Halliday & Hasan, 1976a) Because some clauses were judged by processes instead of logico-semantic relations in Halliday (ibid, p. 7), the logico-semantic system was not complete enough to describe or include every clause complex.

Cheng, X. (2005) disagreed with Halliday's model of clause complex relationships, which focuses more on grammatical and contextual spheres than on logico-semantic aspects. He gave two examples: 1. *John ran away, and Fred stayed behind.* and 2. *John ran away, whereas Fred stayed behind.* And then explained that the former was a paratactic extension while the latter was a hypotactic extension, but their grammatical description could not fully represent semantic and pragmatic differences. Thus, he proposed three problems of Halliday's model: 1. Inconsistency of taxonomy standard of clause complex relationships; 2. Times and spaces do not belong to logico-semantic relationships; 3. A high generation of categories means that complex gradations cannot be fully described and embodied. To solve the dilemma, "semantic relation" was used to describe the relation of clauses that make statements, and "rhetoric relation" (Mann & Thompson, 1987) was proposed temporarily to convey contextual relations between clauses. However, concluding time and special relations as additions were not persuasive and did not prove that these clauses were exclusive of any logico-semantic relations. What is more, Cheng's new model paralleled semantic and rhetoric relations with grammatical and projection, which was also general and did not fully list all semantic and rhetoric clause relations like replacive, subtractive, adversative, concessive, etc. At least, Halliday's model was not perfect and regarded as linguistics law in Cheng's opinion.

Zhang, Y. (2011) generally explained clause complex relations and divided projection into three parts: reports, ideas, and facts for analyzing clause complexes, but there was a lack of any convincing reasons for the theoretical classification method.

From the perspective of Ecolinguistics, He & Cheng (2021) broad two new dimensions (sequence and means of representation) to the taxis system, two dimensions (expansion orientation and origination) to the expansion system, and three new dimensions (projecting roles, projecting markers and projecting orientation) to the projection system, thus constructs extended subsystems for ecological logical analysis of texts with the applicable ecosophy "Diversity and Harmony, Interaction and Co-existence" that non-human beings and physical elements are encouraged to be projectors in natural ecological discourse while human beings and social elements be projectors in social-ecological discourse. The system is different from the traditional clause complex system because it divides clause structures and embedded clauses; moreover, due to the arrangement mode of items, functions are intersected rather than black-or-white selected. Subjective assessment of texts would be disagreed with blended instead of syncretic theories.

This was followed by Wang, Q.'s (2008) research that clause complexes were divided into ten kinds, which was the same as *Fuju* in Chinese grammar. The sameness could be attributed to Halliday's study experience in China, supervised by Luo Changpei and Wang Li.

Wang, Q. (2015) also made a comparison between English clause complex and Chinese *Fuju* (parallel with "clause complex" in Chinese grammar) from the perspective of finiteness. He held the view that non-finite clauses could only be hypotactic extensions and subordination in clause complexes. Without conjunction, the logical relation between the non-finite clause and the main clause would be vague and difficult to define its functions that play in a clause complex. But Chinese *Fuju* lacked finiteness and, hence, could not be regarded the same as clause complex even after translation not only to the term itself but also to what the term refers to.

In examining the differences in the tendency to choose taxis between English and Chinese, Ming, J. (2020) posits that there are two prominent features of clause complexes in English: mutual implication and mutual conversion. For a given logico-semantic relation (LSR), the functional equivalence of LSRs of the same category results in paratactic and hypotactic clause complexes being mutually implicit and mutually convertible. However, these two characters must be based on a clear distinction between the two tactic systems. In English, the distinction between paratactic and hypotactic clause complexes is evident. Firstly, the distinction between the main paratactic and hypotactic markers has been made.

Secondly, finite and non-finite clauses have been separated. In contrast, the distinction in Chinese is relatively vague. In light of these observations, the author puts forth four criteria for the tactic tendency of language systems and offers linguistic evidence to elucidate the paratactic proclivity of Chinese and the hypotactic orientation of English. This discrepancy in linguistic tendencies can be illustrated by the contrasting levels of diversity and extensibility observed in clause complexes across the two languages.

Su, J. (2006b) made a contrastive study of English and Chinese clause complexes from the perspective of cohesion especially. He found more similarities than differences in these two languages regarding semantic cohesion, while great heterogeneity in logical cohesion for their diverse sentence-building patterns. Explicit logical cohesion is usually employed in English clause complexes, while iconicity of the word order and aboutness of the topic contributes more to the cohesion in Chinese clause complexes.

For the automatic identification of clauses, the definition of a “sentence” in Chinese linguistics is not readily operationalizable for two reasons. First, it is not constrained by the norms of linguistic structure and meaning that are typically observed in other languages. Second, its structure and meaning are incomplete. Furthermore, the lack of a comprehensive representation of meaning in the full-stop sentence will inevitably impact the efficacy of language information processing based on this structure. Consequently, the notion of the sentence has yet to be acknowledged as a viable approach for processing language information. The clause complex is defined as “a composite structure of clause sequences that are minimally self-sufficient in terms of logical and semantic relations.” From the perspective of the clause complex, He et al. (2021) selected 12,675 sentences from the Chinese clause complex annotated corpus of Beijing Language and Culture University, including encyclopedia entries, government work reports, news articles, and novels. The automatic recognition of sentence boundaries based on the clause complex theory was then discussed. In particular, the BERT (Bidirectional Encoder Representations from Transformers) boundary recognition model was employed to identify the boundaries of sentences. The experimental results demonstrate that the BERT boundary recognition model exhibits superior automatic recognition efficacy with regard to sentence boundaries in comparison to mechanical segmentation, contingent on the diverse forms of punctuation observed in the corpora. The model also demonstrated superior performance in recognizing sentence boundaries for punctuation sentence sequences with three punctuation marks compared to those with two. Notably, the automatic recognition of sentence boundaries in encyclopedias and novels exhibited greater accuracy than in government work reports and news corpora. The model played an accuracy rate of 91.91%, a recall rate of 88.87%, and an F1 value of 90.36%. It should be noted, however, that this study did not consider the structure of more than three adjacent punctuation sentences, nor did it address the nesting of small sentence complexes.

In examining the extractive machine reading comprehension (MRC) task, Wang et al. (2021) undertook a re-examination of the issue of cross-punctuation sentence reading comprehension. A strategy for integrating the automatic analysis of clause complexes with the MRC task was proposed. The long-distance shared relationship between the Naming and the Telling in the clause complex was employed to provide the model with sentence-level structured semantic information, thereby reducing the difficulty of long-distance answer extraction. Additionally, they proposed an MCR model based on clause complexes and demonstrated the effectiveness of the corresponding method in the required task. Furthermore, the experimental results on the CMRC2018 reading comprehension dataset illustrate that the automatic analysis of clause complexes has a significant impact on the long-distance cross-punctuation sentence question answering problem in these tasks. In comparison to the baseline model, the overall exact match rate (EM) of the MRC model (RoBERTa_wwm_ext_NTC_add_MRC) based on clause complexes exhibited an increase of 3.26%, with the EM of the cross-punctuation sentence question answering problem demonstrating a more pronounced improvement of 3.49%. However, the concept of clause complexes in this study appears to diverge from the concept of clause complexes as defined in English linguistics. The present study focuses on the domain of Chinese machine reading comprehension. In conclusion, the design, research methods, and experimental steps of the fusion model serve as a valuable reference for research on the MRC of English clause complexes in texts.

In investigating the use of the logico-semantic conjunction *and* in clause complexes, Zhang & Ran (2024) investigated the distribution of VP+*and*+Declarative clause structures and their dynamic semantics from the enTenTen18 corpus. In traditional approaches to clause analysis, the *and* conjunction is typically treated as a static element, with the focus being on the interpretation of the constituent clauses. In reality, however, their semantics are subject to change as a result of the influence of pragmatics and context. In addition to the existing categorization of elaboration into exposition, clarification, and exemplification, the function of *and* in extension is further subdivided into experiential statement and interpersonal meaning (including greeting and imperative). Similarly, the function of *and* in enhancement is divided into temporal, causal, result (including experiential statement and interpersonal

imperative), and conditional (including experiential condition, textual topic, and interpersonal imperative). In addition to the aforementioned detailed classification, the study was conducted from the perspective of the three major meta-functions. Of the nearly 500 structures examined, “enhancement” has been proved to be the most prevalent (78.16%), followed by “extension” (16.23%) and “elaboration” (5.61%). In light of the fact that a grammatical structure is capable of embodying a multitude of potential meanings, it can be posited that the dynamic nature of the clause complex represents a presentation of multiple meanings that are collaborative and complementary. These meanings are influenced by the consensus of the language community members on which the clause complex is based and are contingent upon the context in which they are used and the probability of their occurrence. However, with the exception of the VP+*and*+declarative clause, the other structures related to the *and* have not been examined. As the author stated, the VP clause may be a declarative clause with an omitted mood structure, or it may be an imperative mood clause. Consequently, these two scenarios have not been separately discussed statistically.

In a comprehensive examination of the logico-semantic interrelationship between German and Chinese clauses, Zhong, W. (2020) identified that if the embedded clause and the coordinated relationship are conceptualized as a continuum, ranging from weak to strong, the embeddedness of German is discrete, whereas that of Chinese is ambiguous. While the expansion relationship between clauses is explicit and implicit, the implicit mode of German is a marking means, whereas that of Chinese is a universal method. Additionally, there are notable discrepancies between the two with regard to the positioning of logical markers, the proximity of the conjunction of logical markers and clauses, and the syntactic structure of the connected clauses. A quantitative comparative analysis of the projection relationship reveals that the projected information in German is predominantly “reported,” whereas in Chinese, it is primarily “quoted.” The typical mode of mental projection in German and Chinese is that of reporting. Moreover, in comparison with Chinese, there is a higher probability that the projecting clause of German mental projection clause complexes is located before the projected clause. In addition to elucidating the discrepancies between the two languages from a linguistic typological standpoint, the author also elucidates the choice of logico-semantic relationships between German and Chinese clauses from a national cognitive perspective (including category theory, figure-ground theory, and iconicity theory), a thinking perspective (the difference between subject-object separation and subject-object integration), and a cultural perspective. While implicit clause relations have been referenced on numerous occasions, the authors have primarily analyzed them through illustrative examples rather than through a systematic examination and the formulation of a comprehensive theoretical framework. Moreover, the subjectivity of clause relation annotation, particularly with regard to implicit relations, is of considerable consequence in this context. In conclusion, the author examined the embedded structure, provided a summary and quantitative comparison of the projecting verbs, and utilized a substantial corpus for comparison. In light of these findings, the study offers valuable insights that can inform the translation of German-Chinese clause relations.

From the viewpoint of Finite in Zhang, Y.’s (2016) Ph.D. dissertation, taxis in clause complex could be construed by finite interacted with the conjunction, and as “choosing finite means construing a covert projection,” a covert locational projection could be construed by objective finite and a covert mental projection by using subjective finite, which together “made a cline, with the projected clause realized by the modal adjunct between.”

By vertically comparing clause and clause complexes in English and Chinese, Su, J. (2006a) considered that “the syntactic structure of English clause complexes is, in essence, the same as that of English clauses” while Chinese clause complex was more like a sentence group; and by horizontally comparing clause complexes, the semantic-logic relations in English clause complex were mainly conveyed through “inflected forms, connectives, and prepositions,” which made the structure closed and compact and the logic rigid while “half of the clause complexes in Chinese is built through the linear spread of phonetic-semantic chunks,” which made the structure loose and diffusive but the meaning cohesive. In addition, the author pointed out that “lack of inflections and freedom from subject-predicate concord in Chinese, together with the comparatively infrequent use of function words, are the superficial causes for the differences while the dominant thoughts in Chinese traditional philosophy and the philosophical basis of the English language constitute the underlying causes.” Thus, when translating Chinese into English, semantic relations among phonetic-semantic chunks should firstly be clarified, then identify the semantic “fulcrum,” grasp syntactic features, rearrange the word order, divide the clause complex into several semantic units and lastly adjusting grammatical units. The problem lies in that the thesis emphasizes the Chinese clause complex based on the English SFG theoretical framework, which has not contributed to the theoretical framework for comparative analysis in clause complex.

Yang, X. (2008) simply discussed English logical cohesion besides grammatical and semantic cohesion based on Cohesion theory and discovered that “tense and aspect play a more important part in

the cohesion within the clause complex.”

By combining quantitative analysis with statistical analysis, case analysis, deduction, induction, literature retrieval, etc., Liu, D. (2017) came up with two insightful concepts, i.e., Characteristic Matrix of Clause Relations (CMCR) and Correlation of Discourses (COD), “as well as a discourse measurement method based on quantitative analysis of clause relations.” CMCR was a qualitative method to analyze the textual features in a discourse comprehensively, and COD was “a quantitative description of the correlation between different discourse genres from the perspective of statistics” that MATLAB was employed to calculate the Pearson’s correlation coefficients between two CMCRs from a weighted average of semantic, rhetorical and projection. Thus, researchers could “implement the large-scale training for the machine on discourse genres,” and “realize the automatic genre recognition and classification of massive texts.” However, this method relied on the premise that computers could identify and count the various categories of each clause relation accurately, which had not been completely realized yet.

By connecting the clause complex system with the MOOD system in interpersonal metafunction, Zhumu (2018) re-classified clause complex relations and manually annotated a corpus of business domain with tags, then compared these annotating results with automatic clause relation identification performance fulfilled by Conditional Random Fields (CRFs) model to investigate if the new classification is scientific and reasonable. Research results showed that the identification of the complement clauses and adjunct clauses was performed quite well. In contrast, that of the subject clause and the appositive clause were not as perfect as imagined due to the limitations of the corpus size. Moreover, “the omission of the subordinate markers” and “the long distance between the complement clause as well as its dominant clause under certain circumstances” in complement clauses and the “flexible sequence” and the “variability of adjunctive markers” in the adjunct clause are primary reasons in Zhumu’s opinion when identification errors occurred in the performance. Though lexicogrammatical levels, such as markers indicating manner and causal-conditional relations in adjunct clauses, were manually counted to make error analysis, Zhumu did not thoroughly explain the reasons for markers or sub-types distribution diversities, nor did she give explanations of frequency variance by connecting these lexicogrammatical markers with the discourse or the whole context to classify the different functions that they play. Another problem was the “vague boundary between clause relations” in the original logico-semantic relationship classification. Thus, Zhu rebuilt the analyzing model, but in essence, it did not resolve this problem. Throughout the whole research process, the particularity of choosing a business corpus was not prominent for interpersonal metafunction analysis, and what if the condition of combining the logico-semantic relation model with two other metafunctions in the same corpus? Plenty of research points were put up and presented in the thesis, and the main direction was, however, not referred to or emphasized.

By annotating and describing the clause complexes of Chinese, English, and Japanese, Zhang, X. (2022) conducted a comparative analysis of the three languages from three dimensions: the type of naming-sharing relationships, the sentence-generating ability of NT clauses, and the structure of NT clause complexes. The study revealed that Chinese and Japanese are predominantly stack branches, whereas English is primarily characterized as new branch patterns. Additionally, Japanese exhibits a higher prevalence of influx structures compared to Chinese and English. Notwithstanding the presence of non-compliant phenomena in the NT clauses of the three languages, the incidence of such phenomena in Chinese is markedly less than that in the other two languages. Moreover, the mean depth of the three languages within the corpus is situated between 2 and 3, although the depth of the Chinese clause complex is deeper than that of the other two languages. The present article, however, is primarily concerned with the structural difference observable within the corpus. As the definition of “CC” and “clause” at the end of the article demonstrates, the term does not fully align with the connotation of the English clause complex.

Yang, H. (2009) was the first researcher who gave a literature review on clause complexes in China from 1979 - 2008. After reviewing 12 important articles written by 11 researchers from the genre, the register, English-Chinese comparison, sub-system, clause relationship, and projection, respectively, she found out that registers like novels, essays, etc., had not paid attention to clause complex aspects, and functional stylistics and sub-systems of clause complex need further researches.

Discourse Analysis

The research on discourse analysis of clause complexes can be broadly classified into five categories: textbook discourse; English writing texts, including essays/theses/dissertations and their components; stylistic research in novels or poems; news discourse; and advertising discourse.

Still, Chen, A. (2000) first made a clause complex comparison between an extract from Leech’s *Style in Fiction* (Leech & Short, 1981) and Studs Terkel’s (1974) *Working*, to observe how logico-semantic structure differences reflected in diverse Modes determined by Field and Tenor of the

discourse, then discussed Theme and Rheme organizing from clause complexes to discourse. As he considered that the global Theme-Rheme structure could form a dynamic process of textual development through connection, anaphora, and cohesion between the Theme and Rheme of these clause complexes, five modes for analyzing Theme-Rheme structure in adjacent clauses in discourse are put forward i.e., parallel, extension, concentration, intersection, and juxtaposition. Juxtaposition (No relationship can be found in Theme and Rheme between adjacent clauses) is the most common in legal discourse for enumerating information objectively, while extension (Regard Rheme in the last clause as Theme in the next clause) appeared mostly in scientific discourse to argue and ratiocinate.

Li, T. (2020) is interested in Business English Listening & Speaking Courses especially on Sales Review and Performance Reports (SRAPR) as 45 paragraphs from a general English textbook (Zheng, 2001) and 45 from business English materials (Wood et al., 2002a; Wood et al., 2002b; Brieger & Comfort, 1996; Sun, 1998) were compared of their clause complex causal relations that conjunctive of the latter were outnumbered and presented various, professional and more official than the former ones because SPARP discourses not only introduce sales condition and performance of the company, but also teach students to imitate reporters making connection between the company's business decisions and actions, and its sales and performance, leading their listeners make beneficial decisions for the company by using CC causal relations and corresponding conjunctive. The quantitative comparing process and outcome were lacking in the article, making it only a qualitative analysis. As the chosen learning materials were written or edited twenty years ago, this research might not be so useful due to fast changes in language.

In a study conducted by Bai, L. (2021), the potential relationship between the explicit logical connection awareness of English texts and English writing proficiency was investigated. The explicit logical connection categories, the explicit logic connects, and the explicit logical connection awareness of 165 second-year undergraduate students majoring in English were assessed. The results demonstrated a significant positive correlation between these variables and the subjects' English writing proficiency. The elaboration relationship was not identified as a significant predictor of English writing proficiency. However, the enhancement and extension relationships were found to have a significant predictive effect. Moreover, the logical connection methods between sentences, groups, and paragraphs, as well as those within clauses, were also identified as having a significant predictive effect on English writing proficiency. However, the author did not merely examine the interconnections between clauses, nor did she undertake a quantitative analysis of the three distinct logical relationships intersecting with clause complexes. Moreover, the evaluation criteria employed in the English writing proficiency examination were not explicitly defined, making it difficult to determine whether the observed correlation results are influenced by the subjective assessment outcomes of the writing proficiency examination.

In the analysis of logico-semantic relations, Huang, Z. (2015) compared bachelor's theses of the Chinese ESL majors with articles written by ENL scholars from *Applied Linguistics* to uncover diversities of applying explicit logical connectiveness/semantic links and implicit relations based on Halliday & Hasan's (1976a) theoretical scheme which included four kinds of relations: additive, adversative, causal and temporal. The author concluded that: 1. ESL writers had a tendency for explicit logical connectives, "leaving a minimal part realized implicitly confined to the causal and temporal category"; 2. they would overuse or misuse some connectives, lacking diversity; and 3. "Logico-semantic relations can also be expressed by grammatical metaphor (nominalization and verbalization), non-finite verbal clauses or prepositional phrases, which will contribute to objectivity, compactness of text, as well as to flexibility in the use of language." The first query is on the comparability of two corpora as ENL scholars came from universal educating backgrounds with higher academic research levels than Bachelor's students, the outcome might not be persuasive; second, the research was limited in explicit connective (conjunctive) markers without deeply investigating into logical relations within clauses, a more applicable theoretical framework should be established; third, implicit relations were not fully clarified and compared with explicit links, making the research doubtful.

In probing into English paper abstracts utilizing the clause complex theory, Zhong, K. (2023a) selected 21 abstracts from each field of language and literature, as well as computer science, as the subjects of the study. The analysis revealed that self-sufficient clauses constituted approximately one-third of the nearly 1,000 NT clauses observed in the academic texts. Additionally, the new branch pattern and stack pattern both accounted for approximately 25% of the total. The remaining patterns, including single naming, back-position, and influx, were found to account for relatively minor proportions. Moreover, English abstracts of Chinese papers tend to commence with relatively simple clauses. Although some viewpoints were mentioned, they were not elucidated in the text. Moreover, the author did not distinguish between the variations in clause complex features across the abstracts of Chinese-English translated texts and those of native English speakers. The classification and display of disciplines were also not addressed.

Zhong (2023b) conducted another study that compared English abstracts translated from Chinese to English in Chinese Research Articles (RA), English RA abstracts written by Chinese students, and English RA abstracts from soft disciplines (economics, philosophy, language, and literature), and hard disciplines (mathematics, physics, computer science). The findings revealed that the distribution of naming-sharing structure (NSS) in RA abstracts differs significantly from that in other genres. Additionally, there was a notable distinction between ENSs and CNSs in the utilization of stack patterns, with CNSs employing a greater number than ENSs. Furthermore, the utilization of the stack pattern in Chinese-to-English translated abstracts written by CNSs is markedly higher than that in English abstracts directly written by CNSs, which indicates the impact of language proficiency and translation. It is curious to note that the disciplinary character has no bearing on the selection of NSS or the syntactic type of naming and description. This observation calls into question the reliability of the structural analysis. In any case, the classification of the corpus is worthy of emulation.

By combining logico-semantic relations with stylistics, Huang, H. (2017) attempted the last episode in Joyce's *Ulysses*, which is a transcriptive presentation of Molly's monologue in an over-300-sentence text characterized by a lack of punctuation and known as the first "longest sentence in English Literature" until the year 2001. Especially focusing on analyzing the most common connective *and*, which was most frequently used in the episode, the multi-level flow of consciousness in the linear textual progression was highly explicitized. *And* in the episode can be affirmative addition or enhancement, which includes temporal enhancement, spatial enhancement, and causal enhancement. Huang agreed with Feng, Z.'s (2012) argument that the logico-semantic relations conveyed by *and* represented languages' values of indeterminacy and fuzzification. Thus, it was necessary to differentiate covert markers from explicit markers of conjunctions *and* to understand the flow of consciousness in discourse structure.

Liu, Y. (2003) applied clause complexing theory to *Wuthering Heights* and analysed a discourse of Catherine Earnshaw from clausal semantic symbols, clausal numbers' comparison, and textual organization schema to reflect Catherine's different emotions toward Heathcliff and Edgar Linton. The theoretical framework of clause complex was not adequately explained, but analyzing methods especially comparing the schema of textual organization of syntactic units from the discourse, thoroughly coordinated with Peng, X.'s (2000, p. 239) ideas that "Syntactic units refer to ideational structures of clauses and structural organizations of clause complexes. They are linearized and organized into a discourse on the surface layer and then represent semantic relations on the deep layer."

Wu, Y. (2019) compared the original and simplified (Stuart, 2007) versions of *Alice's Adventures* qualitatively and quantitatively to discover their language choices in CC and norms of simplifying language for understanding through analysis of the number of CCs, univariant and multivariant structures, parataxis and hypotaxis, and logico-semantic relations of CC. The main methods of simplifying a CC were found to be the lower frequency of high density, the lower percentage of multivariant structures, and less usage of even multilayer multivariant structures.

Li, C. (2006) first compared five Chinese poetries from the Tang Dynasty with five English poetries from the 16th to 18th century of their clause complex relationships to discover that: 1. a clause complex in English poetry usually includes more than three clauses while no more than two in Chinese poetry; 2. The proposition of hypotaxis in English poetries is even bigger than that in Chinese ones, most of which are projections; 3. More enhancement clauses and embedded clauses were employed in English poetries than in Chinese ones. Then, five more Chinese poetries with their translations from Giles & Fletcher (Lv, 2002) were contrasted that the number of clauses in each clause complex and hypotaxis was higher than those in the original Chinese version, while extension clauses were more translated into enhancement and other forms. All these diversities resulted from thinking habits, means of expression, cultures, etc., which the researcher did not study in-depth. The number of poetries and of clause complexes was not enough to show the panorama of composite textual features of poetries from two languages. (Peng, 2000)

In contrast to other genres, the clean energy news report selected from *China Daily* by Shi & Gu (2023) comprises 64 clause complexes from three texts. These encompass 22 locutions, 13 enhancement relations, 12 elaborations, 9 idea relations, and 8 extensions. According to the authors, locution projective clauses primarily indicate the opinions or observations of an expert or notable figure. They are typically employed to assess the progress and outcomes of energy development initiatives, thereby enhancing "the credibility and objectivity of scientific news reports." Additionally, they assist readers in comprehending the particular circumstances of a scientific undertaking from diverse viewpoints, thereby rendering the project more illustrative in the news report. An enhancement clause "can supplement the content of the preceding clause with information indicating the time, place, and reason of the energy transition and energy development project." It can also be used to express the manner in which a particular technique or goal is achieved. An elaboration clause is employed primarily for the provision of a detailed explication of specific "energy terms and the number of energy facilities," as well as the offering of further elucidation of statements relating to energy. The extension clause serves to augment

“the information conveyed in terms of the emphasis and influence of energy development through the use of conjunctions,” thereby increasing the overall amount of information present in the original sentence. Furthermore, it serves to elucidate the content of the energy project by reiterating the preceding clause, thereby facilitating a more perspicuous expression of the information in question.

After quantitatively comparing distributions and occurrence, and the principal markers of the clause relations between English and Chinese news discourse, Liu, T. (2017) concluded that: 1. the combination of hypotaxis and parataxis is the most common pattern occurred in both English and Chinese news discourse; 2. English news discourse used more enhancement while Chinese employed more extending in regards to expansion; 3. More than half of parataxis in English were marked with explicit markers, while the percentage in Chinese news discourse was around 10%. According to the author, these differences lie in the reasons for both languages and thinking models. The quantity of the data, however, was small, with about only five thousand words in each language, which could be improved.

For advertising commentaries, Mu, F. (2009) collected 200 English materials and analysed them from clause complex theoretical framework (ibid, p. 7; Thompson, 1996a) to give a conclusion that more than 60 percent of the clause complexes are hypotactic relations, and most of the hypotaxis is enhancement. Projection clauses are rare in the data because of their indirectness in conveying information to readers, while hypotactic enhancement performs the function of highlighting, evocation, and persuasion, thus making advertisements more outstanding and motivated for consumers.

Fu, X. (2011) clarified definitions of clause complex terms, from dependency to logico-semantic relations, from parataxis to facts, then analysed an episode from Lincoln’s *Gettysburg Address* under the framework. But firstly, she did not give any analysis of projection because of the excerpt limitation that was put forward in the article; secondly, she mixed embedded clauses with clause complexes, thus could not correctly uncover clause relations when classifying; thirdly, no further explanation about contents, emotions, aims of the episode was demonstrated after construction analysis; and we could not dig out any reason of her choosing Gettysburg Address as an object and making such an analysis.

Qi, W. (2016) also compared the logico-semantic relations, especially subtypes of expansion between Chinese and English clause complexes. He found that the extension took the major part, then the enhancement, and at last, the elaboration in both English and Chinese texts; English texts preferred elaboration, while Chinese texts used more extension and enhancement, and more explicit conjunctive expressions were employed in English to convey logico-semantic relations than in Chinese. Projection was not involved in the theses, however.

Translation

In the process of translating the English text into Chinese, it was observed that the short story *Katu* employs a high frequency of clause complexes to reflect the character traits and the author’s emotional tendencies. This observation led to the development of two translation strategies by Ge, F. (2020), which were based on an analysis of the relationship between clauses. When the taxis or logico-semantic relationship between clauses in the original text is entirely consistent with the meaning expressed by the author, the translation adheres to the original clause relationship by supplementing the clause meaning and incorporating conjunctions. Conversely, when the relationship between clauses selected in the original text does not align with the author’s portrayal of the character image and the expression of emotional attitude, the perspective transformation and interpretation or paraphrase are employed to elucidate, supplement, or even alter the relationship between the original clauses. Nevertheless, as the author observed, the translator must consider the potential consequences of their choices. These include the possibility of losing the original text’s semantic connotations or modifying them in ways that impact the author’s emotional stance and the novel’s theme. Additionally, the translator must contemplate the impact of their interpretations on the narrative’s overall effect.

In examining the capacity of machine translation to convert Chinese-English clause complexes, Xing & Xu (2023) selected 44 clause complexes as samples. These were then annotated and quantitatively compared in terms of naming sharing types (completing, changing, converting, refining) and naming sharing relationships (naming-sufficient clause, new branch, back position clause) exhibited by three translation engines based on neural networks, ChatGPT based on large language models, and expert translations. The study revealed that there is still scope for enhancement in machine translation engines in comparison with expert translations, though ChatGPT demonstrated a closer alignment with expert translations than the initial three. Furthermore, ChatGPT has superior capabilities in “naming refining” compared to expert translations. Nevertheless, it is unclear whether the findings of this study are influenced by the number of samples, whether the level of the corresponding indicators reflects the quality of the translation, and whether text translation should adhere to the principle of equivalence, which dictates that the number of clause complexes in the original text should be reflected in the translation and that the translation should contain an equivalent number of clause complexes. These topics are worthy of further consideration.

Lin, X. (2018) chose around five hundred Chinese clause complexes and compared them with corresponding English translations from 4 online machine translation sources to make clause-complex level error analysis. The results showed that these errors can firstly be classified into “morphological errors, sharing structure errors, and logical relation errors”; secondly, they can indicate grammatical diversities originating from the Chinese and English language themselves; thirdly, they can perform differently with probability in genres that encyclopedia outnumbered social news and political texts. Thus, she advocated that “long sentences should be parsed into several clauses before they are translated for the convenience of identifying naming-telling relationship” and asked for further research into how strong the clause-complex level errors and grammatical differences in the two languages were connected.

By using the same method as Lin, X., Ye, D. (2018) discovered English-Chinese machine translation errors for the clause-complex level into four categories: “context-sensitive errors, naming-telling relation errors, logical relation errors and quotation errors,” in which naming-telling takes the largest proportion; and at last, advocated an establishment for “a formalization model for data training” in machine translation.

2.2.1.2 On Projection

In addition to the aforementioned categories of theoretical research, discourse analysis, and translation, research on projection is also reflected in its application in the field of teaching.

Theoretical

Theoretical research on projection primarily concentrates on the following aspects: the projection system itself, its ideational functional model, conceptual metaphor, projection phenomenon, interpersonal meaning, experiential meaning, projection grammatical metaphor in academic discourse, multimodal analysis of projection in academic discourse, projection analyzing framework of graphic discourse, projection system of academic cartoons, the recontextualization of projection system, factual projection, rankshift of the projection system, recursiveness of projection, and so forth.

In logico-semantic relations, projection, compared to expansion or inter-dependency, is a category with closer attention from Chinese scholars, in which the earliest researcher should be Ding, J. (2000), who inaugurated profound developments for other Chinese researchers from mainly four aspects. Firstly, Ding held the opinion that Halliday’s views on projection were deeply rooted in research on parole and thoughts expressing approaches from traditional rhetoric and narratology, in which the core of focusing on direct and indirect speeches could be dated back to Plato’s *Republic III* and Aristotle’s *Poetics*. As meanings can be increased or lost in projected clauses, Halliday distinguished proposition (statement and question) from proposal (commandment and offering) in projection, in which the former includes *that*-clause and *wh*-clause while the latter is represented by non-finite clauses. Secondly, Ding found Halliday’s (1994a) deficiency in the system, though Halliday maintained consistency with Firth’s (1968, p. 24) poly-systemic methods on language research that language analysis must be poly-systemic, and any language cannot be supported by only one united system to explain all language facts. For the projection system, Ding discussed it from three subsystems, i.e., parataxis, hypotaxis, and embedding, and supplemented with pseudo-projection clauses. Parataxis has two sets of variables: proposition/proposal and utterance/thought, hypotaxis is divided into one more set: finite/non-finite, and embedding is more delicate with head/post-modifier. Thirdly, a mode pattern of disassociation and recombination processes was designed to explain hypotaxis/parataxis and utterance/thought variables split inspired by Nesbitt & Plum’s (1987) semogenesis hypothesis. Though other delicate variables have not been involved, the confirmation of probabilistic grammar and Lemke’s (1984) regarding language as a dynamic, open system by quantitative methods enable systemic linguistics to transfer from theory to practice. Thus, in the last part, discourses from three genres: comment, report, and advertisement were chosen to investigate their projection distribution. The differences are due to mutual predictability between discourse structures and discourse genres that the latter constrain and, at the same time, are constructed by the former. Report discourse prefers projection because of its dialogism; comment discourse is highly written in embedding projection for periodic sentences’ authority and loose sentences’ communication ability; advertisement discourse usually abstains from projection because that projection performs cognitive and interpersonal intrusion for facts and experience essentially, and speaker can be seen as an intruder (Halliday, 1978a, p. 144). In addition, the article emphasizes the mutual relation between the system, discourse, and context.

Zeng, L. (2000a) probed into the relationship between projection and text. In the feminine autobiography, projection helped women reveal their true sentiments and psychological activities. Quotes are more related to narrated genres like fiction or script, while reports are less compatible with communication functions than quotes. Projection verbs and nouns should be more focused when teaching students their different functions in reading texts. Lastly, distinguishing Theme from Rheme in a text would make the text easier to understand by students.

For comparative analysis between English and Chinese projection, Zeng (2000b) was the first researcher who attempted to explore the systemic functional approach to projection clause nexuses in Chinese discourse and their textual functions. Though, in the researcher's opinion, quotation marks were often omitted, especially in proposition ideas in Chinese Mandarin; no conjunction was used in Chinese hypotactic clause complex; and projecting and projected clauses would be departed by full stops, systemic functional grammar could still be a useful tool for Chinese analysis as most of its constructions had equivalence in English. Halliday's interpretation of projection had the foundation that "it takes language as an inherently dynamic phenomenon, as a system is constantly renewing itself through the familiar semogenetic process of recombination of associated variables." (Halliday, 1994a, p. 254)

Projection could also be analyzed of its experiential function and other processes except verbal and mental, according to Zeng (2002), as "speakers have available a range of ways in which they can choose to introduce 'projection' into their texts," a clause, an embedded clause, a phrase or a noun/word could serve as projection message, and a clause, prepositional phrase or adverbial word groups, quotation marks could serve as projection signals, to make experiential function feasible for analysis. But only logical function existed when the relationship between clause complexes was inspected.

According to Halliday & Matthiessen (1999, p. 48), "a phenomenon is the most general experiential category: anything that can be construed as part of human experience. The phenomena of experience are of three orders of complexity: elementary (a single element), configurational (configuration of elements, i.e., a figure), and complex (a complex of figures, i.e., a sequence)." For Zeng (2003), a sequence may also be realized by a clause and even a group in which grammatical metaphor is involved, as it expands the semantic potential of the system. The ideational metaphor reflects a shift from sequence to figure or from figure to participant and thus falls into the ideational function, as its primary effect is ideational. When the referential meaning of a projection clause (complex) cannot be reflected adequately by logical function, the meaning is likely expressed grammatically-metaphorically. Thus, a parallel analysis method is chosen to explain the meaning more "closely" and "congruently." Usually, five distinctive features can be employed to identify projection ideational metaphor: 1. Led by the first-person singular; 2. The simple present tense; 3. Mental process; 4. Hypotaxis; and 5. More spoken forms, being partially different from Halliday and Martin's evaluation criteria that were summarized by Hu, Z. (2002, p. 93-4).

Liang, L. (2004) discussed projection types (direct and indirect speech), projection contents (words and ideas), metafunction (ideational and personal functions), and related semiotic generating processes (nominalization and metaphorization) from the perspective of phenomenon category that projected clauses in direct speech differ from those in an indirect speech on personal, temporal, behavior and spatial reference because of choosing diverse phenomena reference substance; the phenomenon existence of covert participants endows embedded projection with characteristics of "naming materials" in nominalization, and also deliberation in the proposition; trans-phenomena would bring about transfer of processes, explicitness degree diversity, explicitness and recessiveness polarity of projection source, and also the consistency of projection forms; and reflection of phenomena can affect the selection of projection modes. As the phenomenon (it consists of element, figure, and sequence on the semantics level, which correspond to words and phrases, clauses and clause complexes on the lexicogrammatical level) belongs to the experiential category (Halliday & Matthiessen, 1999, p. 49), projection phenomenon reflects projection significance on semantics and inter-dependency relationship on lexicogrammatical, then projection represents a kind of logico-semantic relation. Thus, the angle from the phenomenon gives epistemological and ontological explanations to the projection of how human experience is construed through the meaning of language.

Zeng & Yu (2005) discussed projection from the perspectives of the interpersonal decomposition of grammatical resources (Berry et al., 1996; Halliday & Fawcett, 1987b). By investigating the interpersonal function of "projection signal" from the mood system, especially the modality system, i.e., propositions were analyzed with modalization and proposals were connected with modulation, the interpersonal modality scale of projection was built to demonstrate that different projection signals were chosen to mark the effectiveness or validity of proposition and to increase or decrease the execution of proposals, thus we found four scales which had been mentioned before: explicit subjectivity, implicit subjectivity, explicit objectivity, and implicit objectivity. However, this classification could not include all projection signals, which was also pointed out by the researchers. Diverse projection verbs in academic discourses, however, were listed to convey speakers' attitudes toward projection messages (Thompson & Ye, 1991), which was valuable for further studies.

Based on the theoretical framework of systemic functional linguistics, Liao, Y. (2006) advocated that distinctive methods should be adopted to analyse verbal, mental, and embedded projection. To avoid traditional dispute between transitivity and intransitivity of verbs, transitivity analysis on parataxis/hypotaxis or locution/idea etc. is suitable for verbal projection like *verb + quoting/reporting*; on Infinitival Perception Verb Complements (IPVC) from mental projection, Liao prefers to do

functional syntactic analysis (Fawcett, 1995; Huang, 1996, 1997) rather than do Equi (Rosenbaum, 1967; Fiengo, 1974) or non-sentential (Declerck, 1983) analyses; for embedded projection like *it + be + noun(s)/adj. that*, Halliday (1994a) called it a “predicated theme structure,” while Fawcett’s “enhanced theme” was accepted by Liao and thus was regarded as a “thematic building-up” (Fawcett, 1995) structure. Although several practical implications can be found in this study, the basic theoretical foundation is still consistent with Halliday’s clause complex analysis framework.

Setting Halliday & Matthiessen’s (1994) functional-cognitive frame as theoretical background, Li, J. (2007) systemically investigated the function of projection in human experience, construing on three levels: sequence, figure, and element. On the figure level, projection construes sensing and saying diversely in content, symbolizer, object, and directionality; and on the element level, projection construes four kinds of semiotic abstract (case, chance, proof, and need) things and “the various experience phenomena represented by things which are modified by projection-type qualities.” The most impressive part is on the sequence level, that is, firstly, a figure can project another figure into the semiotic phenomenon level, making it a meta-phenomenon, i.e., meaning or saying, and construes the projected figure as content (the projected idea is content on the semantic level while projected locution is on the lexicogrammatical level); secondly, the projected figure has equal status in quote projections and thus can be construed as a first-order reality (experience) while the projected figure has an unequal status in report projections and thus can only be construed as a second-order reality (meaning); and thirdly, projected proposition is actual and temporarily specific while a projected proposal is not truly existed, non-exemplary, and future-tense-tendency. This affirms what the experience-construing function of language emphasizes: the relationship between categories of experience is constructed initiatively by language and powered by lexico-grammar.

Zhang, S. (2009) generally described the theoretical framework of projection.

In the meantime, Zeng (2007) briefly explored the syntactic and semantic features of projection in academic discourses from the perspective of ideational, interpersonal, and textual metaphors. Substantial metaphor, transitivity process, and self-viewpoints projecting metaphor are mainly focused on ideational metaphor that the second one can emphasize objectivity of projection signals on the one hand and demonstrate researchers’ self-protecting consciousness and avoidance of academic responsibility on the other hand. Projections that are constructed by interpersonal metaphor can help expressers reach the aims of modal meanings: explicit subjectivity or implicit objectivity. However, the researcher did not give a detailed explanation of the research methods applied in this research.

In accordance with the grammatical metaphor theory of systemic functional linguistics and the social semiotics of multimodal discourse, Zeng & Liang (2007) sought to investigate the multimodal “metaphorical” manifestation and functional significance of projecting language in academic discourse. The authors posit that, in many instances, the use of projecting language in academic discourse employs metaphorical techniques to convey specific ideational, interpersonal, and textual-related meanings. A parallel analysis of conceptual metaphors was conducted to examine projecting clauses. It indicates that the conceptual metaphor encompasses three dimensions of transformation: process transformation, functional component transformation, and lexical and grammatical transformation. Furthermore, the authors consider that scholars may utilize metaphorical forms of verbal processes when selecting projecting clauses for the purpose of achieving objectivity, neutrality, or other vague objectives. The pairing of verbal or projecting verbs with impersonal nouns in academic discourse serves to emphasize objectivity. This is achieved by highlighting the scientific and rational nature of the reasoning employed, which is based on objective conditions. Conversely, the use of a passive projecting clause serves to obfuscate the author’s responsibility for the veracity of their judgment. This reflects a recognition of the importance of self-protection while pursuing academic innovation, as well as an awareness of the limitations of subjective perspectives in the pursuit of objectivity. The congruent projection strategy, whereby the author employs personal subjects and verbal verbs, serves to reinforce the interpersonal interaction between the author and the reader. The multimodal “metaphorical” nature of projecting information indicates that, through the foregrounding of projected information, the intentional emphasis on projected information can be placed, or the deliberate blurring of projected symbols, a specific effect can be achieved.

Zeng (2008) further studied the tense of projecting verbs in academic discourse from grammatical metaphor theory. Because mainly three tense modes could be found in academic discourse: present tense, past tense, and mixing tenses, interpersonal metaphor and textual metaphor were thus selected as research objects. The results of the study showed that in interpersonal metaphor, the present tense of projecting verbs summarized recent researchers, emphasized scientificity, objectivity, and immutability of research contents, and then helped realize the interpersonal function of tense, especially in abstracts, literature reviews, and research outcomes; past tense usually introduced research background, described experiment procedures and stressed the ideational function, i.e., the logical and experiential time that

things happened; while the present perfect made a conclusion of research outcomes; as to mixed tenses of present and past, the former could demonstrate that the researcher's confirmation of widespread effects of the research while the latter played the opposite function. With regard to textual metaphor, the same tense would usually be elected to make the text more objective, concise, and coherent. Otherwise, it was the same as those in interpersonal metaphors that mixed tenses could also imply the researchers' reservations about research outcomes or opinions. To sum up, the operation or practice mechanism of tense of projecting verbs connected with grammatical metaphor meanings can be moved forward.

In Zeng & Hu's (2007) review of academic discourse and projection, features such as scientificity, logicity, validity, and stringency rely on projection to the greatest extent because citation or reports on other researchers' achievements or ideas is a necessity in academic discourse, thus from projection structure, Swales (1990, p. 18) found two major sorts: integral structure and non-integral structure; from the function of projection, Thompson (1994b, p. 204; Thompson & Ye, 1991) concluded that: 1. To confirm the author's standpoints and establish his/her theoretical foundation, 2. To strengthen the author's viewpoints, and 3. To indicate the author's erudition; for reasons of applying projection, Swales (1990, p. 18) stated convention, intellectual property protection, genre style, persuasion, authority, etc., briefly and finally regarded it as a tool for building up the researcher's research space; and for frequency of application of projection, Hyland's (2000, p. 24) manifested that "there was far less variation in the ways disciplinary communities refer to sources, with all but philosophy displaying a distinct preference for non-integral structures." Differ from mono-modal discourse, Zeng & Hu discussed multimodal texts from figures and tablets, charts and mathematical formulas, paragraphs, code-switching, and links to generally demonstrate that projections are non-linear in especially scientific discourse and projecting information is separated from projection signals in order to achieve concise, coherent, convenient, economical, visualized, etc. effects.

Developed by Zeng & Hu (2007), Zeng & Yang (2016) constructed a preliminary image-text analytical framework of projection based on SFL and Multimodal Discourse Analysis. The basic conception of image-text projection includes mainly four elements and implementation models: 1. Projection signals consist of the vectorization of participants' situation and dialogue box; 2. Projection message is composed of text or image visual modal discourse, visual text-symbol multimodal discourse, and zero-symbol discourse; 3. Projection signals and projection messages are both inconstant but coordinated with each other that the image contents can be construed not only as verbal projection but also mental projecting process; 4. Text-image projection coexisted with mono-modality and multimodality. Then, in the framework of analyzing text-image discourses, Locution is split into Sayer, Projection Signal, and Projection Message, while Idea is divided into Sensor, Projection Signal, and Projection Message. According to the researchers, Sayer and Sensor can be zero-symbol, animate people/things, and non-animate things; projection signals are ranked by differentiation degrees from high with both dialogue box and participants' verbal situation to low with only the latter. Interestingly, comics discourses are analyzed afterward accordingly under the newly built framework to verify their feasibility and veracity.

For the method of image-text discourse analysis by SFL, Matthiessen (2007b) indicated that the projection relations can be image-projecting-text and also text-projecting-image. Based on Functional Grammar and Visual Grammar, Zeng & Yang (2019) constructed an extensive, adaptive projection analysis system for academic comics and answered two questions (the theoretical foundation of this system and the choosing model of sub-systems) from the case *Understanding Comics: The Invisible Art*. The abstract entering items (functional components and recursion system), the stratified layering relations, and the various projection elements or items (i.e., projection subjects, information, and bubbles in the same layer) made academic comics projection meaning-potential system diversified from traditional systems. It was discovered that visually projecting participants and constituent-analysis speech balloons are the main realization means for verbal projected messages; also, the linear recursion can be repeated more than two times.

In light of the distinction between two realities in the semantic construal of human experience, i.e., "between the everyday reality of our material existence on the one hand and the other hand the second-order reality that is brought into existence only by the system of language" (Halliday & Matthiessen, 1999, p. 106), Liu (2020) posits that projection serves as a bridge that connects two discourses and, in turn, two realities and two contexts, and plays a role in contextual connection in the process of discourse generation. The projection resources that construe the first-order reality have a recontextualizing effect on the projected information from the second-order reality. Furthermore, different projection resources recontextualize the projected information in different modes through functional fractals, producing different projection effects and thus realizing different discourse functions. As the process of recontextualization can be observed in the intertextuality of a given text, the use of elements or content drawn from other texts, and the borrowing of genres or activity types at a more

abstract level (Linell, 2009, p. 248), the author conducted a qualitative analysis of the recontextualization of projection from the three major components of context: field, tenor, and mode. The field dimension examines which type of field requires the implantation of other field activities and which type of field is selected to be implanted in the first-order reality context. The tenor dimension examines which tenor relations in the first reality require the implantation of the tenor relations in the second-order reality and what kind of tenor relations in the second reality context can be projected into the tenor relations in the first reality. The mode dimension examines the manner and strategy by which the projected content of the second-order reality enters the first-order reality, influenced by the configuration of the field and tenor. This is reflected in the configuration mode of the three elements of projection processes, participants, and environmental components in the projection structure. While genres such as the reportage text, narrative novel, and academic paper were referenced on numerous occasions, the author did not offer a delicate theoretical framework for the construction of the projection system across different genres within the three-dimensional context.

On the definition of “projection,” Dong & Li (2011) argued that indirect speech did not truly belong to projection as the latter (projected clause) had already changed in a clause complex according to grammar rules. But their definition of “projection” itself was based only on a Chinese dictionary, which was not sufficient at all to describe or explain the whole meaning scope of “projection.” In addition, according to the definition from Mathematics, “From the most general point of view, a projection may be defined as the representation of any given surface upon any other surface, whether planned or curved, in such a way as to satisfy certain prescribed conditions. In the representation of any on-developable surface (e.g., the sphere) upon a plane of certain errors are of course unavoidable, but any of these errors may be diminished, or even made to disappear altogether, at the cost of increasing some other.” (Thomas, 1882) Thus the change of projected clause is allowed to exist if we borrow the term “projection” from projected geometry, and the argument can be unnecessary in this sense, which will be discussed in detail later in Chapter 3.

Following Zeng, more researchers start to research projection and text. Wang, L. (2013) compared the function of projection verbs between fiction, news report, and academic discourse to re-attest that different genres would affect the employment of projection verbs, and these choices made each genre distinctive in return.

In the projection part, besides reports and ideas, another category is facts. Zeng & Liang (2012) gave a clear definition of “facts” after discussing defects and problems of past research that distinctive characteristics were ambiguously explained; taxonomy was unclear; and analyzing system was incomplete or improper, etc., then pointed out distinctive features of “facts” from non-clause complex, recessiveness or implicitness, metaphorization, and meta-function, classified “facts” into transitivity, projection relationship, and projection message, and at last described a detailed and comprehensive system framework for analyzing “facts” projection. But because “facts” can only exist in clauses as embedded form, it will not be further probed into in this dissertation.

Wang, G. (2014a) however, constructed a disparate analyzing framework on “facts,” in which transitivity processes, verbal functions, and meta-function are three main categories. Verbal and mental processes are divided into personalized and impersonalized projections, while relational and existential are all impersonalized that Zeng & Liang (2012) did not ever consider existential processes. If advantage is discussed, then the meta-function angle that “facts” can convey personal functions like appraisal or polite meaning through the metaphor of modality, ideational and textual functions like experiential or textual meaning through nominalization is a brilliant supplement, and “facts” is seen by the researcher as a kind of metaphorized idea or report in essence. But verbal function can be categorized into verbal processes like Zeng & Liang did.

Connected with grammatical metaphor, Wang, G. (2014b) also discussed the rank shift of the projection system because her viewpoint was that grammatical metaphor directly resulted in rank shift, which was decided by multi-layered and multi-functional features of the language. To be specific, she found that the rank would shift from downgrade to upgrade in interpersonal metaphor either with rank or delicacy, which emphasized and highlighted speakers’ subjectivity and cognitive non-determinacy, reflecting subtle distinction of interpersonal meaning; while it shifted from up to down in ideational metaphor by nominalization or transitivity processes, which could gradually strengthen experiential meaning and recessively project information. There are more metaphorical variants and delicate ranks between clines of interpersonal metaphor, while congruent and metaphorical forms almost correspond to each other in ideational. Rank shifts in interpersonal are to promote social interaction and guarantee communication success, which accords with politeness and desirability principles; while in ideational are to acquire the best cognitive effects by the least effort like classification, condense and extraction, which keeps to economic principles. The conversational space is broadened by interpersonal projection metaphors while it is compressed in ideational. As seldom researchers have ever paid attention to the

rank shift phenomenon in projection, this direction can be another interesting point in the future.

For the recursiveness of projection, Jia & Zhang (2015) divided it into intrinsic and extrinsic recursiveness, which indicated that the former was further categorized into recursion at the same rank and recursion across different ranks, and the latter referred to as embedded and linear recursion. For the projection recursiveness system network, the end of intrinsic projection means the end of the whole projection recursiveness, while the end of extrinsic projection means the start of intrinsic projection, thus making the projection recursion system. According to the researcher, limited recursion would be formed because of kinds of reasons in practical situations, but the reasons were not pointed out. Furthermore, Jia adopted Cardiff Grammar instead of Sydney Grammar to analyze projecting and projected clauses, and hypotaxis in Sydney Grammar was regarded as embedding in Cardiff Grammar. Though Cardiff Grammar is reasonable for explaining clause relationships, the researcher did not provide us with logical reasons for her choosing Cardiff Grammar. In addition, the division of intrinsic and extrinsic recursiveness is, in essence, the same as hypotaxis and parataxis. So, the recursiveness of the projection system network can be deeply investigated.

To give a literature review of the methodology, findings, and study trends of projection in publication, the development of each sub-domain, and also issues and future research gaps for the projection, Xuan & Chen (2020) retrieved related key terms from CNKI, One Search, Web of Science, ProQuest Dissertations and Theses, Language Behavior Educational Resource Information Centre, etc. and synthesized 66 studies in English and Chinese with a meta-analysis approach. Five major domains (theoretical discussions, language description, language education, translation studies, and professional communication) were investigated in which the first part was presented particularly in seven categories: system network, cognitive interpretation, metafunctional considerations, discourse features, syntax, grammatical metaphor, and multimodality. (ibid, p. 6) In the authors' opinion, the methodology for researching projection has not changed too much since 1977, in which more than 90% of the data chose descriptive and qualitative methods that would usually not describe the text samples. As researchers have diverse denotations of the term "interpersonal projection," they defined it as "manifestations of projection through interpersonal resources in lexicogrammar" paralleled with logical projection and experiential projection (ibid, p. 12). The suggestions of building projection networks at the verbal/phrasal rank and the textual level due to a mixture of projection and conjunction relationships were sparking. For further research directions, it was advocated that more attention should be paid to the experiential unit matter, and division should be evident for the borderline of expansion and projection (Lauranto, 2017). Additionally, applying SFL grammatical networks into the pedagogical field and then developing a unique teaching-directed grammar is of great worth. As to how Xuan & Chen reached a reliability of 93% for their coding projection, items have not been explained.

Shortly, projection is conveyed through signals and information in which projection nouns are realized by nominalization, thus creating grammatical metaphor and working for cohesion in texts as the media of transmitting information. Processes of the transitivity system can all be metaphorized, i.e., one process can be metaphorized by another. Specifically, ideational metaphor includes the transformation of processes, functional elements, and also lexis and grammar. As to projection verbs collocated with impersonal nouns in academic discourse, they function as reinforcing subjectivity for scientific and reasonable inferences and also as obscuring the authors' responsibility for self-protection and limitation recognition in passives. According to SFL (Halliday, 1994a; Thompson, 1996a), interpersonal modality is categorized as definite subjectivity, definite objectivity, non-definite subjectivity, and non-definite objectivity, in which the former two are metaphorically realized by clauses, while the latter two are non-metaphorical realized by definite modal verbs or modal adverbs. Thus, projections in academic discourses are presented as "metaphorical projections" to show their ideational (Zeng & Liang, 2007), interpersonal, and textual meanings that analyzing principles and methods of grammatical-metaphorical theory and multimodal discourses in social semiotics can explain the categories and complex realization forms of "metaphorical projection" reasonably and effectively.

Discourse Analysis

The discourse analysis of projection systems is primarily concerned with the examination of a diverse range of textual sources, including academic papers, news reports, news texts, novels, other literary works, marine accident reports, diplomatic speeches, and so forth. Given its distinctive grammatical characteristics, it is frequently employed in the examination of news discourse.

Zhao, L. (2008) expanded the study of projection from ideational function to interpersonal function by building up a theoretical model motivated by Martin & White's (2005) engagement system, Bakhtin's (1981, 1986) dialogism theory, and Brown & Levinson's (1987) politeness theory. As the projection could offer the article author a valid way to appropriately and politely consult with the referenced writers, potential readers, and diverse positions of value that they hold; "persuade readers to gain credibility"; and establish kinds of dialogic relationships between them, the analyzing model was dichotomously split into

Extra-vocalization (projecting “a certain proposition explicitly to the external source”) and Intra-vocalization (“from within the writer’s utterances, the writer refers, invokes, anticipates, acknowledges or responds to convergent or divergent voices” and establishes various degrees’ solidarity with potential readers) that the former was embodied by Endorse, Acknowledge, and Distance while the latter was divided into Pronounce, Objectify, Concur, Entertain, and Directive. Zhao found that the realization of interpersonal meanings through projection in academic texts is a cline degree with reference to “dialogic expansion or contraction,” such as: “seeking external support from the cited author, signaling apparent neutrality, distancing from the projected message, invoking dialogic partners, shortening the distance with readers, contracting dialogic spaces, etc.” Thus, the academic text should be seen as a “dynamic and ongoing dialogue” instead of “objective and voiceless” genre in the author’s opinion. But the article did not pay much attention to quantitative analysis of each category, nor did it investigate diversities in each chapter, i.e., abstract, introduction, literature review, method & data, etc. Although the model is a brief description without further or deeper classification, the advocacy of raising awareness of teaching and acquiring projection (interpersonal aspect), including readership and writers’ identity, deserves to be mentioned.

In order to explore the use of projection in the writing of summaries by college-level English language learners in mainland China, Chen et al. (2022) selected a corpus comprising three levels of students (91 in total) from a university. The students were enrolled in English language majors and were distributed across three levels: 35 in Level 1, 28 in Level 2, and 28 in Level 3. A mixed-method approach was employed, utilizing the UMA corpus tool for quantitative recording. The task-based research comprised an argumentative essay, which was read by the participants, who were then required to write a summary of between 300 and 500 words in a total of 80 minutes. The findings indicate a notable increase in the frequency of both locution and idea from the first to the second year, followed by a slight decline in the third year. The number of verbal and mental processes employed in projecting clauses was notably higher in the second year than in the first year and also higher than in the third year. In particular, the most frequently used words by students across all three years were *suggest*, *argue*, and *say*. However, there was a notable decline in the frequency of *say*, *mention*, and *think* as the level increased. The data indicate that first-year students tend to utilize straightforward projecting verbs, such as *say* and *think*. In contrast, second-year students employ a narrow range of projecting verbs in quotations, exhibiting minimal lexical or grammatical variation. This suggests that second-year learners have a more expansive repertoire of projecting verbs but tend to utilize them repetitively and inappropriately. However, third-year students employed a considerably more extensive range of projection verbs in their summary writing and interpreted projection at varying levels as some sophisticated reporting verbs, such as *harbor* (*the view that*) and *conceive*, which are exclusive to the writing of third-year students. In general, students in their second and third years of study have developed a conceptual understanding of the use of projection in academic writing, while those in their first year of study demonstrate a relatively weaker awareness of this concept. Quantitatively, there is no clear indication of advancement in the use of projection in students’ writing across the three years of university study. Qualitatively, however, it can be observed that as the level of the learners increases, so does the accuracy and appropriateness of their use of projection. However, the conditions under which projection is chosen in each subcategory have not been demonstrated. As the authors have noted, the larger sample size, change in writing task, and longitudinal rather than cross-sectional research may yield different results.

Chen, M. (2005) discussed the projection meaning of English verbs from the angle of the aspect that the meaning of the verbs could affect the authenticity of projected messages and showed speakers’ attitudes toward projected information. Non-finite propositions and proposals could both have impacts on choosing aspects of projection verbs in embedded clauses, and discourses of the same type within diverse contexts would use different projection relations, which correlated with the meaning of verbs to a large extent. To be specific, for example, verbs in proposals preferred exactly to future tense rather than others in propositions; the aspect of β -verb in projected clauses would usually be perfect in both propositions and verbal proposals. If the propositional statement is expressed through non-finite clauses, the aspect of projected verbs would be imperfect; if it is the condition of interrogation, then verbs would be perfect; so would propositional offer, etc.

Marginal research is from Qiu, Y. (2007), who examined scare quotes in the journalistic genre from the theory of evaluation functions. Instead of adopting Halliday’s framework of projection, the researcher first chose Martin & Rose’s (2003, p. 52) classification of projection, which included four primary kinds from the angle of projected elements: projecting clauses, names for “speech acts,” projecting within clauses and scare quotes; then delineated “scare quotes” that quotes were partly originated from other people’s words especially when the expresser wanted to push away the evaluative buck, though the provenance had not been stated. Obvious were the elements of “scare quotes,” which were cited words or phrases, which is not the emphasis of this dissertation. Besides analyzing “scare

quotes” of journalistic discourse from Martin & White’s (2005, p. 134) evaluation system, especially the engagement system, the researcher also inspiringly gave explanations on how to infer projection from the evaluation function by Halliday & Hasan’s (1989, p. 45-7) context theory: the context of situation, intertextuality, and context of culture. The results showed that “scare quotes” enabled the author in the news report to make clear or to disown the evaluation for objectivity and to express attitudes such as acknowledgment (approval, neutrality, and disapproval) and distancing (indefiniteness and doubt).

Based on Ding’s research, Hao & Li (2015) chose three reports on the topic of the US troop deal in Australia in 2011 from *The Guardian*, *China Daily*, and the *New York Times* each to reveal ideology differences between China, US, and Australia in news discourse by analyzing projection clauses. Though the amount of applying projection is no more than one hundred in each discourse, results still show that report objectivity will be opposite from ideology tendency as projection source varies.

Coincides with Xie, L.’s viewpoint that genre and text can mutually restrict each other, Li, L. (2010) compared clause complexes from three metafunctions in news discourse and fiction in *The Times* that the former tend to use hypotaxis and locution from an ideational function perspective and use more proposal and proposition, in fact, projection part from the interpersonal function angel, and multiple themes instead of single themes are more chosen from textual metafunction, to give a conclusion that wider distribution of hypotactic projections makes the discourse more formal and complex, and “provide a good effect on making the news dynamic and dialogic.”

As frequent usage of projection in news discourse strengthens the accuracy, objectivity, impartiality, formality etc., of its discourse, Xie, L. (2011) discussed the relationship between genre and text structure by analyzing verbal and mental processes of projection structures, including experiential, logical meaning, projecting verbs, and syntactic structures to draw a conclusion that “text/discourse restrains the choice of projection clause on the one hand and the choice of projection clause builds the text/discourse on the other hand.” On fact projection in news discourse, the researcher held the view that it enabled writers to be invisible and thus diminish their “responsibility for the proposition.” And by analyzing the thematic structures of these projection clauses, the dominant theme is multiple Themes, in addition, a large number of textual Themes made news discourse “more formal and complex.” As a result, the widespread availability of hypotactic projections means dynamic and dialogical discourses where an active interaction between writers and readers is constructed.

By comparing ten news discourses from *The China Daily* and 11 texts from the *Washington Post* on the topic of the commemoration of the 70th anniversary of the Chinese People’s War of Resistance Against Japanese Aggression and the World Anti-Fascist War, Guo, X. (2016) found that Chinese and American news reports had similarities in choosing more paratactic projections than hypotactic and even embedded projection that the former “describe the event or the public’s opinions vividly and truthfully by retaining the exact wording of the reported on the spot”; verbal projections were more favoured over mental ones that *say* and *think* are principal markers in which *say* implied “what the reporter said without referring to the purpose or the way of speaking”; and propositional outnumbered proposal projections lied in that the primary purpose of news discourse is to inform people of recent information through statements and questions rather than goods or services. But they differed from each other in three aspects as Chinese news reports preferred more overt projecting sources, positive projecting verbs, and unmarked themes than American texts to demonstrate their opposite purposes of dissemination that Chinese media attempt to describe the significant events and “convey their responsibilities for the world peace” while American media incline to make clear of the historical background of the events and represent their “suspect and disbelief for the announcements of China government.” Thus, a thorough analysis of news discourse from three metafunctions of projection was fulfilled within limited samples.

In investigating the choice of projections in English-language reports from *The Navy Times* and *The China Daily*, Ren, M. (2020) observed that while both media outlets included projections in their respective reports on the South China Sea issue, the projections utilized were primarily hypotactic, paratactic, and embedded projections. However, despite the differing projection styles, there was no discernible difference in the selection of projections between the two media outlets. The distinction lies in the differing preferences of the U.S. and Chinese media. The former tends to utilize propositional projection, whereas the latter favors proposal projection. Additionally, the U.S. media frequently employs unmarked Themes, such as ordinary or proper noun phrases, whereas the Chinese media favors marked Themes, including prepositional phrases as adjuncts. In light of these observations, the author concluded that both reports are highly subjective in nature. And it is evident that the Chinese side is seeking to influence readers’ attitudes and behaviors through the strategic use of proposal projection.

Jia & Zeng’s (2021) study, based on systemic functional grammar and combined with the study of news language ontology, examined and distinguished the projection sources in news discourse from three directions: self-nature, specification, and individuality. The author posits that the projection sources can be classified into two categories: those belonging to the “missing” and those concerning the

“self-nature.” The latter category encompasses both the “self” and “others”. The term “self” encompasses both the first-person singular pronoun *I* and the first-person plural pronoun *we*. The pronoun *I* can represent either the reporter or other participants in the discourse, whereas the pronoun *we* can refer to either the “audience-inclusive” or “audience-exclusive.” “Specifiable” and “non-specifiable” are the two categories that distinguish “others” from the perspective of “specification.” This, in fact, implies the projection source. The “specifiable others” are divided into three categories from the perspective of “specification”: “fully specified,” “partly specified,” and “unspecified.” From the perspective of “individuality” or “consciousness,” they are divided into three categories: “conscious individuals,” “conscious non-individuals,” and “products of consciousness.” Nevertheless, the author has not conducted a comprehensive investigation into the circumstances surrounding the absence of projection sources, nor has he addressed the question of why these sources are missing. The author has not addressed the categorization of the situation of “non-specifiable others.” With regard to the pronoun *we*, the author posits that it is employed in contexts that do not encompass the audience. Indeed, it is unclear how the projection source can be excluded from the audience when it is, in fact, the subject. The question thus arises as to whether this is to be understood from the perspective of the projected or from that of the reporter. It is, therefore, worthwhile to engage in discussion regarding this classification. From a structural perspective, this study elucidates the concept of “projection.” From a functional standpoint, it is worthwhile to examine the role of projection sources within clauses and the grammatical and rhetorical functions they serve. Does the impact of disparate projection sources on the objectivity, authenticity, and spreadability of news reports exist on a continuum? What is the function of the stance or attitude expressed by the projection sources? It would be beneficial to ascertain whether this projection source system network is also applicable to other genres. In any case, the author’s classification discussion on clarifying the individual consciousness of others can be used as a reference point for further analysis.

Agreed with Halliday’s (2001a) views that “Discourse is a semantic concept; and a mutual-predictable relationship exists between discourse structure and discourse genre, in which the former construes and is restricted by the latter,” Wang, J. (2008) chose each of ten English discourses from the report and the law to explore their diversity of projection usage and reasons. Wang found that more paratactic and hypotactic projection were applied in English report discourse for panoramically present discourse voice, while verbal projection and embedded structures were preferred in law discourse for pedant and authority, which is similar to Ding’s result. But statistical figures of projection frequency in the article are peculiar with non-projection in both genres, which may affect statistical results as the number of chosen texts has not reached the minimum (30) for testing. In addition, the length of each text is unknown, which makes the result and conclusion suspect.

For projection in literary work, Ding & Shu (2013) inspected Alice’s speech and mental processes in *Alice’s Adventures in Wonderland* quantitatively and qualitatively from three sub-systems: projection levels, modes, and verbal function to “uncover the charm of the childly mentality and discourse as well as the descriptive capability of projecting theory.” A verb cline was found in discursive verbs, thus forming a hierarchy in a fairy tale by investigating projection levels; four intermediate modes together with other conventional ones formed a projecting continuum that childish discourse, the expression freedom of inner thoughts, and the intervention degree of the author were demonstrated; except proposition and proposal, other sub-categories like minor clauses were also calculated to analyze Alice’s verbal features. The research elicited that the author was inclined to use direct projection and multi-modal continuum projection to describe Alice’s verbal and mental activities, which made Alice a vivid and lively image in fairy tales. Based on Halliday & Matthiessen (2004, p. 434 - 4, 465), Halliday (2000, p. 271), and Thompson (1996a, p. 208-9), the researchers built up a conclusive and useful theoretical framework for projection analysis, which enabled a quantitative comparison and confirmed Nesbitt & Plum’s (1987) research outcome. But the meaning continuum of projection verbs was only listed in a diagram instead of adequately explaining.

The theory of projection could not only be applied to analyse fiction, academic papers, and news reports but could also be used to focus on children’s literature. Thus, Chang, L. (2014) chose *Charlotte’s Web*, written by White (1952), as the object and discussed it qualitatively and quantitatively from three metafunctions, i.e., ideational, interpersonal, and textual. It was found that: 1. Paratactic projection occupied the largest proportion, and then hypotactic projection, the embedded projection was least used; 2. Verbal projections were far more favoured over mental projections from the perspective of ideational metafunction that the speakers’ “original words” were kept well and the story was more “vivid” in this way; 3. More propositions were found than proposals regarding interpersonal metafunction in the discourse; and 4. Simple theme projections, which made the story easier to understand, were mostly used, and “constant thematic progression pattern is favoured than derived and simple liner thematic progression, which is characterized by a series of statements with one theme but different rhemes.” It is limited that a parallel comparison could be made between *Charlotte’s Web* and other children’s literature.

By building up a corpus which included 70 texts of *Marine Incident and Accident Reports* (MIAR) from 2002 to 2010 for projection analysis by using computer software and program: Perl (short for Practical Extraction and Report Language, created by Larry Wall in the mid-1980s for string manipulation and text processing) and AntConc by Cao, C. (2011), it was found that hypotactic locutions took a large proportion, which was “in accordance with the stylistic features of MIARs in that the long sentences are largely used and there is no lack of direct quotes of crewmembers’ recalling of the incidents and accidents although they are less than type of hypotactic projection, which however takes a proper portion in order to guarantee the authenticity of what the witnesses say and restore the picture of the accidents as possibly as the narrators could.” In addition, the most frequently used projecting verbs were “inform, state, say, instruct, tell, etc.,” which helped narrators to take a neutral stance toward what were projected for “truth” and “reality” of the accidents. And most of these projecting verbs would choose past tense to “prove and enhance the authenticity and accuracy of the reports” and “to keep align with the tense in the whole marine accidents reports.”

By comparing discourses from news report and marine accident report, Cheng, H. (2014) found that the number of projections in the latter was far less than that in the former, thus giving a conclusion that the employment of projection in their two discourses was affected by their style or genre at one hand; using more projections in news report would enable discourses more vivid and visualized, thus constructed characteristics of diverse genres at the other hand.

Following the former two theses, Tang, L. (2016) randomly chose 30 American Diplomatic Speeches (ADS) culled from the official website of the United States Department of State (<http://www.state.gov>) dated from May 2014 to September 2015 to do projection analysis and found that speakers preferred to simple themes in making speech clear and distinct, helping listeners “easily get point.”

Translation

The translation study of projection mainly concentrates on literary works and usually on comparative analysis of Chinese and English texts.

In studying the projection language in the English translation of *The Analects*, Hu, H. (2024) continued to utilize the methodology of discourse analysis to qualitatively analyze the projection language in Xu Yuanchong’s English translation. This analysis was conducted from three perspectives: lexicogrammatical, meta-functional, and contextual. At the lexicogrammatical level, the original text of *The Analects* is structured as a single unit comprising a projecting clause and a projected clause. In contrast, the translation employs a variety of methods, including the use of *a projecting clause + a projected clause*; *a projected clause (first part) + a projecting clause + a projected clause (second part)*; *a projected clause + a projecting clause*; and *an implicitly projected clause + a projecting clause*. At the meta-functional level, the weakening of the ideational functional significance of the projection structure renders the translation more acceptable to modern readers, and it allows for the construction of interpersonal meaning, thereby highlighting the meaning of the discourse. Moreover, the employment of diverse thematic structures and thematic progressing modes guarantees coherence within the discourse. From a contextual perspective, the multi-projecting structure of the translation represents a modern transformation of tenor and mode while maintaining field equivalence. Accordingly, this deliberate divergence from the source text contributes to the translation’s popularity. However, the study only compared the translated version with the original version in terms of the locations of the projecting clause and the projected clause, as well as the changes in the thematic and rheme structures. The investigation did not extend to an analysis of whether the seven types of projection (locution or idea, hypotactic or paratactic, propositions or proposals, facts, etc.) were still translated equivalently.

Zeng (2016) also discussed the logical functional equivalence of projection between *Lunyu* (*The Analects of Confucius*) and its translation text. As suggested by Fawcett (2010), the theoretical framework for describing forms and meanings of English discourses provided by Systemic Functional Grammar was not yet perfect and needed improvement, especially for those on higher levels, and many researchers had started working on trans-clause complex projection (Halliday & Matthiessen, 1999, 2004; Halliday, 2008b; Martin & David, 2003; Ding, S., 2007), discussing meaning scales such as phase sequences, sequence of description, activity sequence, etc. in discourse, the definition of projection paragraphs was put forward: “a meaningful unit larger than a clause complex with continuously projection features (logico-semantic relation and interdependency).” And four kinds of projection units (sequential combination) that are larger than clause complexes thus were proposed for analysis: 1. Trans-clause complex projection: projected messages were extended to another clause or clause complex and embedded in the same paragraph. 2. Projection paragraph: one paragraph. 3. Trans-paragraph projection: projected messages were extended to another paragraph. 4. Clause complex projection clusters: two or more continuous and complete clause complexes existed in a paragraph. Contrasting the original *Lunyu* with Ku, H.’s (1989) translation text *The Discourses and Sayings of Confucius*, the

functional meaning was not equivalent in each clause or clause complex but in the whole paragraph on a trans-clause complex level. A conclusion was finally drawn based on Halliday's (2001b, p. 17) view that the higher language units were, the more equivalent functional meaning would be in comparing Chinese and English translations.

By employing the corpus investigation tool AntConc, Li, N. (2021) conducted a comparative analysis of the cognitive projecting verbs utilized in the two English translations of the Chinese classical novel *Hong Lou Meng*. This analysis aimed to elucidate the translators' (Yang & Gladys with Hawkes & Minford) strategies and translating styles, as well as to demonstrate the role of context in interpreting the source text and informing translators' choices regarding cognitive projecting verbs in the target language. Among the mental projecting verbs employed in translation, some solely describe the action without indicating what is being thought, and there are also numerous verbs that directly or indirectly convey the characters' thoughts. With regard to the two versions, it can be observed that alternative methods to literal translation have been employed to elucidate the mental projecting verbs in accordance with the context. Furthermore, both Yang and Hawkes have altered some direct thoughts in the source text to indirect thoughts, thereby enhancing the narrative's fluidity.

Teaching

To apply projection theory in teaching English reading texts after Zhao, W.'s (2006) advocating, Wang, X. (2011) put forward a practical model which combined the macro perspective with a micro perspective that, from top-down, communicative purpose determines the text genre and then decide the choice of projection types, which in turn determine the exact genre of each text in addition; from bottom-up, grammars or aspects like "nominal group complex with a that-clause," "verbal group complex," "quote/report distinctions," "the choice of the thematic structure," etc. should also be taken into account for suiting classes with a large number of students with diverse English proficiency levels. As this model is constructed on four texts from two genres only, a more delicate or expanded system can be established and tested on a large corpus with multifarious genres.

2.2.1.3 On Expansion

The expansion research is primarily comprised of theoretical inquiry and discourse analysis.

Theoretical

As Halliday (1985a, p. 213 - 73) classified parataxis and hypotaxis into primary and secondary clauses in which the former included initiating and continuing, respectively, while the latter covered dominant and dependent clauses, Zhu, Y. (2005) considered initiating and continuing to be informative function but dominant and dependent to be grammatical function. Especially on expansion, Halliday's definition that "the secondary clause expands the primary clause" is filled with doubts; thus, Zhu advised a two-way expansion instead of one-way from the perspective of transferring information, which coincided with Hu, Z.'s (2005, p. 199) view that all clause complexes are mutual dependent relations: one clause depends on another or they depend on each other to convey logico-semantic meanings. In addition, connecting with Halliday's Theme Theory, Zhu advocated that clauses can be divided into thematic initiating clause and theme-continuing clause, in which the latter can be subdivided into an initiating clause with a message-linking function and a continuing clause with a message-expanding function in order to solve contradiction in clause classification just as what Chen Xiaotang has investigated above. Though dichotomy may not fully cover all clause conditions, the angle from theme theory can bring forth new ideas for clause complex classification.

Yang, S. (2014) probed into expressions of modality in the clause complex of Chinese to certify that they could serve as "conjunctive markers" and thus manifested logico-semantic functions, particularly of expansion, "expressing modal meanings and simultaneously constructing clauses into a clause complex." Modal verbs of compulsion such as "必须"(must), "得"(have to), "不能"(cannot), "不准"(not allow) could be used to show the relations of exposition between clauses in the clause complex that 'the second clause restates the meaning of the primary one in different wording' because of their polarity diversity; modal adverbs and verbs of possibility like "或许"(perhaps), "也许"(perhaps), "兴许"(perhaps), "可能"(possible/may), "可以"(may) could function as conjunctions like *or* and *alternatively* in paratactic extension and "signalled the alternative relationship"; the modal adjunct of probability "大概"(probably) could be employed repetitively in parataxis to show the additive relation; the modal verb "情愿" (*would rather*) could serve as the marker of replacement with the conjunctive adjunct "也不"(than); and modal verbs of compulsion and expectation such as "必须"(must), "得"(have to), "不许"(not allow), "不准"(not allow), "应该"(should), "要"(will) connected with "否则"(otherwise), "不然"(or) to indicate that the results in the latter clause would not realize without conditions being met in the former one, basing on the fact that "laying an obligation is to make something necessary" (Palmer, 1990, p. 8). Hence, expressions of modality could be regarded as "marked" markers of logico-semantic relations, which

widened researchers' horizons of the "expansion" knowledge.

Discourse Analysis

While a qualitative and quantitative analysis of Chinese prose *Rush's* original text and translation was conducted by He, T. (2022) from the perspective of logico-semantic relations, the source of the translation was not elucidated. Furthermore, given the brevity of the text and the limited number of clause complexes (39 in the original text, 36 in the translation), a quantitative comparison is inadequate for illustrating the nuances of language differences. Secondly, an in-depth discussion of the reasons for the discrepancies between the translation and the original text, as well as an analysis of the specific logico-semantics that differ between the two, is lacking. Furthermore, it is yet to be determined whether this article accurately represents the linguistic characteristics of Chinese prose. Accordingly, ascribing the discrepancy between the translation and the source text to the paratactic and hypotactic characteristics of Chinese and English, respectively, does not fully account for the observed differences. Moreover, the act of translation is contingent upon the translator's perspective and approach.

Being different, Guan, L. (2020) explores the characteristics of the CC (expansion specifically) in Chinese¹⁰ and English (Pau et al., 2005) high school physics textbooks for knowledge construction. She explained that English texts have more hypotaxis (189 vs. 74) while Chinese texts have more parataxis (178 vs. 99) because the latter language is linear, meaning potential with iconicity of word order and relatedness of topics that clauses embodying the same topic are usually connected explicitly or implicitly by logical conjunctive and then are organized as parataxis while the former language is formally inclined that clauses are combined by explicit markers for complete structure and clear logics. More colons would be found in Chinese books for paratactic elaboration; and commas for paratactic extension while more *and* in English textbooks. In enhancement, *if* for conditions, *when* for temporal, and *because* for reasonable functions are highly applied in English physics books; characteristics conveying the same meaning in the Chinese language work correspondingly. Thus, physics knowledge was reified, especially for the time or condition that events happen were recorded; and was deducted and concluded for cause-reason, and manner categories that enhance logic constructions were reinforced.

Connecting qualitative analysis (theory of three expansions) with quantitative ways (SPSS and AntConc), Xue, J. (2011) studied the frequency of these features from six annexes of *MARPOL 73-78 Convention* (marine law). It is discovered that in clauses, enhancement takes the major part, and the number of finite clauses overweighs that of non-finite clauses; while in embedded clauses, elaboration is more than the other two, and non-finite clauses take a higher percentage. However, how this study can help people better "understand and grasp the maritime convention in the integral" was not obvious or insightful; and now that Xue had confirmed Chen, A.'s (2000) viewpoint that "the formality of a text is relevant to the numbers of clauses in the clause complex and that of embedded clauses," how can the relevance be measured or quantized was not unfolded for more discussion. A sparkling and inspiring issue on implicit indicators for logic relationships in clause complexes was freshly proposed.

Based on the model for image-text relations proposed by Martinec & Salway (2005), Zhong, W. (2020) investigated English narrative picture books that have won the *Caldecott Medal*¹¹ from the multimodal discourse analysis to dig out contained image-texts relations, analyzed the influence of image-texts interaction on translated versions, and proposed corresponding English-Chinese translation strategies that elaboration, extension and enhancement relations should be paid attention to instead of cutting images apart from texts. Regarding elaboration, exposition (the meaning of images is the same as that of texts, and they have paratactic relations) and exemplification (the meaning of images is diverse from that of texts, and they are hypotactic relations) can be distinguished for translation. The angle and the genre are of novelty, which can absorb broader theoretical scopes, including projection.

2.2.1.4 Related Research

After the division of two kinds of recursiveness in functional grammar, linear recursiveness and embedded recursiveness (Huang, 1999, p. 166), He, W. (2002) held the view that the former connected

¹⁰ 课程教材研究所,等. (2007). 普通高中课程标准实验教科书物理必修 (1). 北京: 人民教育出版社. 第四章: "牛顿运动定律".

Zitzewitz, Elliott, Haase, et al. (2005). *Physics: Principles and Problems*. USA: The McGraw-Hill Companies, Inc.

¹¹ The *Randolph Caldecott Medal*, frequently shortened to just the *Caldecott*, was named after English illustrator *Randolph Caldecott*. It annually recognizes the preceding year's "most distinguished American picture book for children," and is awarded to the illustrator by the *Association for Library Service to Children* (ALSC), a division of the *American Library Association* (ALA). The *Caldecott* is considered one of the most prestigious American children's book awards, together with *Newbery Medals*. To be eligible for a *Caldecott*, the book must be published in English in the United States first, and be drawn by an American illustrator. (Excerpted from Wikipedia, https://en.wikipedia.org/wiki/Caldecott_Medal)

two units which were in the same rank and were paratactic, however, the latter connected two units in the same or diverse rank in hypotactic relation. Although Thompson (1996a, p. 23) considered embedding as a common principle that a unit was expanded by including other units from the same or higher ranks, which was different from Halliday's group element, and Fawcett (1996, p. 309) & Huang (1999, p. 121-3) tended to categorize dependency into embedding, He considered embedding to be stratified from both word groups/phrases and clauses, and a structure to be regarded as a clause complex, given that the embedded element was a complete clause.

2.2.2 Research of the CC in Other Countries

Given the paucity of research on clause complexes in countries and regions outside of China that does not focus on the clause complex system itself and the rarity of research that focuses solely on the projection and extension of its subsystems, this section mainly reviews previous research from the content aspect, namely theoretical research, discourse analysis, and translation.

2.2.2.1 Theoretical Research

The author's collection of 65 papers comprises only 15 theoretical studies centered on clause complexes and two analyses of projection subsystems.

Clause Complex

The theoretical analysis of clause complexes is predicated on several fundamental principles, including the examination of clause relations, the probability of subsystem category choices, the selection of subcategories based on the recursion level of clause complexes, systematic research combined with corpus analysis, the clarification of related concepts, and the corpus-based theoretical analysis of clause combinations. Additionally, the Theme and Rheme of clause complexes, the experiential realization of clause complexes in court texts, psychological cognition, and systematic research on clause complexes in Chinese grammar are also integral aspects of this field of study.

Most of the relative clause complex is analysed from the relationship between the antecedent (A) and the relative clause (RC) that the restrictive one denotes that the two have semantic scale or cline of restriction, and the non-restrictive clause (complex) means that it does not include the range of denotata that restricted by the antecedent. So Rydén (1974, p. 542-5) proposed a schematization of the inter-clause dependences system of the relative clause complex that three kinds of relations, i.e., the A: R (as clause-linker or a clause complex), A: RC, and AC: RC should be considered on account of clauses closeness and contents boundness. Though the RC complex discussed above is seen as nominal embedding in systemic functional grammar, which is excluded from this research, Rydén's division can be insightful for clause complex relation exploration.

For Halliday (1993/2005b), "the frontier between language and cognition becomes a stratal boundary between grammar and semantics, or wordings and meanings; while the frontier between langue and parole becomes an instantial boundary between the system (of grammar or semantics) and the instance (of wording or meaning)." (ibid, p. 248) Meaning is shaped by three dimensions of historical development: the evolution of the linguistic system, the unfolding of the individual texts, and the growth of the human infant into adulthood (which is referred to as "phylogenesis," "logogenesis," and "ontogenesis"). (ibid, p. 251) In light of these considerations, it becomes evident that there are two principal "historical perspectives on register variation in language." On the one hand, a register may be conceived of as "a distinct sub-system" that evolves, as exemplified by "the language of science." On the other hand, a register may be viewed as "an accumulation of related texts" that unfolds over time, as evidenced by "the discourse of science." (ibid, p. 263)

As Firth observed, the system "gives value to the elements of structure." "It specifies the oppositions, or sets of alternatives, to which a defined place in structure provides the condition of entry." Halliday's objective was to propose that an intrinsic characteristic of "any linguistic system is the relative probability of its terms." (Halliday, 2005c, p. 133) The frequency of occurrence of a given term within a text can be considered an instantiation of the probability of that term occurring within the system under consideration. In order to understand language in probabilistic terms, the grammar must be paradigmatic, that is, it must be capable of representing language as a matter of choice. This is because the probability is the probability of selecting one option over another. The "probabilistic model of lexicogrammar" provides a means of explaining register variation. The two essential vectors are undoubtedly "those of instantiation (the relation between the language as a system and the instance, or language as text) and realization (the relation between strata, relating to semantics, lexicogrammar, and phonology)." This is what distinguished by Saussure as "langue and parole" regarding the former and "signified and signifier" as to the latter. Concurrently, Mathesius proposed the notion of "oscillation," which refers to linguistic variation that is "both diatopic and diachronic." A probabilistic interpretation of the system effectively

“combines the Saussurean and the Mathesian perspectives.” (cf. Halliday, 2005d, p. 60-61)

Therefore, a register can be defined as a tendency to select specific combinations of meanings with particular frequencies. This can be presented as the probabilities associated with grammatical systems, provided that “such systems are integrated into an overall system network in a paradigmatic interpretation of the grammar.” It should be noted that the terms “system” and “instance” are not mutually exclusive; rather, they represent a further aspect of complementarity. (cf. Halliday, 2005a, p. 66) From the perspective of the “instance” observer, the system represents the potential with associated probabilities. Each instance is inherently “unpredictable,” yet the system appears to maintain a consistent state over time. From the perspective of the “system” observer, each instance contributes to the redefinition of the system, albeit in a minor way, either maintaining its current state or “shifting its probabilities” in a particular direction. However, “the transformation of an instance into a system can only be observed through the technology of the corpus,” which allows for the accumulation of instances and the monitoring of “diachronic variation in their patterns of frequency.” (cf. *ibid.*, p. 67)

In introducing the work of Nesbitt and Plum (1988, p. 21), Halliday observed that an examination of a sample comprising 2,733 clause nexuses was conducted to “investigate the internal relationship between two systems within the grammar: interdependency (parataxis/hypotaxis) and logico-semantic relations (expansion/projection, and their sub-types).” The findings confirmed a strong connection between parataxis with locution (“direct speech”) and hypotaxis with “idea” (“indirect thought”). (Halliday, 2005c, p. 141, see also 2005d, p. 56)

Once the probability profile of a given grammatical system has been established, “a measure of the information generated by that system” can be derived. (*ibid.*, p. 67) This may be achieved by employing the formula proposed by Shannon and Weaver (1963 [1949], p. 8-16):

$$H = - \sum p_i \log_2 p_i$$

In which p_i represents the probability of each term in the system when considered separately. The value of H (information) varies from 1 (when all terms are equiprobable, i.e., 0.5: 0.5) to a minimum of zero. To illustrate, when the ratio of two terms is 99: 1, the H value reaches a mere 0.08. While R (redundancy) is defined as $1-H$ (1 minus this information value), a measure of the skewness of a system, it can be seen that as skewness (departure from equiprobability) increases, so does redundancy. This is because as the probability of each term in the system differs, the information value, H , decreases. (Halliday, 2005d, p. 68, see also 2005c, p. 138)

In accordance with Sandy’s methodology regarding typology, e.g., the concept of “clines of categoriality” as an alternative to dichotomies (Hopper & Thompson, 1984) and “the parametric or systemic framework developed in the modeling of transitivity” (Hopper & Thompson, 1980) for dealing with clause combining; explanation of clause combining features, especially “the placement of dependent clauses” termed as the “flow of information” in discourse; “investigation of clause combining” related to the “flow of events” in discourse; and “interpretation of clause combining as the grammaticalization of the rhetorical-relational organization of text,” Matthiessen (2002) adopted “a holistic approach” rather than that based on Cartesian Analysis (cf. Capra, 1996), in locating clause combining “within the total meaning potential of the linguistic system” as well as “specifying this location in terms of the dimensions that define the organization of this meaning potential.” (Matthiessen, 2002, p. 237-8)

The neighborhood of clause combining in strata includes rhetorical organization as it “realizes patterns of rhetorical relations in text” when it is looked at “from above,” contains tone sequence as it is realized by certain sequences of tones when looked at “from below,” and encompasses mainly three aspects when looked at “from around,” which are (i) the textual systems of conjunction in providing “a cohesive, non-structural alternative to clause combining,” and of theme in “managing the flow of information in combinations of clauses,” (ii) the experiential system of transitivity in complementing “clause combining in the construal of sequences by construing the organization of the processes linked in a sequence” and in competing with “clause combining in the construal of circumstantial meanings of time, cause, condition etc.,” and (iii) the interpersonal system of mood in offering each clause that is combined in a clause complex a “dialogic” or “negotiatory status of varying degrees of arguability.” (cf. *ibid.*, p. 239)

As the sample of texts increases in size and registerial balance, the accuracy of estimates regarding linguistic categories is likely to improve. Matthiessen conducted a lexicogrammatical analysis of text, which entailed the counting of taxis and logico-semantic type from a modest sample of spoken and written Australian, British, and American texts. The counting results indicate that the most prevalent combination of clause complexes is that of hypotactic enhancement, while the least prevalent combinations are those of paratactic projection of ideas and hypotactic extension. (*ibid.*, p. 250) Enhancement is consistently a major factor as it exhibits a minimum of approximately 30%, whereas hypotactic extension is invariably a minor one when it demonstrates a maximum of approximately 3%.

Such differences in orders of magnitude are likely to be manifested all along the cline of instantiation. (ibid, p. 257)

Matthiessen (2006) was the first to consider the depth of nesting (related to clause complex recursion) in investigating the choice of taxis and logico-semantic types. From the 6,536 annotated clause complexes, the frequencies of instantiation of parataxis versus hypotaxis at the different levels of analysis were calculated. The results indicate that parataxis and hypotaxis are nearly equally prevalent at the first level of analysis. However, as the level of nesting increases, these two taxis diverge. Parataxis becomes more prevalent, while hypotaxis declines, although it exhibits a slight reversal on level 6. This indicates that the choice of taxis is not uniformly distributed as the levels of nesting become more complex in the development of clause complexes. “The more global the tactic pattern is, the more likely it is to involve parataxis rather than hypotaxis.” (ibid, p. 118) However, it is unclear whether the same result would be obtained if the number of clause complexes was to exceed ten thousand. Does this testing provide an effective index for measuring the diversity of genres and registers in the choice of clause complexes?

With regard to the choice of logico-semantic types and variants in accordance with the depth of nesting, an examination of the figure created by Matthiessen (ibid, p. 119) reveals that:

on Level 1, enhancing is 36%, extending is 22%, elaborating is 20%, locution is 11%, and idea is 11%; on Level 2, enhancing increases to more than 40%, extending decreases to 20%, elaborating is 20%, locution increases while idea decreases; on Level 3, enhancing increases, extending keeps 20%, elaborating decreases, locution increases, and idea increases slightly; on Level 4, enhancing increases, extending increases, elaborating increases but is smaller than 20%, extending decreases and is smaller than 10%, idea slightly decreases; on Level 5, enhancing decreases to around 40%, extending increases, elaborating increases to 30%, locution decreases, and idea decreases; on Level 6, enhancing increases, extending increases to more than 30%, elaborating decreases to no more than 10%, locution increases to more than 10%, and idea increases; on Level 7, enhancing increases to around 60%, extending is zero, elaborating increases to more than 40%, locution is zero, and idea is zero.

Once more, when the number of clause complexes is sufficiently large, does the condition of five logico-semantic types remain consistent with the aforementioned demonstration? What is the outcome of the intersection of taxis and logico-semantic types at each level? What is the rationale behind the choice of, for example, elaboration on changes in a diverse manner at different levels? Does the register of genre specialization have any influence on the analyzing results? As the analysis of the results is based on the manual annotation of clause complexes in the software SysFan, the process is both time-consuming and labor-intensive, particularly given that the recording of these results may also be conducted manually.

With the questions of whether a subordinate clause still upholds the functional meanings (experiential, interpersonal and also textual) like a clause or a CC; whether the hypotactic CCs more agnates with the paratactic ones or the subordinate clauses of hypotaxis more agnates with simple clauses (or even prepositional phrases), Anderson & Holsting (2018) examined expanding clause combining by comparing SFG with Cardiff Grammar and held the view that 1. β -clauses do not differ from prepositional phrases (or embedded clauses) regard to construing figures; 2. β -clauses are functionally similar to prepositional phrases or embedded clauses for their relationship of the move because the negotiation hierarchy between them relates to the finite/non-finite distinction, which means that non-finite β -clauses are less challengeable than finite ones; 3. the idea of the clause-hood seems ambiguous as the hypotactic CC does not constitute a univariate structure, and Halliday and Matthiessen treat the subordinate clause in a hypotactic CC as a non-embedded clause, which gives rise to an inconsistent description of clause complexing; and 4. it is problematic for claiming that parataxis and hypotaxis are agnate because they do not have the same range of choices. A functionally based differentiation between expanding embedded clauses and expanding β -clauses should thus be suggested for systemic description strengthening. The researchers’ assumption that two clauses in a CC should be presented in two messages, two moves and two figures as “complexing is a matter of linking similar functional elements (making up a univariate structure)” (ibid, p. 5) and “the basic consideration has to do with how much textual, interpersonal and experiential semiotic ‘weight’ is to be assigned to the unit: the more weight it has, the more likely it is to be constructed as an interdependent clause in a CC rather than a circumstantial [prepositional] phrase (or adverbial group) augmenting a clause” (Halliday & Matthiessen, 2014, p. 434) was not accepted however, for each clause cannot present three functions at the same time and the linkage of clauses into a CC does not mean that their functions should also be numerically added.

For the connotations and denotations of clause complex, Tãm (2013b) shed light on mainly three aspects: its composition, the location of clause complex in the whole system, and its functional organization. Before comparing a clause complex with a clause, three metafunctions of clauses, i.e., textual (construing the world, at least containing Theme and Rheme in its thematic structure),

interpersonal (enacting social roles, discussing Mood and Modality through the system of Mood), and ideational (presenting information through the system of Transitivity: process, participants, circumstances) were demonstrated. For the distinction of a clause and a clause complex, it is explained that a clause complex is just a combination of clauses, thus, no new systems of MOOD, THEME, or TRANSITIVITY are produced, i.e., those of each clause reserved as they are after combination. In the author's conclusion: "what distinguishes the clause complex from the grammatical units in the rank scale is its openendedness." And its relationship should be judged by its semantic relation, i.e., "a clause complex is more of a semantic unit" instead of a grammatical unit compared to clauses, which provides sufficient explanation for the saying that "a cc is a logico-semantic unit above the clause." Thus, CC is a meaningful linkage of messages, exchanges, or representations in the language, and also "an impressive intricacy emerges" in speakers' language expansion. For the difference between cc and a sentence, the latter can consist of a simple one, a complex one, or a compound one. Thus, a clause simplex, which is the opposite of a clause complex, is also a sentence. Instead of representing a constituent of writing just as a sentence is, a cc is that of lexicogrammar and is more applicable in analyzing spoken language turn, that is, it can happen cross-turn. All in all, the author did not bring forward his own opinions or develop the theoretical framework. The article resembles a literature review of cc or a description of the detailed content of cc in Halliday's grammar system. Horizontal (clause or cc denotation in other scholars' views) or vertical (cc connotation changes through time in Halliday's or Functional system) contrasts were lacking in the article.

To examine the overarching attributes of text organization and to analyze the genre and diachronic distribution patterns as well as grammatical metaphors of diverse clause-combining structures, He, Q. (2019) initially constructed a system network of clause combining based on the relation types of clause complexes and the interdependent relation types of embedding. Subsequently, a corpus analysis was conducted. The corpus is derived from the Crown Corpus, the British National Corpus (BNC), the Corpus of Contemporary American English (COCA), and the Corpus of Historic American English (COHA). By writing corresponding regular expressions, retrieving relevant data from the corpora, and then analyzing the data statistically through manual and automatic operations using data analysis software such as UAM Corpus Tool 2.8.16 and SPSS 21 (cf. *ibid.*, p. 45), He identified the application tendencies of different types of clause combining, summarized the basic features of text organization through corpus-based quantitative research, explored synchronic and diachronic distributions, and investigated the transfers within clause-combining patterns. It was discovered that the various types of logico-semantic relations exhibit differences in genre (including formal Academic, Newspaper, and informal Fiction) distribution. The same is true of the condition of embedding. The research findings indicate that the overall frequency of logico-semantic relations exhibits minimal variation in diachronic distribution over the past 200 years. However, some subcategories do demonstrate discernible shifts. Moreover, the distribution of specific sub-categories of clause complexes varies across genres. The writer's analysis of only 1,200 sentences for a comparison between semantic and lexicogrammatical stratum did not allow for a more nuanced understanding of delicate lexical choices within the sub-categories. Consequently, the conclusion that clause complexes are primarily semantic in nature rather than lexicogrammatical lacks sufficient evidence.

The purpose of any linguistic utterance (conceived as a message) is to express a certain kind of content, and the function of the content being accompanied by additional information is mainly to ease and improve understanding. (Vallauri, 1995, p. 359) In an utterance, the rheme carries the asserted contents, and the theme contains expressions (without asserting) that enable better comprehension of the rheme "by linking it to the shared knowledge, to the context, and to the preceding text itself" (*ibid.*, p. 362). As the old *WH-question* test for thematic structure "can only be applied to utterances that have been separated from their context and displaced" (*ibid.*, p. 358-9) instead of functional sentence perspective, Vallauri (1995) put forward an "illocutionary change test" for distinguishing the rheme and the theme in a clause complex ((a) one main clause together with any subordinate clauses that depend on it; or (b) two or more coordinated main clauses, not separated by a full stop or a colon, of which none has subordinate clauses depending on it. (*ibid.*, p. 369) in English texts. In the formula $T = t_x U$, U can be any surface realization, depending on the analyzing text and our segmentation; t_x can be represented by some interchangeable realizations such as t_1 = "It is not true that"; t_2 = "Is it true that...?"; t_3 = "Perhaps." (*ibid.*, p. 371) But the test cannot be used effectively for 1. projective complexes, because its illocutionary value is always directly established by its projecting predicate; 2. complexes carrying a non-assertive illocutionary value, interrogative utterances, truncated and verbless sentences; and 3. thematic structure and presupposition. (*ibid.*, p. 365 - 7) So, the formula has some additional surface adaptation rules in the procedure of application:

Rule 1: When a non-restrictive relative clause occurs at the beginning of the tested unit, transform it into an independent clause, by replacing its relative pronoun with a demonstrative or personal pronoun.

Rule 2: When the tested unit begins with a coordinating connective, such as *or*, *and*, or *but*, put the t×-string to the right of this connective.

Rule 3: When necessary to obtain a grammatical sentence, restore zero-anaphora with a pronoun. (ibid, p. 374)

For law texts, Setia et al. (2009) summarized from three hundred nexus that the ideological (also include experiential) interpretation such as “[1 + 2] ®Actor ^ Pro.MAT ^ Gol ^ Sirc./TIME: (part: human)]” can be used to depict statement and “[(1 +2)® Senser ^ Pro.MEN ^ Phenomenon + Sirc.ACC: (part.nonhuman)]” can convey rules/authorities. Moreover, the author held the opinion that “the distinction between parataxis and hypotaxis has developed as a powerful grammatical strategy for guiding the rhetorical development of text” but did not prove the rhetorical development of court texts by grammatical comparison.

As “SFL is theoretically tied to the social contexts of language use, it has been a necessary feature for examining language in psychiatric syndromes,” and “detailed linguistic analysis can show what atypical meanings are at stake in given psychiatric disorders and show how those meanings are expressed through discourse, lexis, grammar, and intonation.” (Nagar & Fine, 2013, p. 596) Thus, by seeing how people with or without depression have varieties in using categories of elaboration (restating, exemplifying), extension (adding, contrasting), and enhancement (qualifying with specific details) to code the semantic transitions in their essays, i.e., how information is conveyed when the written text proceeds, 25 currently depressed, 24 formerly depressed and 28 never depressed individuals were investigated in Nagar & Fine’s (ibid) research. In detail, means and ANOCAs of logico-semantic categories and the percentage and ANOVAs of each category out of the total number of clauses in the text are measured for the three diagnostic groups. Then, Pearson and partial correlations between the severity of depression and the average use of the linguistic factors are calculated in each category. They found that the current and lifetime depression individuals would use more elaborations and extensions because of their difficulty in concentrating and self-focus while fewer enhancements, which “provided evidence for the Interacting Cognitive Subsystem approach (Teasdale, 1993) by claiming that different schematic mental models were responsible for interpreting experience” (Nagar & Fine, 2013, p. 611). The register, i.e., texts written by psychiatric disorder individuals, is seldom focused on in clause complex research, which broadens the possibility that SFG can be applied. One problem, however, lies in the narrow scope of grammar categories that wider and deeper items like hypotactic expansion, exposition, alternation, and conditional can also be compared, another is about explanations of these grammatical varieties that have not been shed light on.

In examining the Chinese characteristics of parataxis, zero anaphora and logico-semantic relations, and nominal groups as clauses (Hsu, 2017, p. 151), Hsu (ibid) posits that, in English, “each two clauses within a clause complex should have specific formal markers, whether conjunctive or a semi-colon, to indicate their interdependency.” (ibid, p. 152) However, the situation is markedly different in Chinese. That is to say, it is not uncommon for Chinese to lack conjunction in a clause complex, which can make it challenging to discern “whether a certain relation should be paratactic or hypotactic.” The presence of a comma or period between clauses can further complicate this distinction. (ibid) After proposing “a conceptual difference between English and Chinese” (Hsu & Tsai, 2012, p. 92), namely two dimensions (one is “arrow” and the other is “chunk”) for presentation, they attached the horizontal arrow line to “the clausal linear sequence in Chinese.” This sequence is temporal arranged in sequence, with the clauses occupying a linear order. These observations provide an explanation for “why the dependent clause usually precedes the dominant one in Chinese.” (Hsu, 2017, p. 153) Hsu argues that there is no hypotactic relation ($\emptyset$) in elaboration, nor is there an $\alpha + \beta$ type in extension. It is proposed that parataxis is a more prevalent form of interdependency in Chinese than hypotaxis. The deployment of zero anaphora and its logico-semantic relation indicates that it is predominantly “a paratactic extension or enhancement when the zero anaphoric reference is the Actor or Carrier of the primary clause”; it is primarily “hypotactic elaboration when the zero anaphora is the Goal or Attribute”; and it is predominantly paratactic elaboration when the zero anaphora represents the entirety of the preceding thesis. It should be noted, however, that these observations represent merely a tendency in the logico-semantic relation. (cf. ibid, p. 166)

Projection

Theoretical research on projection encompasses a range of topics, including interpersonal meaning realization, the choice of subjunctive mood, and other related areas.

Because a writer usually holds a specific stance, represents the power relations of writers and authors, and also attempts to construct a kind of mutual consensus in choosing projecting clauses such as *I believe*, *the company believes*, *it is believed*, etc., Forey (2009) reviews projecting clauses in thematic position, discusses their patterns, functions, the importance and relevance in construing viewpoints and identity, and extends the research scope by emphasizing marked Theme in an authentic corpus of memos,

letters and reports (62 workplace texts with altogether 1,486 main clauses). Three kinds of projecting Themes based on Halliday (1994a) and Thompson (1994a, 2004) (i.e., “thematized subjective viewpoint”; “thematized comment” and “thematized obligation/inclination”) are applied to analyze the data, and the whole projecting clause and the Subject of the projected clause are considered as the Theme of the clause complex as suggested by Davies (1988, 1994, 1997), which would influence or frame the interpretation of the projected clauses. Findings have suggested that patterns of choosing projecting clauses as Theme indeed exist: projecting clauses with *I* are present in all three text types, but are far more common in memos, while projecting clauses with *he* or a named person are more used in reports (Forey, 2009, p. 159); verbal projecting clauses (63) rather than mental projecting clauses (14) are more typical in reports, which can be explained by the genre report that summarizes activities and events already occurred and restates related wordings.

In SFG, the projection is considered as “a displacement of a proposition (figure) to another order of reality,” and also “a displacement of modal responsibility, being away from the speaker” (Holsting, 2008, p. 381) when it usually contains a mental or verbal Process, and a Senser or Sayer “representing the source to which the content of the projected clause is attributed” (ibid, p. 382). For how a clause functions as a projection in German, Holsting (ibid) discussed the way that projection and subjunctive mood are applied in different registers and also their interaction. Being different from English (projection can only be realized logically in a clause complex; experientially in a simple clause with a Circumstance of Angle; or interpersonally in a simple clause with a Modal Adjunct), German was found to deploy other grammatical categories (interpersonal metafunctions instead of logical ones) to organize a projection. Thus, assuming a language that can use only one kind of grammatical function is absolutely not right. This combination “enables the speaker to give the listener clause information (i.e., identifying the source by making it Sayer or Senser, to explicate how something is said or thought by choosing more or less specified verbal or mental processes, to give information about time and different circumstances) along with the information that a given clause is projected” (ibid, p. 397), and also prevents the speakers’ being forced to apply for information when the listener had already known.

2.2.2.2 Discourse Analysis

The 46 collected papers on clause complex discourse analysis can be roughly divided into the following categories: research articles, journal articles, abstracts of articles, thesis writing, essays, editorial texts, novels or stories, books, law texts, spoken texts, and learning materials.

Research Articles

Damanik (2018) studied CC in a background study of research papers written by students and concluded that expansion enhancement is dominant in all categories, expansions are used to avoid misinterpretation. But the article is more like a demonstration of theoretical framework in detail instead of research which contains analysis of examples, description of data, data statistics, etc.

To facilitate a comparative analysis of the disciplinary aspects of taxis and logico-semantics systems, a corpus of 40 research articles from the sciences (SRAs) and the humanities (HRAs) has been constructed by Leong (2021). The results indicate that both the SRAs and the HRAs exhibited a comparable distribution of independent and dependent clauses. The ratio of dependent to independent clauses was approximately one to two. The SRAs exhibited a greater prevalence of simplexes and diminished use of parataxis and embedded clauses. Furthermore, the SRAs exhibited a greater propensity for paratactic extensions and hypotactic elaborations while displaying a lesser inclination for paratactic elaborations and hypotactic projections than the HRAs. No paratactic projections were identified in the SRAs. The findings indicate that scientific writing exhibits a relatively straightforward clause structure compared to writing in the humanities. However, notable discrepancies in specific techniques and logico-semantic relations suggest that disciplinary variations do exist. However, as the writer considers the independent and dependent structures, including simplex, parataxis, and hypotaxis, as well as embedded simplex, embedded parataxis, and embedded hypotaxis, the taxonomy raises concerns from both structural and functional perspectives.

Following register-based frequency variables of logico-semantic and taxis combination in Matthiessen (2002), etc. (e.g., Nesbitt & Plum, 1988; Sellami-Baklouti, 2011), Rodríguez-Vergara (2015) selected around four hundred Spanish and six hundred English clause complexes from introduction and conclusion parts of articles in applied linguistics journals, explained the relationship between text and context, confirmed the highest distribution of hypotactic enhancement relations which include “all the adverbial clauses that construe circumstantial meaning in terms of time, space, manner, cause, and condition” and thus is the widest category of all nexus kind in both languages and declared that the outnumber of hypotactic relations can be ascribed to the easier way to manipulate their order that a second clause in a hypotactic nexus can precede or follow the primary one while can only follow the former one in parataxis. The comparison of lexical choices through nexus relations between these two

languages, however, is not made in the article. And how the “function of bringing readers to the textual world in the introduction part and taking them back to the material one in the conclusion part,” as in some scholars’ (Rosenwasser & Stephen, 2009) retrospective realized is also not discussed or concerned by the author.

Journal Articles

Wang, F. M. (2020) searched CC used in scientific journals concerning the *Dutse Journal of Pure and Applied Science* (DUJOPAS) of Federal University, Dutse Ma in Katsina state that thirty-two CC were selected from sixteen papers of the journal of the year 2019. It showed that the most frequent clause type is hypotaxis, in which dependent clauses are almost non-definite, then followed by both parataxis and hypotaxis clause complexes, and the least occurred is parataxis; and expansion outnumbered projection in scientific papers. If parataxis with two independent clauses and hypotaxis with only one independent and one dependent clause are easier for readers to understand the contents and get information, the researcher’s suggestion that complex clause complexes consist of more than two hypotaxis should be avoided in case of misinterpretation is unconventional. Unclear annotations or even wrong divisions, such as embedding clauses, may not affect research results because the data is also quite small, after all.

In offering an alternative approach to enhance academic writing or publication language and providing guidance to those seeking to publish in journals related to the environmental field, particularly in *Applied Environmental Research* (TCI tier 1) and the *Journal of Environment & Development* (SSCI), Jutharat (2019) has selected 45 articles in each journal for social-semiotic textual analysis and investigation of the intersections between taxis and logico-semantic types within clauses using AntConc. The findings indicate that the hypotactic clause type is employed with greater frequency than the paratactic one. Specifically, the hypotactic clause type is used in 90.67% of cases in the SSCI group and 67.84% of cases in the TCI group. In contrast, the paratactic clause type is used in only 9.28% of cases in the SSCI group and 32.15% of cases in the TCI group. In both groups, hypotactic enhancement is the most prevalent, though it is more common in the SSCI group. The prevalence of hypotactic extension and hypotactic projection in TCI is approximately 10%, while in SSCI, it is approximately 5%. The occurrence of paratactic enhancement, paratactic projection, and paratactic elaboration is less than 5% in both groups. However, the SSCI group exhibits a higher prevalence of hypotactic elaboration compared to the TCI group, while the proportion of paratactic extension in both groups is approximately 25%. However, what is the significance of merely examining the discrepancies in grammatical preferences across disparate journal categories? What is the rationale for selecting environmental studies as the subject of investigation? If an author wishes to submit an article to a particular journal, is it necessary for them to modify their writing in accordance with the journal’s established grammatical standards? Has the author conducted a more detailed study of the relevant grammatical subsystems? As the author stated, the study is for Thai and non-Thai writers. However, it would be beneficial to understand whether the author has conducted a comparative analysis of the clause complex system choices of relevant texts written by Thai and English students or scholars.

For journal articles, Charisma (2013) chose two journal articles written by native speakers and two journal articles written by non-native speakers downloaded from *The Internet TESL Journal* 2010 to analyse clause complex and found that native speakers have a higher percentage of CCs than non-native speakers.

Abstracts

In order to ascertain how contextual factors (genre and disciplinary features) influence the probability of choosing a specific clause structure, a corpus of 100 research article abstracts was constructed, representing two distinct academic disciplines: linguistics and medical science. In accordance with the “activation-construal dialectic” proposed by Hasan (2009, p. 170), “the context ‘activates’ semantic choices, which in turn activate lexicogrammatical choices” when seeing downwards. Conversely, “lexicogrammatical choices ‘construe’ semantic choices, which in turn construe contextual ones” when seeing upwards. Sellami-Baklouti (2013) posited that “quantifying structural choices” in a text enables the construing of meanings, which can then construe context on the one hand, and the context, on the other hand, serves “as an activation force,” elucidating the rationale behind “semantic and structural choices.” (cf. *ibid.*, p. 221-2) Before computing the probability of clause structure, it is necessary to distinguish between a local probability and a global probability. A local probability is “particular to one subsystem or text type, or even one body of text” (Halliday & Webster, 2009, p. 252), whereas a global probability “pertains to the language as a whole, in all contexts and registers” (*ibid.*). The local probability is conditioned by two factors (Halliday, 2005d, p. 48): intrastratal conditioning, which concerns probabilities conditioned by factors internal to the system, and interstratal conditioning, which concerns the impact of one stratum on probabilities in another. The latter is the primary focus of the argument presented by the writer. Once more, the corpus was annotated with the aid of the *UAM Corpus Tool*. The study posits that contextual factors, specifically the discipline of research and the RAA

(research article abstracts) genre, influence the probabilities of choosing clause structure within the corpus. The medical sub-corpus (hard discipline), compared to the linguistics (soft discipline) sub-corpus, presents a higher probability of clause simplexes. This can be explained by their differences in research methods and reporting findings. In the former, writers prefer “looser interaction in meaning,” and the medical sub-corpus “contains structured abstracts where the moves are presented under headings” (Sellami-Baklouti, 2013, p. 231). Another result is the higher frequency of hypotaxis in the corpus than parataxis. This can be explained by “the suitability of this type of interdependency to the characteristic features and communicative purposes of the RAA genre.” As hypotaxis “enables integration” by “packing information into an idea unit” (Chovanec, 2003, p. 51), a degree of compactness can then be achieved, and the persuasive purposes required by the RAA genre can be fulfilled. The study thus has confirmed the claim “that the system cannot be studied without reference to text and context; and that text cannot be studied without reference to the system on the one hand and context on the other.” (cf. Sellami-Baklouti, 2013, p. 240)

In order to investigate the choice of taxis and logico-semantic types by native (N) and non-native (NN) writers, a corpus of 20 Applied Linguistics research article abstracts (RAAs) was constructed by Farrokhi & Ghandkaran-Shotorban (2014). The corpus consisted of 10 RAAs written by N English writers and 10 RAAs written by NN English writers on the sub-fields of Discourse Analysis and Language Assessment. A quantitative analysis of these categories indicates that NN English writers exhibit a preference for “hypotaxis” over “parataxis,” “enhancement” over the other two subcategories within the “expansion” category, and “locution” over “idea.” Similarly, native speakers also exhibit a preference for hypotaxis and enhancement while demonstrating an equal inclination towards “idea” and “locution.” However, there is not a significant divergence in the choice of text types and logico-semantic types between native and non-native writers, which may be attributed to the limited corpus size or the similarity of the two sub-fields. Furthermore, the investigation did not fully elucidate the choice of taxis and logico-semantic types at their intersection.

Analysis of aphasic discourse usually involves three aspects: 1. Pragmatic, e.g., turn-taking, initiation of topics, conversational repair; 2. elements of text super-structure (e.g., in a narrative-story setting, participants, complicating action, resolution, etc.); and 3. intra-sentential/intra-clausal phenomena, e.g., length of the clause in terms of words and morphemes, amount of clausal embedding, syntactic error (e.g., omissions of clause elements) and lexical errors. (ibid, p. 261) “Rather than simply focusing on individual clause structures within a text, SFG allows examination of the semantic function of each clause in relation to what precedes it and what follows it, and allows analysis of a discourse basing on meanings that the speaker conveys through the forms s/he chooses rather than simply focuses on the forms themselves.” (ibid, p. 262)

To investigate the quality of clause complexes written by graduate students, particularly in article abstracts, which require a high level of condensed and complex language description, the abstracts of 12 students from the Semarang State University in Indonesia were analyzed by Rukmini (2014). Of the 46 clause complexes examined, only five were identified as problematic by the author. However, the criteria for defining “problematic” and the reliability of the small corpus used to assess the quality of graduate students’ writing are open to question. Since linear correspondence between cultures in translation is impossible and cultural manifestations are hardly realized in texts (Kenny, 1998; Yallop, 2001), equivalence has continuously been a controversial concept. Seeing texts as both artifacts (culturally valued) and specimens (systemically formed) (Halliday & Matthiessen, 2014, p. 3; Halliday & Hasan, 1985) thus significantly contributed to how translations should be upon how originals are culturally valued as positively in their (source) contexts. (Rukmini, 2014, p. 46) The primary purpose of clause complexes in the flow of events as a narrative is to “keep in a quantum of narrative figures (Halliday & Matthiessen, 2014, p. 428) and also to typify characters as agents” (Rukmini, 2014, p. 47), thus like the narrative axis, it is not only the linguistic mechanism used by the narrator “to keep the unfolding of the plot but also to build the characters’ identities” (ibid, p. 52).

As the capacity of chatbots to generate plausible, human-like responses gives rise to inquiries concerning the degree of their resemblance to the original texts, Leong (2023) sought to examine the utilization of clauses and inter-clausal relations between the original and AI-generated (artificial intelligence) abstracts. The corpus consisted of 200 abstracts, 50 of which were published in *Nature*. The remaining 150 abstracts were AI-generated versions of the original *Nature* abstracts, created using *Google’s Bard*, *OpenAI’s ChatGPT*, and *Quora’s Poe Assistant* (or *Poe*). The Welch one-way analysis of variance (ANOVA) test and the Games-Howell post-hoc test was employed to generate a list of results. For instance, “the occurrence rates of main clauses (4.59) and subordinate clauses (1.71) in *Nature* abstracts” were found to occupy a middle position. “The rate of embedded clauses for *Nature* abstracts was 3.35, differing only from that for *Bard* abstracts. With the exception of *Bard* abstracts, the rate of simplexes in the other three abstract groups was approximately 2. The proportion of simplexes to main

clauses in *Nature*, *ChatGPT*, and *Poe* abstracts was between 44% and 52%.” The hypotactic logico-semantic relations in the *Nature* abstracts were more prevalent among the ranking clauses (1.64) than the embedded clauses (0.49). The most prevalent hypotactic logico-semantic relation observed in the ranking clauses of *Nature* and *ChatGPT* abstracts was an enhancement. While the abstracts generated by *ChatGPT* relied predominantly on non-finite clauses to convey circumstantial information, those produced by *Nature* also employed finite adverbial clauses. *ChatGPT* demonstrated a twofold propensity for the utilization of elaborating clauses yet exhibited a sixfold reduction in the usage of finite adverbial clauses. (cf. *ibid*, p. 126-7) While the conclusion is that none of the chatbots matched *Nature* in all categories, the question of whether uniformity is a worthwhile goal is open to question. If AI software is intelligent enough, authors may well choose to use *ChatGPT* for writing directly instead of by themselves. Given that the abstract generated by AI may be extracted from a limited set of words on the Internet according to the initial request and that subsequent requests may result in changes to the content, it is evident that the use of AI is questionable. In addition, it can be argued that the title of the original article may not be entirely suitable for accurately reflecting the content of the original article. Secondly, AI will identify content that is not logical based on the keywords of the title and link it to other content. It is, therefore, necessary to determine how to generate texts suitable for this research purpose. Nevertheless, the comparative approach taken in the article is still relatively novel. Furthermore, the combination of the clause complex sub-categories and the traditional structure categories raises questions about the theoretical classification of the article and the subsequent quantitative analysis of those categories.

A study conducted by Kurnia et al. (2020) on 170 clauses from 20 article abstracts presented at the *International Conference on Science and Education and Technology* 2018 (ISET 2018) revealed that 47% of the clauses were parataxis, while 20.5% were hypotaxis.

In the analysis of research article abstracts, Raputri (2022) focused on the clause complex in only two abstracts from the Indonesian Journal of Applied Linguistics. However, the analysis is merely qualitative, with the framework applied to the texts in question. No discernible results were identified, and the characteristics of the genre remain unaddressed.

Halliday (1983) claimed that language has a function to make meanings influenced by culture and situation in use, and the process of making meaning is done semiotically by choosing. For abstracts of the final project written by undergraduate students, Rukmini (2010) chose 5 in each program (literature and education) from the year 2006 to 2010 to qualitatively (descriptively) discuss the taxis and the logico-semantic relation of CC (Gerot & Wignell, 1994), and the realization of CCs through lexicogrammar. It is evident that the locution projection cannot be found in abstracts as decided by this genre. Hypotaxis is used nearly twice as much as parataxis in these abstracts, and more CCs are applied by students from the Literature program than those from the Education part. As the main function of abstracts is for describing given phenomena and providing information, elaboration is thus the most frequent category in the whole CC network. Writing errors of voice, grammar, and tenses are mentioned later on, while some are not caused by students' inability or inefficiency of using clause complex relations. The reasons for how these errors come out, e.g., the transfer of the mother tongue or the comparison of their languages with English, were not analysed. The right application of CC relations might be related to the consistency of English grammar in the aspects of voice and tense. The literature review of CC in the abstract genre was not comprehensive because of less-collected articles, and only ten abstracts may have influenced the findings due to a lack of students' background information and language habits. Another problem lies in the researcher's undistinguished embedded CCs from non-rankshift CCs.

Thesis Writing

For the analysis of clause complex relations in citations, Jomaa & Bidin (2019b) chose the Literature Review parts from 20 PhD theses written by EFL postgraduates in two disciplines (Information Technology and Applied Linguistics) in a public Malaysian university (10 in each) to quantitatively discuss how they are realized in texts and to what extent their usage are similar or different from each other, as the method “assists in making internal generalizability of the researcher's claims, helps in identifying inapparent patterns, attributing specific features to the varied groups of participants involved in the study, and supports the researcher's interpretation in qualitative studies” (Maxwell, 1992, 2010). The study demonstrated disciplinary effects on choosing Quoting as students from ALs often quoted directly to “show the importance and strength of their arguments” while those from IT “focuses more on facts and methodological procedures rather than arguments” (Jomaa & Bidin, 2019b, p. 168). In projection, the Verbal processes may include several sub-classifications: Findings processes (find and show), Arguing and Stating processes (state, argue, and confirm), and processes for Suggestion (suggest, propose, and assume) that students in IT use more to “create a research gap and justify conducting their research studies” (*ibid*, p. 172). But the quantitative comparison stayed on the surface, and though

expansion relations are indeed included in citations, what the title emphasizes, i.e., quoting and reporting, were not given higher status in the study; and usually, the classification of projection was not the same as that was described by the researcher. No more details or comparisons were made between hypotactic projection and paratactic projection, and inappropriate uses of projections, though some were considered inappropriate by the author, were mentioned instead. The investigation of projection verbs is actually two-sided in that more discussion on verbs separates them from clause complex relations and functions in fact. What is innovative for the research should be the concepts of expanded “elaborated” and “enhanced” projection to denote the mixed type of projection and expansion relations.

In 10 English theses (around 260 clause complexes) written by undergraduate students from Indonesia, Ngongo (2018) reached a conclusion that “the use of conjunction showing hypotaxis relationship has more varieties than parataxis” in these written texts, which is the very reverse in the spoken text; and the use of expansion outnumbers that of projection. In his statement, “the use of logico-syntactic and semantic relationship is so important to determine the coherence and cohesiveness of the text as a unity of structure and texture of the text,” the “coherence and cohesiveness” here should indeed be substituted by “connectedness” as confusion of concepts is of necessity to be avoided for preciseness that the former is more textual and refers to relations within clause complexes, paragraphs, sense groups, and even texts while the latter denotes to that of clauses in a clause complex. Another point is that as the textual theme often functioned in the main or independent clause and then followed by the dependent clause, the researcher indicated that the “main clause is dominant of complex clause (should be clause complex) while dependent clause as its qualifier.” Not to discuss whether the indication is right or not, the dependent clause is not distinguished from the subordinate clause by the researcher that qualifiers signify the latter to describe the Subject in the Theme or Rheme in the main clause, and non-finite clauses are also not qualifiers from its function in a clause complex. Also, “a continuous process of semantic choice in texts” (Halliday & Hasan, 1989, p.10) and how “writers construct solidarity and alignment with potential or target readers through written texts” (Thompson, 2001) by choosing diverse clause complexes and their markers that have mentioned in the article are not touched.

Tam (2013a) compiled 300,000-word corpora of dissertations from 135 articles in 135 journals in English and Vietnamese; randomly selected 2000 clause complexes (1000 in each language) for close observation; manually annotated in SysFan and compared CCs to find the similarities and differences in the realization and operation of logico-semantic relationship. For projection, Vietnamese is specific in exploiting adverbs of manner while English is specific in using the subject-verb in version, and they also differ from each other in the hypotactic combination of CCs and the way projection facilitates the linguistic phenomenon of substitution in the complex. Major differences can also be found in elaboration and extension that further studies are necessary to explain such similarities and differences.

Essays

One study (Trung & Hoa, 2020) investigates academic *IELTS* sample essays (two types of argumentative essays: Exposition and Discussion on topics of education, work, environment, health, social issues, and family) collected from textbooks and relevant *IELTS* websites with high scores or prepared by examiners as examples of very good models which focus on parataxis and hypotaxis. It is apparent that from 1045 clause complexes, *IELTS* writers make use of more clause complexes (51.4%) than clause simplexes (48.6%), in which two-clause complexes, three-clause complexes, and more-than-three complexes account for 34.4%, 12.8%, and 4.2% respectively; and hypotaxis is preferred than parataxis. However, how the findings can promote students’ essay writing was not suggested.

In conducting a longitudinal genre-based course study at a Thai government university, Srinon & White (n.d.) chose three argumentative texts from three students (i.e., nine essays in total), divided them into three groups: low, mid, and high, and calculated clause complex categories from Pre-test, Exposition, and Discussion, to confirm their suspect that a positive trend of developments can be dug out and more conjunctive relations and lexical resources were used in their essays. They approved the viewpoint that “a trend towards more frequent use of logico-semantic relations is likely to also involve positive developments in a writer’s academic literacy, in what they can do more communicatively in these areas.” (ibid, p. 141) For analysis, how the classification of relations as “attempted” (to determine the nature of the relation which was intended) and “achieved” (those aspects of the clause involved in construing the relation needed to be lexicogrammatically “standard” or “idiomatic”) would affect the choice of clause complex is not answered, and the classification of clause relations as “internal” (it operates in the rhetorical world of the text) or “external” (the relation that is presented as existing in the “real world”) is not connected with the relationship between clause complexes. The researchers have at last pointed out that their result was not apparent in some categories because of data limitation, but they did not figure out whether students knew that they were being tested and then used some logico-semantic relations on purpose after their consciousness was raised, and no comparative group was designed which contains those students who had never been exposed to such concepts. Besides this, the relations of projection,

causality, and counter-expectation have not been thoroughly contrasted except from the counterparts of internal and external usages. Thus, whether diversions come from the teaching and learning experiences provided by their course remains unknowable.

To facilitate the writing of texts by Chinese EFL students, two essays have been selected to identify the logico-semantic types of their choices. These are a comparison and contrast essay and an argument essay, both of which are writing samples in the Test for English Majors-Band 4 (TEM-4). Although Zhou & Abidin (2023) have not provided a clear explanation of the denotation of the percentage of research results, it is evident that paratactic extension is a prominent feature of the comparison and contrast essay, while hypotactic, particularly hypotactic enhancement, is more prevalent in the argument essay. This is evidenced by the students' inclination to "present factual information and provide additional supporting details" in the former genre, while in the latter, they tend to "offer reasons to substantiate their viewpoints." (ibid, p. 10) Therefore, it is the genre that determines the selection of these logico-semantic types in the construction of clauses in text writing.

Though Yuniar (2018) had focused on a promising genre, i.e., the analytical exposition written by university students, and accounted for the number of detailed markers that signify clause complex relationships, he did not explain how the number difference of these markers affects meaning expression in text writing. As the number of collected texts is only 14, which produces only 74 clause complexes with only one projection, the query for the credibility of the research conclusion because of insufficient materials is thus inevitable. In addition, why students prefer the enhancement with markers "because" and elaboration with "that" in this genre remains unanswered. If descriptive analysis of clause complex relations in texts can still be accepted, numerous grammar mistakes appear in the article, making readers uncomfortable and the research itself questioned. Suggestions for teachers and students at last, however, were worthwhile.

Kurniawan et al. (2021) examined clause complexes in analytical exposition texts authored by the English teachers at SMKN 2 Palangka Raya. The distinctive aspect of this study is the examination of a set of lexical markers for logico-semantic relations and the analysis of analytical exposition texts, though the results have not been mentioned.

Similarly, in the quantitative analysis of clause complexes from 17 descriptive texts written by 17 students of 1 Muzik class in SMK Bandar Sungai Buaya, Malaysia, a total of 214 clause complexes were identified by Anggraeni et al. (2018). Unsurprisingly, hypotaxis outnumbers parataxis. Enhancement is the most frequently occurring clause relation, at 41%, followed by extension (33%), elaboration (16%), locution (6%), and idea (4%). However, it is notable that the number of taxis and that of logico-semantic relations in descriptive text do not align.

To examine the language of written text as a dynamic entity, the software Inputlog (Leijten & Van Waes, 2013) is integrated with SFL theory as postulated by Bowen (2018). This approach allows for an investigation of how student writers influence the linguistic characteristics (or potential for meaning-making) of their texts in real-time. The analysis was conducted on four essays (the students' final drafts) from two female UK undergraduates (aged 19), each of which is approximately 1,600 words in length. In particular, the software generated three types of detailed information: linear analyses, which provide information on mouse movements and keys pressed; a revision matrix, which shows deletions and insertions made by the user; and S-notation, which indicates the locations in the text where amendments were made (Kollberg, 1998). Therefore, a comprehensive representation of the "linguistic realizations absent from the synoptic texts" can be reconstructed. Following the categorization of extension into addition, variation, accompaniment, and possession; elaboration into apposition, clarification, and role; and enhancement into place, time, manner, cause, and contingency; as well as the projection into matter, angle, and so forth - that is, intersecting the experiential and logical relations - it was found that hypotactic enhancing relations and paratactic extending relations are the most common interdependency relations. Furthermore, the unfolding revision activities of expansion and projection, the total number of revisions in each essay, the final relative count after revision, and the uptake ratio (the division of the latter two) were calculated. Although a list of figures of "choices in these systems throughout the revision activity" of each essay was provided, the small number of the corpus and participants made the quantitative analysis unrepresentative. The combination of two functions (experiential and logical) for analysis, namely expansion and projection, which "permeate all levels of the content plane of language," led to the consideration of not only relations within clauses but also relations between clauses. As the corpus is relatively small, it was not possible to calculate each subcategory. Furthermore, the writer has not elucidated the relationship between the revision of writing and genre. Additionally, the impact of an individual's writing habits on revision processes remains unaddressed. Another question arises concerning the contextual effects of revision. Accordingly, the quantitative analysis of this study is open to question. However, subtle differences between discussions, which tend to include "more factual information," and expositions, which often "recount information,"

have been identified. The research methods, which include a detailed qualitative analysis with a research design, are worthy of note. Furthermore, the selection of Inputlog allows writers to identify their progress and evolution in the writing process.

Editorial Texts

Focusing on the expansion relations of CC in editorials of a Ghanaian daily newspaper (the *Daily Graphic*) and the realization of those CCs through the use of conjunctive signals in the lexicogrammar, Adjei & Opoku (2017) qualitatively (descriptive survey and textual analysis) investigated ten purposively sampled editorials (with 158 CCs and 462 clauses in total), and found that these texts employ explicit conjunctions far more than implicit ones to enable writers to narrate the incidents, challenge, inform and also entertain the readers, and facilitate readers' understanding. Also, the more frequent use of extension-addition helps the text to afford more information in organization, and enhancement relations provide readers with more "necessary circumstantial information (most temporal) to understand the sequence of events in editorials" (ibid, p. 25). But problems exist in dividing a clause complex into a clause and also a clause into other lower ranks; the signals indicating rank in a CC were not fully marked or noted, thus leaving signalling vague and creating wonders for understanding; moreover, the researchers made a mixture of embedding with normal clauses, and also a clause with phrases or word groups. As the study analysed editorial texts, how the realization of these CCs influence readers' reading experience was not thoroughly answered, in the researchers' words: "the study would also be comprehensive and meaningful when it is extended and linked with the total consideration of the ideational metafunction to the interpersonal metafunction as well as speech acts which are concerned with establishing and maintaining the interaction between the speaker and the hearer." (ibid)

News Texts

In the conventional comparison of two news articles on Russian intervention in Syria from three metafunctions of clauses, clause complex relations, register analysis, and nominalization, Eid (2016) verified that "the idealized function of the paratactic structure is to represent the wording; whereas with hypotaxis the idealized function is to represent the sense or gist." (Halliday & Matthiessen, 2014, p. 520) and also Eggins' (2004, p. 338) claim: "Hypotaxis is generally more common in written text because dependency relations require more care by the writer to construct and more effort by readers to interpret than parataxis. The hierarchic organization of information demanded by hypotaxis also offers the writer a resource for offering readers a more closely controlled logic between events." But the writer did not make adequate comments on questions such as "which article better depicts the event," "are the two articles adopt a critical or positive attitude toward the event," "are they really objective in describing the whole process," "do they convey ideological differences through lexical and grammatical choices" etc.

Noviandari & Yuwono (2020) chose 5 *CNN* news texts to discover their dominant type of logico-semantic relations in a qualitative approach and content analysis designation. After documentation, reduction, display of the data, and conclusion verification, they found that more than three-fourths of the relations are Expansion; almost half are Elaboration, Locution accounts for one-fourth with no Idea (this is confusing whether they really make clear what is Locution and the scope of Idea); Enhancement and Extension take up to 16% and 14% respectively. But research on CC in the News genre is merely counting the number of relations or only demonstrating categories in chosen texts without explaining the advantage of using more CCs or any specific category in the News genre in enabling listeners' understanding and getting information.

In the same news genre, Andriani (2019) studied *VOA* news taken on 28th April 2019 and 18th June 2019 and found more projections in 58 clause complexes (33 projections account for 56.90%). But it was just a discourse analysis instead of a genre comparison or fixed theme in news investigation.

Additionally, Noviandari (2019) calculated 201 clause complexes derived from *CNN* news reports, in which 76.10% were identified as expansion and 23.90% as projection. The author's findings differed from those of previous studies, with expansion (elaboration) identified as the most prevalent logico-semantic relation across all categories, occurring 92 times or 45.80% of the total. The second most prevalent type was projection (locution), which occurred 48 times or 23.90%. The third frequent type was expansion (enhancement), which occurred 33 times or 16.40%. Meanwhile, the category of expansion (extension) was identified 28 times, representing 13.90% of the total number of clause complexes. The final category was projection (idea), which was not represented in the data set. In instances where the number of clause complexes exceeds 1,000, the resulting data may differ. This is because the choice of the majority of elaborations and the absence of ideas in news reports may not fully align with the specific characteristics of the register.

Novel/Story

By choosing three extracts from Oscar Wilde's novel *The Picture of Dorian Gray*, its translation (by João do Rio), and adaptations (by Clarice Lispector and Claudia Lopes) into Brazilian Portuguese, Rodrigues-Júnior (2018) comparatively discussed the role of expansions for construing the characters'

identities and personality traits and exemplified how the original text (English-speaking cultures) was linguistically transposed to the Brazilian Portuguese-speaking context within two registerial possibilities. Because of the difference between readerships, concepts of omission (Munday, 2012) and rewriting (Munday, 2008) in the literary translation context confirmed the points that (re)textualizations may create negative cultural impacts on the target-text reader. According to Bastin (1998, p. 8), “the controversy surrounding the supposed opposition between adaptation and translation is often fuelled by ideological issues.” One explanation is that translators re-evaluate the uses of clause complexes just ‘to simplify the linguistic constructions of characters’ identities and behaviours as a kind of linguistic economy that informs the style of this specific mechanism of text rewriting’ (Rodrigues-Júnior, 2018, p. 45). But the qualitative only research with three extracts was not persuasive. Interestingly, it is “how clause complexes function as spaces for manoeuvre” (ibid, p. 53) when translators are deciding to choose the type of clause in translating the narrative axes.

In mixed-taxis, paratactic extension plus hypotactic enhancement is dominant (Wulansari, 2017) through analysis of the text *The Snow of Kilimanjaro*. But what is the condition of that in other texts or genres that deserves attention?

A logico-semantic analysis of the short story *The Witch’s Brew* written by Penelope Lively, conducted by Sarip & Hidayat (2019) revealed that in all 158 clauses, the parataxis accounts for 37% and the hypotaxis accounts for 25%. In all instances of parataxis, the extension occupies 50% of the total, with elaboration accounting for 17.2%, idea 13.8%, locution 12.1%, and enhancement 6.9%. In all hypotactic constructions, enhancement accounts for 35% of the total, as does idea. Elaboration constitutes 25%, while locution accounts for 5%. This indicates that no hypotactic extension was identified. However, the writers have not demonstrated the genre uniqueness of the short story. Furthermore, the rationale behind selecting this particular story as an object of study, as opposed to other stories, has not been provided. In addition, how the research results are presented and explained is unclear. The results themselves do not indicate the characteristics of the genre, which is a notable omission.

As Halliday (1971, p. 341) stressed that “stylistic studies are essentially comparative in nature,” Huang H. (2020) examined the system of CC and explicated its relation to readers’ understanding of literary meaning (characterization specifically) by adopting alternative structures for comparing them with the original CC in the case study of James Joyce’s short story *Two Gallants*, to demonstrate the embodied meanings, reveal their hidden thematic motivations, and show the shaping of choices into a coherent pattern throughout the text. The process of how Lenehan’s pretence (giving readers a flat and minor character) from the beginning gradually changed to a more sophisticated image in later narrative paragraphs is presented, just like what Leech & Short (2007, p. 170) had noted, linearity is an overriding property of texts that is above all a matter of sequence. And an important authorial resource to realize this, among other things, is clause complexing.

Books

Sriniwass (2009) outlined a methodology consisting of a heuristic-grounded theory- document analysis-qualitative approach (Patton, 1990; Creswell, 1994; Halliday, 1996; Biber et al., 1998) that was employed in recent cross-disciplinary doctoral research for the genre of the textbook for encoding experiential meanings. The CC system was analysed in 3 analytical chemistry textbooks (Rubinson & Rubinson, 1998; Skoog et al., 2000; Christian, 2004) used on the chromatography subject at the undergraduate level in the University of Malaya, and principal markers were identified as 57 coding categories remodelled according to Halliday & Matthiessen’s (2004) conceptualization. Data analysis included constant instances comparison of one category with another to refine emerged patterns. The research demonstrated “how theoretical resources such as the coding decisions, interpretation of complex grammatical constructions and association patterns, overlapping categories, and other indeterminacies were resolved by using a manual (clause to clause) analysis to construct specialized knowledge in chemistry” (Sriniwass, 2009, p. 131-2). Similar to the observation on geography texts from Wignell et al. (1993), technical terms in chemistry were proved to be explicitly marked orthographically by being boldfaced (Sriniwass, 2009, p. 139). In all, the article specifically showed the stages of annotating and categorizing clause complexes in texts, but the scope of statistics is quite narrow, leaving lexical patterns not fully listed with such a small amount of data. “57 coding categories” mentioned were not described, though they had been explained in the researcher’s unpublished PhD thesis.

Li, Y. (1991) comparatively investigate explicit and implicit clause linkage in Chinese and English, i.e., the ways of using coordination and subordination conjunctions to express logico-semantic relationships to predict possible difficulties for Chinese learners in mastering English. The Chinese data comes from *Wandering in the Garden, Waking from a Dream - Tales of Taipei Characters* written by Pai Hsien Yung, and the English data comes mostly from a translated version edited by George Kao and other CCs. It is assumed that coordination and subordination categories that are present in one language but not

the other and those that were found in both languages but were diversely distributed should be given special attention by English textbooks and classroom teachers. Besides structures that Lakoff (1984) called *pure paratactic* and *near hypotactic* in both English and Chinese data, English also uses other two labelled as *mixotactic* and *pure hypotactic*. While both languages permit reduction clauses by factoring out common elements from one clause or the other, English has more widespread practices and can thus affect structures that cannot be applied to in Chinese. The biggest difference found in the research could be implicit connectives in both coordinate constructions (English) and subordinate constructions (Chinese). At last, pedagogical implications for teaching English as a second language to Chinese students are given in detail from seven aspects: 1. The “pregnant” clause; 2. Conjunction and reduction; 3. The parataxis and hypotaxis continuum; 4. Copy-cleft transformation; 5. Connectives linking restrictive and non-restrictive adverbial clauses; 6. Explicit subordinate clause connectives; and 7. Implicit subordinate clause connectives.

To better comprehend clause complexes and offer new information used in English textbooks written by native and non-native speakers, Yanuar (2013) was motivated to study the Interchange Books.

Law Texts

As a distinct genre, Fki (2018) has investigated the interrelationship between contextual elements and the lexicogrammatical resources utilized in international treaties (ibid, p. 402). Adopting an SFL perspective, the study scrutinized the nexus between context and systemic choices in the legal genre of international treaties, with a particular emphasis on the deployment of hypotactic enhancing clauses. (ibid, p. 419) The study employed a bottom-up approach, beginning with “a quantification of the lexicogrammatical choices of hypotactic clauses” and progressing to a more detailed examination “of these choices in relation to the communicative purposes and sectional features of treaties,” a corpus of ten international treaties was analyzed using the *UAM Corpus Tool* (O'Donnell, 2008) to classify enhancing clauses according to a predefined scheme. (Fki, 2018, p. 408) “The analysis of the use of this structure in the Preamble and Core Text sections of treaties have provided further empirical support for one of the major SFL claims,” namely that “to explain why specific patterns of wordings appear rather than any other, one must appeal to the meanings which, being relevant to the context, activated those wordings” (Hasan, 2009, p. 170). In conclusion, it appears that the drafters of the treaty Preamble consider a range of resources while formulating their justifications, with the aim of persuading the audience of the merits of the forthcoming treaty. This proclivity towards promotionalism and argumentation demonstrates the impact of the communicative objectives of the Preamble section on “the choice of the hypotactic enhancing relation of reason.” (Fki, 2018, p. 413)

Because “the complicated structure of clauses realized in a text creates the confusion towards the readers because of the inability of analyzing and understanding the content of text itself” (Anggara & Sunardi, 2015, p. 2), Anggara & Sunardi (ibid) portrayed the logico-semantic relations of 21 CCs from *AFTA* agreement (signed on 28 January 1992 in Singapore) by using the framework of Eggins (2004, p. 254 - 295). The reason for more CC (51.2%) can be explained as applying for additional information and making the CC more effective for easier understanding. There is no doubt that hypotaxis outnumbers parataxis, elaboration has the highest percentage for providing details, extension ranks second for adding information, and enhancement provides circumstance for an explanation. This genre (agreement) focuses more on presenting the process and the actor. But one dilemma should be non-sense in a sense, though examples were analysed at length, as the data is small that only the comparison of such a small number of CCs is not representative, especially when the text is old, and no vertical (old version vs. new version) or longitudinal (this agreement with others) comparisons were conducted.

An analysis of the logico-semantic relations in the agreement text on the common effective preferential tariff (*CEPT*) scheme for the *ASEAN* free trade area by Sunardi (2017b) indicates that more clause complexes are selected than clause simplexes (46%); the majority of the clause simplexes are embedded clauses (63%); and the proportion of hypotaxis outnumbers that of parataxis (19%). However, it remains unclear how this genre differs from others in its selection of grammatical categories.

To investigate particular patterns and functions of projecting clauses found in the thematic position in the memos, letters, and reports with their extents, Forey (2002) sees projecting as a presentation of the gist, as a fact, and as a Theme. A corpus of 42 workplace texts was analyzed, and 203 projecting clauses in the thematic position were identified. The study is based on research from a number of sources, including Davies (1988, 1997), Nesbitt and Plum (1988), Thompson (1994a, 1996a), Iedema (1995), and Harvey (1995). The research has identified three categories of projection: thematized subjective viewpoint, which includes the first-person pronoun *I*, the personal pronoun *you* or *we*, and common nouns; thematized comment, which Thompson (1996a, p. 129) defines as “serving to set up as a starting-point of the message the speaker’s own comment”; and thematized obligation/inclination. The analysis revealed that the “thematized subjective viewpoint” is the most prevalent realization of projection across all three text types, accounting for 70.9% of instances. Letters exhibit the highest

frequency of thematized subjective viewpoint (76.2%), followed closely by memos and reports (71.4% and 68.7%, respectively). In the three text types, the realization of thematized comment appears to be limited to a similar extent in memos and letters, at 22.4% and 21.4%, respectively. "Thematized comment is more frequently realized in reports (31.3%)." The least frequent of the three types of projection is thematized obligation/inclination, occurring in just 6.1% and 2.4% of the cases in memos and letters, respectively, and not at all in reports. (cf. Forey, 2002, p. 188) Furthermore, a significant majority of projecting clauses in thematic position are verbal projections, whereas mental projections in reports are notably less prevalent and are typically manifested "through the use of a second person pronoun" (ibid, p. 208). The study has confirmed the function of a report as outlined by Harvey (1995, p. 196-7), namely, "introducing, informing, describing, stating, appraising, asserting, reasserting, challenging, contending, assuming, estimating, warning, exhorting, suggesting, and recommending." However, more nuanced analyses have yet to be demonstrated. Moreover, when an individual or group of individuals are composing texts within a professional setting, the author establishes their identity, status, and authority within the text (Fairclough, 1992; Ivanic, 1998).

Sunardi (2017a) explored 207 clauses found in *the 1945 Constitution of the State of the Republic of Indonesia* that there are 17 clause simplexes with embedding (takes up 19,6%) and 88 clauses in hypotaxis (73,3%) for explaining other clauses to make the information clearer and understandable; 54 clauses have elaboration (which has a higher percentage than that of other expansion and projection processes) for elaborating clauses in detail, giving restatement or detailed information for reader comprehending the context of the constitution. A thorough description of the structures of clause complexes and even clauses is presented in the article, with an inspection of clauses with hypotactic & paratactic patterns (44 clauses can be discovered).

Spoken Texts

In Armstrong's (1992) study, the ways that an aphasic speaker FP connected clauses into a text over six months after his stroke was documented, and his data was compared with those from a normal speaker ED and also research results from Nesbitt & Plum (1988). In practice, thirty-minute samples of FP's oral discourse were audiotaped on three occasions - at 1-, 3- and 6-months post-onset, and then he was given two psycho-linguistically-based tests - the Boston Diagnostic Aphasia Examination and the Boston Naming Test (Kaplan et al., 1983).

The research found that FP had no difference from normal people in choosing the categories from the scope of the clause complex system, but differed in the frequency of each pattern that occurred in the discourse. With regard to tactic relations, FP has a higher proportion of paratactic and lower hypotactic percentage compared to ED, and the former would increase as time passed by because of the subject's adjustment strategy that it is "easier to simply add independent clauses rather than to manipulate what would probably have to be increasingly sophisticated structural dependencies" (Armstrong, 1992, p. 272). The same condition would be the application of elaboration that FP chose more in his discourse to re-state and give more information, to emphasize a viewpoint, to correct an incomplete or problematic clause, and to be more specific in conveying, which is called the correction strategy. (ibid, p. 273) The reason for FP's less use of enhancement can be his difficulty in constructing hypotactic relations where enhancement normally occurs. But discourse from just an aphasic people compared with normal people and also Nesbitt's research data was also deficient of stringency. The variant results may derive from personal language habits.

A transcript of students' causal conversation (approximately 1,500 words) by Ifadloh & Nufus (2017) indicates that the proportion of hypotaxis and parataxis are almost equivalent, though the choice of clause complexes is approximately one-third in all sentences. The proportion of expansion exceeds 90%, while that of projection is less than 10% in both hypotaxis and parataxis. In projection, the locution accounts for approximately 57% of the total, while the idea accounts for approximately 43%. In expansion, elaboration accounts for approximately 40% of the total, with extension accounting for approximately 35% and enhancement accounting for approximately 25%. The elevated ratio of elaboration in comparison to enhancement may be attributed to the nature of the discourse in question.

Oral presentation is a kind of classroom-speaking activity commonly used in language teaching and learning contexts, which is often prepared in advance where students discuss a specific topic (Harmer, 2008). (Sulistyaningrum & Rasyid, 2015, p. 42) To examine how students develop their ideas presented in slides through their taxis and logico-semantic relations and to present their patterns of clause complex complexity, Sulistyaningrum & Rasyid (ibid) chose four female (group A) and four male students (group B) from the class K-12 in Acceleration Program of SMA Labschool Jakarta, recorded their observation videos, transcribed and descriptively analysed non-embedded CCs (Gerot & Wignell, 1994, p. 21). From 169 (group A) and 68 (group B) clause nexuses, it is found that almost all kinds of relations are used except hypotactic extension in the former and locution in the latter; girl students (68.05%) use more parataxis than boys (55.88%) with hypotaxis in reverse, and elaboration is dominant in both groups. As

to the saying “the students of acceleration program have a good complexity in the production of their language,” this cannot be proved as research results were not compared to those from normal students, and “they have language ability that may be higher than other students in the same age” is not precise or scientific at all. One advisable point is that the meaning of each clause nexus, such as exemplification, proposition, etc., in a paragraph was classified, which enabled the representation of meaning progress in presentation; and another is gender comparison of choosing CCs in this genre.

For spoken texts, Panggabean (2020) studied the video taken from the collection of TEDx Talks entitled *The First 20 Hours How to Learn Anything* by Josh Kaufman. In a total of 130 clauses, parataxis is the dominant construction (accounts for 62.30%), and 97.63% are expansion types. But why this text rather than other texts is chosen was not explained, i.e., how unique the video is, has not been presented by the researcher.

There have been different terms to describe elaboration as a logico-semantic relation, such as “explanation” (Kong, 2006; Van Leeuwen, 2005); “exposition” (Martinec & Salway, 2005; Unsworth, 2007); “exemplification” (Ifadoh & Nufus, 2017; Gerot & Wignell, 1994; Panggabean, 2020); “specification” (Djonoy, 2005; Ifadoh & Nufus, 2017; Sulistyaningrum & Rasyid, 2015), etc. For analyzing “how coordination enhances cohesion in EkeGusii (a Bantu language) spoken texts and how the conjunctive elements enable the elaborative logico-semantic relation in EkeGusii” (Isaac, 2017, p. 224), Isaac (ibid) investigates the ideational metafunction (the logical metafunction specifically) to discuss the building of grammatical units into complexes by conjunctions, and the textual metafunction in realizing grammatical complexity of the texts in twelve monologues. After recording and transcription, constructions with conjunctive forms from the recorded texts were annotated based on the F-unit concept (Lieber, 1981, p. 93-6).

The reasons for EkeGusii speakers’ choosing conjunctive forms were found to be the meaning they intended to elicit in the listeners and their judgment of the listeners’ understandings; and some particular forms were employed to help listeners understand the subject matter and then achieve better communicative goals. Eggins’s (2004) and Thompson’s (2004) arguments were confirmed by the research result that the interdependence among clauses is usually signalled by conjunctions. Moreover, these forms in the corpus are usually “equipped with various elaborative logico-semantic relations, i.e., corrective, distractive, particularizing, summative, verificative and resumptive functions (cf. Halliday, 1994a; Gerot & Wignell, 1994)” (Isaac, 2017, p. 226), and they were “realized indirectly in expressing certain meanings which presuppose the presence of other components in the discourse and constructions that they have been used” (ibid, p. 232). But the relation of F-unit annotation structures with SFG has not been explained.

Learning Materials

For promoting multimedia, designers and teachers comprehend meanings through semiotic modes and fulfil their communicative functions in teaching, a hermeneutic framework (Vorvilas et al., 2011) which describes logico-semantic relations in components of multimedia materials (usually includes two main semantic units: items and clusters) following a list of scholars (Djonoy, 2005; Kong, 2006; Martin & Rose, 2008; van Leeuwen, 2005) is firstly put forward and then implemented into examples from Wisc-Online and Flexible Learning Toolboxes repositories by Vorvilas (2014). Items (phonic, musical, visual, and linguistic components) refer to communicative acts that offer, ask or demand information or goods and services, while clusters (local groupings of components with particular communicative functions on the multimedia layout (Baldry & Thibault, 2006)) can be interpreted as an instantiation of a genre type (Vorvilas et al., 2011, p. 27). Generally speaking, the theoretical framework of multimedia analysis is not quite the same as that of texts in which the former still has “augmentation” and “divergence” in addition, and “identification” and “specification” in elaboration, representing more functional than meaning-potential in classification. What if the relations between clusters can be understood as more than one or may cause ambiguity remains unanswered, and the interpretation of logico-semantic relations within texts can also be complemented by multimedia materials?

2.2.2.3 Translation

As part of the *Covert Translation* project, Becher (2011) conducted a diachronic corpus study of paratactic and hypotactic enhancement in translation and hypothesized that parataxis may replace hypotaxis as the preferred structure in German written discourse. To achieve comparability with the project, Bisiada (2013) compiled two corpora: a translation corpus (TC, English management and business articles published in the Harvard Business Review with their German translations) and a comparable corpus (CC, comparable German non-translated articles published in the Harvard Business Manager), both of which include two sub-corpora of texts from the year 1982-3 and 2008 (55 and 65 in TC respectively). It is discovered that a combined development from hypotaxis to parataxis in causal CCs in both translated and non-translated German texts indeed exists, such as the frequency of the

conjunction *weil* in the TC has decreased significantly to around 40%, which is the same as that throughout the CC. The frequency of hypotaxis also decreased in the TC, while that of parataxis increased (which is no more than the former though including situationally determined cohesive devices); being diverse from Becher (2011) with results that parataxis has become the most favoured method of building concessive clause relationships in popular science articles.

Translated and non-translated causal CCs differ noticeably with regards to preferable conjunctive patterns as well as the place of sentence boundaries due to enhancement clauses in diverse genres on the one hand, as Sellami-Baklouti (2011, p. 521) argues that abstracts with the aim of persuading readers by argumentation show a high frequency of hypotaxis; while texts with experimental data to convince readers show a higher percentage of clause simplexes, and have a greater amount of parataxis than hypotaxis. Another explanation for the similarity between the causal CCs may be that “there are stylistic differences within the group of enhancement clauses, which would mean that concessive clauses exhibit a different syntactic trend to causal clauses” (Bisiada, 2013, p. 23). Causal clauses, then, would still be changing from hypotaxis to parataxis, possibly because hypotactic causal CCs are easier than hypotactic concessive CCs for understanding. In addition, cohesive devices available in German seem to be used differently when contrasting translations with non-translations.

In investigating “taxis and logico-semantic relations in an Arabic translation of an English text (target text, TT) from the genre of comparative religion,” Othman & Al Qutob (2022) compared it with the source text (ST) and a sample of non-translations (five Arabic books) from the same genre. From each text, 200 clause nexuses (in total 1,400) were selected for qualitative and quantitative analysis using Chi-square or Fisher Exact tests. The objective was to assess the TT in terms of its alignment with the established norms and conventions governing the register. As a list of problems in defining the notion of the sentence in Arabic can be faced, it should be noted that “punctuation marks are not strictly rule-governed” and that there is an “indeterminacy regarding the classification of Arabic conjunctions as paratactic or hypotactic markers.” In order to determine sentence boundaries, the author has adopted “the syntactic and semantic criteria proposed by Kohlani (2010, p. 201)” and has also followed his definition of a *sentence*. Based on this, the authors have proposed definitions of “paratactic constructions” and “conjunction” in the Arabic language. In Arabic, there are five main types of hypotactic constructions: “non-defining relative clauses; conditional clauses; conjunctive phrases ending with the complementizer *that*; adverbial clauses; and clauses initiated with non-finites, prepositions, and prepositional phrases that explicitly mark the logico-semantic relations.” (Othman & Al Qutob, 2022, p. 9) The research revealed that the prevalence of paratactic over hypotactic constructions in Arabic writing, in general, resulted in “an over-representation of hypotaxis and an under-representation of parataxis in the target text (TT) compared to the non-translations.” (ibid, p. 11) The approximately equal distributions of the two taxis modes in the target text (TT) bring it “closer to the English source text (ST)” than to the non-translations in Arabic, indicating an overuse of hypotactic relations relative to the conventions of the genre, as well as the influence of the source language (SL). (ibid, p. 14) Furthermore, elaboration is more frequently observed in paratactic constructions (81%) than in hypotactic ones in non-translations. This can be attributed to “the high frequency of the Arabic paratactic conjunction /fa-/.” (ibid, p.11) While the target text (TT) manifests contrasting patterns and different distributions, hypotactic elaboration accounts for 67% of the total. In addition, the target text (TT) employs a greater “use of hypotaxis in construing extending relationships.” (ibid, p. 12) Consequently, the construal of expansion relationships in the target text (TT) is inconsistent with the target language (TL) or the genre conventions of TL; and the TT is slightly more similar to the ST than to the non-translations. Although this resemblance between TT and ST dates back to “the translator’s literal approach to translation at the clause complex level” (ibid, p. 15) to some extent, it is important to consider the impact of English structures on Arabic translations.

2.2.3 Summary

While there is a greater number of papers on Chinese writing or Chinese research than on research in other countries or regions outside China, this does not necessarily indicate that the former has more research than the latter in total. On the one hand, the research on the projection and expansion subsystems has been partially referenced in section 2.1. Furthermore, the author’s ability to search for foreign literature is constrained. For instance, the research databases of countries (districts) such as the Middle East, Australia, and New Zealand have not been consulted. It is, therefore unnecessary, from a regional perspective, to simply compare the numbers. The earliest article in China to study clause complexes was published in 1997. Given that the concept of clause complexes was initially proposed by Halliday and subsequently presented in a paper in 1974, it can be reasonably inferred that the earliest research on this topic can be traced back to 1989. There have likely been several peaks in the study of clause complexes in China, including 2000, 2007-2008, 2011, 2014-2016, and the present. Outside China, the field has

seen a surge in interest since 2017, except for the peaks in 2005 and 2013.

In terms of research content, in addition to the structural classification (clause complex, projection, and expansion), 39, 45, and 6 articles were collected in China, respectively, while 58, 4, and 3 articles were collected outside China. Furthermore, theoretical exploration, discourse analysis, and translation teaching were identified in China, with 47, 35, and 9 articles, respectively, while 17, 46, and 2 articles were collected outside China. Indeed, when considering the papers that have addressed the distinction between projection and projection verbs, expansion, and conjunction, it becomes evident that scholars in China tend to engage in qualitative and quantitative research on projection structures and the overarching system of clause complexes. In contrast, scholars based outside of China tend to prioritize quantitative research on the clause complex system. There is a paucity of studies on the expansion subsystem. Furthermore, Chinese scholars are more inclined to engage in theoretical exploration, whereas scholars in other countries are more likely to conduct discourse analysis. Specifically, regarding the text type of discourse analysis, Chinese scholars have a particular predilection for literary works and news texts, whereas scholars outside China, in addition to legal texts and speech reports, tend to focus their attention on academic texts, including journal articles, academic papers, classroom writing texts, and paper abstracts. Moreover, Chinese scholars tend to focus their attention on advertising texts, academic comics, graphic texts, and marine accident reports. In contrast, foreign scholars often direct their attention toward editorial texts, workplace texts, religious texts, and other areas.

From the perspective of research purposes, the most common direction for Chinese scholars, in addition to general comparative studies, is the comparative study of English and Chinese. In contrast, the vast majority of comparative studies conducted by scholars outside China do not involve the comparison between languages. The practical applications of these studies include the examination of the logic employed by the speaker or author in their writing, the enhancement of students' awareness of clause complexes, the improvement of writing skills (not limited to students), the advancement of teaching methodologies, the promotion of theoretical development, and so forth. Concerning research methods, Chinese scholars tend to employ qualitative research (58 articles), yet they do not neglect the significance of quantitative analysis (33 articles on qualitative-quantitative synthesis). However, quantitative analysis typically culminates in a comparative analysis of numbers or frequencies. In contrast, scholars outside China exhibit a more balanced approach. In addition to pure qualitative analysis (25 articles), 17 articles employ pure quantitative analysis to compare grammatical phenomena. Furthermore, 21 papers adopt a combination of both approaches. In addition to corpus data analysis, traditional qualitative and quantitative mixed methods, discourse analysis, and text comparative analysis, scholars outside China employ a range of other scientific research methods, including survey methods, experimental methods, interviews, questionnaires, and so forth.

From the perspective of the researchers, a total of 24 master's theses, two doctoral dissertations, and 65 journal articles were collected in China, while a total of 6 master's theses, four doctoral dissertations, 41 journal articles, and 14 book chapters were collected outside China. Therefore, except for the general attention paid by master's students to clause complexes outside of China, scholars have consistently maintained their focus on this topic and have continued to produce scientific research results. Consequently, in countries or districts outside of China, except for the general attention paid by postgraduate students to clause complexes, scholars have not diminished their focus on them and have been engaged in the generation of scientific research findings. In essence, professors tend to prioritize theoretical discourse and sporadically engage in comparative analysis between texts. Young and middle-aged scholars evince enthusiasm for clause complexes, yet their research output is limited, frequently manifesting in the examination of projection phenomena or projection verbs in academic texts and quantitative comparisons between texts. In contrast, graduate students primarily assess the quantity of categories within the system and offer qualitative interpretations of examples.

About innovation or the expansion of theoretical practice, the level of research is variable, and it is challenging to achieve breakthroughs in theoretical research. The majority of these studies are situated within the framework of the clause complex subsystem. Some studies are based on the research ideas of predecessors. With a specific research objective in mind, they select relevant textual materials, conduct clause complex annotation analysis on them, compare the statistical data, draw conclusions from quantitative analysis, supplement these conclusions with corresponding example analysis, and briefly discuss the significance of the research in the summary. On occasion, some studies challenge specific elements of the theoretical framework. However, it is unclear whether these critiques are widely accepted.

Diachronically, research on clause complexes in China has undergone a series of shifts. Initially, the focus was on theoretical exploration. This was followed by an emphasis on discourse analysis, language teaching, and machine translation. Additionally, there was a shift in focus from single discourses to multiple genres. Moreover, the scope of research was narrowed from a single direction to a comparative

analysis. Furthermore, qualitative analysis was superseded by a combination of qualitative and quantitative approaches. Ultimately, there was a transition from projection to an examination of the entire clause complex and its expansion subsystems. Similarly, research on clause complexes conducted outside of China has also undergone a shift, moving from theoretical research to discourse analysis. However, other shifts are not readily apparent. The majority of this research focuses on academic discourses, with multiple discourse types progressing in parallel.

A review of previous studies reveals several shortcomings:

1. Many studies are comparative in nature, and some appear to lack sufficient empirical evidence.
2. The manual annotation and coding of clause complexes is a challenging process, particularly given the complexity of the theoretical system. This results in a significant investment of time and resources. Consequently, the majority of studies select a limited number of texts, resulting in a correspondingly reduced number of CC and its subsystem categories. These data are not suitable for quantitative comparative analysis. Additionally, the quality of the selected texts is questionable, and their representativeness (in terms of both horizontal representation of writers and vertical representation of time development or variation) is debatable. Consequently, the qualitative conclusions drawn from these texts warrant further scrutiny. It is not necessary to conduct comparative studies on some texts, and the representativeness of some texts is not high.
3. Some researchers focus on identifying differences between data sets without adequately explaining the reasons and significance of these differences. Furthermore, the methodology employed in these studies for data comparison is rather rudimentary, relying primarily on simple numbers or frequency analysis. Consequently, there is a dearth of longitudinal or latitudinal comparisons between different authors, language genres, or levels, which hinders the identification of potential research avenues for other scholars to address existing research gaps.
4. In light of existing research, it is worth questioning whether research questions can truly be addressed through the examination of clause complexes. It must be acknowledged that the scope of problems that CC can solve is still limited. It is also relevant to consider whether some studies truly necessitate an answer derived from an examination of clause complexes. This illustrates the extensive scope of CC's applicability. It is thus pertinent to inquire as to the scope of problems that can be solved by analyzing the choice of clause complexes.
5. It remains unclear whether the clause complex system framework is insufficiently delicate or overly complicated when employed for discourse analysis. It would be beneficial to ascertain whether any categories are not covered by the system network. Which category boundaries are the subject of debate? What factors contribute to the difficulty of annotating some clause complexes? If theoretical exploration is conducive to theoretical development, it would seem reasonable to posit that past theoretical achievements should have had some impact on the general framework of clause complexes. However, this does not appear to have been the case. (Not considering the Cardiff Grammar)
6. What contributions can the study of clause complexes make to the fields of text translation and machine translation?
7. The primary issue is that a considerable number of studies have not engaged in a comprehensive examination of the lexicogrammatical aspect of CC. This entails an analysis of the linguistic choices at the grammatical level and an elucidation of the corresponding lexical choices. Consequently, the majority of research is focused on a comparison between the first (taxis and expansion/projection division) and the second level (locution/idea; elaboration/extension/enhancement) of the system. There is a paucity of studies that transition from the third level of the system (addition/variation/alternation; cause/temporal/conditional, etc.) to lexical analysis. An analysis of grammar that does not consider lexis will be incomplete; similarly, an analysis of lexis that does not consider grammar will lack depth. To provide a comprehensive and reasonable explanation for the text variation, it is necessary to consider the choice of lexis and grammar simultaneously.
8. Although the recursion of clauses represents a particularly significant grammatical phenomenon, few scholars have conducted theoretical exploration and qualitative-quantitative analysis on this topic.
9. There has been a paucity of in-depth exploration of the connection between clauses, particularly within the three delicate categories under the expansion subcategory, across different language genres.
10. To date, no scholar has proposed or researched the automatic annotation of the logico-semantic relations between English clause complexes.

2.3 Conclusion of Previous Research

This chapter presents a review of the existing literature on the topic of academic English writing and clause complexes. The analysis reveals that academic English writing necessitates an examination of the text from the perspective of clause complexes. Furthermore, the analysis of clause complexes has not

been sufficiently explored from the genre of academic discourse. This is particularly evident in the qualitative analysis of thesis writing, which has not sufficiently explored some key areas, including the choice of logico-semantic relations between clauses, the interpersonal meaning embodied by projection, the recursion pattern of clause complexes, the distinction between explicit combining clauses and implicit clause complexes, and so forth. From a quantitative perspective, the following aspects require analysis: differences in lexical and grammatical choices in academic text writing at different English proficiency levels; commonalities and characteristics of recursion patterns; and measurement methods of text complexity. Accordingly, these two areas will be integrated in subsequent research endeavors to address the diverse research questions posed in Section 1.3.

Chapter 3 Theory

3.0 Introduction

This chapter is essentially divided into four main parts. The first part is also the most important part of this chapter, namely the presentation of the relevant theories of Systemic Functional Linguistics. The reasons for the choice of the theory will be roughly explained. Specifically, the distinction between the concepts of clause and clause complex will be made, and the connotation and denotation of the components of clause complex, i.e., expansion and projection will be demonstrated. The theoretical framework of the clause complex at the end of Section 3.1 will be the core framework for the annotation of corpora in this thesis. The second part is related to English academic writing. In detail, its role in this study will be answered, that is, considered as a register, and its intersection with SFG, i.e., the content of the Appraisal Theory will be combined to build a part of the framework for the qualitative analysis of projection in chapter six. The third part will be an examination of the impact of CL on SFL. In doing so, CL will be seen as a methodology, and the question of where CL is going in specific applications will be suggested, although it is not a panacea. Finally, based on the relevant theories of SFG, AW, and CL, a comprehensive theoretical framework will be constructed to serve this study to guide subsequent qualitative or quantitative analysis.

3.1 SFG: Background, Originality, and Comparison

Background

Since Plato and Aristotle, there have always been two major linguistic traditions: one is based on anthropology and sociology, which attaches great importance to language meaning, to the relationship between language and social culture, and between language and linguistic environment, to the variability and irregularity of language, and the communicative ability, and takes the acceptability or the use of discourse as the ideal standard; the other is on philosophy, logic, and psychology that it emphasizes the form of the language, the interior of the language system, regular phenomena and language ability, and takes grammaticality as the ideal standard. These two traditions can also be called the humanistic tradition and the scientific tradition, respectively, in the history of Western linguistics; the latter has been dominant most of the time, especially in the era from Bloomfield to Chomsky. It was not until the rise of the systemic functional grammar school that the opposition between the two major schools emerged, namely functionalism and formalism.

Originality

As a moderate functional theory, Systemic Functional Linguistics (SFL) is an approach to language that originated from the main intellectual tradition of European linguistics that developed after the work of Saussure. By not only inheriting the functionalist ideas and theories of the founder of the London School in England, J. R. Firth, but also referring to and incorporating various schools of European functionalism such as the Prague School and the Copenhagen School, French functionalism, the anthropological research of the British anthropologist Malinowski, and the linguistic relativity theory of the American linguist B. L. Whorf, etc., SFL was created and mainly developed by Firth's student M. A. K. Halliday and a member of scholars of the London School in the late 1950s.

While many approaches to linguistic description highlight structure and the syntagmatic axis, SFL foregrounds Saussure's paradigmatic principle and regards it as the point of departure in making clear how language works, as suggested by Halliday (2004, p. xi):

"I wanted to foreground the paradigmatic dimension, whereby a language appears as a meaning making resource and meaning can be presented as choice. This paradigmatic principle had been established in semiotics by Saussure, whose concept of value and of terms in a system, showed up paradigmatic organization as the most abstract dimension of meaning (1966, Part 2, Chapters 3-5, pp. 107-27 [French original 1915])."

Hjelmslev's (from the Copenhagen School) views on the structure presenting syntagmatic relations while the system for paradigmatic relations and the system is the bottom layer of structure then lay theoretical foundations for the systematic grammar. But it was J. R. Firth who defined the technical term "system" in comparing it with "structure" and then characterized systemic theory as the theory in which the system takes priority that the most abstract representation of language at any level is in paradigmatic terms:

"Structure consists of elements in interior syntagmatic relation and these elements have their places in an order of mutual expectancy. The place and order of the categories set up are recognized

in structure and find application in renewal of connection with the sources of the abstractions. System of commutable terms or units are set up to state the paradigmatic values of the elements. The statement of structures and systems provides, so to speak, the anatomy and physiology of the texts.” (Firth, 1957, p. 30)

The excerpt from Halliday’s interpretation of “system” also confirms its significance:

“The thing that distinguishes SFG is that it gives priority to paradigmatic relations: it interprets language not as a set of structure but as a network of systems, or interrelated sets of options for making meaning. Such options are not defined by reference to structure; they are purely abstract features, and structure comes in as the means whereby they are put into effect, or realized.” (Halliday, 1994a, p.15-16)

The name “systemic” comes from the fact that “the grammar of a language is represented in the form of system networks” (Halliday, 2014, p. 23). As languages would evolve instead of being designed, and “evolved systems cannot be explained simply as the sum of their parts,” the systems’ thinking enables us to “seek to understand the nature and the dynamic of a semiotic system as a whole” (Halliday, 2014, p. 20; Matthiessen, 2007a). And “a language is a resource for making meaning, and meaning resides in systemic patterns of choice” (Halliday, 2014, p. 23), the systemic approach, therefore, allows language users to take a “path” through the network and combine all the structural rules to make meaningful choices for their communicating purposes without considering about the related structure in realizing it. The grammar of a certain language is accordingly made up of a set of choices that some depend on others and then organize as a network. The characteristic of the systemic approach being comprehensive, according to Halliday (2014, p. 20), is that “it is concerned with language in its entirety” so that the discussion of one aspect can not only “be understood always with reference to the total picture” but also “contribute to the total picture” as well as “recognize where everything fits in.” Hence, the significant concept of the system network for language description, which originated from Saussure and Hjelmslev, developed by Firth and then perfected by Halliday, makes the discussion of certain grammatical categories from a trinocular perspective, i.e., below, roundabout, and above possible and lays the foundation for the improvement of the systemic functional linguistics theory.

The term “function,” in fact, denotes “meta-function” in SFL as the former only presents the “purpose or way of using language” (Halliday, 2014, p. 31) and does not have any implication of analyzing the language itself (cf. Halliday & Hasan, 1985, Ch. 1; Martin, 1991) though functions indeed are performed on the structure and organization of language at all levels when it “evolves in the human species.” To distinguish the “functionality” that “is intrinsic to language” or along which line the whole system of language is organized, the concept “metafunction” is thus chosen by SFG linguists and regarded as an indispensable part of the whole theory. Following Karl Bühler’s (from the Prague School) Organon model of communicative functions, i.e., expressive function (*Ausdrucksfunktion*), representative function (*Darstellungsfunktion*), conative function and referential function (*Appellfunktion*, i.e. appealing function), in addition to the textual function that reflects characteristics of the language itself, three metafunctions are identified by Halliday as ideational (experiential, logical) function, interpersonal function and textual function that the last one takes the Theme-Rheme structure and the information theory from the Prague School as the main content.

Besides the inheritance of the “system” from Firth, the social context for behaviors of human beings, especially their linguistic behavior of communicating with each other in the social environment, is also adopted to reveal how language plays as a form of behavior potential excluding cultural factors. As Firth (1957) considers that meaning in language (that is, the function of the language item in use) and the social context in which discourse is produced are of significance, he attempts to establish a linguistic theory based on Malinowski’s “context of situation,” which is initially forged in German as *Situationstheorie* by Wegener (1885, p. 21-23). For Firth (1950, p. 43), “context of situation” should better be “used as a suitable schematic construct to apply to language events” and regarded as “a group of related categories at a different level from grammatical categories but rather of the same abstract nature” that it relates to fixed categories in linguistic works: “A. The relevant features of participants: persons, personalities including (I) The Verbal Action of the Participants, and (II) The Non-verbal Action of the Participants; B. The Relevant Objects; and C. The Effect of the Verbal Action.” As “a convenient abstraction at the social level of analysis and forms the basis of the hierarchy of techniques for the statement of meanings” (ibid, p. 44), Firth further provides an explanation of the relationship between meaning, functions, and the context of the situation in a linguistic perspective: “Meaning, then, we use for the whole complex of functions which a linguistic form may have. The principal components of this whole meaning are phonetic function, which I call a ‘minor’ function, the major functions - lexical, morphological, and syntactical [...], and the function of a complete locution in the context of situation or typical context of situation, the province of semantics.” (Firth 1969 [1935], p. 33) Therefore, it is crucial for Halliday to use the perspective of situational context to establish a linguistic theory that can clearly

explain the categories in language and the relationships between them.

The influence of American anthropological linguistics and traditional as well as modern linguistics developed in China on SFL is self-evident when Halliday drew on the work of linguists such as Boas, Sapir, and Whorf; expressed his gratitude toward Whorf and regarded him as the person “who the most profound effect” on his thinking as Whorf “showed how it is that human beings do not all mean alike, and how their unconscious ways of meaning are among the most significant manifestations of their culture,” put forward the notion of the cryptotype, and proposed the “conception of how grammar models reality”; and listed Luo Changpei and Wang Li as two scholars from whom he gained “new and exciting insights into language” that the former provided him with a diachronic perspective and a profound understanding of another language family and the latter informed him of “research methods in dialectology, the semantic basis of grammar, and the history of linguistics in China.” (Halliday, 1985c, p. 188) On the one hand, Halliday and Wang Li hold the same opinion on the sociability of language, the legitimacy of grammar, the relationship between the universality and particularity of grammar, and the relationship between grammar and semantics, etc. in terms of language ontology; on the other hand, Wang Li was earlier than Firth to instill the ideas of treating discourse as a research object, combining spoken and written language, treating clauses as the main grammatical unit, regarding the language as a multi-layer system and adopting a probabilistic view of language studies into Halliday’s mind in terms of language researching methods. In addition, Halliday’s cohesion theory (such as concepts of anaphora, ellipsis, substitution, connection, and lexical collocation) and grammatical categories (such as modality and mood, passive voice, division of speech categories, and the transitivity of verbs) have all been discussed in works from Wang Li.

Hence, with inspirations on “structure and system,” paradigmatic vs. syntagmatic relations, “three communicative functions,” “the context of situation” vs. “the context of culture,” the socio-linguistic perspective and the grammatical description of the Sino-Tibetan family from these scholars, SFG and the theory of SFL gradually improved and without which the investigation of any topic under the theoretical framework cannot be separated from mentioning them.

Comparison

One reason for taking a systemic functional perspective for this research comes from three main aspects, i.e., research objects, purposes, and research methods of linguistics that are clearly diversified when compared to the critical structuralism and the transformational-generative linguistics, and it was on some insufficiencies of the latter two that Halliday’s functional theory of language developed.

When transformative-generative linguists hold the opinion that language is a psychological phenomenon and the generation and comprehension of language are inseparable from studying human psychological activity processes, Halliday considers language as a social phenomenon that the study of language is inseparable from the influence of external factors such as the social environment and cultural traditions. For the former, syntax is seen as independent and self-sufficient that the description of it is of no necessity to consider contextual factors but only to find the explanation of syntactic rules within the language; while for the latter, the interpretation of language should better depend on the related social and cultural environment in which the language is produced. Regarding language ability, it is innate and inherent for scholars in the former group as they tend to believe that people are born with a language acquisition mechanism. But this is not for scholars in the latter group, as language ability must be cultivated and developed in a language environment that cannot be inborn at all; hence, the ability to use language is more required.

For Chomsky et al., the main task of linguists is to find out rules for sentence generation and transformation and answer questions such as what kind of structure the language has and according to what rules these structures are generated and transformed. As they think that only the language used by ideal speakers conforms to the rules of grammar, they focus more on this kind and language ontology, especially the syntax, while ignoring complex contextual factors. What is important for them is to discover the universal rules of human language through the study of one fixed language.

Halliday, however, opposed limiting linguistic research to the language itself. He made it clear that linguists should study both *langue* and *parole* and pay attention to the real language (both spoken and written language) that denotes not only language phenomena complying with grammatical rules but also those that cannot, while be understood, accepted, and used by people rather than the so-called ideal and imagined languages. The task of linguists is to investigate tendencies or principles of using these languages; make clear the relationship between contextual factors and semantic expressions; study various varieties of the language with their characteristics and differences in meaning expression; explain why language has such kind of structure and dig out what other factors that affect structural changes except for being restricted by language rules themselves; and emphasize on not only people’s “language ability” but also their “communicative ability,” that is, the ability of appropriate application of language, which requires the combination of language structure and its producing environment.

The difference in research methods within the two perspectives resides in four directions: the starting point of conducting research, the research scope, the methods of choosing and handling with corpora, and the processing, analyzing, and performing of data.

For the starting point of conducting research, while scholars from American structuralism and transformational-generative linguistics stick to form and the internal structure of language, those from systemic functional linguistics persist in meaning and function as they deem that the meaning of language exists in a specific process of using it and would be difficult for comprehension without considering about the environment, they advocate studying the structure and function of language in related to specific contexts.

For the research scope, transformational-generative linguists generally take sentences as the analyzing object and research the generation and transformation rules of sentences, but systemic functional linguists stress discourse and suppose that the structure and meaning of clauses largely depend on the structure and function of the entire discourse. Therefore, they not only analyze clause structures but also take on discourse analysis.

Regarding the methods of choosing and handling corpora, because transformational-generative linguists study the relationship between the surface and the deep structure and generalize the generation and transformation rules of sentences, all the knowledge a native speaker needs to know in producing all grammatical sentences can be revealed. But the interference of materials choosing makes the analyzing results of language rules skeptical as scholars would often create their sentences and judge whether they are grammatical or ungrammatical. Systemic functional linguists pay more attention to the language people practically use instead of filtered and edited language materials, let alone sentences made up by people at will. Therefore, they attach great importance to the observation and collection of actual, especially spoken discourse, building various corpora according to research purposes, discovering language patterns based on the corpora, counting the frequency of diversified language phenomena, and digging out language tendencies in specific genres and registers.

In processing, analyzing, and presenting the data, transformational generative linguists usually concentrate on why people can use limited forms to create infinite sentences. Hence, they perceive language to be a logical deduction system, adopt mathematical and logical dealing ways, and employ numerous formal means to describe the structural and transformational rules of language. In exploring the construction and function of language from aspects such as the information organization of discourse, the cohesion and coherence of discourse, the relation within discourse and register, etc., scholars taking a functional perspective believe that there is often no clear boundary between grammatical and ungrammatical clause complexes and the so-called language rules are just presented as probabilistic tendencies that are difficult to be formalized, so they do not particularly pursue the formalization of language analysis.

3.1.1 SFG as a Theoretical Base

Some basic questions will be answered in this section about why SFG is chosen as the theoretical background for this research, what differences can be discovered compared to other Functional or related grammars, what essential characteristics of SFG lead to the best choice for text analysis in academic writing genre, and how does the socio-functional focus of SFG influence the perspectives of meaning construal in texts. Then, significant notions and dimensions of the theory contents would be explained from eight groups: constituency; structure, rank, and unit; system with delicacy; function and metafunction; stratification and realization; the cline of instantiation; text, context, and register; and probabilistic and markedness, for the aim of drawing forth the emphasizing system of this study, i.e., the clause complex system, and providing theoretical bases for further analysis afterward.

3.1.1.1 Choosing SFG

Another reason for taking a systemic functional perspective for this research resides in the comparison of SFG with other related functional theories that the discussion of this section starts from the introduction of functionalism to which SFG belongs. Holding a functional view on the nature of language and concerning principally on functional relations of diverse levels in the construction of language by scholars, and intending to be a theory that can practically be applicable to analyze various aspects of language and language application, functionalism is unlike traditional linguistic theories when it pays close attention to the resource rather than the rule and is designed to present the overall system of grammar rather than only fragments. Having been developed by a list of functional grammarians such as Dik (1978), Halliday (1985a, 1994a), Eggins (1994), Bloor (1995), Lock (1995), Thompson (1996a), Martin & Matthiessen (1997), VanValin and LaPolla (1997), Matthiessen (2002), etc., functional theories can be divided into three groups: extreme, moderate and conservative that all of them reject the

syntactocentric view from formalists but recognize the significance of communicative factors or cognitive factors or even both in analyzing languages including Functional Grammar (Dik, 1978, 1997a, b), Systemic Functional Grammar (Halliday, 1994a), Role and Reference Grammar (RRG; VanValin and LaPolla 1997), Lexical Functional Grammar (LFG; Bresnan 2001), etc., in which the former three belong to the moderate functionalism or “external functionalism” as described by Newmeyer (2000, p. 13-14).

In general, SFG differs from the other two moderate functional theories, i.e., Functional Grammar (FG) and Role and Reference Grammar (RRG), in at least three aspects in terms of what the theory is attempting to account for and why the description is like this. Theoretically, according to Butler (2003a, p. 200), though SFG also refuses to regard “pragmatics as a separate component” as the other two, in fact, present stronger support for “pragmatic adequacy” after Dik compared with the others practically that the language system and language use are “intimately and inextricably linked” together for SFG so that the whole system can be a meaning potential for speakers and writers to choose from in producing clauses or in Dik’s words can “reveal the instrumentality of language with respect to what people do and achieve with it in social interaction” (Dik, 1989, p.3; 1997a, p.3). While FG “hives off much of the complexity and richness of language use into an unspecified and largely unresearched theory of verbal interaction” (Butler, 2003a, p. 200) and “does not intend to describe the pragmatic rules” to achieve pragmatic adequacy but “only wants to be compatible with them” (Nuyts, 1983, p. 383), and RRG does not see the explanation of “how speakers use language in different social situations” as a primary concern (Van Valin & LaPolla, 1997, p. 15). In mutually rejecting the generativist argument over the autonomy of syntax regarding semantics and pragmatics, three theories differ from each other by degree, though all of them accept the claim that “the syntactic properties of languages do not constitute a self-contained system but are correlated with and motivated by the meanings which the forms convey.” (Butler, 2003a, p. 200) Methodologically, SFG theory is advanced by increasing new materials or creating specific concepts and systems to original ones instead of focusing on former works as suggested by Halliday (1994b, p. 4507):

“A feature of systemic work is that it has tended to expand by moving into new spheres of activity rather than by reworking earlier positions. This reflects an ideological perspective in which language is seen not as unique or *sui generis* but as one aspect of the evolution of humans as sociocultural beings.”

Hence, systemic functional linguists advocate for a new approach rather than emphasize works that rigorous methodological standpoint play in developing the theory. And just as what Halliday & Fawcett (1987a, p. 3) have considered: “Often the major steps of progress do not come from painstaking methodology - important though this is in its place - but from what has been called ‘the creative imagination,’ the process of developing SFG theory can be described as a pattern that “a new insight for the overall model is followed by a period of testing, adjustment and a relative tightening up of the model; to be followed in turn by various applications (a very important stage); then possibly new insights and a reinterpretation of the existing model; and further testing and tightening up.” (Halliday & Fawcett, 1987a, p. 4-5) This kind of methodology that moves into new aspects of activity sets up explicit hypotheses, and then conducts rigorous tests is not favored by most of the systemic theories, in fact. And applicably¹², SFG attaches vital importance to the applicability in developing its grammar, which is not supported by either FG or RRG as for Halliday (2013, p. 128), an applicable linguistics theory intends to solve problems that would be raised and answer questions be asked by not only professional linguists but also other people who apply language in daily life and concern with language including educators, translators, legal and medical specialists, computer scientists, students of literature and drama, etc., to clarify “what sorts of questions can usefully expect to be asked, and whether or not there is any hope of coming up with an answer.” Apart from paying attention to four main aspects, i.e., literature, foreign language learning and teaching, machine translation, and the broad issue of undervalued languages and varieties, SFG is also developed as a theory for text analysis, which is “very far removed from the concerns of practitioners of FG and RRG” (Butler, 2003a, p. 201).

In specific, as has been mentioned in the former section, the paradigmatic characteristics of SFG, i.e., the system networks presenting the options of meaning potential provided by the discussed specific language and structures originating from sets of systemic properties in the process of grammatical-lexical realization, makes it unique from other linguistic theories by focusing on systematizing language differentiations and enables the applicability of it in research fields such as stylistics, educational linguistics, computational linguistics, machine translation, etc. Regarding the

¹² The word “applicable” is chosen instead of “applied” or “applicable” has been explained by Halliday in “Pinpointing the Choice: Meaning and the Search for Equivalents in a Translated Text” (2012, p. 144) in the book *Halliday in the 21st Century* (Webster, J. J. edited) that “applicable” indicates a unification of “applied linguistics” and “linguistics” to “give the sense of something that is capable of - having evolved in the context of - being applied” while “applicable” only suggests “some particular sphere of activity or other.”

relationship within levels of linguistic patterning, according to Butler (2003b, p. 247-8), though traditional designations are replaced by the concept “lexicogrammar” with a “partially specified ‘higher’ semantic level,” i.e., discourse, the research scope of SFG “ranges over phenomena which in other theories would be regarded as syntactic, semantic or pragmatic” although no differentiations have ever been made in these three aspects on the function that the language play in meaning expressing and comprehending. FG and RRG, on the other hand, not only follow labels “syntax,” “semantics,” and “pragmatics” but also divide the language into these three levels without making a clear distinction within them. For SFG, the lexical level is seen as the most delicate one, and lexical items are the products of choices from the less delicate grammatical realization level, i.e., system networks instead of the choosing objects at the beginning of the producing process, the “top-down” approach is therefore adopted by most scholars. In comparison, FG and RRG linguists both have a “bottom-up” tendency to conduct research on clause structures from the information of predicates. With metafunctional levels, the representational level of FG can be regarded as almost the same as the ideational, especially the experiential metafunction of SFG, and the interpersonal metafunction of the former corresponds to that of the latter. But while FG and RRG take the relationship between layers as hierarchical, i.e., one can contain another at a lower status, SFG sees any metafunctional layer as independent and parallel to each other. (See *ibid*, p. 249)

The reason for not using the term “syntax” in SFG is twofold: one is to make the fact that “syntax and vocabulary are part of the same level” (Halliday, 1994b, p. xiv) in presenting the language evident, hence the term “lexicogrammar” is chosen instead when “syntax” can only be assigned to part of the language grammar except vocabulary and morphology; another is to interpret the language “as a system of meanings, accompanied by forms through which the meanings can be realized” rather than “as a system of forms, to which meanings are then attached” in the proceeding direction of syntax, i.e., meaning and meaning potential of language is more favored by Halliday. Another unique character of clause structure that distinguishes SFG mostly from FG or RRG is the functional relations of clauses within a clause complex, as Halliday treats them as a constituency of the logical function of the grammar. The distinction of projection by not only verbal and mental processes but also propositions and proposals, for example, connects the logical function with the interpersonal function of language; and the changeable clause order in hypotactic clause complexes can influence the Theme or Rheme from the textual function of language. The biggest diversity within three theories stems from the division of units relations in complex sentences that FG chooses coordination and embedding to describe clause relations, RRG deems them to be coordination, subordination and cosubordination, while SFG (Sydney direction not Cardiff one) categorizes them into tactic relationships of parataxis and hypotaxis except for embedding, and logico-semantic relationships of expansion and projection, in which parataxis and expansion can be paralleled with coordination in FG. In detail, adverbial clauses are categorized into hypotactic expansion: enhancements for Halliday’s SFG instead of being embedded to satellite positions for FG; complement clauses are hypotactic structures in the former rather than embedded ones in the latter; though restrictive relative clauses are both regarded as embedded structures in two theories, non-restrictive relative clauses in SFG, however, are treated as hypotactic elaboration structures but not as embedded ones within a nominal group, and having relations with the whole dominate clause rather than with the nominal group itself. As SFG always focuses on language users’ actual behaviors in employing the language, it is no wonder that the logico-semantic relations, i.e., the division of expansion and projection, can be found in corresponding categorizations in FG or RRG. (See Butler, 2003a, p. 296-7)

In a word, being different from FG and RRG that takes more advantage of layering model (prediction vs. proposition, core vs. clause) in analyzing complex sentences and focuses more on fields like scoping (*ibid*, p. 301), general aspects of SFG such as the theoretical characteristics of supporting for “pragmatic adequacy” and rejecting the autonomy of syntax regarding semantics and pragmatics, the methodology of expanding the theory by moving into new spheres of activity and developing the theory by a testing-adjusting-applying-reinterpreting pattern, and the applicable distinction of solving practical problems by language users from and in various disciplines or directions make the trans-disciplinary and cross-sectional analysis of language under the SFG framework possible, and specific aspects like the system networks presenting the options of meaning potential, the relationship within levels of linguistic patterning, the “top-down” in place of “bottom up” approach from grammatical to lexical choices, the concept of “lexicogrammar” rather than “syntax” or “semantics,” the functional relations of clauses within a clause complex, and the division of units relations in complex sentences, etc. provide guarantees for delicate and systemic analysis of how meaning are realized by lexes and how language functions are performed through grammatical choices and then lexical realizations are influenced. With the research object being language ability that is cultivated and developed in social and cultural environments, the interdependent relationship between language and context can be described as the former being applied

in specific contexts and the latter predicting and expecting specific languages. Then, for the form of language, Halliday (1994b, p. xiv) argued that “A language is interpreted as a system of meanings, accompanied by forms through which the meanings can be realized. The question is rather: “how are these meanings expressed?” This puts the forms of a language in a different perspective: as a means to an end, rather than as an end in themselves.” Hence, form in SFG is rather seen as a means of expressing meaning and the embodiment of meaning and function, through the choice of which the purpose of constructing experience and expressing communication can be achieved. With the research methods of taking the meaning and function of language as the starting point for conduction, clause structure and discourse instead of sentences as the research scope, the collection of practical language use and the construction of corpora for language tendencies or probabilistic calculation as operation processes, the purpose of probing into language choices by different people from diverse educational levels, studying language varieties from kinds of social background, explaining language structure with its changes, discussing language usage and idiomaticity but not “grammaticality,” revealing communicative and interpersonal competence but not generative capability, etc. can all be fulfilled, and the investigation of this research is of deep theoretical and practical foundation and significance.

3.1.1.2 Socio-function for Academic Writing

Except for the reasons listed above for choosing SFG as a theoretical background for studying texts in academic writing, the socio-functional perspective of SFL is also significant in helping us understand and interpret meanings construal from texts, which takes language as a social semiotic system that socio-cultural factors, as well as the context of the situation that texts are produced due to variants of field, tenor, and mode, cannot be ignored. And the most fundamental reason for all these operations lies in the applicability of the theory in solving practical problems.

3.1.1.2.1 Language as Social Semiotic

In Section 3.1, the influence of works from Malinowski and Firth on Halliday’s theorization of connecting the grammar and meaning of language to context has been mentioned that Malinowski’s (1923, 1935) research on the Trobriand Islands in the South Pacific prompted him to treat the utterance produced in a particular context as the foundation for linguistic description to make clear of what the utterance actually means, especially when this kind of contexts of situation are encompassed in the whole context of culture; then Firth referred to the concept of “context of situation” to elaborate socioculturally related languages that “the meaning of any particular instance of everyday speech is intimately interlocked not only with an environment of particular sights and sounds, but deeply embedded in the living process of persons maintaining themselves in society” (Firth in Palmer, 1968, p. 13), and “meaning is a property of the mutually relevant people, things, events in the situation” (ibid, p. 14). The sociocultural tendency of the SFG theory was virtually developed by Halliday when two essential questions in the field of language research were proposed (Parret, 1974, p. 119-120): “one is intrinsic and the other is extrinsic. The intrinsic question is about the self-exploration of language: why is language as it is? Mankind could have evolved an enormous number of different semiotic systems; why did they evolve a system that has these particular properties that language has?” And the extrinsic question is about the social attributes of language: “The second question concerns language and the transmission of culture: how is it that the most ordinary, casual, informal, everyday uses of language, without any kind of instruction and even any kind of explicit understanding behind them, so efficiently transmit to the child the fundamental patterns of the culture, systems of knowledge, social structure, value systems and the like?” For Halliday, language is a social semiotic, a resource that people deliberately make use of to express meanings in a specific context. Hence, language should be interpreted within the socio-cultural context. As the culture itself is accounted for by semiotics, language can thus not only describe social structures and systems but also symbolize them. “A social reality (or a “culture”) is itself an edifice of meanings - a semiotic construct. In this perspective, language is one of the semiotic systems that constitute a culture; one that is distinctive in that it also serves as an encoding system for many (though not all) of the others.” (Halliday, 1978a, p. 2). In taking a “socio-semiotic” perspective for language research, “the shifts, the irregularities, the disharmonies and the tensions that characterize human interaction and social processes” are all meant to be interpreted, and besides trying to “explain the semiotic of the social structure, in its aspects both of persistence and of change, including the semantics of social class, of the power system, of hierarchy and of social conflict,” scholars also attempt to “explain the linguistic processes whereby the members construct the social semiotic, whereby social reality is shaped, constrained and modified - processes which, far from tending towards an ideal construction, admit and even institutionalize myopia, prejudice and misunderstanding.” (Halliday, 2009a, p. 200) Therefore, the study of academic writing discourse can not only be interpreted from lexicogrammatical

choices under the SFG framework but also be explained from socio-cultural factors of these choices when the language is regarded as a social semiotic in meaning transmission from the SFL perspective, though not all aspects of social structures, classes and social reality mentioned above would be presented.

3.1.1.2.2 SFG as an Applicable Linguistics

Another trait of SFG is the endeavor of treating the functional orientation theory to language “as a problem-solving enterprise and trying to develop a theoretical approach, and a theoretical model of language, which can be brought to bear on everyday activities and tasks.” (Halliday, 2006b, p. 19) This means that SFG theory is to be developed as “a comprehensive and theoretically powerful model of language” that can be applied to solve both theoretical and practical problems always being encountered by “many groups of people in our modern society who are in some way or other having to engage with language.” (Halliday, 2010, p. 3-4) Besides foregrounding choices for exploring differences “in the search for equivalence in translated texts,” offering researchers a semiotic system so that they can “improve teaching and learning in a systemic and principled way,” being “extendable to multimodal relations,” including “the intrinsic relation to social context” that can broadly encompass kinds of concepts like genre, register, affiliation, etc. (ibid: 11), SFL theory also provides a model for text or discourse analysis that be conducted from both qualitative and quantitative angles in especially the academic genre; hence its applicability and usefulness are overwhelmed in this research.

3.1.1.2.3 The Context of Situation

Meaning on the one hand is demonstrated by three metafunctions in SFG, and on the other hand resides in the text being constituted by three types of situational context, i.e., field, tenor, and mode that field is about the social action or “a complex of acts in some ordered configuration” and it usually describes something that is “going on” by taking use of the text including the “subject-matter” structure; tenor indicates the role characters or to be more specific, kinds of “socially meaningful” relationships within participants “that are specified to the situation” as well as their “permanent attribute,” such as “the speech roles, those that come into being through the exchange of verbal meanings,” scales of familiarity, power relations, etc.; and mode is “the symbolic organization,” which signifies “the particular status that is assigned to the text within the situation,” can include “the channel or medium” in spoken or written form, “the rhetorical mode,” and “the degree to which language is self-sufficient, rather than supplemented by other codes, in the interaction.” (Butler, 2003b, p. 243) (See Halliday, 1978b, p. 142-143) What is more, Halliday (ibid, p. 117) considered that the field has a tendency to make the scope of meaning affirmed as content in which the language plays an ideational function as an observer, tenor attempts to decide the sphere of meaning as participant in which the language plays an interpersonal function as an intruder, and mode as texture that language plays the textual function and has relation to the environment being described. Or the other way around, ideational meanings “realized in thing-names, transitivity structures and the like” reflect the field in the situation of context; interpersonal meanings “realized in moods and modalities, expressions of comment, attitude and so on” indicate the tenor or the relationship within people “who-all are taking part”; and textual meanings “realized as patterns of cohesion and the organization of discourse” present the mode or “what part the exchange of meanings is playing in the total unfolding scene.” (See ibid, p. 121)

For Master’s theses and doctoral dissertations written by Chinese English major students that are chosen for this research, before investigating how meanings are realized and how students’ writing experience are construed through choices of grammatical-lexical structures of clause complexes in meaning expressing, the context of situation that these texts are included in has to be clarified that the field is the representation of the discussion process for some objects by students who wish to obtain their academic degrees through demonstrating their master of English major field related knowledge, methods for conducting research, critical thinking for raising up questions and solving problems, abilities of referencing scholars’ works, making clarification and conclusions, etc.; the tenor includes two aspects: one is the relationship within students and their research objects that whether the latter are objectively described and explained in theses or not is of great significance, another is the relationship within students and readers that when the latter are supervisors or other professors in guidance and revision, an unequal status exists as students have to abide by suggestions from professors and then make modifications or sometimes act against their wills in revising and even referring to works from authoritative scholars only to cater to supervisors, but when the latter are also students or unprofessional researchers, the equal relationship exists within them for an unpublished thesis while the relationship turns to be unequal again for a published or a fixed one as the thesis can no more be revised and the writer has the right to refuse any suggestion from other people after publication; the mode is of course a written form that each part of the thesis is delicately planned and structured within a fixed writing framework

such as introduction - literature review - theoretical foundation - research methods - analysis - discussion - conclusion, and the thesis can be published in paper or online according to writers' acquirement. Hence, the content of the text is decided by the field that the ideational function of the language is performed by clarifications, discussions, explanations, references, conclusion, additions, etc., of, on, in, from, or even by the research object, in which the logico-semantic relations within clauses to fulfill these performances are the focus of this research; the status of the relationship within readers and writers is demonstrated by the tenor that the interpersonal function can be realized through the moods and modalities of proposals on the one hand and choices of projecting verbs or referencing methods on the other; and the textual structure is decided by the mode that the textual function can be presented by parataxis and expansion relations in clauses except for cohesions in the discourse.

3.1.1.3 The Contents and Main Components of SFG

When the question "what does SFG mainly contain in describing a language" is raised up, some basic principles would have to be included, i.e., to treat language as a text and system; as sound, writing and wording; as structure for configurations of its parts, and as a resource for choices among alternatives, hence dimensions and components in SFG for investigating language from aspects above are significant and should be clearly demonstrated: the first one is the term "constituency," which reveals the configurational or structural feature of language; then is "structure" in presenting how language is organized, with connection to its "rank" and "unit" that organizes it; followed by the essential term "system" and "function" in exploring why this grammar is called the "systemic functional grammar" with the descriptive term "delicacy" for the former and three contrast but unified "metafunctions" for the latter; as an essential character for "structure" and "system," the term "stratification" in dividing language into form and content planes clarifies where this study stands, with the relationship within strata being "realization"; the notion "instantiation" indicates the practical beginning point for research in probing into language, the cline of which would not exist without considering about the relationship within "text," "context" and "register"; the last pair can be "probabilistic" and "markedness" when any category in its system is applied in texts within fixed register in the context of situation, it must be varying from low probability (marked) to high probability (unmarked), displaying users' choice and preference for the category.

3.1.1.3.1 Constituency

When a structure is applied in describing experiential meanings, it can "be a configuration, or constellation, of discrete elements, each of which makes its own distinctive contribution to the whole" (Halliday, 1979, p. 203), this kind of structure would thus be linguistically presented as a "functionally labelled constituent structure" (ibid) that "larger units constituted out of layered clusters or bracketed strings of smaller units, each part having its own specific function with respect to the whole" (ibid, p. 204), hence the term "constituency" indicates a certain of "syntagmatic principle of compositional organization based on the part-whole relationship between a unit whole and its constituent parts" (Matthiessen et al., 2010, p. 76) that it usually relates to the concept of "rank" when the constituent of a unit at a rank can be regarded as unit of the next (below) rank. Though the term "interdependency" rather than "constituency" is more emphasized in this research as the degree of the former is clearly distinguished within clause relations, it is the comparison of the two that makes the denotation of the former clearer and more understandable. To be more specific, in SFL, these two terms are usually seen as complementary pairs for depicting syntagmatic organization when constituency tends to be selected for presenting multivariate structures such as textual waves, interpersonal prosodies, experiential configurations, etc., while interdependency is for univariate such as taxis which will be discussed later.

3.1.1.3.2 Structure, Unit, Rank

Having defined "constituency," the term "structure" in SFL signifies the syntagmatic ordering or regularities in language, i.e., what structure can be linked together to express the meaning as suggested by Halliday (2014, p. 22) or "denotes the syntagmatic patterning of a contextual or linguistic unit represented as a configuration of (structural) functions such as *Mood + Residue*, or *Ictus + Remiss*" (Matthiessen et al., 2010, p. 207). Though both are related to syntagmatic patterning, "structure" differs from the term "syntagm" as the former refers to "a configuration of functions" while the latter tends to be "an arrangement of classes in sequence" (Halliday, 1966, p.108) such as nominal groups or verbal groups. As each "metafunctional mode of meaning" (Matthiessen et al., 2010, p. 208) in the language would produce a unique type of expression, the structure for describing each metafunction would hence vary from each other.

The term "constituency" for denoting the part-whole relation indicates the existence of "the whole"

when constituent parts are discussed, and this “whole” for the “organic constituency structure” is what SFL scholars called the “grammatical units.” (Matthiessen & Halliday, 2009, p. 26) Being applied in depicting “the stretches that carry grammatical patterns” (Halliday, 1961, p. 247), “unit” is also defined as the “domain of systemic and structural organization ordered by the rank scale of a stratum from the most extensive to the least extensive” (Matthiessen et al., 2010, p. 234). One essential character of units is the “syntagmatic integrity” that they would often be thoroughly uniquely described by their structures, being diverse from other units, and the relation among these units is each including one or more than one of the units in the next (smaller) rank when they are inspected from top (largest) to bottom (smallest) in the system. Another trait of a unit is its possibility of being identified in functional terms as grammatical units are usually the entry points of system networks on the one hand and “function as constituents in their entirety” on the other since they can be functionally decided when a rank-based type of constituency is reached to (Matthiessen & Halliday, 2009, p. 26) or in Matthiessen et al.’s words, “units are the domain of realization of structure forming realization statements associated with terms in these system network: they are carriers of structures” (Matthiessen et al., 2010, p. 234).

Being different from the compositional aspect of the language “constituency,” the concept “rank” indicates the compositional layers structured by the belonging relationship (Halliday, 2014, p. 21) that “rank orders units into a hierarchy according to their constituency relation” (Matthiessen & Halliday, 2009, p. 26) or composition that “the highest ranking units consist of units of the rank immediately below” (ibid), and then include units of the next rank below until to the lowest rank where no more internal constituent structure can be found. Though rank is a universal grammatical concept for the distribution of units in any language, the number of it in a specific language is quite few, for example, the grammatical rank scale in English is “clause - group/ phrase - word - morpheme,” meaning that a clause is composed of groups or phrases, than of words, and of morphemes, and only four ranks are considered to exist.¹³ A basic feature of the rank scale is that it signifies realization patterns as the structural element of the units is realized by units of the rank immediately below. Another comes from the unique concept of “rankshift,” that a unit at a certain rank can be shifted to a unit at a lower rank, and only downward rankshift is allowed for in theory, but not the upward rankshift. In addition, “only whole units can enter into higher units” (Halliday, 1961, p. 248), i.e., a unit can be involved in another unit from a higher or the same rank instead of a unit from a lower rank even in more than one degree or a part of any unit.

3.1.1.3.3 System, Delicacy

As has been described in the former section, Halliday’s theory of the concept “system” originates from Saussure’s views on language when the abstract “langue” was seen as a system of structures as opposed to “parole,” the actual spoken discourse and language semiotics were treated as being organized by two basic relationships in the language system, i.e., paradigmatic and syntagmatic relations, in which the former was established upon his rough notice to corresponding paradigms within signifying forms and signified values. And then “system” developed from the Copenhagen School that Hjelmslev distinguished between “system” and “process” in regarding the bottom layer of the former as the paradigmatic relationship while that of the latter as the syntagmatic relationship, and the London School as Firth called the paradigmatic relationship “system” and the syntagmatic relationship “structure.” For Halliday, though the system was paradigmatic, it should not be deemed as standing for a list of signifying forms relating to a bundle of signified values but being organized as a system network which people can make multiple options from. Halliday believes that language is not only one of many meaning or semiotic systems but also a special one because of its possibility of expressing meanings in addition to presenting other meaning systems, so it is a useful tool for integrating these various meaning systems and for people to comprehend and transform the world.

Being four fundamental categories for the theory of grammar in Halliday’s early theoretical work on language, i.e., unit, structure, class, and system (Halliday, 1961, p. 41), the former two have already been explained; “class” is a group of members of a given unit defined by operation in the structure of the unit next above and thus describes a paradigmatic relation (ibid, p. 49); and “system” is rather a network of features for options that Firth described it as contrasting options in value realized by those in form, which (options) are features of the entire form and the entire value, i.e., linguistic “system furnish values for elements of structure, and the ordering of systems depends upon structure” (Firth, 1968, p. 183). In addition, the system is “restricted to a set or sets of paradigmatic relations between commutable units or terms which provide values for the elements of structure,” (ibid, p. 184) and is thus defined as an “ordering on the other axis” being contrasted with “structure” to depict those “patterns in what could go instead of what” (Halliday & Matthiessen, 2014, p. 22), account for “the occurrence of one rather than

¹³ In SFL, clause complex is at the same rank as clause, which will be explained in the next section.

another from among a number of like events” (Halliday, 1961, p. 256) and present paradigmatic organization at any stratum, i.e., phonological, semantic or grammatical, of language. By thinking in systems, the description of any research object should be based on its relations to other units in the hypothesis of the theory, which means that a category or a semantic component needs to be depicted by adopting a trinocular perspective: “from the standpoint of its realization in the lexicogrammatical system (i.e., “from below”); from the standpoint of the functions of the linguistic system in relation to some higher-level semiotic that is realized through the linguistic semiotic (i.e., “from above”); and from the point of view of the organization within the semantic system itself (i.e., “from the same level”)” (Halliday, 1978a, p. 130-1). Hence, a complex or comprehensive system is what SFL scholars intend to investigate in describing the language organization, and grammatical or other features of language should thus best be understood from the abstract description of sets of options (systemic features).

According to Halliday (2003, p. 209), the essence of systemic grammar is that the language system is “a paradigmatic set of alternative features, of which one must be chosen if the entry condition is satisfied.” To be specific, language is a complex system in which many subsystems at different ranks are included, such as the system of the clause complex. As “a language is a resource for making meaning, and meaning resides in systemic patterns of choice, the grammar of a language is represented in the form of system networks, not as an inventory of structures” (Halliday, 2014, p. 23) that the system network is constituted by multi-level systems which can provide people with choices between two or more features or items, and the product of this kind of ongoing selection in a system network becomes a text. In making various choices, the first element for selecting items is logic, i.e., the relationship between the items being selected and the included item for selection must be logically affiliation. Otherwise, these items would not be included in the system network. For Halliday, “the grammar of language is that level of linguistic form at which operate closed systems, i.e., sets of terms with following characteristics: (a) The number of terms is finite: they can be listed as A B C D, and all other items E . . . are outside the system. (b) Each term is exclusive of all the others: a given term A cannot be identical with B or C or D. (c) If a new term is added to the system this changes the meaning of all the others.” (Halliday, 1961, p. 244-5) The upper level in such kind of network system is the entry condition that items having semantic characteristics of this upper level conform to the entry condition, and further choices can then be made within these items. Usually, “the entry condition is a simple feature or a feature complex and these features are terms in other systems.” (Matthiessen et al., 2010, p. 211) When choices in the system are made, the final item or combination of items available for selection is the “choice expression” (selection item), and when selection items in the system no longer serve as an entry condition for a more delicate selection system, the item has reached to the end, and at this time the selection is terminated.

Delicacy indicates the degree of refinement or, as Halliday (2002a, p. 58) defines: “delicacy is the scale of differentiation, or depth in detail.” Being a cline from general to more specific, delicacy at the one end is limited to “the primary degree in the categories of structure and class”; at the other end to “the point beyond which no further grammatical relations obtain: where there are no criteria for further secondary structures or systems of secondary classes or formal items.” (ibid) In a system network, it relates to the arrangement of systems from left to right according to diverse entry conditions and is analogous to rank in the structure. In this research, the delicacy of the clause complex system in thesis writing by Master students and PhD candidates from China will be demonstrated and compared to choices of specific items, the terms “system” and “delicacy” are hereby of necessity for clarification.

3.1.1.3.4 Function, Metafunction

Function, as has been mentioned before, can be dated back to Malinowski’s functional approach in anthropology and the development of pioneers from the Prague School. Being one of the characteristics of language as an integrality, it indicates the structural principle presented through the whole system. The reason language is functional is that it “evolves together with the social or ecological environment and then develops in the individual together with the environment” (Matthiessen et al., 2010, p. 101). The “function” of language in connotation can be closely related in two aspects: the extrinsic function, which signifies the use of language, and the intrinsic function, in referring to the internal organization of language that for SFL, the latter one would usually be adopted in two various but directly connected folds: one is the integral construction of language as far as the aspect of diverse modes of meaning in different phases of language development is concerned; another is the “local organization of the structure of a unit” (ibid, p. 102) at any level or rank of language. In describing the construction of language from the aspect of diverse modes of meaning, Matthiessen, et al. (ibid) considered that function can be divided into three kinds or three phases from the angel of language development or evolution, i.e., microfunction, macrofunction, and metafunction, in which language develops from organizing protofunctional elements into some microfunctional ones such as regulatory and interactional structures that are strongly influenced by contexts of use at the first phase (infancy language); to generalizing microfunctions to

macrofunctions, i.e., the learning function or the mathetic macrofunction, and the doing function or the pragmatic macrofunction at the second phase (post-infancy language); and then to transforming macrofunctions into metafunctions, i.e., ideational, interpersonal and textual functions at the third phase (adult language) that these three metafunctions are complementary and simultaneous in language system, can thus be mutually instantiated and realized structurally by diverse modes of meanings. And in syntagmatically being organized into a unit, each element of structure performs at least one structural function, thus representing its contribution to the organic integrality of the unit that it belongs to.

The same viewpoint can be found in Halliday's distinction of "functional" senses for systemic grammar when it is depicted from three aspects: applications that systemic grammar was developed to support practices in connecting with language use, the functional or register variation of language, hence are considered as important and necessary, corresponding to the extrinsic function of language; explanations as the grammar itself and the specific forms for describing the grammar are explained from the angle of functions that language developed to perform, showing the intrinsic function of language; and representations that "a structure is an organic configuration of functions" instead of being "a tree with nodes labelled as classes," relating directly to the structural organization of a unit that language intrinsically functions as. (See Halliday, 1994b, p. 4506) Another confirmation is from his statement that systemic grammar is functional when the interpretation of texts, the system, and the elements of linguistic structures are taken into account: as texts (no matter written or spoken) are adopted to keep a record of language use in specific context, and language evolves to fulfil human needs, the organization of language is of necessity being functional to satisfy the description of these uses, the functional grammar is hence designed to explain everything, especially is developed to describe "how the language is used," to perform the extrinsic function of language; as language meaning are essentially functional, languages are structured for expressing two kinds of meanings: the "ideational" or reflective meaning, and the "interpersonal" or active meaning, hence language performs the ideational function to enable people to learn about the environment and the interpersonal function to suggest the interaction within persons, in addition to the metafunction that is manifested by the linguistic system, i.e., the textual function which connects two former functions together in an organized text; as the element of a language is interpreted according to its function in the whole linguistic system, the systemic grammar is functional in construing all language units "as organic configuration of functions," to fulfil the structural organization related intrinsic function of language. (See Halliday, 1994a, p. xiii-xiv)

It has been demonstrated from above that the function of language can not only indicate the purpose or method of using language in certain contexts, i.e., from the extrinsic perspective, but also denote the whole organization of language that is structured along functional lines, i.e., from the intrinsic perspective as functions of language are essential when it evolves or develops in human species. For referring to the integral component or organization within the systemic language theory that language performs in its development, the concept of metafunction was adopted, which distinguishes it from the macro- or micro-function in earlier phases. For Halliday, language can name things, construe human experience into categories, and then taxonomies by choosing more names that the latter two would be correspondingly changed via different languages. Language hence enables the existence of a theory that part "resources of the lexicogrammar of every language are dedicated to this function," i.e., the ideational metafunction, which can then be divided into the experiential and the logical metafunctions. The clause of the grammar for this function "is a figure," revealing processes such as "doing or happening, saying or sensing, being or having" with numerous participants and circumstances when language is seen as a reflection. Besides construing, language also presents our personal and social relationships with other people and conveys more active meanings that "the function of the grammar is 'language as action,'" the concept "interpersonal metafunction" is hence introduced to express the interactive and personal functions of language. The clause of the grammar for expressing this kind of function can be a proposition or a proposal, help us "inform or question, give an order or make an offer, express our appraisal of and attitude towards" people being addressed and things being discussed. The third metafunction is textual, which correlates to "the construction of text" and enables or facilitates the former two functions in "building up sequences or discourse, organizing the discursive flow, and creating cohesion and continuity" during the producing process of discourse. (See Halliday & Matthiessen, 2014, p. 30-31)

3.1.1.3.5 Stratification, Realization

The stratification of language is attributed to Hjelmslev, who proposed that language contains three levels, i.e., content, expression, and substance, as "the sign is a two-sided entity, ... and with effect in two respects: "outwards" toward the expression-substance and "inward" toward the content-substance." (Hjelmslev, 1953/1961, p. 58) Language, therefore, is not exactly a symbol system mentioned by Saussure but rather a system of signs with interrelationships between various stratifications, which is the

category that organizes “language in context into subsystem” (Matthiessen et al., 2010, p. 205) based on the abstraction extension of symbols. When traditional grammar is divided into categories such as “pronunciation,” “orthography,” morphology, syntax, etc., language is confirmed to be a “complex semiotic system” that has various levels, or *strata*. For SFG, “grammar and vocabulary are not different strata” but are at two ends of a line, being organized into “lexicogrammar,” as do syntax and morphology, which are on the cline of grammar. (Halliday & Matthiessen, 2014, p. 24) As grammar has to deal with “what goes on outside language: the happenings and environments of the world, the social processes people participate in”; as well as “structuring the construal of experience, and the enactment of social processes,” language thus has two strata for the content plane: one is for experiential and interpersonal relationships being transformed into meaning, which suggests the stratum of semantics; the other is the meaning above being transformed into wording, which is the stratum of lexicogrammar, providing for the expansion of the meaning potential of language. (ibid, p. 25) Another important condition that enables other conditional changes of SFG is the division of the expression strata into phonology (the system of sounding) and phonetics.

Being contrasted with the “mutation” of the transformational grammar, the “realization” of SFG is also a basic principle for underlying grammatical theories (Gleason, 1965, p. 106) that the existence of “realization” comes from the fact that texts have to be “realized in sound or writing” when they are accessed by people, which means that language instances from higher strata cannot be immediately obtained no matter they are selections from meaning or wording. (See Halliday & Matthiessen, 2014, p. 33) According to Halliday & Matthiessen (ibid, p. 25), “the relationship among the strata, i.e., the process of linking one level of organization with another” is defined as “realization” that each strata realizes the next one: lexicogrammar realizes semantics, phonology realizes lexicogrammar, and phonetics realizes phonology. But the relationship between the content plane and the expression plane, and between strata being described above that constitute these planes is inter-strata, which is analogous with “expression” or “coding,” that the former, i.e., between lexicogrammar and phonology is more conventional or arbitrary, while the latter, i.e., between semantics and lexicogrammar or between phonology and phonetics is natural. For the intra-stratal realizational relationships, i.e., “the inter-rank relationship between a higher rank and the next lower rank on the rank scale of a given stratum, and the inter-axial relationship between the paradigmatic axis and the syntagmatic axis of a given rank” (Matthiessen et al., 2010, p. 171-2), the most regular way of presenting the modeling of the realizational relationship of language in a specific context for SFG can be the realization statement, which comes out related to a category in a system and then includes one realizing operator and at least one realizational operand. As suggested in Figure 3 (ibid, p. 172), the organization of diverse kinds of realization statements has started from one statement and then broadened into an operator with one or more operands, but the delicate explanation of each type would not be offered here.

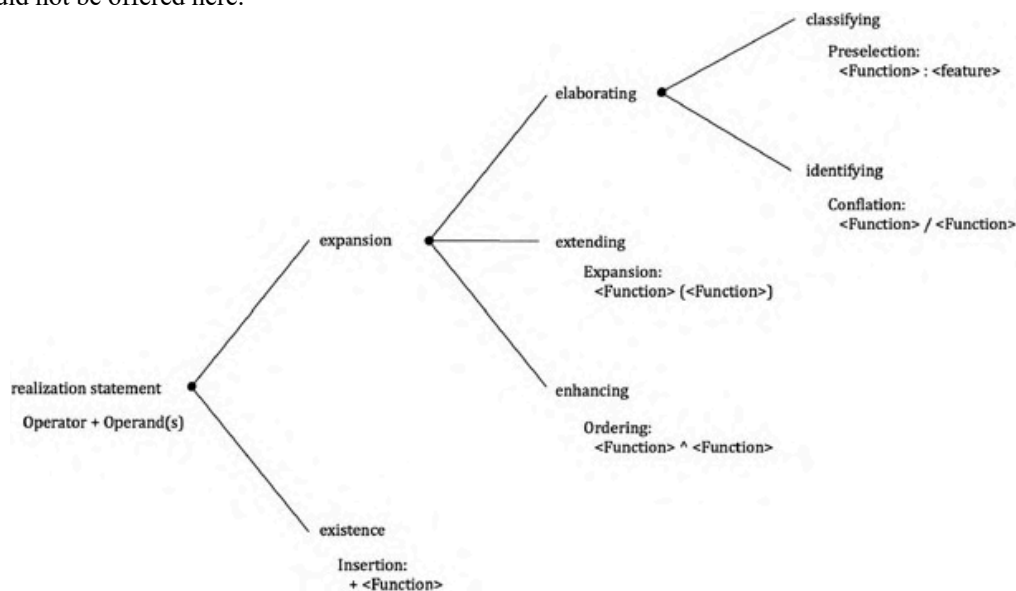


Figure 3 The organization of different kinds of realization statements in SFG

3.1.1.3.6 Instantiation

Except for “the hierarchy of stratification,” which orders levels of patterning according to strata abstraction, and “the spectrum of metafunction” which classifies patterns into diverse modes of meaning, “the cline of instantiation” is also “one of the global dimensions” in organizing language in context,

which suggests a continuum of patterns on aspect of generalization, (Matthiessen et al., 2010, p. 122) or the cline on which the relationship within system and text can be defined. As suggested by Halliday (2014, p. 27-8), system and text are two poles of a cline that the former is “the underlying potential of a language” when seeing language as “a meaning-making resource,” while the latter is “an instance of an underlying system.” As language is presented via real instances in the system, the system of language is not a gather of all inclined texts but a “theoretical entity” with which certain characteristics can be endowed and “considerable explanatory power can be invested,” system and text are therefore closely connected on the cline of instantiation, within which are “intermediate patterns,” i.e., subpotential or sub-systems been seen from the system pole or instance types from the instance pole. (ibid, p. 28) The cline of instantiation ontogenically appears gradually as a child begins to express meanings, especially “when instances of potentially symbolic behavior become systemic.” Though it is an organizational norm that functions in all physical, biological, social, and semiotic systems, as to language, its instance pole concerns more about “what speakers mean on a given occasion,” presenting “their acts of meaning,” and its potential pole is about “what speakers can mean.” (Matthiessen et al., 2010, p. 121-2) The detailed intersection of the cline of instantiation and the stratal (from the content plane) has been demonstrated in the following figure (cf. ibid, p. 125) that the system is at the potential pole of the cline, and the instance is another pole, the register is hence the subsystem at both the semantics and the lexicogrammar strata. The significance of the cline of instantiation, according to Matthiessen et al., (ibid, p. 123) lies in its prescribing the research sphere of observing, analyzing, describing, and theorizing the scientific processing of language from both methodological and theoretical aspects that systemic functional scholars can investigate the phenomenal domain of language through inspecting, sampling and interpreting instances, i.e., texts in contexts of the situation at the instance pole, then “move further up towards the potential pole” to make generalizations of some sampling texts that are typical of a higher point at the cline such as “a text type or a register,” or of semiotic systems from the potential cline. As the systematically chosen texts based on certain criteria are defined as corpus, the generalizations of “points along the cline of instantiation” via annotating corpora, categorizing them into patterns, and explaining analyzing results become feasible and reasonable. More significance, explanation, and application of the cline of instantiation can be referred to in Section 3.3.4.

3.1.1.3.7 Text, Context, Register

The relationship within *text*, *context*, and *register* can be clearly demonstrated by Figure 4.

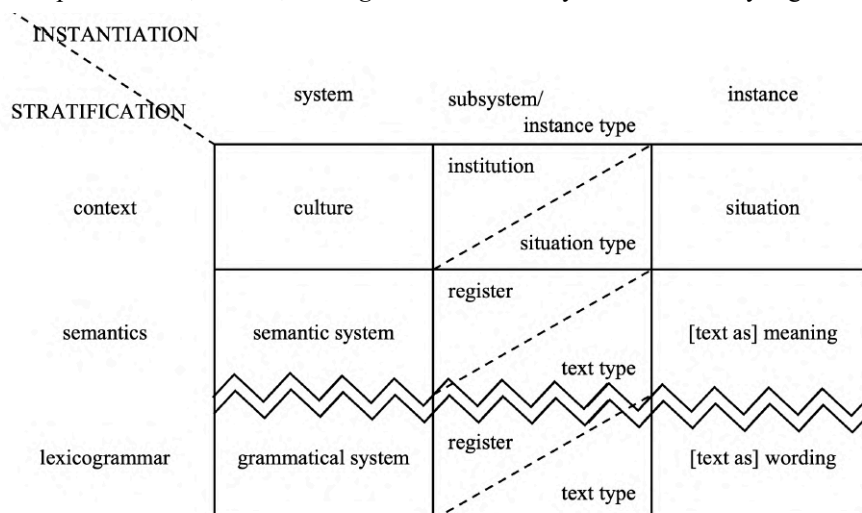


Figure 4 The intersection of the cline of instantiation and stratification

When the cline of instantiation is moved from the instance pole to the system pole, i.e., when a single text is connected with others by the same criteria (no matter from semantics or lexicogrammar strata provided that they are “systematic and explicit”) to investigate common linguistic patterns and describe them as a text type because texts would change systematically based on the essence of the contexts they are applied in that all kinds of text types are from the application of language in diverse contexts, the register is then formed in denoting the instantiation patterns of the whole system related to a specific “type of context (a situation type).” (cf. Halliday, 2014, p. 29)

Text, as has been described above, can be the “highest unit on the rank scale of semantics” or lexicogrammar functioning in a context of the situation at the cline of instantiation, is hence closely related to the context that it is usually structured based on “the contextual structure (generic structure,

schematic structure)” being projected onto it. Though the contextual structure can also influence the development of social activities and presentations by making use of semiotic systems, the degree that is instantiated by language as texts lies on “the mode of the context,” or specifically, on the division of work that the systems functioning in the context. Being a semantic unit, a text does not possess any structural closure and can hence either be seen as a process developing through time that adjacent choices in a text indicate lexicogrammatical patterns of wording in construing the meaning from a logogenetic perspective or as a product after a while. Usually, a text can be “variable in size and nature, ranging from a line” to longer or even a macro-text. For the difference between “text” and “discourse,” though both refer to the same entity in SFL, the latter is a text viewed from the sociocultural context, whereas the former is discourse being treated “as a process of language.” (See Matthiessen et al., 2010, p. 218-9)

The notions of “context of situation” and “context of culture” again date back to Malinowski, who believed that a language is primarily rooted in the custom of the culture and social life of the people who speak the language through his observations on the language used by indigenous people on the Trobriand Islands. As it is difficult to fully understand the language without considering the context, its importance for interpreting texts is therefore reckoned and investigated. Being “a higher-order semiotic system above the linguistic system,” context can be both “first-order” context and “second-order” context as a semiotic model that the former is of “social processes” (the first-order field) and “social roles and relations” (the first-order tenor); and the latter is of “linguistic and other semiotic processes” of the aspects of experience (the second-order field) and “the speech roles and relations” (the second-order tenor) produced by semiosis, as well as the function of “language and other semiotic systems in context” (the second-order mode). Though context was once theorized and depicted primarily from the perspective of stratification in the “genre model” that it was organized by ideology-genre-register (Martin, 1992a), from Fig. 3.1.1-2, it is quite obvious that context later on is described from two semiotic dimensions, i.e., the stratification and the instantiation that it “extends along the cline of instantiation from the potential pole (context of culture) to the instance pole (context of situation),” with the intermediate points being subpotential (institution) and instance type (situation type). (ibid, p. 77)

Register for Halliday (See Thibault, 1987, p. 610) is not the same as the “context of situation” when the former is seen as “a setting of probabilities at the semantic level,” i.e., a functional variety of language depended on some specific features of the context and by what the user is socially acting, that register variables indicates the quantitative re-ordering of these systemic probabilities; while the latter is “an interface” that include categories of “field, tenor, and mode,” register in consequence is defined as “the configuration of semantic resources that the member of the culture associates with a situation type. It is the meaning potential that is accessible in a given social context.” (Halliday, 1978a, p. 11) In Halliday’s (2013, p. 88) opinion, Martin, however, intends to make “a stratal distinction” between the notion register and the genre that the latter is a level above the former, develops from the field but explains “the nature of the activity” on purpose or intentionally. Therefore, in interpreting two communication planes of context, register (context of situation) is regarded as a situation type, or “the expression form of genre,” being structured by field (social action), tenor (role relationships), and mode (symbolic organization), reflecting metafunctional varieties in its expression form, i.e., the language, “leaving genre (context of culture) to concentrate on the integration of meanings engendered by field, tenor, and mode as systematically related social processes.” (See Martin, 1992a, p. 495)

3.1.1.3.8 Probabilistic, Markedness

Halliday adopted “approximative” or “probabilistic” ideas from the information theory and considered that one of the inherent characteristics of language is probability: “A grammar is an inherently probabilistic system, in which an important part of the meaning of any feature is its probability relative to other features with which it is mutually defining.” (Halliday, 2002a, p. 400) This “degree” phenomenon in language exists not because some forms are ungrammatical but because of the basic characteristics of language as it is primarily conventional and then constantly evolves during application. Hence, the “quantitative feature of the whole system of language (or, more generally, of other semiotic systems) or of a local given registerial subsystems” indicates the “distillation of relative frequencies in text.” (Matthiessen et al., 2010, p. 163-4) As few people have applied the analogy of probabilistic principles to the description of grammatical systems, Halliday considered that probability and corpus can be related to each other: when an item is chosen from the grammatical system, it is not just a choice from *a* to *b* or *c*, but from the probabilities that *a* or *b* or *c* entails, which are in fact derived from the corpus. Or in Halliday’s (2003, p. 23) words: “The frequencies that we observe in a large corpus represent the systemic probabilities of the language; and the full representation of a system network ought to include the probability attached to each option in each of the principal systems,” although these global probabilities are limited to various conditions: they may vary “locally with the register,” “transitionally throughout the

(logogenetic) unfolding of a text,” and “conditionally according to selection in other systems.” (See Halliday, 1995/2003, p. 410)

The term “markedness” is closely related to “probability” when it is described by approximate values from 0 to 1 that “marked” is contrasted with “unmarked,” the latter of which is used to refer to items with higher approximations in a system. As the differentiation is embodied in many systems, names of any term would be reflected obviously by being depicted as “unmarked ...” or “marked ...” The “markedness” is mainly realized threefold: for the “hierarchy of axis,” the “marked” term in a system is inclined to have a more important realization on the syntagmatic axis for “do something,” while the “unmarked” term may be absent of a “syntagmatic marker” for “do nothing,” and the former tends to enable a more delicate system in differentiating diverse kinds of marking; regarding “the cline of instantiation,” the marked term is usually chosen significantly much less than unmarked one in texts, the contrast of which can be represented by a skew in probability actually from zero to one; and with regard to “the hierarchy of stratification,” unless there is a specific reason for choosing the marked one from the above stratum, otherwise the unmarked term would often be selected. (See Matthiessen et al., 2010, p. 236-7) The probabilistic theory and the importance of markedness will be further explained in Section 3.3.4, without which the quantitative analysis of comparing categories and summarizing linguistic patterns of clause complex recursion in this thesis would be impossible.

3.1.1.4 Drawing for the Clause Complex

Having clarified the specialty of SFG from other functional or related grammars in describing systems of language, construing meanings in daily life use, treating texts from socio-functional perspectives, being applicable for texts analysis, emphasizing the influence of the context of situation in producing texts, pointing out three meta-functions that language play in meaning expression, and adopting probabilistic angle of language choice to demonstrate the unequal status of items from language system in application, the theoretical bases for the focus of this study, i.e., the clause complex system have been built that without which the analysis of categories’ choices in the system by diverse students in academic writing register from our corpus would be impossible and segmented.

3.1.2 Clause & Clause Complex

Before getting straight to the point of this research, the basic connotation and denotation of the concept “clause” has to be elucidated, of its location in the whole grammatical system of language, its definition by different scholars, three meta-functions that it plays in transmitting knowledge to, providing information for, and getting meaning transformed to people, and of its organization. The notion of “clause complex” then will be explained from both in and around the SFG theoretical frame, of its formation of and diversities from clauses, its location from the perspective of metafunction, rank, and stratification, its definition, differentiation from the traditional term “sentence,” its organization, and then the sub-system of taxis and the logico-semantic relations (expansion and projection). This section will end with an example of a complex clause complex chosen from the corpus to linearly and horizontally present the intersection of taxis and logico-semantic relations, as well as introduce the following sections on expansion and projection.

3.1.2.1 Definitions of a *Clause*

The location of the *clause* is quite clear when it is defined as a “grammatical unit of the highest rank on the lexicogrammatical rank scale” (cf. Tomori, 1977; Osisanwo, 1999; Halliday, 1985a; Saragih, 2007, etc.), hence when a clause is speculated from the structure of language stratification, it is a unification of diverse “metafunctional strands of meaning,” which realizes a move (a proposition or a proposal from the interpersonal metafunction), a message (from the textual metafunction) and a figure (from the experiential metafunction). (See Matthiessen et al., 2010, p. 72) As Halliday (1994a, p. 31) noticed, in the second half of the nineteenth century, grammarians made the distinction between three types of “Subject”: the grammatical Subject (indicates something being predicated); the psychological Subject (“the concern of the message”); and the logical Subject (“doer of the action”), he (ibid, p. 32) replaced the “grammatical Subject” by Subject, “psychological Subject” by Theme, and “logical Subject” by Actor that these three different strands of meaning in the clause result to three types of function: the Subject “functions in the structure of the clause as an exchange”, the Theme “in the structure of the clause as a message,” and the Actor “in the structure of the clause as a representation” (ibid, p. 34), which are later on normally referred to in SFG as metafunctions, for the aim of distinguishing it from functions or “functional roles” such as Actor, Subject or Theme in traditional grammars. The “clause as exchange” performs the interpersonal metafunction, which is defined as “enacting social relationships”; the “clause

as message” fulfills the textual metafunction, “creating relevance to context”; while the “clause as representation” is called the experiential metafunction for “construing a model of experience.” There is still a fourth metafunction, i.e., the logical one, which concerns “constructing logical relations” and enables us to combine clauses (and also groups, etc.) into complexes. (ibid, p. 36)

The reason for not using the term “syntax” by Halliday comes from two folds: one is that “syntax” is not separated from “vocabulary” as the latter is just seen as the most delicate end on a cline of grammar; another is to refrain from a mixture of the same primary term in the Formal grammar when meaning, however, has the biggest significance in functional grammar that “a language is interpreted as a system of meanings, accompanied by forms through which meanings can be realized.” (Halliday, 1994a, p. xiv)

From the lexicogrammar perspective of language, a clause is “a point of entry or domain or a number of simultaneous systems” (Matthiessen et al., 2010, p. 72) to perform three metafunctions, i.e., interpersonal by the Mood system, textual by the Theme system, and experiential by the Transitivity system that diverse realization of structures lead to the variation of languages. As language is a tool for people to carry out social communication, one of its functions is to reflect the relationship between people. Specifically, it expresses the role relationship, status, and intimacy between communicators, such as questioning and responsiveness, the tone and attitude of the communicators, judgment, and evaluation of the possibility and frequency of the occurrence of things, the fulfillment of stuff by ways of role interactions, etc. This function is called the *interpersonal* function, embodied by the mood and modality systems. In performing this function, “social relations can be established and maintained, social groups are delimited, and the individual is identified and reinforced.” (cf. Halliday, 1970, p. 143)

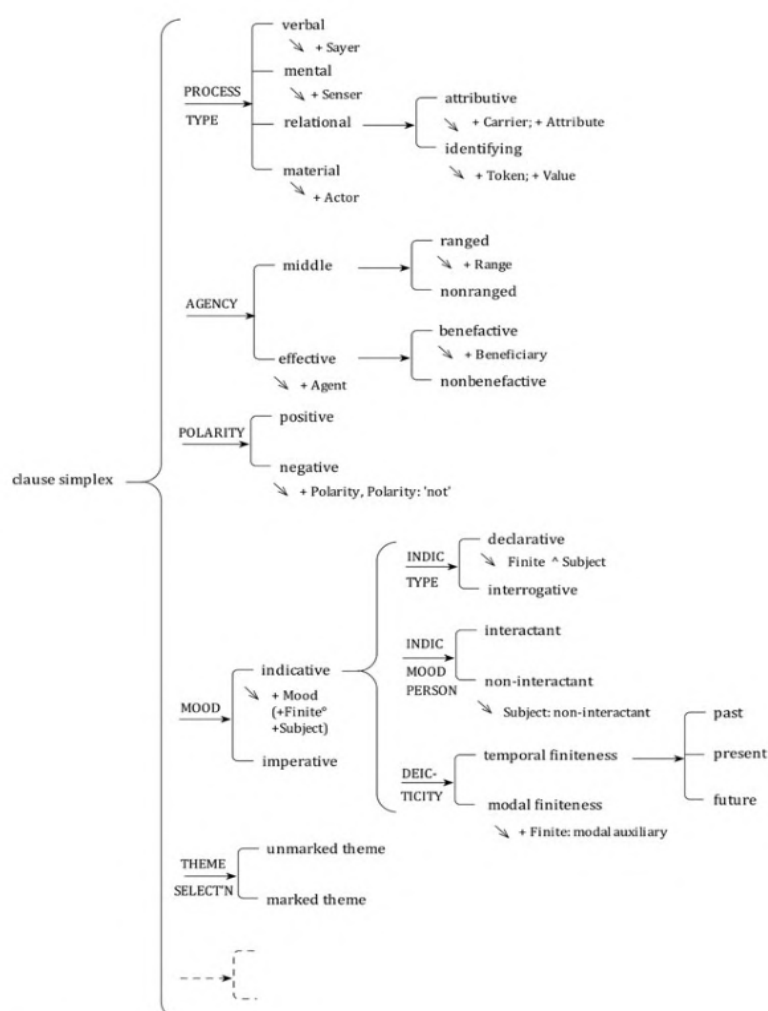


Figure 5 The system network of the clause complex in English (ibid, p. 44)

Textual function is the speaker’s organizing ideational and interpersonal functions into discourse. It makes the text a semantic whole, connects passages of discourse with the situation, and enables the text to be an integral part of the context that speakers can only generate a discourse consistently and commensurately with the situation, and listeners or readers can distinguish a text from a random group of sentences. Thus, Halliday believes that the basic unit of the language in actual use is not a word or a

sentence but a relatively complete “text” for expression. What embodies the textual function is primarily the information unit of a clause, the main function of which is to convey information. From the perspective of information content, the information of a clause generally contains two parts: known information and new information. From the information form perspective conveyed by a clause, it can also be divided into topic (theme) and predicate (rheme). In addition, the textual function encompasses the cohesion system, denoting the anaphoric relationship between words and various connection means.

Language, being a reflection of processes and things that exist in the subjective and objective worlds, also performs the function of what is “about” or “discussed,” i.e., the “experiential function.” Another function is “logical,” which appears in linear recursion structures expressed by hypotaxis and parataxis. Since both are “the speaker’s experience of the real world, including the inner world of his own consciousness” (ibid), they are collectively called the overarching “ideational function,” which is exactly embodied by the transitivity system, including actors, processes, goals, environments, etc. Six different “processes” are realized in the system: material process, psychological process, relational process, behavioral process, verbal process, and existential process. Because different types of processes would be constructed by various participants, goals, environments, etc., these processes are mainly decided by the predicate verb of a clause. In performing ideational function, language can “give structure to experience, and help to determine our way of looking at things.” (ibid)

For the organization of a clause, it can include units of the rank directly below, i.e., the rank of group or group/phrases in many languages (though it can be the rank of a word in some languages) from the lexicogrammar of a language, which is demonstrated by the automatic parsing of a list of classes of units at the lower rank in the machine; or is indirectly realized by a tone group from the phonological system of the language, i.e., information units, being coextensive with clauses in the unmarked case, are realized by tone groups. (cf. Matthiessen et al., 2010, p. 72) In the system network of the clause in English (see Figure 5), except for clause complexes, the clause simplex can be selective of process type (including verbal, mental, relational and material processes), agency (middle or effective), polarity (positive or negative), mood (indicative or imperative) and theme (unmarked or marked theme) in diverse systems in performing three metafunctions mentioned above, details of which would not be spread here because of space limitation.

3.1.2.2 Connotation of the *Clause Complex*

In describing the rhetorical relations of the semantic organization of text from the Rhetorical Structure Theory, Thompson & Matthiessen (1988, p. 301) came to realize that “clause combining in grammar has evolved as a grammaticalization of the rhetorical units in discourse defined by rhetorical relations.” As rhetorical relations of projection and expansion (unmarked by structural conjunctions or marked by non-structural, cohesive conjunctions) form rhetorical complexes and then organize into a network with each other, rationally extending either to “the global organization of the whole text” or to “the local organization of units (figures/ moves/ messages) realized by clauses” semantically, the local rhetorical complexes may be realized by clause complexes. (cf. Matthiessen, 2005, p. 798-9) Hence, clause complexes share the same scope of logico-semantic relations as rhetorical complexes, i.e., the division of expansion and projection, or in Halliday’s (2014, p. 430) words: “Semantically, the effect of combining clauses into a clause complex is one of tighter integration in meaning: the sequences that are realized grammatically in a clause complex are construed as being subsequences within the total sequence of events that make up a whole episode in a narrative.”

In detail, when a clause can be a textual message (a large quantity of information in the information stream), an interpersonal move (exchanges during conversational interaction), and an experiential figure (changes during events), the clause complex is a list of messages, moves and figures that textually, clause complexes suggest the progression or movement of themes in a text as each clause is internally thematic but mutually related in sequentially organizing into a text, and provide continuous new information for receivers as each clause encompasses an information unit and is structured as given information at the beginning and new information at the end in clause combination, i.e., the original clause in a clause complex would either be paratactically combined or hypotactically complexed by other clauses, with the former being the core of the message and the dependent or added clauses being the peripheral information for the reinforcement of the message; interpersonally, clause complexes weaken the mood element of dependent or added clauses though each clause has its unique modal elements; and experientially, clause complexes have to consider the transitivity system or the figure in each clause and combine them according to clause relations.

As an element in a clause, such as the circumstantial one, “contains only a minor” but not a major process, it cannot “construe a full figure,” “enact a proposition/proposal,” or “present a message” like a clause does. While the clause-hood unit in a clause complex performs the same function as a clause (i.e., the experiential, interpersonal, and textual system) when the former is connected to the latter by

expansion or projection, circumstantial elements hereby are only “part of the ‘configurational’ structure” of a clause, but a clause is a structure for “forming part of a chain-like or serial structure” of a clause complex. In the production of a text, whether to enhance a clause “internally” by choosing a circumstantial element or “externally” by choosing another clause to form a clause complex depends highly on “how much textual, interpersonal and experiential semiotic ‘weight’ is to be assigned to the unit”: the more weight indicates the higher possibility of being “constructed as an interdependent clause in a clause complex” instead of an element in a clause. (cf. Halliday, 2014, p. 434)

As an element in a clause, such as the circumstantial one, “contains only a minor” but not a major process that it cannot “construe a full figure,” “enact a proposition/proposal,” or “present a message” like a clause does; while the clause-hood unit in a clause complex performs the same function as a clause (i.e., the experiential, interpersonal and textual system) when the former is connected to the latter by expansion or projection, circumstantial elements hereby are only “part of the ‘configurational’ structure” of a clause, but a clause is a structure for “forming part of a chain-like or serial structure” of a clause complex. In the production of a text, whether to enhance a clause “internally” by choosing a circumstantial element or “externally” by choosing another clause to form a clause complex depends highly on “how much textual, interpersonal and experiential semiotic ‘weight’ is to be assigned to the unit”: the more weight indicates the higher possibility of being “constructed as an interdependent clause in a clause complex” instead of an element in a clause. (cf. Halliday, 2014, p. 434)

Location

The relation of projection or expansion can either be realized by two clauses that are cohesively connected or by a single clause taking the “prepositional phrase or adverbial group” as its “circumstantial element” except for being realized by two clauses in structurally organizing into a clause complex, the latter of which therefore is an intermediate condition that on the scale of grammatical integration or interdependence degree, one clause can be dependently connected to another in a clause complex, being nearer to the circumstantial augmentation pole and having unequal status; or independently connected to another, being nearer to the cohesive structures pole and having equal status. Hence, regarding metafunction, the clause complex is structured by logical relations from the ideational metafunction, which is diverse from the circumstantial augmentations of the clause in performing the experiential metafunction or cohesive sequences for the textual metafunction; with regards to rank, the clause complex is at the same rank of the grammar as the clause does, i.e., the highest rank as it only connects clauses by complex logical relations instead of taking them as the experiential constituency; and for stratification, the clause complex has been explained to be the grammaticalization of semantic organizations of projection or expansion from above and the realization of organized tones in speech or structured sentences in texts from below. (cf. Halliday, 2014, p. 434-5)

Definition

As the clause complex system functions together with “the experiential structures of the transitivity system” in offering “language users with structural resources to construe logical relations within experiential events” (Egins, 2004, p. 256), the ideational meanings are thus conveyed in transforming daily life events into texts; and the understanding of the formation of English language logics into clause complexes enables deeper comprehension of the basic complexing process that language can be made use of its creative potential in developing meaning from any grammatical unit, the research of clause complex rather than the clause is more urgent and challenging. But before giving a thorough definition to the term, one point is to make clear the “unit simple” vs. “unit complex” that the former denotes “a single exponent of a unit” from a specific rank, but the latter includes “a coherent grouping or configuration of two or more units” (Morley, 2000, p. 27) at the same rank though sometimes also combines a unit of the rank next above. Hence, “clause complex” concerns two or more clauses while “clause simplex” is the single clause or a sentence organized by only one clause; phrases “complex clause” and “simple clause” are just the comparison of complexing degree within a clause of its grammatical constituency, will not be mentioned in this thesis.

In short, a clause complex is a grammatical and semantic unit that is developed serially through logico-semantic relations of projection and expansion and is organized clause nexus by clause nexus, i.e., one pair of clauses tactically combined in certain systematic and meaningful ways by coordination and/or subordination at a time, being interdependent on one another. In realizing a sequence of figures in the semantics of a text, a clause complex is usually ended with a full stop in either spoken or written language. (See Egins, 2004, p. 255; Halliday, 1985b, p. 66, 79; Halliday, 2014, p. 432; Halliday & Matthiessen, 1999/2006; Matthiessen, 1995, p. 121–185; Matthiessen, 2002; Matthiessen & Thompson, 1988; Thompson, 2014, p. 22-3)

Difference from a “Sentence”

When the clause complex is a grammatical and semantic unit that can be described for spoken and written language, the sentence is more idealized for written texts but not spoken sequences. The relation

of clause complex and sentence relating to other units or unit complexes in the system is demonstrated by Table 3 that the former is more stretching on the lexicogrammatical rank, while the latter is “the highest unit of punctuation on the graphological rank scale.”

Traditionally, sentences can be sorted into simple and complex sentences (Radford, 1999), simple, complex, and compound sentences (Quirk & Baum, 1987), or simple and non-simple sentences (van Valin, 2001) that the non-simple sentences can be further divided into either semantic relations (coordinate or subordinate) or syntactic relations (complex or compound sentences including conjunctive/correlative structures/adjacency). Normally, a simple sentence can be a single clause or an independent clause with at least an embedded one; a complex sentence can be the independent clause with at least one subordinate clause being its element(s); and a compound clause includes at least two coordinate independent clauses with equal status. Hence, the sentence can absolutely be replaced by a clause simplex denoting the single clause and a clause complex. Though a graphological sentence indeed is regarded as a clause complex in written language that the single clause sentence is a specific case, the separation of these two concepts, on the one hand, helps us to distinguish ungrammatical clause complexes from sentences split by punctuation (see Thompson, 2014, p. 186), and on the other hand enables a better construal of logico-semantic meanings in spoken language when the adoption of the long-standing traditional notion “sentence” may thwart a further interpretation of speech in practice while the influence does not exist by choosing “clause complex” and regarding it as a unit at the same rank as a clause but not from the above rank.

Semantics	Lexicogrammar	Graphology	Phonology
text	—		
(rhetorical) paragraph	—		
sequence	clause <i>complex</i>	sentence	tone groups [in tone sequence or concord]
message/proposition (proposal)/figure	clause	sub-sentence	tone group
element	group/phrase	—	—
	word	orthographic word	—
	morpheme	—	—

Table 3 The location of clause complex and sentence in the stratification of ranking units (From Halliday & Matthiessen, 2014, p. 436)

Organization

The relationship within clauses in a clause complex is defined from two systems, i.e., taxis and the logico-semantic relation, which is suggested in Figure 6:

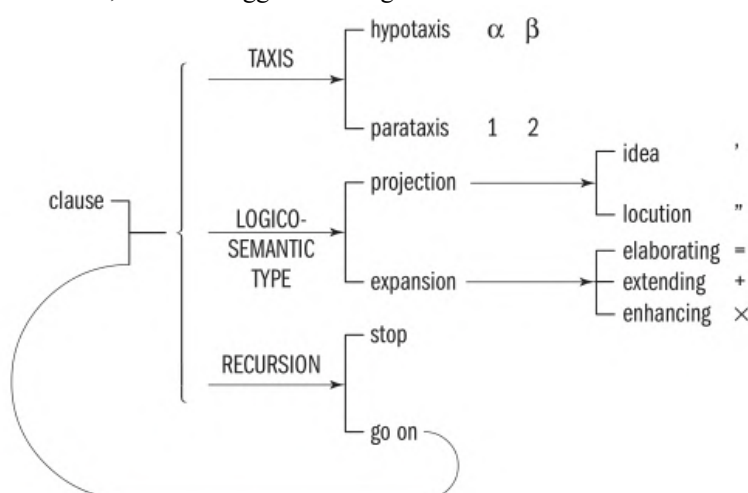


Figure 6 The general clause complex system from SFG (Halliday & Matthiessen, 2014, p. 438)

Before explaining the whole system with each category, one term deserves notice, i.e., “any one pair of clauses related by interdependency” (ibid, p. 441) is called a “clause nexus,” which is overwhelming in the system and important for representing a clause complex and investigating its recursion.

Taxis

According to the Oxford English and Spanish Dictionary, “taxis” is a Greek word denoting “arrangement.” In linguistics, it is narrowed into “the systematic arrangement of linguistic units in a linear sequence.” In the clause complex system, as all clauses connected by a logico-semantic relation are interdependent, i.e., one unit is interdependent on another in the relational structure “clause complex,” “taxis” is therefore used to mean the “degree of interdependency” (ibid, p. 440), which is then be separated into parataxis and hypotaxis¹⁴ according to their diverse degrees that the former indicates the relation between two like elements (one initiating and the other continuing) of equal clause status, and the latter suggests the relation between a dependent element and its dominant, i.e., the element being dependent.

Though parataxis and hypotaxis are identified as a continuum by Lakoff (1984, p. 481 - 491) where four kinds of relations (clauses especially) are mentioned: 1. Pure parataxis or juxtaposed clauses in which no explicit markers can be found to sign the relationship; 2. Mixed type or mixotaxis of clauses in which coordinate conjunction “and” is chosen for the relationship; 3. Near hypotaxis, i.e., one clause is subordinate to another with explicit clause relationship and semantic essence of the relationship such as conditional, temporal, causal, manner, concessive, etc., and 4. Pure hypotaxis that the subordinate clause is absolutely of no identification at all, the last one is normally regarded as an embedded clause relation, which is out of the discussion scope of this research, the third one had not taken hypotactic projections into account, the second one is also incomplete when the adversative or alternative relation is considered, leaving the first one being the “implicit clause complex,” which will be explained elaborately in Section 7.3 in this thesis. But the idea of locating parataxis and hypotaxis on a cline provides us with a complementary or unified rather than contrary perspective for studying clause relations.

For the ordering of a clause nexus, as no dependency is observed within two clauses in parataxis, it is just presented by the sequence of two clauses as the primary (initiating) and the secondary (continuing). In hypotaxis, the ordering is highly determined by the dependency within clauses but not the sequence that the dominant clause can either proceed or be followed by a dependent clause and can either encompass or be encompassed in a dependent clause, the latter of which is defined as an embedded relation.

In presenting the clause relations in a clause nexus, numerical notation such as 1 (initiating), 2 (continuing), 3 (continuing)... would be used to mark paratactic structures, and Greek letter notation such as α (dominant), β (dependent), γ (dependent)... would be applied in marking hypotactic structures. When a clause is logico-semantically related not by a single clause but by a “subcomplex,” i.e., another clause nexus in a change of taxis, the clause complex would be presented by a string of relations “with the internal nesting factored in for each clause” that the relation each clause is attached to is demonstrated clearly, e.g., $1 \wedge 2 \alpha \wedge 2 \beta$, or be presented by explicitly choosing brackets “with the internal nesting factored out” so that the hierarchic relation of the nested clause nexus is emphasized, e.g., $1 \wedge 2(\alpha \wedge \beta)$. (See Halliday, 2014, p. 450)

As clauses in parataxis are of equal status that the initiating and the continuing clauses are interpersonally free and each can be a functioning element alone, the connecting marker for clauses with such a “symmetrical and transitive” logical relation is usually linkers such as *and*, *but*, *or* for expansion clause complex; and clauses in hypotaxis are of unequal status that the dominant clause is free while the dependent is not or bound, the connecting marker for clauses with such a “non-symmetrical and non-transitive” logical relation are binders such as *that*, *whether*, *if* for projection, *when*, *because*, *since*, *if*, *although*, *as*, etc. for expansion, *after*, *before*, *because of*, *despite*, etc. for conjunctive prepositions, and some specific non-finite clauses. Hence, which marker is chosen for the logico-semantic relation, how many are chosen for each marker, and how many mixed markers are applied for each clause complex not only indicate the specific genre of a text but also present the preference of writers for the text. (See ibid, p. 452/454)

The delicate description of characteristics of parataxis and hypotaxis intersected with the clause sequence of logico-semantic relations (expansion and projection) can be referred to in Table 4 that the primary clause is called an expanded one while the secondary is the expanding one in expansion relation regardless of taxis, which means that the dominant clause or the clause with a projecting verb is expanded and the dependent clause or the clause being projected is expanding.

¹⁴ They are relationships that denote element complexes at any rank, i.e., clause complex, group/phrase complex, word complex, etc.

		Parataxis	Hypotaxis
'projection' & 'expansion'	status	equal status: initiating + continuing	unequal status: dominant + dependent
	(i) symmetry	symmetrical [except for projection]	non-symmetrical
	(ii) transitivity	transitive	non-transitive
	sequence	1 $\wedge$ 2	$\alpha \wedge \beta$, $\beta \wedge \alpha$, $\alpha <<\beta>>$, $\beta <<\alpha>>$
'projection'		quoting: two free-standing clauses	reporting: report dependent on verbal/ mental clause
	primary	projecting [if 1 $\wedge$ "2] / projected [if "1 $\wedge$ 2]	projecting dominant
	secondary	projected [if 1 $\wedge$ "2] / projecting [if "1 $\wedge$ 2]	projected dependent – determined by dominant
	conjunction	—	secondary: binder [<i>that; whether, if</i>]
	characteristic tone	[see Table 7-30]	[see Table 7-30]
'expansion'	primary	expanded	expanded
	secondary	expanding	expanding
	conjunction	secondary: linker [but some extending correlatives: <i>either ... or, neither ... nor, both ... and</i>] ⁶	secondary: binder, conjunctive preposition (with some non-finite clauses) [but some enhancing correlatives – binder + (cohesive) conjunction: <i>if ... then, although ... yet, because ... therefore;</i> binder + binder: <i>as ... so; the the + the</i> construction]
	characteristic tone	tone 3: 1 [tone 3] $\wedge$ 2	tone 4: β [tone 4] $\wedge \alpha$

Table 4 The properties of parataxis and hypotaxis intersected with logico-semantic relations (ibid, p. 453)

Logico-Semantic Relation

Being different from the configuration of diverse functional relationships within elements that can be described “as segmental from an experiential” angle, “as prosodic from an interpersonal” perspective, or “as culminative from a textual one” in the multivariate structure, the “iteration of the same functional relationship” that is specific to “the logical mode of meaning,” especially in the clause complex is depicted as being structured by logico-semantic relations in a univariate structure. (cf. ibid, p. 451) Though being seen as “semantic domains, i.e., semantic systems that are realized by grammatical systems” functioning in the whole grammar (ibid, p. 666), expansion and projection are characters that belong not only to the clause complex, other ranks, however, are not the focus of this research.

Expansion, which includes the traditional categories of apposition, coordination, and “adverbial clauses,” in clause complex means the relation that “the secondary clause expands the primary one” by either elaborating, extending, or enhancing it. Though being an ideational concept, expansion connects phenomena from the same experiential order and horizontally functions in the text by connecting “the frames that make up the strip” as well as events in the frames. Elaborating means “being equal to” (annotated as the equal sign =) that the secondary clause restates the contents of the primary one in other words, specifies more details, comments on the contents or lists an exemplification of the former, can thus be explicitly presented by markers such as *i.e., e.g., viz., etc.* to show the elaboration “on” a clause; extending means “being added to” (annotated as the plus sign +) that the secondary clause adds new information to the primary one, provides an exception or an alternation for it, can thus be presented by common markers such as *and, or, but, etc.* to reveal the extension “beyond” a clause; and enhancing means “being multiplied by” (annotated as the multiple sign x) that the secondary clause usually qualifies the primary one with circumstantial characters from time, space, condition, cause, concession, etc., can be presented by markers like *because, so, as, if, though, etc.* to demonstrate the embellishment “around” a clause. (cf. ibid, p. 443-4)

Except for applying languages in describing the non-linguistic phenomena, the report or quote of the

linguistic phenomena is also what language functions, thus allowing for clauses being reported or quoted as a projection in the combination of clauses, which together form up a clause complex. Being different from expansion, which focuses on relational clauses, projection is more concerned with verbal and mental clauses, as the secondary clause is usually projected by the primary one in either a locution or an idea in a clause complex. As projection connects phenomena from one experiential order in the methods or events of saying or thinking to the semiotic phenomena from a higher order in the actual action or contents of saying or thinking, it surpasses the sequence of the events connected by the expansion. Locution is saying, which constructs wording and is annotated as double quotes (“”), while the idea is thinking, which constructs meaning and is annotated as single quotes (‘). (ibid)

Normally, clause complexes are organized by clause nexuses with diverse taxis, various kinds of logico-semantic relations or the intersection of both, which is clearly demonstrated from the example below (E.g., 3.1.1) that linearly, the whole clause complex is $1 \wedge x_{211} \wedge x_{21+2} \wedge x_{2x2\alpha} \wedge x_{2x2}\beta\alpha \wedge x_{2x2}\beta=\beta\alpha \wedge x_{2x2}\beta=\beta\beta_1 \wedge x_{2x2}\beta=\beta\beta+2 \wedge x_{2x2}\beta=\beta\beta+3$, in which the relation within the 2nd and the 3rd is paratactic expansion (extension: addition); within the 7th, 8th, and the 9th is two paratactic expansions (extension: addition); within the 1st and the 2nd, and the 2nd and the 4th is paratactic enhancement (cause: reason and cause: result respectively); within the 5th and 6th is a hypotactic expansion (elaboration: exposition); within the 4th and 5th is a hypotactic projection (locution); and within the 6th and 7th is a hypotactic projection (idea), and hierarchically, the logico-semantic relation can be demonstrated from Table 5 that the relation within the 1st and 2nd clause is on Level 1; the relation within the 2nd and 3rd, and within 4th and 5th are both on Level 3; and that within the 7th, 8th, and the 9th are at the highest Level, i.e. Level 6 when the clause complex is analyzed further through recursion¹⁵.

E.g., 3.1.1 ||| **1** Amir blames himself badly, || **x₂₁₁** for he betrayed Hassan || **x₂₁₊₂** and planned to frame Hassan, || **x_{2x2 α}** so in Clause 7 he said || **x_{2x2}**“ **$\beta\alpha$** ” he “was a liar, a cheat, and a thief,” || **x_{2x2}**“ **$\beta=\beta\alpha$** ” which intends || **x_{2x2}**“ **$\beta=\beta\beta_1$** ” that he keeps the dirty secret, || **x_{2x2}**“ **$\beta=\beta\beta+2$** ” deceives his father || **x_{2x2}**“ **$\beta=\beta\beta+3$** ” and conceals the truth. ||| ---- CMAC15_05_590

	7	6	5	4	3	2	1	Clause
1								1 Amir blames himself badly,
2						1	1 x	2 for he betrayed Hassan
3					+	2		and planned to frame Hassan,
4						α	x 2	so in Clause 7 he said
5					α	“	β	he “was a liar, a cheat, and a thief”,
6				α	=	β		which intends
7			1	‘	β			that he keeps the dirty secret,
8		+	2					deceives his father
9		+	3					and conceals the truth.

Table 5 The hierarchical presentation of a clause complex (with the clause ID CMAC15_05_590)

More delicate grammatical categories of the expansion and projection systems will be exposed in the following sections.

3.1.3 Expansion

In the former section, it has been mentioned that the expansion relation can either be elaboration, extension, or enhancement, that the secondary clause can elaborate the existing first clause, extend it by addition or replacement, and enhance the environments of the former. (Halliday, 2014, p. 460) For example:

E.g., 3.1.2 ||| **x β 1** Since pragmatic meaning is complicated || **x β +2** and is realized at different levels (Culpeper and Haugh, 2014; Huang and Yu, 2016:138), || **α 1** some analysis will combine the investigations of different categories to yield interpretation, || **$\alpha=2$** for example, 4.2.2. Chapter Four Relations in Family This chapter examines the quarrels between parents and children. ||| ---- CPhD17_0002_1262

In the first and the third clauses, an enhancement (causal exactly) clause relation can be marked as the former clause explains the content of the latter by using the binder *since*, and the former cannot stand without the existence of the latter, their relation is dependent, i.e., a hypotaxis, which is annotated as $x\beta^{\wedge}\alpha$; in the third and the fourth clause, a paratactic elaboration relation is found when two clauses are of equal status, and the latter clause gives an example to the former by choosing *for example*, can thus be annotated as $1=2$; and in the first and the second clause, the latter only adds more information to the

¹⁵ See Section 7.4 for the explanation of the recursion of clauses on each Level.

former through the choice of *and*, is annotated as 1+2. As the intersection of taxis and logico-semantic relations can be various not only because of finiteness but also due to subtypes, the contents of this section will be centered around these three types, on the paratactic and hypotactic, and on finite and non-finite structures in each type with examples and clarification, and on the classification of each type, i.e., elaboration including exposition, exemplification, and clarification; extension encompassing addition, variation and alternation; and enhancement covering temporal, special, manner, causal and conditional relations in clause complexes.

3.1.3.1 Elaboration

Except for the paratactic elaboration in E.g., 3.1.2 where the fourth clause provides an example for the contents of the third clause, elaboration can also be hypotactic as suggested by E.g., 3.1.3:

E.g., 3.1.3 ||| **α1** The original meaning has a sense of kindness, || **α+2** but it is satirized as bad luck by Xiaoxian, || **=β which*** is where humor arises spontaneously. ||| ---- CMAC13CU_27_341

The choice of *which*, in fact, refers to the whole former clauses in further specifying the origin of humor described in them semantically instead of referring to specific former words, and it limits the domain of the following clause that the third one has to be dependent on former two clauses. So, elaboration indicates specification or description by the secondary clause of the initiating one. One way to distinguish elaboration from other expansion relations is to identify the tone concord in two clauses in a clause nexus that each tactic form of elaboration *is realized by a tone group*, which has the same tone. (ibid, p. 461) Though meaningfully, paratactic and hypotactic elaboration are “complementary” in that the former includes exposition, exemplification, and clarification, and the latter covers a descriptive aspect of elaboration (ibid, p. 462), hypotactic elaboration differs from paratactic elaborating clauses in two folds: one is for “the special grammar of relative clauses” when a relative element referring “back to the domain of elaboration” always existing in the former while no marker can be found in the latter (though the paratactic variant can “correspond to a non-relative anaphoric reference,” it does not have to include a “reference to the domain of elaboration” but takes a lexical expression way for hypotactic agnation); another is for the non-defining relative clause when “an anaphoric reference item, i.e., the relative” is obligatory in finite clauses in hypotactic elaboration while no special grammatical determiners are involved in the clause of the latter. (ibid, p. 461-2)

3.1.3.1.1 Paratactic Elaboration

Paratactic elaboration (annotated as 1=2) can be of three kinds: exposition, which indicates that the secondary clause repeats the contents of the primary one in other words, demonstrates the gist from another perspective or emphasizes the information being expressed; exemplification, which suggests that the secondary clause provides more specific information for the content of the primary one by frequently referring to examples; and clarification, which means that the secondary clause supports the content of the primary one with illustration, explanation or explanatory comments.

E.g., 3.1.4 ||| German functionalist scholars, <<attempting to break the shackles of source language from an absolutely new perspective of target language, [[during which “Skopos theorie” comes into being]], >> argue || that the translation purpose justifies the translation proves || and the end justifies the means (Nord 2001, p. 124), || **that is to say***, what goal we want to achieve determines what kind of strategy we adopt || when translating a text. ||| ----CMAC07CU_35_127

In E.g., 3.1.4, the description of the clause following the phrase *that is to say* clearly presents the exposition of the content of the former clause as one is the citing words from the scholar Nord, and the other is the repetition of the gist from the writer. Besides the mentioned phrase, some conjunctive expressions such as *or*, *rather*, *in other words*, or the acronym *i.e.*, can explicitly signify the exposition clause relation, or some words in the secondary clause being regarded as lexico-semantic links for repetition or synonymy in emphasizing the information expressed in the former. (ibid, p. 463)

E.g., 3.1.5 ||| But if taken into the contextual factor, || tag questions can fulfill various purposes, || **for instance***, requesting confirmation or disconfirmation of the statement from the addressee. ||| ---- CPhD17_0002_1060

The phrase *for instance* in E.g., 3.1.5 explicitly demonstrates the exemplification relation within two clauses that the latter clause, which states *requesting confirmation or ...* is an example provided for *tag questions* which can fulfill some purposes described in the former clause. Except for conjunctive such as *for example*, *in particular*, or the acronym *e.g.*, some implicit cohesive lexis indicating hyponymy or meronymy in clauses can also signify the exemplification relation. (ibid)

E.g., 3.1.6 ||| At the same time, Nord emphasizes || that “Loyalty Principle” is quite different from “fidelity” in equivalence: || the latter shows a relationship of equivalence held between the source language and the target language || while the former is part of the interpersonal relationship between the

translator, the source-text sender, the target-text addressees and the initiator. ||| ---- CMAC07CU_35_121
 Though in E.g., 3.1.6, there are no explicit lexis or phrases for denoting the relationship within the second clause and the third clause, the content after the colon is, in fact, explaining how different “Loyalty Principle” is from “fidelity” and what exactly the “Loyalty Principle” is through comparison, the clause relation is, therefore, clarification instead of exposition though these two are easily mixed in the annotation. Normally, words or phrases like *in fact*, *actually*, *indeed*, *at least*, *to be precise*, or the abbreviation *i.e.*, and *viz.* can be used to convey clarification relation and the transfer of polarity from negative to positive or the opposite direction within clauses is its deviation from exposition semantically. (ibid, p. 464)

3.1.3.1.2 Hypotactic Elaboration

Hypotactic elaboration (annotated as $\alpha=\beta$) is the structure traditionally called “non-defining relative clauses,” which is clearly demonstrated in E.g., 3.1.3 that usually “background information,” “characterization” or some specific points of the subject being depicted in the primary clause with corresponding “evaluation” would be provided or stated in the dependent clauses, (ibid) though sometimes this kind of elaboration for the primary clause would appear for more than one time. (ibid, p. 465) Usually, hypotactic elaboration can be finite or non-finite clauses according to the choice of Subject of the dependent one, in which the former is realized in three types:

1. when the dependent clause begins with *which* to signify all the contents of the independent clause or some of it (not only a nominal group);
2. when the dependent clause begins with *which*, *that*, *who*, or *whose* to denote the nominal group in the independent clause; and
3. when the dependent clause begins with *when* or *where* to indicate the temporal or abstract spatial expressions in the independent clause, which is agnate with the saying *which is when...* or *which is where...* (ibid, p. 465-466)

E.g., 3.1.7 ||| Their translation is really successful || as it not only retains the fidelity to the original || but also achieves functional equivalence, || **which*** can be almost fully reflected by the English version of those cultural terms in *The Scholars*. ||| ---- CMAC07CU_35_4

E.g., 3.1.8 ||| Apparently, all the 100 LRs include reporting verbs, || **which*** allows for interpersonal intervention in these texts. ||| ---- CPhD06_0004_1287

The beginning of the dependent clause after the comma in E.g., 3.1.7 *which* refers to the subject of the former clause complex *their translation*, and in commenting on how the translation is *successful*, the adverb *fully* is adopted to describe *The Scholars* corresponding to the *translation*. In E.g., 3.1.8, the word *which* in fact means the *inclusion of reporting verbs* rather than the nominal group *reporting verbs* itself, which then is elaborated by the present singular verb *allows* instead of only *allow* in the dependent clause, and is agnate with the non-finite clause *allowing for ...* though the latter is obscure for its subject without judging from the lexical-semantic expressions. Except for the adverb or the verb above after *which*, usually adjectives can be found in the “attributive relational” elaborating dependent clause that the evaluation of the contents being mentioned in the independent clause or the attitude toward it from the writer is demonstrated. Hence, this kind of structure is of significance in discourse analysis, especially in sociolinguistics.

E.g., 3.1.9 ||| This phenomenon was, in fact, proved by Purves (1986), || **who*** claims || that “the differences among national communities will be sharper as they (students) progress further through the educational system.” ||| ---- CPhD05_0003_1636

E.g., 3.1.10 ||| Miller et al. (1997), for instance, found || that speakers of two French dialects utilize dissimilarly of durational cues || when perceiving vowels, || **whose*** results lead us to believe || that speakers of different dialects also differ in their perceptual systems in the essence. ||| ---- CMAC15_04_54

The pronoun *who* in E.g., 3.1.9 and *whose* in E.g., 3.1.10 in dependent clauses represent *Purves* and *Miller* in the dominant clauses, respectively, without which elaborating information for them would not be provided and how *the phenomenon was proved by Purves* and how important Miller’s founding for the writer would then be unknown and unembodied.

E.g., 3.1.11 ||| The Scholars dwells on the feudal intellectuals’ life in the Ming Dynasty, || **when*** officials were selected, appointed and promoted only through the testing of their paku essays in the prefectural examination, || (having) nothing to do with their comprehensive abilities such as common sense, knowledge, talents, characters and so on. ||| ---- CMAC07CU_35_313

E.g., 3.1.12 ||| This long-nurtured question seems to have found the answer in Li Ao with his program *Li Ao Has a Word on Phoenix TV* on May 15th, 2005, || **when*** the paper was being written. ||| ---- CPhD05_0003_1295

The interrogative pronoun *when* in the expansion: hypotactic elaboration clause relation can be easily

mixed with that in the enhancement: hypotactic temporal relation in annotation formally, but the former is agnate with *which is when* in pointing to the exact time that things happen being described in the dominant clause and is thus being emphasized by the writer as it is important for further description; and the latter is more focusing on what happens at a specific time, i.e., contents being stated in the independent clause. Hence, in E.g., 3.1.11, *when* refers to *the Ming Dynasty*, and this time is reinforced for *officials being selected, appointed and promoted through the testing of paku essays*; and in E.g., 3.1.12, *when* signifies *May 15th, 2005* exactly for the time that Li Ao's program was written. In short, it is the time for events rather than the events themselves that the writer intends to highlight by choosing hypotactic elaboration relation within clauses.

E.g., 3.1.13 ||| The other is, in Chapter Four, the analysis on presumably the most basic minimum discourse pattern (i.e., problem-solution) in both English and Chinese texts, || **whereby*** convergent features are explored. ||| ---- CPhD05_0003_18

E.g., 3.1.14 ||| The emphasis tends to shift from teachers to students in a constructivist classroom, || **where*** the teacher does not pour any more knowledge into passive students. ||| ---- CMAC14RE_15_271

Similarly, the distinguishing of the hypotactic elaboration relation from the enhancement: hypotactic spatial relation marked by *where* is to check whether the writer is paying attention to the space that events happen in or the events themselves. In E.g., 3.1.13, *whereby* means *by which place* and denotes *Chapter Four* specifically as it is in this chapter that *the convergent features are explored*; and in E.g., 3.1.14, *where* is *which is where* in referring to *the constructivist classroom* that *teacher pour less knowledge into passive students*.

One differentiation between the non-defining relative clause and the defining one is in punctuation in writing, as the former can often be marked by commas or led by a dash, while the latter is absent of any punctuation mark within clauses. (ibid, p. 467) Another distinction is between the elaboration and the extension relation on possessives from *whose*, *whom*, or *which*, etc., that the former only describes the subject of the independent clause, but the latter discusses another related subject possessed by that of the primary clause. (ibid, p. 468)

For the non-finite clauses of elaboration relation, as the Subject of the dependent clause is omitted, the identification for the subject, though sometimes can be regarded as the same as that of the independent clause, can cause obscurity that both the main content of the dependent clause and the semantic relationship within it and the independent one are implicit, and the meaning being expressed hence becomes more abstract. (ibid, p. 469)

E.g., 3.1.15 ||| First, it puts teachers in their places, || **emphasizing*** the individuality of students || and reducing the compulsion [[we sometimes feel to control classroom activities]]. ||| ---- CMAC07LI_20_552

In E.g., 3.1.15, it seems that the subject of the dependent clause is *it* from the dominant one, and then the relation turns to be an extension: paratactic addition. But the non-finite clause starting with the verb *emphasizing* indicates that the action of *it* putting *teachers in their places* reinforces the importance of *individuality of students*, the real subject is clearly the whole primary clause, and the following clause is agnate with *which emphasizes ... and reduces...* for denoting the hypotactic elaboration relation.

E.g., 3.1.16 ||| There appears to be some very good reasons for this, || **having*** to do with the use of point of view to help create suspense. ||| ---- CMAC06LIT_31_394

The Theme of the clause in E.g., 3.1.16 is *there*, but the subject of it is actually *some good reasons*, leading to its description in the dependent clause and the agnate clause of it can be *which has something to do with* The clause relation of the clause nexus is, therefore, a hypotactic expansion: elaboration.

E.g., 3.1.17 ||| In conclusion, every theory is considered to be one array of light, || **casting*** light on each corner of the "dark house" of personality. ||| ---- CPhD10_0001_180

The clause complex in E.g., 3.1.17 can be rewritten as *... every theory is ... that it casts light on each ...* that the subject of the dependent clause is *every theory* and the function of the dependent clause is to provide more characterization for the subject. The choice of the non-finite relative clause for a hypotactic elaboration relation seems to shorten the distance between two clauses by removing the common domain.

E.g., 3.1.18 ||| First, 62.5% of group leaders were always considered comfortable and confident (shown in Item 4), || **attempting*** to get members to listen and respond to each other (shown in Item 6). ||| ---- CPhD11_0007_1434

The clause complex in E.g., 3.1.18 is agnate with *62.5% of group leaders..., who attempt to get ...*, which means that the subject of the dependent clause is indeed 62.5% of group leaders, and the relation within two clauses is expansion: hypotactic elaboration instead of others as the secondary clause states the intention of the subject from another perspective.

In a word, elaboration clause relation can be paratactic, which includes exposition, exemplification,

and clarification; or be hypotactic, which can be finite non-defining relative clauses (dependent clauses led by *which/ that, who/ whose/ whom*, and *when* or *where*) or non-finite ones with the dominant being deleted in dependent clauses. No matter which kind is chosen by the writer, one point is for sure, as tone concord is the only way for the identification of elaboration relation. (ibid, p. 470)

3.1.3.2 Extension

The extension has been explained as the addition of information or description or characterization in the secondary clause or the dependent one to the primary clause or the independent one, i.e., clauses are either paratactically or hypotactically connected by meaning addition. As the extension can be divided into addition, variation, or alternation according to clause function, the intersection of the taxis and function is thus various and will be discussed in detail in the following part.

3.1.3.2.1 Paratactic Extension

The paratactic extension (annotated as $I+2$) has been mentioned in E.g., 3.1.2, though the two clauses are not logically connected. Traditionally, paratactic extension denotes the coordination within clauses, which is usually presented by *and, or, but*, etc. For the additional relation, i.e., one clause is adjoined to another instead of other relations, it can be a positive one marked by the word such as *and*, a negative one by the word such as *nor*, or an adversative one marked by *but*.

E.g., 3.1.19 ||| An examination of reporting verbs in LR shows || that using an appropriate reporting verb in a particular situation is a delicate choice || as it is a crucial means [[of **both** situating one's work appropriately || **and*** communicating with one's peers effectively]], a way [[of engaging with colleagues || **and*** have interactions with them in the academic community]]. || ---- CPhD06_0004_1548

Though two paratactic extension: positive additional clauses marked by *both...and...* and the word *and* in E.g., 3.1.19 are embedded in the noun *means* and *way* respectively, both indicate the positive additional clause relation of the same status. The same relation is also demonstrated in E.g., 3.1.7 when the phrase *not only... but also...* is chosen by the writer to explain why the translation is successful in two perspectives. The common ground for these two examples is the consistency of tense and voice of verbs after the linkage.

E.g., 3.1.20 ||| Without such links academics could **neither** justify their arguments by connecting their research activities to significant work in the field, || **nor*** use this disciplinary knowledge to establish the novelty of their position (Gilbert 1976; Berkenkotter and Huckin 1995). || ---- CPhD06_0004_1137

The paratactic extension: negative additional relation comes from the choice of negative linkage in connecting two clauses, as indicated by E.g., 3.1.20 when the conjunction *neither... nor...* is adopted to mean that academics could not *justify...* and not *use ... without such links*. Two clauses in the nexus are of the same status and do not have any other relation except for an additive one.

E.g., 3.1.21 ||| In the mid-term examination, the passing rate of the experimental class is only 67.86%, || which is lower than that of the controlled class (71.43%), || **and yet***, in the final examination, the passing rate reaches 85.71% || which is much higher than that of the controlled class (78.57%). || ---- CMAC07LI_20_487

E.g., 3.1.22 ||| Therefore, teachers should make it clear to the students || that CSs contribute to L2 communication and learning || **but*** they should make selective use of CSs for learning. || ---- CPhD10_0018_1283

The clause relation in E.g., 3.1.21 seems to be a normal additive one when only the word *and* is considered, in fact, is an adversative one that the third and the fourth clauses described a controversial condition compared to that in the first two clauses. Semantically, the primary clause nexus is on the passing rate in the mid-term examination, while the secondary clause nexus is on that in the final examination. The choice of *and yet* additively linked two clauses together but with converse contents. The most frequent word for indicating the adversative relation is *but*, which is presented in E.g., 3.1.22 when the contribution of CSs is confirmed by the writer while its use should be cautious, i.e., of selection in the writer's opinion.

For the variational relation, i.e., when one clause totally or partially replaces the original one semantically, it can be replacive for the contrary meaning or be subtractive presented by words such as *only*.

E.g., 3.1.23 ||| In foreign countries, the researches of PPA are **not** discussed separately, || **but*** are commonly conducted within the overall study of the whole tense-aspect system, such as the study of situation aspect hypothesis (Shirai & Kuroono, 1988). || ---- CMAC10SY_9_48

By choosing the phrase *not...but...* in E.g., 3.1.23, the expansion: paratactic extension: variational: replacive relation within clauses is showed that instead of *separately discussing*, the *research of PPA* is, in fact, conducted in a whole system. Hence, the meaning or the way that the subject is dealt with of the

primary clause is varied or replaced by that of the secondary one.

E.g., 3.1.24 ||| Leech' Politeness Principle is similar to Gu's principles in essence; || they **only*** differ in the wording of the principles. ||| ---- CMAC05PR_24_354

E.g., 3.1.25 ||| As for the distribution of its tense, **only*** 22 cases are in the simple past tense, || the other 87 are in the simple present tense, || as is shown in Table 5.7. ||| ---- CPhD06_0004_1793

If replacive relation is the total variation of the meaning or the content, then subtractive relation is only the partial variation that some part of the subject of the primary or dominant clause is changed by the description of the secondary one, such as *only* in E.g., 3.1.24 and 3.1.25 that both can be agnate with the phrase *except that...* The former depicts the similarity within Leech and Gu's theory but subtracts the wording part; the latter describes the situation that most cases would adopt for the tense choosing and again excludes 22 cases from the whole corpus. Instead of denying the whole of the subject, a specific scope is disapproved of by the usage of words such as *only*, *but*, *except*, etc.

The paratactic extension: alternative relation, which is often presented by *either... or...* within two clauses, means the choice of the contents from them, and this choice must be performed.

E.g., 3.1.26 ||| In the author's standpoint, the previous studies are **either** based on empirical practice with regard to the respective advantages and disadvantages in translated works, || **or*** remain superficial exploration of specific translation techniques, || yet objective translation criteria are still rather inadequate for the comparative study on translations of Chang Hen Ge. ||| ---- CMAC11PR_10_129

E.g., 3.1.27 ||| **Either** IL system is penetrated by rules or forms of the NL [[not usually evidenced in its speech forms]], || **or*** an internalized TL rule or form is improperly generalized or distorted in some way. ||| ---- CPhD08_0005_707

As part of the *previous studies on translations of Chang Hen Ge* are based on *empirical practice...* and another part of them is *superficial exploration of ...* in E.g., 3.1.26, the writer's research belongs to neither part of them, the rationality and the necessity of the writer's study are explained. The clause relation within the primary and the secondary is alternative that *the previous studies* in the writer's eyes must belong to either kind of them as described in two clauses. Being different from the example above, the subject of the primary clause in E.g., 3.1.27 is transformed from *IL system* to *an internalized TL rule or form* and the choice of *either...or...* for the connection of two clauses provides two kinds of situations to readers.

3.1.3.2.2 Hypotactic Extension

The hypotactic extension (annotated as $\alpha+\beta$) resembles paratactic extension functionally on the one hand as it is divided into three kinds: addition, variation, and alternation, and has the same condition as hypotactic elaboration structurally that the dependent extending clause can either be finite or non-finite. The discussion, hence, will also be the intersection of both functional and formal variants, though this type may rarely appear in common corpora.

Finite clauses for hypotactic extension relation can be additional, presented by conjunctions such as *whereas* and *while*. Though the additive: positive relation and the adversative relation are common for finite clauses, the distinction of them in specific clause complex always brings troubles to researchers in the annotation. The additive: negative relation does not exist.

E.g., 3.1.28 ||| A bare reflexive, such as *ziji*, has neither inherent ϕ -features nor inherent reference, || **whereas*** an inflected reflexive like *taziji* has inherent ϕ -features || but lacks inherent R-features. ||| ---- CPhD15_0001_102

E.g., 3.1.29 ||| Many efforts have also been taken to analyze some Chinese traditional novels—sometimes Chinese version of foreign novels as well—and especially the four Chinese classics, || **while*** few efforts have been taken on the English version of *The Scholars*. ||| ---- CMAC07CU_35_43

The clause led by the word *whereas* in E.g., 3.1.28 cannot exist alone and is therefore dependent on the whole clause complex, and the relation within the first and the second is a hypotactic one. As *an inflected reflexive has inherent ϕ -features*, which is the opposite from that of the *bare reflexive* structure according to the writer, the clause relation turns to be a hypotactic extension: addition: adversative one and *whereas* equals to *but* paratactically. The clause relation in E.g., 3.1.29 denoted by the word *while* seems to be a positive additive one in linking two clauses together, is indeed an adversative for the writer compares the studying condition of *Chinese traditional novels*, *Chinese version of foreign novels* with that of the *English version of The Scholars*, laying the foundation for his/her further research. In this circumstance, *while* is agnate with *yet* paratactically.

Finite clauses for hypotactic extension relation can also be variational, indicated by phrases such as *except that...* or *but that...* for subtractive relation, leaving out replacive ones.

E.g., 3.1.30 ||| For the participants in the comparison group, the procedures were the same || **except that*** they did not receive any prompts about language forms || when they retold the story of the parts and the episode. ||| ---- CPhD08_0012_971

The principle for making comparison within two groups in E.g., 3.1.30 is obviously the receiving of *prompts about language forms... for participants* according to the writer by the choice of the phrase *except that* in the clause complex that one group receives while the comparative or the subtractive group does not. Sometimes, the taxis within finite clauses connected by *whereas*, *while*, *except that*, etc., obscure. But as the extending clause in a hypotaxis can be moveable, i.e., the dependent clause can either follow or precede an independent one, the clause complex would be annotated as $+\beta^{\wedge}\alpha$ or $\alpha+\beta$. Instead, $+2^{\wedge}1$ is strange since two paratactic clauses are of the same status, their annotation is always $1+2$ no matter which clause is in the primary place.

For the paratactic extension: alternative finite clauses, the most frequent connection is the phrase *if...not, then...*, leading the dependent clause in front of the independent one.

E.g., 3.1.31 || It is easy to see || that SH informants do perceive much better than XH informants do, || and Flege has pointed out || that **if** L2 feature is **not** signaling a phonological contrast in L1, || **then*** this feature is not very likely to discern in L2. || ---- CMAC15_04_478

E.g., 3.1.32 || **If** the sender is **not** familiar with the current language, its codes and its culture || **then*** he or she will not be able to express anything at all. || ---- CPhD10_0015_1369

The choice of *if...not, then...* in E.g., 3.1.31 describes two kinds of situation, i.e., L2 feature would either *signal a phonological contrast in L1*, which leads to *discern in L2*, or not according to *Flege*. This situation for sure depends on some criteria that have not been mentioned by the writer. The same alternation is suggested by E.g., 3.1.32 when the familiarity *with the current language...* is regarded as the condition of being *able to express anything* that the dissatisfaction of the condition would bring out the opposite result.

Except for the negative additive relation for non-finite clauses, which resembles finite ones, the alternative relation is also absent for non-finite hypotactic extending clauses. Hence, only additional (positive additive and adversative) and variational (replative and subtractive) relations presented by conjunctive prepositions or prepositional groups such as *beside*, *apart from*, *without*, *instead of*, *rather than*, etc., can be discovered in most clause complex cases.

E.g., 3.1.33 || At the same time, **beside*** emphasizing the native culture, || English major students should also realize the importance of other cultures || so as to avoid the phenomenon of the ethnocentrism. || ---- CMAC14CU_8_393

E.g., 3.1.34 || **Apart from*** being the brand name, || “more” suggests || that the quality of More cigarette is more terrific than any other kind || ---- being longer, better and milder probably. || ---- CPhD07_0011_708

The hypotactic extension: positive additive relation for non-finite clauses is usually suggested by the advancement of annotation markers in leading to a dependent clause and then independent ones in emphasizing the additional contents being described in the latter, though the dependent clause seems to be more important. With the subject being omitted in the dependent clause, the reinforcement of the Rheme in the independent clause is self-evident, as showed in E.g., 3.1.33 when *the realization of the importance of other cultures* is regarded as a more important point than *emphasizing the native culture* by the writer, the former of which has probably been ignored. The same is presented in E.g., 3.1.34 when “more” is given further explanations in the independent clause in addition to *being the brand name* mentioned in the dependent one.

E.g., 3.1.35 || Some speech acts, such as invitations, refusals, and apologies often take place between individuals, || so learners might not have the opportunity to observe such language || **without*** being directly involved in the conversation. || ---- CMAC05PR_24_519

The hypotactic extension: adversative relation marked by the word *without* can be easily mixed with the hypotactic enhancement: negative conditional relation as suggested by E.g., 3.1.35 that the clause complex can be interpreted as *learners might not ... if they do not directly involved in the conversation*, meaning that the action by learners would not fulfill without the condition of *being directly...* If the clause relation is an adversative one, it needs to be written as: *learners have the opportunity ... without taking it (the opportunity) ...* that the object or the Rheme of the independent clause remains the same in the dependent one.

The hypotactic extension: variational relation for non-finite clauses is different from that for finite ones as the replative relation exists for former structures, and usually the dependent clause is agnate with the paratactic one introduced by the phrase *but not...*

E.g., 3.1.36 || It is conceivable || that, given the extent of specialized mutual knowledge, || the writer may exploit it thematically || **instead of*** relying on information [[provided in the preceding Rheme]]. || ---- CPhD09_0002_1425

E.g., 3.1.37 || Although Zhao’s study on Swain’s output theory in the writing class was conducted empirically, || it was primarily to test and verify || that output did work its functions in subjects’ producing, || **rather than*** applying this theory to the teaching practice. || ---- CMAC05WR_20_352

The choice of *instead of* in E.g., 3.1.36 means the preference for the contents being described in the independent clause by the writer, and those in the dependent one, though it has been mentioned, cannot replace those in the former. So, “replative” denotes to the dependent clause that it can be replaced by the independent one, and the whole clause nexus is agnate with the paratactic one as *the writer may exploit but not rely on information* Similarly, in E.g., 3.1.37, *applying the theory to the teaching practice* can be replaced by *testing and verifying Swain’s output theory* by choosing the phrase *rather than* for non-finite clause connection, and the clause nexus equals to *it was primarily to test and but not apply this theory to ...*

E.g., 3.1.38 ||| The Cooperative Principle and its attendant maxims by Grice help exchange information maximally effectively and efficiently, || better convey and understand what the speaker says, || **other than*** giving rise to misunderstanding. ||| ---- CMAC14LIT_25_213

E.g., 3.1.39 ||| **Other than*** segmenting experiences of the world into sequences and entities, || Halliday and Matthiessen (2004, p. 29) point out || that language is always enacting personal and social relationships with other participants. ||| ---- CPhD13_0010_419

Besides the phrase *except for*, the hypotactic extension: subtractive variational relation for non-finite clauses (dependent exactly) can also be introduced by the phrase *other than*, and the location for two clauses in a clause nexus is changeable, i.e., the dependent clause can either be in front of or follow the independent one as demonstrated in E.g., 3.1.38 (follow) and 3.1.39 (in front of). But no matter where the dependent clause is, it is the contents of the independent clause that the writer intends to reinforce, which means that *The Cooperative Principle...by Grice...* has a list of effects but not *give rise to misunderstanding* in the former example and what really points out by Halliday and Matthiessen in the latter example is the aim of choosing the reference and mentioning in thesis by the writer, but not *segmenting experiences of the world ...*

The summary of the extension relation can be demonstrated in the table below:

	Category	Meaning	Paratactic	Hypotactic	
				finite	non-finite
(i) addition	‘and’, additive: positive	X and Y	(both ...) and; not only ... but also	while, whereas	besides, apart from, as well as
	‘nor’, additive: negative	not X and not Y	(neither ...) nor	—	—
	‘but’, adversative	X and conversely Y	but, (and) yet	while, whereas	Without
(ii) variation	‘instead’, replative	not X but Y	but not; not ... but	—	instead of, rather than
	‘except’, subtractive	X but not all X	only, but, except	except that, but (for the fact) that	except for, other than
(iii) alternation	‘or’	X or Y	(either ...) or (else)	if ... not (... then)	—

Table 6 The categories of the extension relation in diverse clause connecting structures with relative markers (Halliday, 2014, p. 471)

No matter what the clause combining structure is, i.e., paratactic or hypotactic (including finite and non-finite clauses) clause nexus, the relation within two clauses for extensional relation can either be addition, variation, or alternation, in which additional relation is further divided into positive additive, negative additive and adversative one that negative additive relation is absent from hypotactic clause nexuses; variational relation includes replative and subtractive that the former cannot be found in finite clauses; alternative relation is absent from non-finite clauses. The comparison of choosing each marker for the extensional relation of clause nexus in our Master and doctoral corpus will be clearly explained in Section 5.5 in this thesis.

3.1.3.3 Enhancement

The enhancement relation means that one clause qualifies another one either paratactically or hypotactically by some means: temporal, spatial, manner, causal, conditional, or concessive. Hence, this section will start from paratactic enhancement classified by the mentioned categories, then go to hypotactic clauses with the same relations from finite or non-finite structures respectively in demonstrating how clauses are combined and what the logico-semantic relations within clauses really mean.

3.1.3.3.1 Paratactic Enhancement

Paratactic enhancing (annotated as *1x2*) relation in clauses is normally regarded as coordination with circumstantial characteristics that can be presented by conjunctions like *then*, *for*, *but*, *yet*, etc., conjunction groups such as *and then*, *and there*, *and thus*, *and therefore*, etc., and conjunctive like (*and*) *at that time*, (*and*) *meanwhile*, (*and*) *in that case*, etc. in order to present the temporal, spatial, manner, causal, and conditional qualification toward primary clauses normally.

The paratactic enhancement: temporal relation can indicate either the same time or different time (earlier or later) when conjunctions such as (*and*) *meanwhile*, *when* or conjunctive with *and* like *and before that*, *and after that*, *and afterward*, (*and*) *then*, etc. is chosen as the connection for clause linking. E.g., 3.1.40 ||| The child-wife is likely || to be aware of her husband's hidden dissatisfaction, || **and meanwhile*** she also feels disturbed by his constant criticism. ||| ---- CMAC10LIT_7_301 conjunction group

E.g., 3.1.41 ||| Firstly, they adapted the construct to L2 communicative situations, || **and meanwhile***, identified several additional influences on L2 communication. ||| ---- CPhD13_0003_370

The conjunction group *and meanwhile* in E.g., 3.1.40 and E.g., 3.1.41 denotes that the event described in the secondary clause happens at the same time as that in the primary one, i.e., the likeness of *the child-wife's awareness of her husband's ...* comes to the wife at the exact time when *she disturbed by the constant ...* in the former example (though two ideas are hardly to co-exist in reality); and the adaptation of *constructing L2 communicative situations* and the identification of *additional influence on ...* in the latter example are conducted by *them* at the same period. With the order of the clause being exchangeable in the clause nexus, and each clause can be independent, the relation of two clauses connected by *and meanwhile* is thus paratactic enhancement: temporal.

E.g., 3.1.42 ||| As has already been discussed in 4.1, || when Chinese speakers want to describe the existence and presentation of a new entity, || to reduce the abruptness of the first mention of the new entity, || they will first present to the hearer a scene, || **then*** introduce the entity into the scene. ||| ---- CPhD07_0008_931

E.g., 3.1.43 ||| Daisy first says || she does not love Tom || **and afterwards*** she admits || she loves Tom, || so her words are implicit and in disorder, || which leads to ambiguity. ||| ---- CMAC14LIT_25_575

The paratactic enhancement: temporal relation for a later time in a clause nexus is usually presented by time sequence that the event described in the secondary clause happens after that in the primary one with almost no exception. In E.g., 3.1.42, the choice of the conjunction *then* comes from *first* in the fourth clause to reveal how Chinese speakers *describe the existence and presentation of a new entity* sequentially; and in E.g., 3.1.43, the choice of the conjunctive *afterwards* plus an *and* has the same situation though *first* is redundant, demonstrating Daisy's transferring attitude toward Tom from her sayings through time, the comparison of which thus leads to the writer's conclusion that *her words are implicit and ...*

Though paratactic enhancement: spatial relation exists and can be introduced by the conjunction group *and there*, as no example is discovered in the annotated corpus in choosing this clause relation, this part will be skipped.

The paratactic enhancement: manner relation includes the means that things can be solved, which can be signified by a conjunctive with *and* such as *and in that way* and a conjunction group with *and* like *and thus*; and the comparison that events or things described in two clauses are compared through the word such as *similarly*, *thus* or the conjunction group *and so*.

E.g., 3.1.44 ||| By employing two rhetorical questions "How is he my master? Am I a servant?", || Jane violates CP || and implements the FTA (off-record politeness) to the servant || **and in that way***, she defends herself || and shows her anger. ||| ---- CMAC10PR_19_458

E.g., 3.1.45 ||| According to Ellis (1999), universal grammar lays a specific grammatical foundation for any human language, || **and thus*** makes an overall frame for language learning (Gass 1980). ||| ---- CPhD11_0006_267

In E.g., 3.1.44, through the *violation of CP and the implements of the FTA to the servant* after two questions, according to the writer's explanation, Jane can *defend herself and show her anger*, hence the conjunctive *and in that way* is adopted to link two clauses paratactically though the last clause more seems a conclusion or a result of taking the strategy. The conjunction group *and thus* in E.g., 3.1.45 is more likely to state the reason for the universal grammar's *making an overall frame for ...*, but it is the way of *laying a specific grammatical foundation for ...* that makes the result possible and feasible, so *thus* is referred to the manner that Ellis adopted. Logically, the event that happens in the secondary clause must be later than that in the primary clause, and the clause order in the nexus is irreversible.

E.g., 3.1.46 ||| The study of rhetoric keeps a range between linguistics and literature; || **similarly***, the pun, << though relying heavily on metaphorical or figurative meanings of words for its effect >>, is excluded by many books on either linguistics or literature. ||| ---- CPhD07_0011_968

When the following clause in the enhancement: manner: comparison relation is very long with a lot of information, the marker for hypotactically connecting two clauses with unequal status would not be so suitable compared with that for hypotactic ones as the former tightens clauses structurally. Thus, in E.g., 3.1.46, the word *similarly* is used for comparison that *the study of rhetoric* and *the pun* have the same character, i.e., fall in between linguistics and literature or is even absent in both fields. How different two subjects of clauses are highly dependent on the semantics of the linking elements.

Clauses that are paratactically combined for an enhancement: causal-reason relation are also irreversible because of the limiting of linking elements logically that phrases such as *and therefore*, *and so* indicate a cause-effect relation while word such as *for* present an effect-cause relation.

E.g., 3.1.47 || He believes || that learners [[who begin second language learning at an older age || **and therefore*** whose CALP is better developed,]] would acquire cognitive/academic L2 skills more rapidly than learners [[whose second language learning began at an earlier stage]]; || however, this would not necessarily be the case for those aspects of L2 proficiency unrelated to CALP (i.e., L2 BICS). || ---- CPhD13_0003_1434

E.g., 3.1.48 || This is determined by the sequence of cognition of human beings, || **for*** things [[human beings first cognize]] are those concrete, solid, and of a certain shape around them. || ---- CMAC06MO_18_172

When the result of an event is reinforced by the writer, the cause-effect order would be chosen as showed in E.g., 3.1.47 that learners with CALP being better developed is due to their late start of second language learning and the result is a character of specific learners that the writer intends to describe or later compare. The choice of *and therefore* clearly suggests the paratactic enhancement: causal relation within two clauses. When the reason for an event is more focused on by the writer, the reason would be explained in the secondary clause with the primary discussion about the effect, just as what E.g., 3.1.48 presents. Though the word *for* can be replaced by *because* or similar ones, *because* leads to hypotactic clauses that the dependent one is reversible orderly; *for* however, introduces paratactic clauses that the secondary one is irreversible, making the clause nexus a fixed effect-cause order.

The paratactic enhancement: conditional relation usually includes positive, negative, and concessive relations in which the concessive one can either be a concession followed up by a consequence or follows the consequence in a clause nexus.

E.g., 3.1.49 || Moreover, we know || that in English reading process, students may encounter many difficulties in comprehension || so the mistakes may occur frequently, || **and in this case*** it is advisable for teachers tolerate the students' classroom errors. || ---- CMAC13RE_2_439

The advice from the writer for teachers *tolerating the students' classroom errors* in E.g., 3.1.49 is on the condition that *students may encounter.... in English reading process*, without which the advice is nonsense or has no practical foundation for being raised. The conjunctive *and in this case* paratactically combined two clauses in which the *case* refers to former clauses or the condition and the whole clause nexus has the agnate *in English reading process, if students... it is advisable for ...* that two clauses are more closely related.

E.g., 3.1.50 || His free translation with annotation provide target readers further illustration on the cultural implication, || **or else*** the analogically lavished royal love and care of the monarch centered on Lady Yang cannot be perceived by those [[being devoid of relevant background knowledge]]. || ---- CMAC11PR_10_405

E.g., 3.1.51 || Students must observe the rules strictly; || **otherwise*** he or she may ultimately be expelled from the group. || ---- CPhD11_0007_963

The negative condition means that something cannot be fulfilled without the condition being mentioned in primary clauses, hence, phrases like *or else* and *(or) otherwise* are often applied in the clause nexus for paratactic enhancement relation as shown in examples above. In E.g., 3.1.50, the perceive of *the analogically lavished royal love...* by people without *relevant background knowledge* would be impossible if *his free translation* does not have any *annotation of illustration on ...* provided for target readers, i.e., the condition described in former clauses makes the realization of contents being valued in following ones, which is emphasized by the writer through the choice of the paratactic linkage *or else*. The same reinforcement comes from E.g., 3.1.51 when the condition of *observing the rules strictly* is paratactically linked to the result of not *being expelled from the group* in choosing the word *otherwise* for a negative conditional relation within two clauses. Slight differences within the selection of *or else* and *otherwise* can be referred to in Section 5.7.

The concession-consequence order for paratactic enhancement relation seems to be possible; however, it has not been discovered in our corpus as it seldom would mention the concession firstly by using a paratactic linkage except for referring to the consequence resulting from this concession. So, the concession clause cannot separately exist, and its advancement confirms the dependent status that the consequence, i.e., the independent clause, is what the writer attempts to focus on.

E.g., 3.1.52 ||| The background of *Pride and Prejudice* is so common || —never rising above the level of ordinary life, || leaving nothing to imagination or invention at all, || **yet*** it satisfies readers with its wit, its truth, and its profound comic power. || ---- CMAC06LIT_31_808

The consequence-concession order presented by E.g., 3.1.52 is signified by the conjunction *yet*, which introduces the concessional clause in the secondary place, placing the consequential one in the former, suggesting the writer's emphasis on the commonness of the *background of Pride and Prejudice*. If it is the satisfaction for readers *with wit, truth, and power* that the writer intends to pay more attention to, the clause complex should be *yet the background of ..., it satisfies readers with...* in which the linkage *yet* is brought forward, and the whole structure is a concession-consequence order. Another problem appears as *yet* is not deemed to be chosen for a hypotactic relation, which should have been replaced by words such as *though, although, etc.*

In a word, besides the spatial relation and the condition: concessive relation with a consequence-concession order, other possible relations within clauses can all be discovered in our research corpus, leaving the cause: purpose and the cause: result relation for paratactic clauses impossible. Sometimes, the paratactic relation is hard to tell as only the word *and* is used for enhancing, i.e., the secondary clause seems to be added to the primary one but brings some circumstantial information. This kind of implicit structure, though rare in the corpus, will be explained in Section 7.3 because of its importance.

3.1.3.3.2 Hypotactic Enhancement (Finite & Non-finite)

Hypotactic enhancement (annotated as $\alpha\beta$) is traditionally called the “adverbial clauses” that the dependent clause is usually taken up by some qualifying information to the contents of independent ones hence, the place that the independent clause resides in is moved far away by the dependent one in a hypotactic clause nexus, which is totally different from the condition for paratactic ones as the enhancement can move forward the discourse in paratactic clause nexus such as that works in narratives. E.g., 3.1.53 ||| α It can be seen || $\beta 1x\beta$ that, **despite*** representing two different searching processes, || $\beta 1\alpha$ these two models are used to realize similar cognitive purposes, || $\beta=21$ that is, English speakers tend to use TtFG constructions, || $\beta=2+21$ whereas Chinese speakers tend to choose GtF constructions to help their hearers to locate || $\beta=2+2x2$ and **finally*** focus on the Figure in the Ground [[1 where it exists || +2 or is presented into]]. || ---- CPhD07_0008_1299

The clause relation within the last two clauses in E.g., 3.1.53 is a paratactic enhancement: temporal one marked by the conjunctive *and finally* (1x2) to present the cognitive process of Chinese speakers by choosing GtF constructions model in comparison with that of English speakers, the former clause *whereas Chinese speakers tend to ...* is hence broadened by the latter clause; the relation within the second and the third clause, however, is a hypotactic enhancement: concessive one marked by the preposition *despite* ($\alpha\beta$) that *two models* are used for the same purpose regardless of their diverse searching processes, the dependent clause provides some information for the subject of the independent clause, is in fact ignored later on through the unfolding of the meaning expressed by clause nexuses.

As hypotactic enhancing clauses can be either finite (connected by conjunction binders) or non-finite (combined by conjunctive prepositions or conjunctions or by *WH-* binders) structurally, detailed clause relation will be separately discussed in this section.

Finite

E.g., 3.1.54 ||| Although in English culture today the bride wears white dress, || for which Queen Victoria is often credited with inventing the fashion, || to show her purity and loyalty to the gods, || the TL readers may well understand the Chinese culture || **while*** reading the descriptions about a traditional red Chinese wedding || and picturing a happy and lively scene. || ---- CMAC07CU_35_383

E.g., 3.1.55 ||| Just as the infants play with speech sounds by babbling, gurgling and spluttering || before they gain control over their speech organs || and produce their LI sounds and utterances, || foreign language learners have to practice their vocal organs repeatedly in pronunciation learning || **until*** they can coordinate and manipulate the movements of their vocal organs subconsciously and with ease. || ---- CPhD10_0009_1392

The word *while* in E.g., 3.1.54 denotes not the addition: adversative relation but a hypotactic enhancement: temporal: same time relation in leading to a finite clause with the subject TL readers being deleted that the *understand of the Chinese culture* by TL readers should exist at the time when they are *reading the descriptions about ...*. The word *until* in E.g., 3.1.55, though also connects two hypotactic clauses with a temporal relation together, the event depicted in the independent clause happens before that in the dependent one, which is determined by the semantics of the binding word. The hypotactic temporal relation for finite clauses can also convey later time, being the opposite from that expressed by *until* above, which is usually marked by the word *after, since, etc.*

Almost most of the spatial relation exists in hypotactic-finite clauses that can be used to refer to a

point marked by the word *where*, the extent signified by the phrase *as far as*, and the spread of the mentioned place indicated by words such as *wherever* and *everywhere*.

E.g., 3.1.56 ||| They found that creating, editing, and sharing messages was easy on this platform, || **where*** they could comment on the reading materials more freely, without anxieties in the comparison with the offline environment. ||| ---- CMAC14RE_15_425

The word *where* in E.g., 3.1.56 is chosen for the replacement of the word *platform* in connecting two clauses. Hence, the spatial relation comes from a concrete place as the platform exists in reality, and a list of actions, i.e., *creating, editing and sharing messages*, are really performed at the place.

E.g., 3.1.57 ||| A typical example is from Shakespeare's *Much Ado about Nothing*, || **where*** Beatrice and Benedick vie to see which can outdo the other in the use of asteismus. ||| ---- CPhD07_0011_320

Being diverse from the example above, *where* in E.g., 3.1.57 presents an abstract place though the relation within two clauses is also a hypotactic enhancement: spatial one as *where* is chosen for referring to the book *Much Ado About Nothing* written by Shakespeare. As the book has many editions, no exact edition or fixed paragraph is made clear by the writer; the place is abstract.

E.g., 3.1.58 ||| **As far as*** the Chinese culture is concerned, || the Chinese people usually have a tendency for literature, || although there existed a military examination system, || which the readers can find in Chapter 26 of the novel. ||| ---- CMAC07CU_35_517

Another condition comes from the semantic extending of words or phrases such as the application of *as far as* in E.g., 3.1.58, that a vector starting from a point *the Chinese culture* can be drawn if the image schema of the phrase is considered. But as the subject is an abstract conception, this kind of binder presents the hypotactic: enhancement: spatial relation as the abstract place is shaded into matter with a point being broadened into a scope.

For the hypotactic enhancement: manner relation for finite clauses, besides the means or comparison disparity when semantics of words or phrases are taken into account, the third one comes from a quality that things may happen by chance or in a specific way.

E.g., 3.1.59 ||| Kellerman (1983) makes psychotypology, the learner's perception of language typology, central to his perspective on transfer, || **whereby*** the learner's recognition of congruent forms between the native and target languages either facilitates or interferes with L2 acquisition. ||| ---- CPhD08_0005_481

The clause complex in E.g., 3.1.59 is agnate with *By taking the means that Kellerman makes psychotypology... central to his perspective on transfer, the learners' recognition of ... either ...or ...* that the word *whereby*, by hypotactically binding two clauses together, leads to a dependent clause depicting an event or action that depends on the means the subject from the independent clause adopts. If the means that the subject takes does not exist, the following description or explanation turns out to be in vain.

E.g., 3.1.60 ||| Huebner (1983) adduces data from a Hmong speaker learning English, || finding || that the learner initially treats subject NPs in English as always definite, || **as if*** they were a syntactic topic. ||| ---- CPhD08_0005_789

The hypotactic enhancement: manner: comparison relation for finite clauses is demonstrated in E.g., 3.1.60 when the phrase *as if* is selected as a binder for two clauses that *subject NPs in English* used by Hmong speaker English learners are compared with *a syntactic topic* though the latter has not been explained by the writer clearly and whether the comparison can be made or not remains unknown to readers.

E.g., 3.1.61 ||| **As*** shown in the two figures, || the parameters of pseudo passive, existential constructions and null elements seem to show the tendency [[to decline gradually with the increase of the proficiency level]]. ||| ---- CPhD08_0005_972

The word *as* in E.g., 3.1.61 indicates the exact way that *the parameters of ...* are presented, i.e., *in the two figures* in hypotactically connecting two clauses for an enhancement: manner logical relation. It may be regarded as a comparison, but the object being compared is absent. Hence, this quality relation more resembles the means one.

The hypotactic enhancement: causal relation for finite clauses has been mentioned before to have three kinds: cause: reason, cause: purpose, and cause: result.

E.g., 3.1.62 ||| **Since*** culture is an extremely complicated concept and enormous subject || and may concern itself with everything in the world: material or spiritual, || cultural terms may be defined as that [[which has elements peculiar to the natural environment, institutions and cultures of its language area]]. ||| ---- CMAC07CU_35_167

The word *since* in E.g., 3.1.62 is a common binder for hypotactic enhancement: causal relation that the dependent clause introduced by it usually plays the reason function, leaving independent clauses the result function logically. The order of clauses in a clause complex highly depends on the writer's intention of reinforcing that the initiating clause is always reinforced, no matter whether it has a logical binder or not.

E.g., 3.1.63 ||| And nobody missed the first class of the semester || and the teacher required || that all the students in both of the experimental and control class should not be absent in the week of 18, || **so that*** all the students took part in the tests with no one missing. ||| ---- CMAC12LI_6_503

The cause-result relation seems to be transitional with the cause: reason relation by changing the place of the binding element or just the connection marker. In fact, this kind is rare as the most common phrase for binding clauses is *so that*, which can be easily mixed with that for indicating a cause-purpose relation in hypotactic finite clauses. One way for distinguishing them and thus annotating is to check whether the clause is agnate with that begins with the expression *for the result that*, as suggested by E.g., 3.1.63. As the whole clause nexus can be rewritten as *all the students for the result that they took part in the tests ...*, it tends to be regarded as a cause-result relation but not a cause-purpose one, though the latter is also reasonable, the verb chosen by the writer is would then be *take* instead of *took*.

E.g., 3.1.64 ||| Therefore, the researcher asked some teachers of English working in those universities for help || and the researcher also made to them detailed explanation of the research and the questionnaire || **in order that*** they could administer the questionnaire in a correct way. ||| ---- CPhD10_0018_717

E.g., 3.1.65 ||| Just as Biber et al. (1998) said, || they provide a large empirical database of natural discourse, || **so that*** analyses are based on naturally occurring structures and patterns of use rather than intuitions and perceptions, || which often do not accurately represent actual use. ||| ---- CMAC11WR_009_321

The cause-purpose relation in hypotactic enhancement for finite clauses is easy for the identification when it has the agnate *for the purpose that...* and common phrases such as *in order that* and *so that* are selected for clause binding, just as shown in E.g., 3.1.64 and 3.1.65. The former example explained the researcher's *detailed explanation* to some teachers *of the research and the questionnaire*, for their being able to *administer the questionnaire in a right way*; and the latter example clarified Biber et al.'s action of *providing a large empirical database of natural discourse* is to make sure that their analyses are *based on naturally occurring* materials instead of intuitional ones. Normally, the dependent clause follows the independent one with an abbreviated subject or a new one, and the sequence is irreversible unless another binder is chosen.

The conditional relation for hypotactic finite clauses resembles that for paratactic clauses when it includes positive, negative, and concessive relations in which the negative one can only be presented by the word *unless* while the other two are flexible in the choice of binders.

E.g., 3.1.66 ||| The range of exercise types and activities compatible with a communicative approach is unlimited, || **provided that*** such activities enable learners to attain the communicative of curriculum, || engage information sharing in communication, || and require the use of such communicative processes as information sharing, negotiation of meaning, and interaction. ||| ---- CMAC05ME_21_204

E.g., 3.1.67 ||| **As long as*** they can figure out the meaning of every word encountered || and make clear rules [[that appear to be inconsistent]], || they will feel emotionally comfortable and secure. ||| ---- CPhD13_0003_1143

The positive condition means that events or things can happen only by the condition provided by the statement in dependent clauses hence, *the range of exercise types and activities... is unlimited* on the condition that those *activities enable..., engage..., and require...* in E.g., 3.1.66 when the phrase *provided that* is chosen; and *they will feel ...* on the condition that *they can figure out... and make clear rules...* in E.g., 3.1.67 by selecting the phrase *as long as*. In these two examples, the marker *provided that* and *as long as* are interchangeable, and so is the sequence of the dependent clause and interdependent clause in each clause nexus, though the semantic and pragmatic emphasis would always be the initiating clause.

E.g., 3.1.68 ||| Actually no language can exist || **unless*** it is steeped in the context of culture || and no culture can exist || unless it holds the structure of a natural language at its center. ||| ---- CMAC07CU_35_140

The word *unless* is a typical binder for connecting two finite clauses with the hypotactic enhancement: negative condition relation that the event or things being described in the independent clause cannot happen or be fulfilled without the condition being mentioned in the dependent one. As suggested in E.g., 3.1.68, the language can only exist on the condition of its *steeping in the context of culture*, and the culture can only exist on the condition of its *holding the structure of a natural language ...* that other conditions are not able to support for the existence of the language and culture.

E.g., 3.1.69 ||| However it is also noted || that this unknown status is valid || **even if*** it is unknown in the Ground [[where it exists or is presented into]]. ||| ---- CPhD07_0008_1294

E.g., 3.1.70 ||| This situation leads them to learn in a passive way || **even though*** they encounter some difficulties during the learning process. ||| ---- CMAC11ME_16_421

E.g., 3.1.71 ||| **While*** a great deal has been learnt about writing from studying [[how native speakers acquire composing skills]], || the emergence of L2 writing as a sub-discipline has opened new theoretical

perspectives, research methods and pedagogical strategies. ||| ---- CPhD10_0015_1575

The concessive relation for hypotactic finite clauses means that facts or the consequence of an event depicted in the independent clauses would not be affected under the circumstances or the condition described in the dependent one. In E.g., 3.1.69, the choice of *even if* followed by the condition *it is unknown in the Ground...* does not affect the statement *this unknown status is valid* before it. So does the situation in E.g., 3.1.70, that students still *learn in a passive way* no matter if they *encounter some difficulties during the learning process* or not. Strange is that the attitude of the writer toward the subject remains unchanged as the negative mood is adopted in both clauses. Another example comes from the choice of the word *while*, though this can be manifold in introducing clauses for diverse relations, agnates *although* here in E.g., 3.1.71 for a concessive relation that *the emergence of L2 writing ... opened new theoretical perspectives...* although *a great deal has been learnt...* This means that studying writing in history does not change the fact that *the emergence of L2 writing as a sub-discipline* has been fundamental.

Non-finite

Hypotactic enhancement relation for non-finite clauses can be marked explicitly by structural conjunctions such as *when, while, although, if*, etc., and conjunctive prepositions like *before, after, since, without*, etc. to present temporal, causal, manner, conditional and concessive functions; or implicitly without any obvious marker, which would then be really hard to identify.

E.g., 3.1.72 ||| They approach the closest to human's expectations or anticipations || **in the process of*** being formed into a concept. ||| ---- CMAC10SY_9_599

E.g., 3.1.73 ||| **After*** handing out the paper of questionnaire and spot dictation to the participants, || the teacher played the tape with inspection on the situation through the whole procedures. ||| ---- CMAC12MO_3_285

E.g., 3.1.74 ||| In addition, there are a few principles [[that guide teachers in the classroom]] || **before*** introducing the learning plan to students. ||| ---- CPhD11_0007_1054

The phrase *in the process of* in E.g., 3.1.72 can be interpreted as the word *when* for a hypotactic enhancement: temporal: same time relation that the event in the independent clause happens at the same time with that in the dependent one, though the dependent clause seems to be embedded into the noun *process*. The choice of the word *after* in E.g., 3.1.73 indicates the temporal: later time relation that the event described in the independent clause follows that in the dependent one, and this temporal sequence is interchangeable though the clause sequence can be turned by turning the marker *after* into *before*. That is, the clause nexus is agnate with *before the teacher played the tape with ..., they handed out the paper of ...to the participants*, the emphasis of which by the writer changes from the contents of the dependent clause to that of the independent one. Being opposite to this example, the choice of the word *before* in E.g., 3.1.74 presents the temporal: earlier time relation for non-finite clauses that by post positing the dependent clause, the subject of the independent clause, i.e., *few principles that guide teachers* is stressed. The commonness for three examples of this kind is the omission of the subject in the independent clause in each clause nexus, which is the characteristic of non-finite clauses.

E.g., 3.1.75 ||| They are less likely to get involved in classroom communication for fear of being ridiculed || **because of*** making mistakes || or not communicating properly. ||| ---- CPhD13_0003_1105

E.g., 3.1.76 ||| Then, on account of the author's limited language ability and scope of knowledge, all her suggested English translations in the thesis are far from perfect || and are only for the sake of reference || **with the aim of*** improving the quality of the tourism translation in Hangzhou. ||| ---- CMAC09PR_28_577

E.g., 3.1.77 ||| As some of the subjects said, || they tried to use [[what they were familiar with]] || **for fear of*** making mistakes and losing scores. ||| ---- CPhD10_0018_1090

The hypotactic enhancement: causal relation for non-finite clauses can be reasonable as suggested by E.g., 3.1.75 when the preposition *because of* instead of the word *because* is applied in the clause nexus for introducing to the dependent one and omitting the subject of the independent clause; or be purpose as showed in E.g., 3.1.76 and 3.1.77 when phrases *with the aim of* and *for fear of* are used respectively. The latter clause complex can be understood as *they tried to use for not making mistakes and losing scores*, which is obviously a hypotaxis marked by the purposeful preposition *for*. The clause *improving the quality of ...* in the former example, though it refers directly to the noun *aim*, can also be replaced by prepositions such as *for, to*, and *in order to* for purpose relation within clauses. In fact, the purpose for conducting something or the performance of subjects usually provides explanations for these actions and can thus be a specific kind of cause: reason relation.

E.g., 3.1.78 ||| Several pairs of students are chosen at random, || **by*** drawing names from a hat, || to perform the same tasks (without scripts) in front of the class. ||| ---- CMAC05PR_24_672

E.g., 3.1.79 ||| Perspective, according to Hyland (2002a), allows the writer to both engage with the disciplinary literature in a continuing debate || and to negotiate the value of his own contribution || **by***

opening a discursive space in that literature, || building on what has gone before. || ---- CPhD06_0004_1482

The hypotactic enhancement: manner: means relation for non-finite clauses is usually introduced by the preposition *by* or *by means of* as suggested by E.g., 3.1.78 and 3.1.79 that both examples place the dependent clause being led by the word *by* afterward, leaving the result or consequence of events being described in the independent clause because of the methods that are taken in the dependent ones more emphasized. Hence, it is the fact that *several pairs of students are chosen at random* and *perspective allows the writer to both... and ...* in two examples are more important, and this can be realized through the means chosen by the writer.

E.g., 3.1.80 ||| Language testers can reduce the bad washback effect || through exploring new items || and improving the contents and forms, || but **if*** considering || the improvement of washback effect on language testing is the language testers' responsibility, || it is too simple to solve this problem. || ---- CMAC13ME_11_91

E.g., 3.1.81 ||| The second one is local independence of items, || which means || that items are not related || **unless*** measuring the same trait (Mislevy, Rupp & Harring, 2012). || ---- CPhD14_0002_132

E.g., 3.1.82 ||| **Without*** exploring wider intertextual chains, || we cannot understand particular texts or the significance of these for the participants. || ---- CPhD10_0015_1438

The hypotactic enhancement: conditional relation for non-finite clauses also clarifies the condition that is inevitable for an event as that for finite clauses, which can also be positive or negative introduced by the conjunction *if*, *unless*, or the conjunctive *without* as demonstrated by examples above. In E.g., 3.1.80, the relation is positive conditional presented by the word *if* that on thinking about the condition of *considering the improvement of ...*, it is not enough for solving the problem mentioned by the writer before; in E.g., 3.1.81, the conditional relation within clauses are marked by the negative conjunction *unless*, indicating that *items are only related* when the condition of *measuring the same trait* is taken into account; and in E.g., 3.1.82, the preposition *without* for the negative conditional relation suggests that the understanding of *particular texts or the significance ...* can only be possible by *exploring wider intertextual chains*. No matter where the dependent clause is, the condition introduced by conditional markers is always the focus of the writer in meaning transferring.

E.g., 3.1.83 ||| **Although*** being exposed to the huge challenges by the traditional media of radio, television, newspapers, etc., || the network will continue to be the most powerful new medium in the present-day society. || ---- CMAC11LI_9_296

The concessive relation for hypotactic non-finite clauses states the characteristic of the subject in the independent clause (as will be deleted in the dependent one) that facts being described in the former would not change no matter whether the condition in the latter clause changes or not. Therefore, *the network continues to be the most powerful ...* no matter whether people are *exposed to the huge challenges* by ... or not, as indicated by the conjunction *although* in E.g., 3.1.83; and *these two models are used to realize similar cognitive purposes* except for the fact that they *represent two different searching processes* as shown by the preposition *despite* in E.g., 3.1.53.

Usually, the dependent clause is placed in front of the independent one in a clause nexus for a concessive relation with the subject being omitted; its contents would not have effects on those afterward.

The hypotactic enhancement for non-finite clauses is sometimes implicitly expressed for temporal, cause: reason, cause: purpose, and cause: result relations that no obvious marker is adopted for clause connection, and the relation can only be interpreted by the logic that the writer conveys through semantics and the context that the clause nexus is in. As implicit clause nexuses are not frequently discovered in our corpus, they will be concentrated on in Section 7.3 in this thesis.

In short, the hypotactic enhancement relation within clauses can be temporal, spatial, manner, causal, conditional, and concessive though the manner: means relation is absent from finite clauses, and the spatial relation cannot be found in non-finite clauses. The intersection of the enhancement relation with taxis presented by principal markers for each category is clearly demonstrated in Table 7 that paratactic clauses differ from hypotactic ones in each category, and the latter are also diversified according to the finiteness of clauses in choosing specific expression in meaning conveying. The exact choice of these markers by students in thesis writing from our corpus with their similarities and differentiations will be clarified in Chapter Five for qualitative analysis.

	Category	Meaning	Paratactic	Hypotactic		
				finite	<u>non-finite:</u> <u>conjunction</u>	<u>non-finite:</u> <u>preposition</u>
(i) temporal	same time	A meanwhile B	(and) meanwhile; (when)	[extent] as, while	while	in (the course/ process of)
				[point] when, as soon as, the moment	when	on
				[spread] whenever, every time	—	—
	different time: later	A subsequently B	(and) then; and + afterwards	after, since	since	after
	different time: earlier	A previously B	and/ but + before that/ first	before, until/ till	until	before
(ii) spatial	same place	C there D	and there	[extent] as far as	—	—
				[point] where	—	—
				[spread] wherever, everywhere	—	—
(iii) manner	means	N is via/ by means of M	and + in that way; (and) thus	—	—	by (means of)
	comparison	N is like M	and + similarly; (and) so, thus	as, as if, like, the way	like	
(iv) causal- conditional	cause: reason	because P so result Q	[cause ^ effect] (and) so; and + therefore			
			[effect ^ cause] for; (because)	because, as, since, in case, seeing that, considering		with, through, by at, as a result, because of, in case of
	cause: purpose	because intention Q so action P	—	in order that, so that	—	(in order/ so as) to; for (the sake of), with the aim of, for fear of
	cause: result			so that	—	to
	condition: positive	if P then Q	(and) then; and + in that case	if, provided that, as long as	if	in the event of
	condition: negative	if not P then Q	or else; (or) otherwise	unless	unless	but for, without
	condition: concessive	if P then contrary to expectation Q	[concession ^ consequence] but; (and) yet, still; but + nevertheless [consequence ^ concession] (though)	even if, even though, although, while	even if, even though, although, while	despite, in spite of, without

Table 7 The categories of the enhancement relation in diverse clauses connecting structures with relative markers (Halliday, 2014, p. 477-8)

3.1.4 Projection (without Embedding)

Being another significant parallel section of the logico-semantic relationship as the expansion does, projection indicates the clause being “a representation of a linguistic representation” rather than a “representation of non-linguistic experience” (ibid, p. 508) in meaning conveying, which can usually be adopted to refer to sources in news reporting, express viewpoints in scientific discourse, organize dialogues in narratives, and raise questions in conversations (ibid, p. 509), especially be frequently found in referring to studies in the literature review part and in listing dialogues in the descriptive process for novel plots from our corpus.

E.g., 3.1.84 ||| However, just like what LU Xun once **said***, || its thrust is at “scholars,” and the thorough portrayals of other kinds of characters properly form the background of the scholars. ||| ---- CMAC07CU_35_30

E.g., 3.1.85 ||| The saying originates from Analects, the book [[which mainly recorded the words and deeds of Confucius and his disciples]], || **meaning*** || that the name is not correct || and what is said does not accord with reason || and that it is not justifiable. ||| ---- CMAC07CU_35_408

E.g., 3.1.86 ||| Due to the low difficulty of the explicit rules of maintaining agreement between subject and predicate, || the majority of the participants **thought*** || they had a good mastery of the rules || as being expected. ||| ---- CPhD08_0012_1007

The verb *said* in E.g., 3.1.84 is a verbal one, introducing sayings or words from LU Xun directly, though the quotation mark seems to have been forgotten by the writer. As the clause being referred to or projected by the verb *said* does not contain any request for people, it is a proposition being paratactically quoted by the former clause. The verb *meaning* in E.g., 3.1.85 and *thought* in E.g., 3.1.86 are both mental ones that the projected clauses are not words or sayings originally but are viewpoints from the writer formed in events or reported ideas from others that these indirectly referring words are hypotactically organized into the projecting clause with principal markers, making up a sharp contrast with that in the former example. Again, the projected clauses are propositions that the truth value of them can be judged instead of proposals for any object (personal).

Here comes the classification of projection that, according to Halliday (ibid, p. 509-511), three systems of differing can be taken into account: the level of projection, the mode of projection, and the speech function. As the clause being projected can either be the representation of the content of a verbal clause said by the subject (of the projecting clause) or a mental one thought by the subject, the former is thus called a “locution,” and the latter is an “idea” from the perspective of the Level. For the Mode, when the structure of the clause nexus, i.e., the interdependency of clauses in combination, is considered like those in expansion, a clause can either be paratactically or hypotactically projected into the former one as a quote (two clauses have the same status) or a report (one clause depends on another with unequal status), which is traditionally called the “direct speech” or the “indirect speech.” Though a clause can also be a constituent of a word for an embedding relationship, it will not be discussed here in detail as the relation within clauses is what this thesis emphasizes. The speech function of the projection involves the characteristics of the clause being projected that it can be a proposition with truth value or a proposal requesting other people to do something, not to mention the minor clause like greetings or exclamations. By the former, the dialogic specialties are transformed into written or illustrative utterances to the greatest extent that the clause nexuses are more complex and organized. Therefore, given these three dimensions of the classification for projection and their intersections, detailed types of projecting processes are demonstrated in Table 8 with annotations for each type.

Type of projecting process	Speech Function	Paratactic (direct, quoted) 1 2	Hypotactic (indirect, reported) α β
Locution-verbal “	Proposal	1 “2 He asked her, “you do the task quickly”.	α “ β He asked if she could do the task quickly.
	Proposition	1 “2 He said, “I do the task quickly”.	α “ β He said (that) he did the task quickly.
Idea-mental ‘	Proposal	1 ‘2 He willed her ‘do the task quickly”.	α ‘ β He wanted that she could do the task quickly.
	Proposition	1 ‘2 He thought, “I can do the task quickly”.	α ‘ β He thought (that) he did the task quickly.

Table 8 Projecting types of the speech function, level, and mode intersection with examples and annotations

The following section will start from paratactic quoting with the Level and speech function intersection, and then the hypotactic reporting with Locution/Idea and Proposal/Proposition categories with examples and key verbs in detail to provide explanations for the classification and demonstrate how each category can be distinguished and interpreted in specific clause complexes.

3.1.4.1 Quoting (“Direct Speech” - Parataxis)

The quoting clause nexuses, traditionally called the “direct speech,” are always paratactic ones in considering the dependency of clauses that two in a nexus are of equal status, as can be shown from the written texts when quotation marks (either double or single ones) are applied in signaling the clause being projected. By quoting, all the interactive characteristics of the projected clause being an exchange remain, including the “mood potential, vocatives and expletives, tone selections, and (textual) continuatives.” (ibid, p. 512)

E.g., 3.1.87 ||| After being told || that Mrs. Sloane does want Gatsby to come, || Tom **responds***, || “She has a big dinner party and he won’t know a soul there. I wonder where in the devil he met Daisy. By God, I may be old-fashioned in my ideas, but women run around too much these days to suit me. They meet all kinds of crazy fish.” || ---- CMAC14LIT_25_625

The projected clause complexes after the verb *respond* in E.g., 3.1.87 keep originally that Tom’s response toward the listener is totally “quoted” by paratactically presenting them in a quotation mark. The clause nexus is a verbal one as *respond* introduces a locution rather than an idea semantically, and the projected clause is a proposition that facts are stated with truth value, but not a proposal for giving offers or demands toward anyone. Hence, the projecting process is a verbal one for representing a saying from a person.

Normally, quoting of sayings is frequently found in eyewitness accounts in news reports, dialogic passages in narratives or literature works, references in scientific theses or articles, etc., for the authenticity of meanings expressed by witnesses, characters, authors, or significant scholars. Another kind of quoting is the projected “wording” in the clause depicts a phenomenon, and the lexicogrammatical relation within clauses depends on the contents of the wording being projected in a verbal process.

	Proposition	Proposal
(1) general member	say	say
(2) verbs specific to speech function: (a) giving (b) demanding	(a) statements: <i>tell</i> (+ Receiver), <i>remark</i> , <i>observe</i> , <i>point out</i> , <i>report</i> , <i>announce</i>	(a) offers: <i>suggest</i> , <i>offer</i> ; <i>threaten</i> (‘offer: undesirable’), <i>vow</i> (‘offer: sacred’), <i>promise</i> (‘offer: desirable’), <i>agree</i> (‘offer: in response’)
	(b) questions: <i>ask</i> , <i>demand</i> , <i>inquire</i> , <i>query</i>	(b) commands: <i>call</i> , <i>order</i> , <i>request</i> , <i>tell</i> , <i>propose</i> , <i>decide</i> ; <i>urge</i> (‘command: persuasive’), <i>plead</i> (‘command: desperate’), <i>warn</i> (‘command: undesirable consequences’)
(3) verbs with additional circumstantial feature: (a) (b) manner specifying connotation	<i>reply</i> (‘say in response’), <i>explain</i> (‘say in explanation’), <i>protest</i> (‘say with reservation’), <i>continue</i> (‘go on saying’), <i>add</i> (‘say in addition’), <i>interrupt</i> (‘say out of turn’), <i>warn</i> (‘say: undesirable consequences’)	[see (2) above]
	<i>insist</i> (‘say emphatically’), <i>complain</i> (‘say irritably’), <i>cry</i> , <i>shout</i> (‘say loudly’), <i>boast</i> (‘say proudly’), <i>murmur</i> (‘say sotto voce’), <i>stammer</i> (‘say with embarrassment’), <i>enthuse</i> (‘say with approval’), <i>gush</i> (‘say effusiveness’), <i>rave</i> (‘say with enthusiasm’)	[largely the same as for propositions] <i>blare</i> , <i>thunder</i> (‘order imperiously’), <i>moan</i> (‘plead whiningly’), <i>yell</i> (‘order vociferously’), <i>fuss</i> (‘order officiously’)

Table 9 Main verbs for introducing paratactic quoting clauses in the verbal process according to the speech function

As quoted clauses are verbal processes for most of the time, which can be proved from the quantitative

comparing results from our corpus in Section 6.2, the main verbs for leading to paratactic locutions summarized by Halliday (ibid, p. 514) are presented here, though more than half of them cannot be seen in theses writing especially in the researching corpora because of genre and contents limitation. Besides the most common verb *say*, which can both introduce proposals and propositions, other verbs for proposals mainly provide offers or commands toward locution receivers, and those for propositions make statements or ask questions to listeners, both of which fulfill the “giving” and “demanding” functions in communication. Still, several verbs contain specific connotations in either adding circumstantial characters to the verb *say* or suggesting the manner or way of “saying” in daily life, which are seldom seen or rare in academic writing texts. (See Table 9)

Verbal-Proposition

Verbal propositions quoted by writers in thesis writing are primarily statements, as seen in E.g., 3.1.91 and 3.1.92 that the essence of “meanings” is described in the former example, and a real example from Chomsky is displayed in the latter.

E.g., 3.1.91 ||| Kress (1995) talks about meaning making in relation to the context of culture and intertextual context, || and **comments***, || “meanings find their expression in text though the origins of meanings are outside the text—and are negotiated (about) in the texts, in concrete situations of social exchange” (p.18). ||| ---- CPhD09_0002_203 remark

E.g., 3.1.92 ||| To illustrate this point, || they **quoted*** a famous Chomskyan example (ibid. p. 150): || Children with English speaking parents readily adopt that “The man is here.” becomes a question by reversal of noun and verb “Is the man here?” ||| ---- CPhD10_0009_244

The verb *comment* resembles *remark* for giving an opinion on the subject, though this comment is on a carrier of a book written by Kress. The same condition goes for the verb *quote*, which, as a verbal word, introduces contents also from a book written by Chomsky instead of a conversation or a record of an interview.

Mental-Proposal

The Mental proposal category for quotes is hard to identify or find in natural language as well as in our corpus as the proposal, being a piece of advice or a request for someone, if not transmitted by locution but by mental ideas, would be almost impossible for realization.

E.g., 3.1.93 ||| David at last **decides***: || “better to be naturally Dora than anything else in the world”. (DC: P765) || since he can’t adapt Dora to himself, || he resolves to adapt himself to Dora. ||| ---- CMAC10LIT_7_299

One special case comes from E.g., 3.1.93, when the projecting verb *decide* is selected. It is the subject of David’s decision, i.e., an idea that is projected, but the contents are proved to be a proposal for David himself rather than for others as David thinks that *Dora is better to be natural than anything else in the world* and the object for accepting the idea is also David. If the object turns out to be other people, the clause nexus would be impossible logically or become a metaphor. This is why mental proposal quotations are hardly found in academic writing discourse, though quite a few are chosen in plot descriptions from literary works.

Mental-Proposition

The mental proposition clauses are usually quoted as a statement or a question by verbs such as *think*, *wonder*, *reflect*, etc.

E.g., 3.1.94 ||| While it is incredible in the novel for helpless Oliver || to get help from his benefactors by coincidences || and finally changes his wretched fate, || it just **reflects*** Dickens optimistic attitude towards humanity: || “the principle of good surviving through every adverse circumstance and triumphing at last.” ||| ---- CMAC10LIT_7_143

The proposition being projected in E.g., 3.1.94 comes from a reflection of Oliver’s miserable experience Dickens in writing his novel and can be agreed with or opposed to by any reader, i.e., *the principle of good can survive through every adverse circumstance and triumph at last* or not according to reader’s views. Hence, this is a classic example of a mental proposition quoting relation.

3.1.4.2 Reporting (“Indirect Speech” - Hypotaxis)

The quoting has been discussed about its essence as construing the phenomenon of experience as either “saying” or “wording,” the reporting, however, indicates the process of construing the phenomenon of experience as “meaning” that the latter process is directly fulfilled whereas the former should take the “meaning” as an immediate that experiences are firstly construed as meaning and then as a wording. That is why in SFG, the single quotation ‘ is used to mean the projected meaning or the reporting relation in projection, while the double quotation “ is applied in signifying the projected wording or the quoting relation. (Halliday, 2014, p. 518) In paratactic projection, i.e., the quoting, any dialogic characters of what is talking about can be presented or projected, including vocative elements or minor speech functions, though the deixis has to be transferred to the orientation of the sayer of the projected clauses;

while in hypotactic projection, i.e. the reporting, those dialogic elements cannot be projected and the deictic orientation of the projecting clause remains. (ibid, p. 519) Hence, reporting is usually found to represent the speaker's thinking to conduct an assessment of contents being projected and the addressee's or the sayer's thinking to query for information in dialogues, the character's consciousness or awareness in narrative or novels, the spokesman's opinions or attitudes in news reporting, the scholar's views or beliefs in scientific materials, the writer's perspectives in scientific articles for presenting his/her reasoning processes, etc. (ibid, p. 515) This is why reporting is also of significance for research in genre such as thesis writing beside the quoting clause relation, and its interpretation follows the structure in that for quoting, i.e., the intersection of verbal or mental verbs with proposals or propositions being projected.

Verbal-Proposal

When proposals are hypotactically projected by "verbal" verbs as "indirect speech," the subject of the projected clause would normally be deleted and need to be judged by the receiver of the reporting clause as the projected clause, possessing a passive attribute, then takes the receiver as its subject no matter the reported proposals are commands, offers or suggestions. (ibid, p. 524, 526)

E.g., 3.1.95 ||| In the first talk, Sartorius wants || Trench to write his family || to **ensure*** || his daughter to be treated fairly. ||| ---- CPhD17_0002_1299

Despite the verb *want* in marking a mental proposal reporting relationship, the choice of the verb *ensure* introduces a verbal clause that *his daughter* is asked by Sartorius *to be treated fairly* by Trench. Here, the theme of the projected clause *his daughter* turns out to be Sartorius' daughter judging from the projecting clause, and the rheme, though missing, in fact, is the receiver of the projecting clause, i.e., Trench, and at the same time is the real subject of the projected clause as it is Trench who is demanded to take action and who needs to treat Sartorius' daughter fairly. So, E.g., 3.1.95 is a typical verbal proposal clause when the verb *ensure* with its projection is considered.

E.g., 3.1.96 ||| All the nine subjects were shown two series of simple drawings and an illustration || and then **asked*** || to describe the illustrations in both their native language and English. ||| ---- CPhD10_0018_443

The same verbal proposal reporting clause nexus is the E.g., 3.1.96 that the verb *ask* is obviously a verbal verb as judged by its semantics and the proposal is *to describe the illustrations* ... with the subject being deleted, which can be inferred from the subject of the projecting clause that *the nine subjects* are proposed to act; then the receiver of the projecting verb *ask* is the subject of the projected clause, with the subject of the projecting clause being the opposite from it, i.e., it might be a researcher who conducts the research and then takes notes down of the researching process.

E.g., 3.1.97 ||| Communicating the purpose and significance of the survey **encourages*** the participants || to answer the questions sincerely || and eliminate their suspicion [[of being identified by the information they provide]]. ||| ---- CPhD08_0012_593

Another example comes from the application of the verb *encourage* that the receiver of the projecting verb and, at the same time, the subject of the projected clause, i.e., *the participants*, are proposed by the researcher through the action of *communicating the purpose... to answer the question ...*, the subject of the projecting clause, though has been hidden by the writer, can hence be the researcher when the verbal verb *encourage* is adopted.

The best way of distinguishing reported proposals from causatives is to identify the quoting equivalent by using the same verb that the possibility indicates the projection attribute of the original structure, which can be proved by the equivalents from three examples above: *Sartorius wants Trench ... to ensure: "my daughter can be treated fairly"*; All the nine subjects were then asked: *to describe the ...*; and ... *encourages the participants: "to answer the questions sincerely..."* Normally, what is projected in the proposal may not happen at all, in fact, and any verb signifying a speech act is assumed to be a choice for projection. (ibid, p. 527)

Verbal-Proposition

The verbal proposition clauses being reported are usually statements and questions introduced by verbs with complex semantics for elaborating speech functions such as *imply*, *remind*, *hypothesize*, *claim*, *maintain*, *enquire*, etc., except simple ones like *say*, *ask*, and *tell*. Being different from those being quoted, reported verbal propositions would change the mood of the original meaning and obscure the speech function, which has to be made explicit then through the reporting verb. And verbs for interpersonal and/or behavioral characteristics of the speech event, especially those for quoting in narrative contexts, are expelled from the reporting relation as the meaning of "say" does not exist in these verbs. Another way of presenting the reported verbal proposition is the verb *express* followed by a noun with some kind of sense like *belief*, *hope*, *concern*, *disappointment*, *anger*, *regret*, etc. and then an optional nominal group with the following *that* clause. (ibid, p. 521-2)

E.g., 3.1.98 ||| In order to **state*** explicitly || that the probability is subjective, || or alternatively, at the other end, to **claim*** explicitly || that the probability is objective, || the writer construes the proposition as

a projection || and encode the subjectivity (I believe, I think) or the objectivity (probably, it is likely), in a projecting clause. ||| ---- CPhD06_0004_2264

The verbs *claim* and *state* in E.g., 3.1.98 are both verbal projecting verbs with their semantics more than a *say* that clauses being projected are statements or claims (the noun of the verb) given by the writer to inform readers of the information that *the probability is subjective* and *the probability is objective*. The *that* clause is obviously an indication of hypotactic clause relation; hence, the clause complex entails two reporting verbal-propositions.

In a word, no matter whether the reported clause is a proposal or a proposition, the verbal verbs for introducing to them can be classified as giving information to the receiver or demanding something from the receiver of the verb that the former turns propositions into statements and questions, and the latter makes proposals as offers and commands. Some specific verbs serving the verbal reporting clauses are summarized by Halliday (2014, p. 523) as follows:

		Proposition	Proposal
(1) general:	(a) giving	[statement] <i>say, tell</i>	[offer] <i>offer</i>
	(b) demanding	[question] <i>ask</i>	[command] <i>tell, ask</i>
(2) elaborated speech function:	(a) giving	[statement] <i>insinuate, imply, remind, hypothesize, deny, make out, claim, maintain</i>	[offer] <i>promise, threaten, undertake</i>
	(b) demanding	[question] <i>enquire, ascertain</i>	[command] <i>command, demand, persuade, forbid, encourage, recommend, implore, plead (with sb), cajole (sb into v-ing), suggest discourage (from v-ing)</i>

Table 10 Main verbs for introducing hypotactic reporting clauses in the verbal process according to the speech function

Mental-Proposal

Proposals reported by mental verbs, traditionally called “indirect thought,” are “exchanges of goods and services,” indicating the process of desire or hopes for “states of affairs.” (ibid, p. 524) They can either be non-finite clauses (primarily perfectives) or be finite ones like declarative, being “modulated with a modal auxiliary of obligation” such as *should, must, have to, might, could, would*, etc. (ibid, p. 525) Especially when the process of the independent clause is about desire with the dependent clause being a future declarative or an agnate with it, the clause nexus highly turns out to be a projection. (ibid, p. 527-8) Being contrasted with the reported verbal proposal, the subject of the projecting clause is made clear and takes part of the subject of the projected clause in the reported mental proposal relation. (ibid, p. 526) E.g., 3.1.99 ||| Sometimes when I met a tough question [[that I cannot deal with it by my own]], || **I wish*** || teacher can be shown in the classroom at any time. ||| ---- CMAC17_02_459

The verb *wish* in E.g., 3.1.99 is obviously a mental one for introducing a projecting clause, and as the word *that* within it and the projected clause is omitted, the hypotactic reporting relation is hidden. Being a dependent clause with a modal auxiliary of obligation *can*, which signifies a future declarative in the clause nexus, the finite clause being projected is a proposal that the subject *I* has a desire for the *teacher* and hopes that the *teacher* can provide some goods and services for *me* or for students. So, the subject of the projecting clause is explicit and has a certain kind of relation with that of the projected clause, though this reported or projected mental proposal may not happen at last.

E.g., 3.1.100 ||| If non-native writers **want*** their RAs || to be accepted by international journals, || they need to know || not only about the culture [[of which the article forms part or the rhetoric of the article]] || but also how to use grammar and lexis || to construct sentences appropriate for the RAs. ||| ---- CPhD09_0002_59

Another example of the reported mental proposal relation is E.g., 3.1.100, when the mental verb *want* is adopted. As the clause *to be accepted by...* is a dependent one with a passive voice and a future tense, which is agnate with the clause *can be accepted ...* in choosing a modal auxiliary *can*, it is projected then by the verb *want* and suggests a desire from the subject of the projecting clause *non-native writers* for the receiver of the theme of the projected clause, i.e., the journalists who accept RAs.

Mental-Proposition

The reported mental proposition, according to Halliday (2014, p. 516), can be “either an indirect statement or an indirect question” and the distinguish between them, being diverse from that of verbal projection in considering the speech functional tendency “of giving or demanding information” to receivers, depends highly on how valid the information is. Hence, a statement would usually be presented “by an indirect declarative clause” with the beginning of *that* in preserving the polarity and transitivity

elements; while a question is conveyed through “an indirect interrogative clause” led by *whether* or *if* for yes/no relation with an open polarity, or “an indirect *WH*-interrogative clause” introduced by *who*, *which*, *when*, *where*, etc. with at least one transitivity elements, the latter of which cover meanings such as *wondering*, *doubting*, *checking*, *finding out*, *contemplating*, etc. expressed by corresponding specific verb *wonder*, *doubt*, *check*, *find out*, *ascertain*, etc.

E.g., 3.1.101 ||| Based on the annotation of the selected data, it is clear to **know*** || that almost all the business email messages are the hybrid of the format of business letters and the format of email genre. |||
---- CMAC15_03_360

E.g., 3.1.102 ||| Thus, hypotheses **predicting*** || that language learning strategies (Cognitive, Metacognitive and Affective strategies) mediate the relationship between personality traits (Extraversion and Conscientiousness) and language performance, is supported. ||| ---- CPhD10_0001_915

The verbs *know* in E.g., 3.1.101 and *predict* in E.g., 3.1.102 are mental verbs semantically, and the projected clause *almost all the business email messages ...* in the former and *language learning strategies mediate ...* in the latter are all declarative propositions (statements) that truth value can be endowed with. The tense is retained in two projected clauses, with the active voice being introduced by diverse subjects compared with that in its projecting clause, making the clause nexus a reported mental proposition without any doubt.

E.g., 3.1.103 ||| Testing refers to the explicit attempt to **check*** || if one is still progressing towards the learning goals through paraphrasing, || trying to think of new examples, || schematizing searching an analogy, etc. ||| ---- CPhD11_0007_567

E.g., 3.1.104 ||| Then she **doubts*** || if it is because of Trench’s self-pride, || and this time she receives a dispreferred answer “Yes, Blanche: I am too proud”. ||| ---- CPhD17_0002_1933

These two examples, though are also reported mental propositions when the mental verb *check* and *doubt* are considered, present the condition of projecting indirect questions that the choice of *if* makes the proposition *one is still progressing ...* and *it is because of Trench’s ...* doubtful and whether the content being described is true or not needs further verification. In fact, the clause nexus in E.g., 3.1.103 is agnate with *...attempt to check: “Does anyone still progressing toward the ...?”* for covering a quoting mental proposition relation; and that in E.g., 3.1.104 can be interpreted as: *Then she doubts: “is it because of Trench’s self-pride?”*, presenting the same interrogative feature of the projected clause. There is no doubt that the answer to these two questions is a yes/no type, proving the broad scope of propositions being mentally projected.

Similarly, the main verbs for reporting mental processes in clause nexuses are listed:

		Proposition		Proposal
		Statement [indirect declarative clause: (that) ...]	Question [indirect interrogative clause: <i>whether/if; who, which, when...</i>]	[perfective non-finite clause, or modulated indirect declarative clause]
Cognitive	“like” type	<i>believe, guess, think, know, imagine, doubt, remember, forget, dream, predict</i>	<i>wonder, doubt; consider; find out, ascertain, check; determine, judge; predict</i> [interrogative/negative clause:] <i>know, remember [= the answer to the question] know</i>	
	“please” type	<i>strike, occur to, draw on</i>		
Desiderative	“like” type			<i>want, would like, wish, intend, plan for, hope for</i>

Table 11 Main verbs for introducing hypotactic reporting clauses in the mental process according to the speech function (adapted from Halliday, 2014, p. 517)

Except for the division of speech function, the character of these verbs is either cognitive or desiderative as the former refers mainly to those leading to propositions while the latter is suitable for those introducing proposals, though part of the cognitive verbs are “please” type instead of the “like” type, which mostly depends on the relation within the verb and the subject that the passive subject is more preferred by the “please” type, leaving the active subject a “like” one.

3.1.4.3 Summary

Therefore, no matter whether a clause is directly, paratactically quoted or indirectly, hypotactically reported, the projecting verb can be either a verbal or a mental one according to its semantics, and the projected clause is either a proposal for offering something to the receiver or demanding something from the receiver or a proposition for making a statement or raising a question. Normally, the quoting clause complex is frequently chosen in narratives such as literary works or in news reports for directly presenting words or sayings from speakers, and so does the literature review part in thesis writing that words or sentences are excerpted from scholars' interviews or books without any change, making the contents reliable and intuitional; the reporting clause complex, though is more welcomed in theses writing in diverse sections, indeed works as a carrier for writers in expressing their ideas through projecting authors' or scholars' meanings, which will be particularly explained in Sections 6.3 & 6.5.

Still, quite a few clause nexuses combine characters of quoting and reporting that structurally they resemble paratactic quoting with the quotation mark “ ” is deleted and replaced by a comma usually, but functionally are paratactic reporting that wording is represented as meaning by verbal verbs (like hypotactic locution) and meaning is represented as wording by mental verbs (like paratactic ideas) in this kind. The intermediate structure within direct and indirect speech is traditionally called “free indirect speech,” with the mood of the projected words remaining but time and person being changed. Though this category is unique, it seems to be ignored by students in thesis writing and will not be discussed with examples here. (ibid, p. 531-3)

3.1.5 Conclusion of SFG

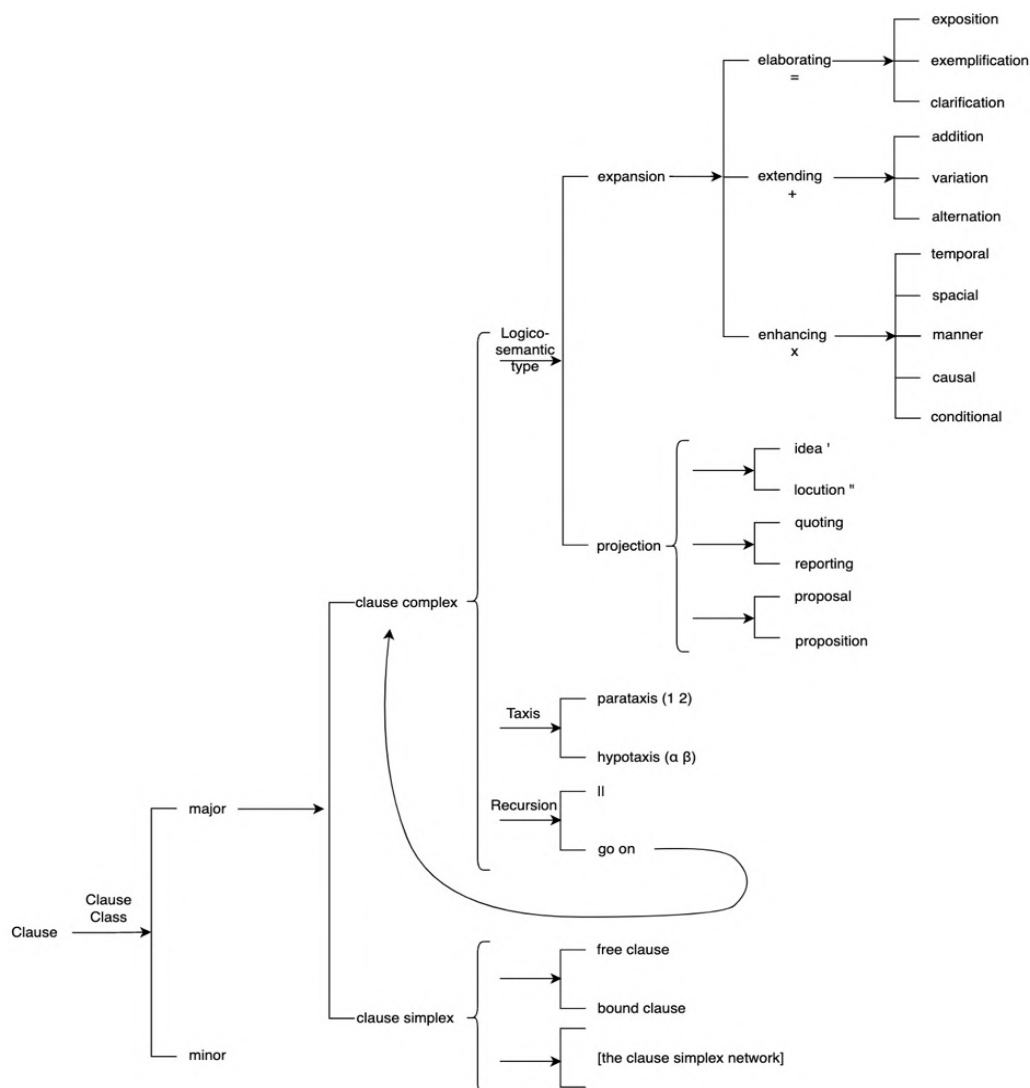


Figure 7 The analyzing framework of clause complexes in corpora for this thesis (from the SFG perspective)

After stating the originality of SFG with its contents, interpreting the reason why SFG is chosen as the theory foundation compared to other functional or related grammars, connecting the socio-functional effects of SFG on academic writing, and then elaborating on the main contents of SFG from eight groups of key concepts, i.e., Constituency; Structure, rank, and unit; System and delicacy; Function and metafunctions; Stratification and Realization; Instantiation; text, context and register; probabilistic and markedness, the focus of this thesis, i.e., the clause complex is drawn forth. This section also clarified the location of the clause and the clause complex in and around the SFG theoretical frame, explained the definition of two concepts, presented their organization, mentioned three meta-functions of the clause, and then made differentiation within them. Besides separating the parataxis from the hypotaxis, the logico-semantic types, i.e., the expansion and the projection, are delicately presented in their location, definition, organization, and application in examples in order to elucidate how texts are decomposed on the hierarchy of clause complexes and how clauses are combined or connected through kinds of relations into clause complexes, then to provide possibilities and feasibilities for discourse analysis in this thesis from the academic writing genre. The theoretical framework for analyzing the clause complex in the corpus is demonstrated in Figure 7 that excluding minor clauses, clause simplexes, and embedding clauses, the clause complex is annotated not only from the division of logico-semantic types but also from taxis, not only their intersections are considered, but also the recursion of clause complexes are discussed in following sections.

3.2 Academic Writing & SFL

The literature review of academic writing research in Section 2.1 has mainly revealed four areas: the focus on written products or materials by conducting text analysis, error analysis, contrastive analysis, contrastive rhetorical analysis including the critical oriented, corpus analysis, etc.; the focus on writing processes from cognitive operation models building, organizing strategies discussing, individual differences of learners exploring, periodic changes finding during the process, computer-assisted writing and research investigating, etc.; the effects of social and cultural perspectives on writing such as the social structure, register knowledge from learners, writing motives and needs of writers, native language transfer on writing, etc.; and the attention to teaching-related aspects such as teaching processes, teaching links in the classroom, online writing courseware or software development, writing tests on learners, learning strategies designing, promotion to learners' language levels, education on writing-related teachers, etc.

Corresponding theories for supporting these studies are comprised of 1. the “scaffolding” theory in writing teaching, which develops from Vygotsky’s (1978) “the zone of proximal development” theory and emphasizes that the teaching for writing should be assisted by teachers, textbooks, model essays, etc., and writing tasks should be assigned by stages and divided into small tasks, so that students can gradually get rid of the auxiliary “scaffolding” and learn to write independently; 2. the register theory in teaching writing, which is based on SFL and emphasizes the role of contextual knowledge and three dimensions for writing: field (determines that writing is based on common core vocabulary and structures), tenor (distinguishes oral from written language) and mode (suggests the mood and attitude in writing); 3. genre-based research on texts as genre, being a recognizable communicative event with distinct internal structures and a high degree of convention, would constrain the construction of discourse, hence the writer has to convey personal intentions or communicative purposes within the framework of a specific genre; 4. the social-cognitive perspective on second language writing, which proposes social-cognitive models for teaching writing, focuses on the dynamic interactive relationship among writers, subjects and readers, and regards the writing process as a cyclical nonlinear process as well as a persuasive activity; 5. as writing is a process of meaning discovering, which includes different stages with each stage being interconnected and intertwined, Flavell (1979) put forward the metacognitive writing strategy in investigating the regulation of metacognition on writing activities from three aspects, i.e., subject knowledge, task knowledge, and strategy knowledge.

Before investigating academic English writing from the systemic functional linguistics perspective, related research will be discussed from three distinct angles, namely genre, register, and interpersonal relationship. As this paper regards academic English writing as a register rather than a genre, the second section will specifically discuss the connotation of register, explain the reasons for considering AW as a register, and clarify what it means to treat AW as a register. The third section of this thesis will focus on the interpersonal relationship and construct a relevant framework from the Appraisal Theory (AP) and SFL for jointly analyzing the projection clause complexes qualitatively.

3.2.1 Combining AW with SFL

In general, the etic perspective would be adopted in the SFL theoretical paradigm for text analysis to systematically describe the choices of genre structure or language forms in texts with universal significance and explain them from relevant context by taking methods such as researchers' observation, interviews, texts analysis, etc. (Coffin & Donohue, 2012; Gardner, 2012) As the academic writing research under the SFL perspective aims to examine the textual features of the academic discourse community and discover the commonalities and differences of disciplines, three sub-paths are usually considered: the genre-related potential structure of academic discourse, the feature of register-related research, and the interpersonal meaning research.

The genre-related research mainly explores the macro-structural characteristics of various sub-genres in academic discourse, as well as the structural characteristics of different parts within the discourse. Specific issues to be investigated include the differences in genre structure in various disciplines, the complexity of genre structure in academic discourse within the same discipline; the effects of the writer's language and cultural background on the similarities and differences in genre structures, etc. Research objects of genre structure normally include the macrostructure of academic texts, almost every chapter of manuscripts and papers, and various genres, such as book reviews and editors' notes. Related research in China mainly focuses on specific parts such as the introduction, literature review, discussion, acknowledgment, etc., in theses.

The register-related study pays attention to the lexical and grammatical forms used in English academic writing and whether these language forms are unique to the discourse community or have disciplinary commonalities. Detailed forms for research can be academic vocabulary, shell nouns, result adverbs, linking adverbs, word clusters, chunks collocation, theme and rheme of clauses, reported clauses, conditional clauses, voice, tense, stance markers, and the overall language characteristics of the discourse.

The study of interpersonal meaning in academic discourse emphasizes the social, interpersonal, and dynamic meaning of language use and attempts to reveal interpersonal meaning characteristics containing the author's identity and stance attitude. Related questions that would often be asked can refer to how writers put forward their viewpoints in thesis writing; how to deal with the relationship between authors and readers; whether and how authors' different native language backgrounds, second language proficiency, and writing experience affect the interpersonal meaning of their academic texts. For detailed analyzing practices, although there are many theoretical paradigms and research methods for evaluation studies, is widely welcomed by discourse analysts because of its comprehensive and systematic analytical framework. Hood (2006) analyzed the expression of evaluation meaning from the attitude and graduation subsystems of the Appraisal Theory (Martin, 2000) in the introduction section of theses, and supplemented the implicit attitudinal meaning of the theory. At the same time, Hyland (2005b) proposed a framework for analyzing stance and engagement in discourse by examining how the writers confirm their new ideas through using language resources such as hedges, boosters, attitude markers, and self-mention and how the relationship between authors and readers in thesis writing is handled from the forms of reader pronouns, directives, questions, shared knowledge, and personal asides. Except for stance, engagement, self-mention, etc., in the Appraisal Theory, other forms include transitivity, metadiscourse, citation, and evaluative language features such as evaluation (overt or covert) strategies, rhythmic patterns, critical attitudes in book reviews, evaluation in manuscript reviews, editorial notes and editorial languages, etc. are also investigated in presenting the interpersonal meaning of academic discourse.

3.2.2 Seeing AW as a Register

In this thesis, academic writing discourse is approached as a register rather than a genre for text analysis. This section addresses three key questions: firstly, what is meant by the term "register" in the SFL perspective; secondly, why "academic writing" is treated as a register; and thirdly, what is meant by "taking academic writing as a register."

3.2.2.1 The Definition of *Register*

In Halliday's (1978a, p. 111) words,

"A register can be defined as the configuration of semantic resources that the member of a culture typically associates with a situation type. It is the meaning potential that is accessible in a given social context. Both the situation and the register associated with it can be described by varying degrees of specificity; but the existence of registers is a fact of everyday experience - speakers have no difficulty in recognizing semantic options and combinations of options that are "at risk" under

particular environmental conditions. Since these options are realized in the form of grammar and vocabulary, the register is recognizable as a particular selection of words and structures. But it is defined in terms of meanings; it is not an aggregate of conventional forms of expression superposed on some underlying content by “social factors” of one kind or another. It is the selection of meanings that constitutes the variety to which a text belongs.”

The definition is threefold: 1. Human beings and their related situations are connected by semantic resources, and specific social contexts indicate the potential for meaning, that is, the range of possible meanings that can be conveyed in the situation; 2. The register is realized by semantic options and the combination of these options in everyday experience, that is, the choice of grammar and vocabulary; and 3. The lexicogrammatical choice is determined by meaning in a specific situation. These come to the concept of *register*, which is therefore “conceptualized in terms of functional variation.” On the one hand, the language changes according to the context in which it is used, that is, “a given register is a functional variant of a language”; on the other hand, a given language is nothing more than the sum of its registers at a particular time in its development. (see Matthiessen, 2019)

Except for the influence of the context and lexicogrammatical choices in a specific situation, the variation of the register is also affected by the meaning potential of language and the instantiation that conveys the real-time meaning in texts in corresponding contexts of the situation. Hence, Matthiessen (ibid, p. 19 - 20) concluded that:

“In terms of the hierarchy of stratification, register variation is semantic variation according to context of use in the first instance; but since the relationship between semantics and lexicogrammar within the content plane is a natural one (rather than an arbitrary or conventional one), register variation is, by another stratal step, also variation in wording. In terms of the cline of instantiation, register variation is located within the mid-region between the two outer poles of the cline, the potential pole and the instance pole.”

This can be demonstrated by Figure 8 when two clines are crossed with the register variation being the intersection:

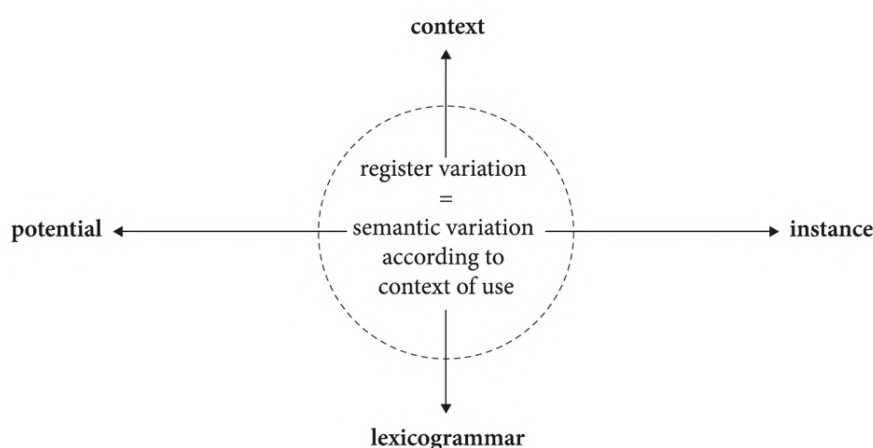


Figure 8 Register varies along the hierarchy of stratification within semantic variation and the context of use (functional variation) crossed by the cline of instantiation between the meaning potential and the instance pole

As variation indicates the probability of occurrence, register variation thus provides possibilities for quantitative analysis when the context changes or individual varies in a specific context of the situation.

3.2.2.2 The Reason for Seeing AW as a Register

The reason for seeing AW as a register originates from the “semogenesis” (the creation of meaning in language) or the three main processes of “semohistory,” “by which meanings are continually created, transmitted, recreated, extended and changed.” According to Halliday & Matthiessen (1999, p. 17-18), the three processes that change along the time refer to:

1. the phylogenetic, i.e., the development or the evolution of language(s) over time;
2. the ontogenetic, i.e., the development or the growth of the individual speaker’s language over time;
- and 3. the logogenetic, i.e., the development or the unfolding of the act of meaning over time.

As *logo* means “discourse,” “logogenetic” indicates “the instantial construction of meaning in the form of a text”, that is, “the potential for creating meaning” is constantly modified based on what has previously come although certain selections are “restricted or disfavored” while others are probable and “opened up” during the process.

Three processes are interrelated to each other as each step provides an environment for the happening of the “next” one (from 1 to 2 to 3), meaning that “the system of the language (the meaning potential of the species) provides the environment in which the individual’s meaning emerges, and the meaning potential of the individual provides the environment within which the meaning of the text emerges.” Conversely, each step “provides the material” for the construction of the former (from 3 to 2 to 1), denoting that “(finite) instances of text” construct into “the individual’s (transfinite) meaning potential”; and “(finite) instances of individual ‘meaners’” construct into “the (transfinite) meaning potential of the species.” (see *ibid*, p. 18)

Taking the creation of the content of this chapter as an example, on the one hand, the meaning of the content unfolds in a specific text, and the text is produced in the individual environment (me and this chapter) and in the environment of the evolution system of the English language (the register of academic English writing); on the other hand, the content of this chapter is an instantiation of the English academic writing, in providing the meaning potential for individual speakers (you, me, him and our works) and then contributing to the register of academic writing.

Therefore, by investigating texts and corpora produced in specific contexts of situation, the logogenesis, ontogenesis, and phylogenesis are probable to be discussed. Or just as Hoey (2005, p. 8) suggested: “As a word is acquired through encounters with it in speech and writing, it becomes cumulatively loaded with the contexts and co-texts in which it is encountered, and our knowledge of it includes the fact that it co-occurs with certain other words in certain kinds of context. The same applies to word sequences built out of these words; these too become loaded with the contexts and co-texts in which they occur.” In a word, the study of academic English text writing cannot be separated from considering the discourse as a register in a specific context of situation, to explore the differences in systems, individuals, and lexical and grammatical choices of the discourse.

3.2.2.3 The Meaning of Seeing AW as a Register

As discussed above, registers are described in terms of their “typical linguistic (lexical and grammatical) features,” i.e., being described from linguistic features that may appear “in any variety but are more common” in a specific register. But register is also influenced by the context of a situation, for example, whether it is produced in a spoken or written form, whether it is interactive, what the main communicative purpose is, etc. Since linguistic features tend to appear in registers for particular purposes and contexts of situations, another important “component of the register description is the functional analysis.” (see also Biber & Conrad, 2009, p. 6)

In the SFL framework, genre and register are considered to lay on different “semiotic planes” (Martin, 1985). Genre is regarded “as a social process in which participants in a culture” organize language into “predictable sequential structures” for certain communicative purposes; (Biber & Conrad, 2009, p. 22) just like what Couture (1986, p. 80) refers to as “conventional instances of organized text.” Register, however, is seen as the “expression-plane” of genre (Martin, 1985), concerning more about the typical linguistic features in diverse genres, which result from the variation of the context of situations being conceptualized as “field (the type of activity in which the text has significant function), tenor (the status and role relationships involved) and mode (the symbolic mode and rhetorical channels that are adopted)” (Halliday, 1978a, p. 122) in SFL. The genre perspective emphasizes traditional characteristics of the text as a whole, whereas the register perspective focuses on the “variation in the use of linguistic features” (see also Ferguson, 1994).

As academic English is a functional variety or a register variation of “scientific English” broadly, it can also be used to “denote a semiotic space within which there is a great deal of variability at any time as well as continuing diachronic evolution” (Halliday, 1988, p. 162). Although the diatypic variation has been summarized by Halliday (*ibid*) as

“in field, extending, transmitting or exploring knowledge in the physical, biological or social sciences; in tenor, addressing to specialists, to learners or to laymen, from within the same group (e.g. specialist to specialist) or across groups (e.g. lecturer to students); and in mode, by choosing phonic or graphic channel, most congruent (e.g. formal “written language” with graphic channel) or less so (e.g. formal with phonic channel), with variation in rhetorical function - expository, hortatory, polemic, imaginative and so on,”

seeing academic writing as a register, in fact, means writing a thesis or a dissertation for obtaining corresponding degrees as well as presenting what has been thought about or learned in several years in “field”; demonstrating the relationship within student (as a writer) and student fellows / teachers / supervisors / professors / professional scholars, etc., in “tenor”; and conducting literature reviews, theory clarification, methods explanation, data analysis, results discussion, limitation present, etc. through the written “mode.”

In correspondence with the situational characteristics of registers and genres from Biber & Conrad

in analyzing academic writing (see Figure 9) particularly in theses writing for getting an educational degree, the participants typically include writers and readers that the addressor is a single student from a university, with an age range of approximately 20 to 30 years old, an educational level of either graduated post-graduated or doctoral level, and a profession that is not specified; the addressees are described as “plural” and “other,” as readers may include fellow students, teachers, professors, or others.

-
- I. Participants**
 - A. Addressor(s) (i.e. speaker or author)
 - 1. single / plural / institutional / unidentified
 - 2. social characteristics: age, education, profession, etc.
 - B. Addressees
 - 1. single / plural / un-enumerated
 - 2. self / other
 - C. Are there on-lookers?
 - II. Relations among participants**
 - A. Interactiveness
 - B. Social roles: relative status or power
 - C. Personal relationship: e.g., friends, colleagues, strangers
 - D. Shared knowledge: personal and specialist
 - III. Channel**
 - A. Mode: speech / writing / signing
 - B. Specific Medium:
 - Permanent: taped / transcribed / printed / handwritten / e-mail / etc.
 - Transient speech: face-to-face / telephone / radio / TV / etc.
 - IV. Production circumstances:** real time / planned / scripted / revised and edited
 - V. Setting**
 - A. Is the time and place of communication shared by participants?
 - B. Place of communication
 - 1. Private / public
 - 2. Specific setting
 - C. Time: contemporary, historical time period
 - VI. Communicative purposes**
 - A. General purposes: narrate / report, describe, exposit / inform / explain, persuade, how-to / procedural, entertain, edify, reveal self
 - B. Specific purposes: e.g., summarize information from numerous sources, describe methods, present new research findings, teach moral through personal story
 - C. Factuality: factual, opinion, speculative, imaginative
 - D. Expression of stance: epistemic, attitudinal, no overt stance
 - VII. Topic**
 - A. General topical “domain”: e.g., domestic, daily activities, business / workplace, science, education / academic, government / legal / politics, religion, sports, art / entertainment, etc.
 - B. Specific topic
 - C. Social status of person being referred to
-

Figure 9 The framework for analyzing situational characteristics of registers and genres (Biber & Conrad, 2009, p. 40)

The relationship between the writer and their supervisors is intricate when the writer seeks to engage with his/her supervisors who possess shared specialist knowledge and a relatively higher status in determining the quality of his/her written texts. The mode is, of course, writing, though a defense (speech) is a necessary component in obtaining the degree, and it is the permanent channel of printed products that the knowledge can be saved and spread. The production circumstances are planned, revised, and then edited before publication. Usually, the time and place of communication would not be shared by participants as the place for writing is private, and the time lasts for three or more years for an individual writer. The communicative purposes of thesis writing are multiple, as listed by Biber & Conrad, general purposes include narrating or reporting research results, describing the research methods or experimental data, expositing/informing/explaining the diversity of data outcomes, persuading readers of the reasonability of designing the research, demonstrating the experimental procedures, and edifying other researchers to theoretically or practically improve their works, etc.; specific purposes include summarizing information and presenting new research findings. In experimental research in hard disciplines, writers must be factual in their descriptions and research designs, speculative in their results, and imaginative in their analysis of peculiar results; but in theoretical research in soft disciplines, writers may either agree or disagree with the views of previous scholars, particularly when sources are cited based on their epistemic modality, attitudes, overt or covert stance, and so forth. In general, the topic of the thesis is educational or academic with scientific data analysis, though these are suitable for every discipline, and the social status

of the person being referred to is normally higher than that of the writer, as the person must be a scholar who has published articles in a related research field.

3.2.3 The Qualitative Analyzing Framework for the Projection CC (from AP & SFG)

From the perspective of the interpersonal meaning, the influence of SFL on academic English writing research can be discussed from the Appraisal Theory (AP). Before describing its specific analytical framework, the definition of “appraisal”, the originality, and the purpose of AP theory will be briefly introduced. Although the theoretical framework is quite applicable, only a part can be borrowed to study clause complexes in this study, especially for the projection part. In taking the particularity of projecting verbs into consideration, after referring to Thompson & Ye’s (1991) study on the evaluative meanings of projecting verbs, a new analytical framework will be proposed in this section to examine the complexity of expressing the interpersonal meaning in projection clause complexes in academic English writing. Research suggests that through the choices of Theme/Rheme and projecting verbs in projection, writers can show their attitudes toward the projected propositions or proposals on the one hand; and can construct corresponding relationships with authors and readers on the other hand.

The rest section will be published in the following journal article:

Projecting Verbs, and the Corresponding Analysing Framework Construction in the Appraisal Theory: from the *Chinese Academic English* (ChAcE) Corpus

To summarize, the whole framework is organized by an author-writer-reader cline, crossed by Clines 1-4, as shown in 错误!未找到引用源。 , in which Cline 1 and Cline 3 are related to each other in practice, and Cline 2 and Cline 4 are affected with each other in practice.

By taking the writer’s perspective as the center, this framework can comprehensively analyze the writer’s choice of projecting verbs in English academic writing discourse to present the construction of interpersonal relationships with the author and the reader. In practice, projections are firstly categorized from the analyzing framework of the clause complex system; and then from the “cline-crossing” framework described above, at the same time considering the influence of subject/tense/voice/modality on reporting verbs. Detailed examples can refer to Sections 6.3 & 6.5.

Although the analyzing and classification or categorization in practice would also be subjective in a sense because of the unclear integration of some basic linguistic features such as tense, voice, modality, etc., into the framework; the hard judgment or ambiguous boundary of subjectivity from objectivity sometimes; the dynamic instead of static classification of projecting verbs into four clines; and the dichotomy of trichotomy of the taxonomy of each cline, the framework is applicable in general.

3.3 Corpus Linguistics & SFL (Probabilistic)

As previously stated, in taking language as a paradigmatic system, SFL is “a set of choices for each instance from which a speaker must” take one. This implies that it is “inherently probabilistic” in nature, whereby “in each situation,” various categories are “more or less likely” to be chosen by the speaker. (see McEnery & Wilson, 2001) In light of these considerations, Halliday proposed that priority be accorded to studying naturally occurring language and counting corresponding frequencies. In the SFL system, a corpus is denoted as “the result of sampling text systematically according to explicit criteria,” which is “always located at the instance pole of the cline of instantiation.” (Matthiessen, 2006, p. 104) This is why Halliday posits that “the transformation of instance into system can be observed only through the technology of the corpus,” allowing for the accumulation of instances and the monitoring of “the diachronic variation in their patterns of frequency.” (Halliday, 2005a, p. 67) Consequently, SFL is distinctive in that “it provides an explicit, multilayered, and detailed model for explicating a set of texts in relation to the social context they construe.” (Hunston, 2013) The corpus-related approach to grammar, however, dovetails with part of the SFL theory in that corpus linguistics facilitates the quantification of linguistic data across a vast corpus of texts. Furthermore, it elucidates the pivotal role of lexis and phraseology in text organization, informing “models of language organization and meaning” that diverge from those proposed by SFL (Hoey, 2005; Sinclair, 2004; Teubert, 2003). (see *ibid*)

This section mainly examines SFL research from the perspective of CL. After a brief description of the history of the overlap between the two fields, the corresponding connections and differences will be clarified. In addition, the general methods used in the research and some related directions are mentioned. Since certain relevant concepts, such as “instantiation,” “choice,” “probability,” etc., are unavoidable in SFL research from a CL perspective, their connotation and extension are explained. Although CL’s

methods of dealing with corpora can be absorbed by SFL research, some issues, such as the representativeness of the data, the implication of implicit grammar in the data, and the annotation of higher-level grammatical linguistic features in the corpus, require more time and effort to solve. Therefore, CL is mainly considered as a research perspective in this project to support the theoretical framework of the SFL system in analyzing language features.

3.3.1 A Brief History of the Intersection of CL & SFL

Historically, SFL has been employing “textual materials and corpus-based methodology in investigating language and language use” for a long time.” (Wu, 2009, p. 128) This practice can be traced back to Halliday’s early work in 1949 when he built a corpus from natural text samples written in Cantonese for “studying the grammar of the dialects of the Pearl River Delta in southern China” (see Halliday, 1992a, p. 76). However, the concept of “collocation,” proposed by J. R. Firth (1957 [1951], p. 11) from the famous saying “you shall know a word by the company it keeps,” provided a new perspective for corpus research. Subsequently, in his doctoral dissertation *The Secret History of the Mongols*, Halliday (1959) presented “a corpus-based description of word classes, syntactic constructions, collocations and colligations” of Chinese grammar. This description elucidates “the relations between words and grammatical patterns of their use,” thereby providing a foundation for the first systematic approach to grammatical analysis (Halliday, 1957). (Sharoff, 2017, p. 534) “Starting with the analysis of Chinese and English grammar” by manually calculating “very small samples” (Halliday, 2005a, p. 64; 2005c, p. 132-5), Halliday has been modeling the probability of language. He weighed the “feature choices in networks” (i.e., assigned probabilities to grammatical choices) to represent how often “they occur in the text.” This approach allows the establishment of “the probability profile of any grammatical system” (Halliday, 2005a, p. 67) that “can be undertaken for the overall language system (“global probabilities”) or different registers (“local conditioned probabilities”) (ibid, p. 70),” where register variations are viewed as “systematic variation in probabilities” (ibid, p. 66). (see also Bednarek, 2010)

John Sinclair’s (Halliday’s doctoral student) project on English lexes, which began in 1963, pioneered new developments in corpus studies and turned in the 1970s for a study of large corpora for lexicographics. Since then, a series of studies that apply corpus research methods to SFL have emerged, such as Plum & Cowling (1987) adopted “(non-representative, elicited) interview data to study the probabilistic variation of tense and recursion depending on contextual features such as class, age, gender”; Nesbitt & Plum (1988) used “123 narratives elicited from sociolinguistic interviews” to analyze “probabilistic realization and context” in “clause complex relations”; Halliday & James (2005, originally 1993) selected “an 18 million word sub-corpus of the Bank of English to investigate polarity and primary tense”; (Bednarek, 2010) other related studies like Hasan & Cloran (1990); Souter (1990); Halliday (1992a; 2002b); Tucker (1998); Matthiessen (1999), etc. With the advancement of corpus technology, an increasing number of systemic linguists “(e.g., Pagano et al., 2004; Herke-Couchman, 2006; Matthiessen, 2006; Patpong, 2006; Teruya, 2006; Wu & Fang, 2007) have begun to take full advantage of available computer-readable textual data, and use them in a wide range of areas such as language description, discourse analysis, language acquisition, translation studies and natural language processing.” (see Wu, 2009, p. 128)

3.3.2 The Connection of CL & SFL

According to Halliday (2006a, p. 293), there is “a natural affinity between systemic theory and corpus linguistics.” Since SFL relies on naturally occurring language, it constructs language description modes based on language instances in the form of texts (Halliday, 2009b, p. 63) and probability (ibid, p. 69), which are also concerned about in CL (McEnery & Wilson, 2001). As to the corpus, as a text, “[R]ealization comes in because what becomes accessible to us is the text as realized in sound or writing. We cannot directly access instances of language at higher strata - as selections in meaning, or even in wording” (Halliday & Matthiessen, 2004, p. 33). Which means that some corpus searching is “restrictive in the sense of being tied to formal realisation” (Hansen-Schirra et al., 2007, p. 246). However, by automatically, semi-automatically, or manually preprocessing the data in CL, corpus researchers can have access to strata of language, just as Matthiessen (2006, p. 109) described, and then modified by Bednarek (2010, p. 240):

by searching for “either single words or expressions involving combinations of words” realized in distinct graphological units in a “raw” (i.e., unannotated) corpus (graphological stratum);
by searching for closed-class lexicogrammatical systems, either in raw corpora or in part-of-speech tagged or grammatically parsed corpora (lexicogrammatical stratum); and

by searching for “discourse semantic systems in previously semantically annotated corpora (discourse semantic stratum).”

Therefore, “corpus studies underpin the general principle of functional variation in language; they make it possible to quantify the lexicogrammatical differences among different registers, and to interpret this kind of variation as a redistribution of probabilities.” (Halliday, 2006a, p. 294)

In analyzing texts specifically, there is a genetic connection between SFL and CL. The initial aspect relates to the emphasis on social semiotics in SFL. As “language is a social phenomenon,” it is reflected “in the production of speech and text” (Halliday, 1978a). In light of this, SFL advocates the system-process-product approach, which posits that “each text and each linguistic unit in the text is the product” of various text production processes, and is an instantiation of multiple choices that may appear in the English system. The selection of these options is contingent upon their utilization within the context of the academic chapter register. The second aspect is the focus on variation. When speakers utilize language to fulfill their “communication goals,” “the lexicogrammatical properties of their speech would vary according to the context of their use,” thus leading to the concept of “register” in denoting this context-related variation. Another change comes from the variation of users in choosing lexicogrammatical patterns. Identically, the variation of frequencies and patterns is also core in CL, as well as the comparison of these two uses for bringing in illuminating results. (cf. Sharoff, 2017, p. 534)

The lexicogrammatical choices of grammar under SFL framework in annotating corpora and the lexicogrammatical patterns in CL by only mentioning related “delicated grammar” in SFL framework indicate the diversity within two perspectives, as the former is treated as “doing SFL using corpora” while the latter is considered to be “doing corpus linguistics.” (see Hunston, 2013, p. 619) Another diversity is that SFL focuses more on paradigmatic (choices) rather than syntagmatic (structural) relations (Nesbitt & Plum, 1988, p. 7), and its system networks mainly show the selection of the former; while CL treats the two as equally, as Sinclair (2004, p. 170) suggests: CL “works on the assumption that meaning is created on both axes; . . . that they contain equal meaning potential. There is no reason why one should have a priority in meaning potential over the other.”

3.3.3 Related Research and Methods of CL & SFL

In light of the aforementioned similarities between SFL and CL (the study of textual variations, the examination of the clear connection “between the linguistic configuration of a given text or group of texts and the situation of their production and reception”; the view of “register depends on the relative frequency of linguistic features”) (Hunston, 2013, p. 619) and the differences (“SFL prioritizes grammar” while CL prefers to lexis), related research can be roughly divided into three directions: studying “a common topic (register) but along different lines”; studying the “interaction between grammar and lexis” by using methods that either diverge from or complement to each other; unifying the two perspectives and “using methods and/or concepts from each.” (see *ibid.*, p. 636)

An increasing number of studies indicate that by connecting the quantitative information received from software in investigating corpus with the powerful explanation under SFL theory, CL and SFL are complementary (see Gardner, 2008, p. 30; Bednarek, 2010, p. 249). Notable research includes genre-based studies conducted on the BAWE corpus (e.g., Gardner, 2008; Gardner & Holmes, 2009; Nesi & Gardner, 2012); studies concentrated on participant and process (e.g., Goatly, 2004; Moreton, 2007; Holtz, 2007; Mulderrig, 2011); on Appraisal Theory (e.g., Flowerdew, 2003; Miller, 2006; Bednarek, 2010); (see Hunston, 2013, p. 623) others such as: Coffin & O’Halloran (2006); Thompson & Hunston (2006); Kaltenbacher (2007); Bednarek (2008) with respect to interpersonal meaning), etc.

For CL, the categories constructed by SFL serve for “describing the co-text of key terms” (e.g., Mulderrig, 2011; Holtz, 2007), or “providing a theoretical framework” for pattern research (Bednarek, 2009). Conversely, as Halliday & Matthiessen (2004, p. 34) posit: “[t]he corpus is fundamental to the enterprise of theorizing language.” The impact of corpus methods on SFL research is twofold: “annotating corpora with SFL categories to obtain relative frequency data (e.g., Matthiessen, 2006), or using word lists and concordancing software to identify specific processes and participants and their co-texts (e.g., Goatly, 2004).” The influence of corpus linguistics findings, particularly in the domains of “lexis and phraseology”, on SFL research is evident in “the incorporation of phraseological information in computer models of SFL (e.g., Tucker, 2006), Halliday’s (2008a) discussion of complementarity,” and other notable instances. (cf. Hunston, 2013, p. 618-9) Moreover, scholars engaged in theoretical studies relating to the combination of CL and SFL have sought to “incorporate notions of co-selection into SFL modeling through several novel concepts including: ‘pre-selected’ paths through system networks in the Cardiff Grammar (Tucker, 2006); through the intersections of simultaneous systems (Matthiessen, 2006), particularly on ‘collocational patterns’ (Matthiessen, 1998); and through notions such as ‘bonds’ (Knight,

2008) and ‘couplings’ (Zappavigna et al., 2008), ‘syndromes’ (ibid) and clustering (Bartsch et al., 2005).” (Bednarek, 2010, p. 240)

3.3.4 Related Key Concepts

Influenced by the CL perspective, some concepts that are bound to be mentioned in SFL research, such as *lexicogrammar*, *instantiation (cline)*, *choice*, *probability & frequency*, *potential*, etc., although some of these concepts have been elucidated in Section 3.1.

Lexicogrammar

One perspective holds that lexis is a more delicate form of grammar (Hasan, 1987), that is, “the grammatical systems specify more general configurations, while the lexical choices are specified in more delicate systems following the grammatical choices.” (Sharoff, 2017, p. 537) Sinclair (1991) puts forth a view of lexical grammar that asserts the absence of a distinction between lexis and grammar, which means that lexis and grammar shall not be effectively studied separately because of their intertwinement, as specific lexical items are associated with particular patterns, and certain patterns encompass some kinds of lexical items. SFL research conducted from the lexical end starts with “lexical patterns and generalizes them into more grammatical phenomena.” Investigating a corpus from the lexical end is much easier than that from the grammatical end because “word forms can immediately accessible to a corpus queries,” which makes the creation of “concordances, frequency lists, and collocation lists” probable. When related information is obtained, “any regular patterns of use” can be categorized. Despite the inherent limitations of POS tagging, particularly its lack of reliability sometimes in producing rough “models of grammatical functions, basic colligation patterns” can still be examined when “relatively simple queries” are asked. “Statistical corpus-based” research from the grammatical end is challenging due to the reliance on “large annotated corpora (in which grammatical phenomena have been automatically annotated),” while a considerable part of grammar cannot be automatically annotated because of various restrictions. (cf. Sharoff, 2017, p. 537)

Instantiation (cline)

As “[T]he system of a language is ‘instantiated’ in the form of text” (Halliday & Matthiessen, 2004, p. 26), the relation within system and text is mutually dependent: while a text ‘is an instance of an underlying system, and has no meaningful existence except as such’ (ibid), language as system can only be observed by text; it is “a virtual thing; it is not the sum of all possible texts but a theoretical entity to which we can assign certain properties and which we can invest with considerable explanatory power.” (ibid, p. 27) When observing language, “we can locate ourselves at any point along the line of instantiation.” This does not imply that the action is confined to a particular area, rather, it suggests that we are free to some extent “to move around” and make use of “resources along the way.” The object of the study can be a single text, which allows for an examination of “how a particular language system is instantiated” in that text; or that “beyond a single text” for being investigated to identify “recurrent patterns” in a bundle of texts with the aim of characterizing a text type, or the whole linguistic system, which is a typical approach for corpus-based research. (cf. Wu, 2009, p. 131) Just as what Matthiessen (2006, p. 105) summarized:

Systemic profiles of particular texts reveal logogenetic patterns of texts as process-instantial patterns in the creation of meaning through the instantiation of the system, and also the accumulation of patterns of texts as a product.

Systemic profiles of text types reveal patterns that are intermediate between instancial patterns in text and systemic patterns in the overall potential of language. These patterns can be interpreted either as macro-logogenetic or as micro-phylogenetic.

Systemic profiles of the overall systemic potential reveal patterns inherent in the linguistic system itself - patterns that evolve as part of phylogenesis.

Corpora, which comprise “systematic collections of textual instances,” is “located at the instance end of the cline of instantiation.” (Wu, 2009, p. 131) Consequently, they can be moved along the cline of instantiation from specific instances to generalize groups of instances within a register, patterns of linguistic systems, or even the whole systemic potential.

Choice

The connotation of *choice* is twofold: one is an “option” in a system, and the other is the act of choosing any option in a system network, i.e., “selection.” Being an “option,” its connotation and denotation are usually determined by the system: “(1) what the option realizes (“from above,” its signification), (2) what the option is realized by (“from below”), and what other options the option contrasts with (“from roundabout,” systemic value).” (Matthiessen et al., 2010, p. 69) While being a “selection,” choice partly describes the process of traversing a system network, whereby selections are made along the network. In the framework of SFL, the fundamental units for quantifying analysis are

the choices within each system and their realizations by lexicogrammatical features. As a corpus typically demonstrates information at the level of linguistic forms, the choices on delicate levels of the system can thus be readily identified; while with regard to those beyond the level, annotation must be conducted on the corpus in order to investigate generalized choices.

Probability (systemic) & Frequency

In the system network, when a system has no marked choice, the probabilities of these choices are roughly equal; and when it has one unmarked choice, the probability of this choice of course has a considerably higher probability than the other one (marked choice). As Tucker (2006, p. 101) points out, “[t]here is . . . a mixture of obligatory and optional co-selection, and, where there are options, they are highly restricted, and probabilistic.” Halliday (1992a, p. 87-88) makes a similar distinction, i.e., between “transitional” and “conditional” probabilities, in which the former “are linear, with the first choice always constituting the environment for the second”; and the latter indicating that “the two choices are being made simultaneously, and so either can be treated as the environment for the other.” In giving an example with marked probabilities (“polarity”: positive (0.9) / negative (0.1)), Halliday (ibid, p. 65) expresses that a 9:1 “distribution of the probabilities of marked and unmarked choices is relatively common in the language system because it allows for an appropriate balance of processing effort in terms of the information-theoretic entropy principles” (Sharoff, 2017, p. 535):

“The skew value of 0.9/0.1 seemed rather an unmotivated artefact of decimalism, until I noticed that a possible explanation for it could be found in information theory...A system of probabilities 0.5/0.5 (“equiprobable”) is of course minimally redundant. The values 0.9/0.1 incorporate considerable redundancy; but this is just the point at which redundancy and information balance out. In a binary system, H (information) = R (redundancy) = 0.5 when the probabilities are 0.89/0.11. It seems plausible that the grammar of a natural language should be constructed, in outline (i.e., in its most general, least delicate categories), of systems having just these two probability profiles; rather than, say, having all systems equiprobable, which would be too easily disrupted by noise, or having systems distributed across all probability profiles from 0.5 / 0.5 to 0.99 / 0.01, which would be practically impossible for a child to learn.” (Halliday, 1992a, p. 81)

Regarding the relationship between probability and corpus, Halliday (ibid, p. 82) stated: “the larger the corpus from which we obtain our picture of the frequencies, the more accurate our account of the system is going to be.” And “frequency in the corpus is the instantiation (note, not realization) of probability in the grammar.” (ibid) When looking specifically at system probabilities in the corpus, the relevant questions for investigation may be:

“The overall relative frequencies of the terms in a set of low-delicacy grammatical systems; the same grammatical systems with the frequencies broken down according to register; whether, and if so to what extent, the probability of selecting one term in a given system is affected by previous selections made within the same system; the tendency for complexity to increase in the course of the text; seek from the corpus information about the degree of association between simultaneous systems; the investigation of conditional probabilities can provide insight into historical linguistic processes; the choice between choosing and not choosing,” etc. (ibid, p. 83-91)

In general, the importance of probability lies in “the frequencies observed in text both reflect the system (at a given point in time) as well as shape the system (over time).” (Plum & Cowling, 1987, p. 284)

Potential

The term *potential*, as usually seen in “meaning potential,” concerns the possibilities that a speaker/writer can do with language, as opposed to the “actual” in the sense of what the speaker/writer actually does with language. In terms of the relationship between potential (system) and instance (text) on the cline of instantiation, “meaning potential is instantiated in text as instantialised potential.” (Nesbitt, 1994, p. 54)

3.3.5 Disadvantages of SFL in CL Perspective

“Work based on corpus studies has already begun to modify our thinking about lexis, about patterns in the vocabulary of languages; and it is now beginning to impact on our ideas about grammar. In my view, this impact is likely to be entirely beneficial. Corpus linguistics brings a powerful new resource into our theoretical investigations of language.” (Halliday, 2005c, p. 130) From the grammatical perspective, corpus linguistics enhances the ability of researchers to systematically study language variation in large numbers of texts. By identifying common and uncommon choices, patterns that are typical and untypical in particular contexts can be revealed. As Biber et al. (2004, p. 376) explain, “The usefulness of frequency data (and corpus analysis generally) is that it identifies patterns of use that otherwise often go unnoticed by researchers.” In corpus-based grammar research, explanations of frequency-related analysis can depend on various aspects of linguistic theory, including judgments

about the factors that influence language choices and the unconscious motivations of speakers/writers in making those choices. Thus, corpus analysis allows us to describe grammar not only from a structural perspective but also from a probabilistic perspective, in “describing the typical social and discourse circumstances associated with the use of particular grammatical features.” (O’Keeffe et al., 2007)

Nevertheless, the results of the statistical analysis conducted from a CL perspective cannot be considered entirely reliable, as the representativeness of the data, its composition, and the statistical methods employed will inevitably exert a certain influence on the outcomes. Moreover, some minor or even negative linguistic features (which may be significant for the text) are not given sufficient consideration. This may result in individuals failing to acknowledge the significance of these factors based solely on the numerical data presented, which may not fully capture the nuances and complexities of the situation. This also refers to the obtaining of evidence from corpora. As “corpora do not provide negative evidence,” namely, confirm that a certain language phenomenon has not occurred, “negative evidence can be inferred only from a non-significant number of examples if positive evidence on a large corpus is overwhelming.” (Sharoff, 2017, p. 542) While positive evidence demonstrates the existence of a phenomenon, it is nevertheless essential to consider its usage conditions and to make comparisons with those from other contexts of use. That is, the number of positive examples itself is not a criterion; it is only meaningful when compared with another value. Consequently, in addition to comparing explicit language features, some implicit features that cannot be statistically analyzed at all need to be manually investigated and discussed.

While CL may assist in resolving issues relating to lexis or phrases in SFG, it is unlikely to offer significant insights into grammatical or even higher-level analysis. If a suitable measurement of grammatical complexity exists, corpus queries of complex linguistic features in a multitude of articles can serve to test this hypothesis. Tucker (2006, p. 102) posits that the combination of corpus linguistics and SFL provides a supplementary lens for interpreting social semiotic processes, as it clarifies linguistic patterns across the corpus. Furthermore, the annotation (manual, semi-automatic, or fully automatic) of corpora and the subsequent “querying and interpretation of the results” offer additional sources of linguistic information in supporting and refining language models in SFL. (see Holz, 2011, p. 39) However, since corpus-based phraseology and collocation studies are less dependent on complex structure-related search mechanisms, they have progressed at a considerably faster rate than corpus-based grammar studies. As Gilquin (2002) observed, the dearth of accurately annotated treebanks has had a detrimental impact on corpus grammarians, who tend to gravitate towards topics that can be studied through relatively straightforward corpus queries, rather than those necessitating a high degree of abstract structure. Nevertheless, a considerable number of grammatical studies conducted using corpora rely on the identification of semantic or pragmatic features, which frequently necessitates the manual annotation of individual texts by grammarians. (see Leech, 2015, p. 149) This process presents significant challenges to research. Just as Matthiessen (2006, p. 141) posits, “the more significant a system is in the organization of the lexicogrammar of a language, the harder it will be to automate the analysis.” In other words, “analysis can be automated at the stratum of graphology, but the automation of the analysis becomes increasingly difficult as we ascend the stratal organization (from graphology to lexicogrammar and then) towards semantics.” (ibid, p. 109). Teich (2009, p. 115) similarly points out that while there are tools to support manual SFL analysis, “a fully automatic annotation with systematic functional categories is not possible at this stage” as “computational analysis tools cannot yet cope with the combination of rich analysis and a flow of registerially unrestricted text.” (Matthiessen, 2006, p. 141) Teich & Fankhauser (2005) observe that while there are numerous tools for lower levels of analysis, automatic annotation is often infeasible or only exists for specific higher-level analyses (such as rhetorical structure or cohesion). Additionally, they note that there are numerous unresolved problems in processing corpora at the text level. This represents a significant challenge for those engaged in corpus-based research in the field of SFL, particularly in relation to the investigation of higher-level meanings. Consequently, SFL research has only partially employed corpus linguistics methodologies, and thus far, there have been few studies utilizing extensive corpora.

Although other modes of syntactic extraction except for annotation, such as the development of small-scale customized programs for investigating specific grammatical features in texts (see Biber et al., 1999), or “extending the functionality of search software by providing the ability to use patterns or templates built based on regular expressions and/or CQP query syntax (see Hofmann et al., 2008)” (cf. Leech, 2015, p. 150) or machine learning methods such as “supervised machine learning” (starting with a certain amount of data with the desired annotation level as an example to annotate more similar data to generate “automatic annotations of sufficient quality”) or “unsupervised machine learning” (text without the desired annotation level, but is still used of “some more basic linguistic features for

inferring statistical regularities at the desired annotation level”) can provide more annotated corpora for SFL studies by annotating texts more finely than the traditional POS tags and lemmas in corpus studies, improve the availability of annotated texts at the linguistic level, and thus offer statistically significant evidence for the distribution of choices and its realization in language, (cf. Sharoff, 2017, p. 543) the specific practical operations involved are highly complicate and require a very high level of theoretical expertise and computer practical knowledge from linguists. Consequently, more endeavors are imperative to attain these objectives.

3.3.6 Summary

To summarize this section in Halliday’s words:

“A language is a meaning potential, one that is open-ended; the grammatics has to explain how this meaning potential is exploited, and also how it can be enlarged. And this is where I see a complementarity between systemic theory and corpus linguistics. This is not a complementarity of theorising and data-gathering: systemic linguists have always tried to base their descriptions on observable data, while some corpus linguists have proclaimed themselves ‘mere data-gatherers’ (not without a touch of disingenuousness since I do not think they were really disparaging their own work!), data-gathering is never theory-free, and collecting, managing and interpreting corpus findings is itself a highly theoretical activity.” (Halliday, 2006a, p. 295)

In this thesis, corpus linguistics is considered as a way of looking at language data and a perspective for processing language data. A large number of texts are manually annotated under the theoretical framework of SFL to conduct perform statistics and mutual comparison, explore the similarities and differences between texts, and find the significance of language choice and realization.

3.4 Theoretical Framework for the Study

Through the explanation of the reasons for choosing the SFL theory for this research, the distinction between the concepts of clause and clause complex, the description of the connotation and denotation of the components of the clause complex: expansion and projection, and the presentation of the core annotation framework of the clause complex in the first section; the discussion of the role of English academic writing related contents in this research, the investigation of its intersection with SFL, i.e., the Appraisal Theory, and the construction of qualitative analysis framework for projection in chapter six in the second section; and the clarification of the advantages and disadvantages of SFL research and its solutions from the perspective of CL in the third section, this chapter constructs a comprehensive theoretical framework based on the relevant theories of SFL, AW, and CL to serve this paper:

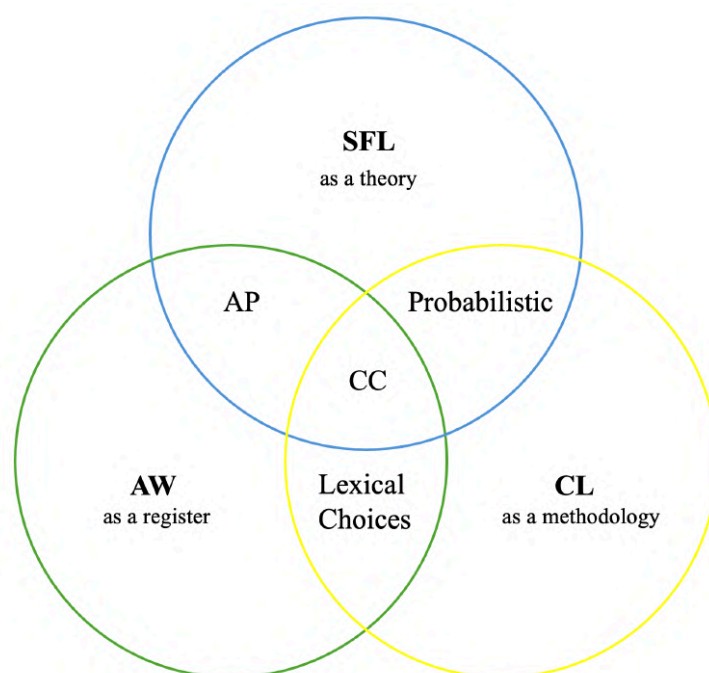


Figure 10 The theoretical framework for analyzing the clause complex system in this thesis under relevant theories of SFL, AW, and CL

As Figure 10 demonstrates, SFL is mainly seen as a theory in which the theoretical framework of the clause complex system is absorbed in this project for corpus annotation, and part of the theories, being relevant to the grammatical-lexical description of “expansion,” are applied in qualitative analysis in chapter five. AW is regarded as a register for text analysis, and CL is considered a methodology for corpora building, corpus annotation, and comparative analysis of data from the corpus. The intersection of SFL and AW, i.e., the Appraisal-related studies, is used for constructing a qualitative analyzing framework for the projection part in chapter six; SFL under the CL perspective pays attention to the probabilistic regularities for conducting quantitative analysis for grammatical features of the clause complex as well as qualitative analysis for lexicogrammatical patterns that have been rarely focused on; and AW under the CL perspective will be centered on “choices,” i.e., the qualitative and quantitative analysis of lexical linguistic features in the corpus. Therefore, the theoretical support for the study of clause complexes in this thesis is inseparable from Systemic Functional Linguistics, English Academic Writing relevant theories, and the perspective of the methodological improvements in Corpus Linguistics.

As Halliday (2005a, p. 67) argues: “the transformation of instance into system can be observed only through the technology of the corpus, which allows us to accumulate instances and monitor the diachronic variation in their patterns of frequency.” Corpora on the one hand, is always located at the instance pole, allowing researchers to identify patterns, or “systemic profiles” (Matthiessen, 2006, p. 104) at different points along the cline of instantiation; on the other hand, allow us to infer meaning potentials along the cline of instantiation, i.e., to describe “systemic profiles of particular texts, systemic profiles of text types” and even “systemic profiles of the overall system potential.” (ibid, p. 105) Thus, this is another important point for the emphasis of the corpus in choosing different kinds of instances, ranging from the individual text to a particular corpus and then to a general corpus (being maximally representative of the language as a whole), to make predictions about the meaning potentials of texts, registers/genres and even the whole language system.

Chapter 4 Data & Methods

4.1 Data (ChAcE) Description

Under the advice and guidance of my supervisor, Prof. Dr. Schmied, and with help from two colleagues in our English for Academic Writing research group, Xinlei Zhang and Jessica Dheskali, the *Chinese Academic English (ChAcE) Corpus* was built up at the beginning of the year 2016 as a database for AW research. The collection work should be attributed to Zhang, who spent more than a quarter downloading MA theses and PhD dissertations from websites such as *CNKI*¹⁶ and *Duxiu*¹⁷.

Data Collection Processes



Figure 11 The homepage of the Chinese version website CNKI¹⁸

After opening the homepage and clicking the MA/PhD button, a keyword such as “Semantics” can be input into the column, and then a list of theses and dissertations (a total of more than 20 thousand results can be found) is shown underneath:

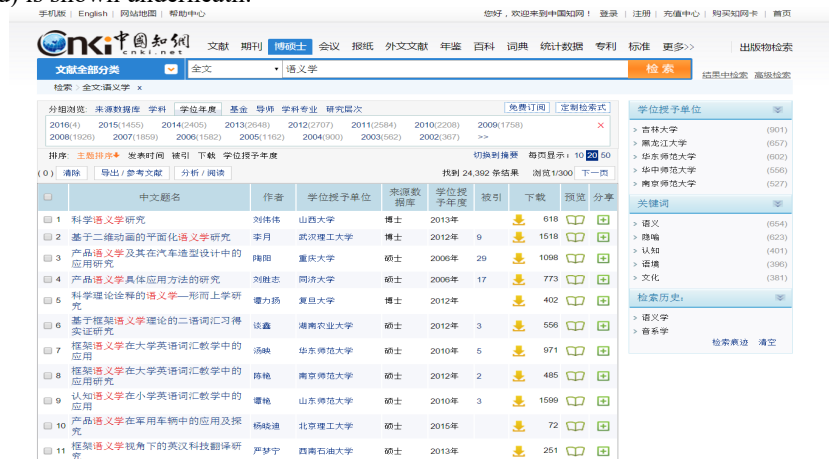


Figure 12 The page that shows search results with presumptive keywords or phrases

¹⁶ An important comprehensive China Integrated Knowledge Resources System (including journals, theses, dissertations, newspapers, E-books, etc.) under the instruction of Tsinghua University, supported by the PRC Ministry of Education, PRC Ministry of Science, etc. It was first launched in 1996 in CD version (China Academic Journals full-text Database) and then has developed an online version since 1999. Now it has become the largest and most used academic online library in China. (Description on <https://en.wikipedia.org/wiki/CNKI>, 2016.07.01)

¹⁷ Being a huge knowledge system composed of a large number of literature resources such as books, periodicals, newspapers, conference papers, dissertations, etc., the Duxiu Academic Search Engine (<http://www.duxiu.com>) is a platform that can conduct in-depth retrieval of literature resources, and their full-text content can be transmitted within Chinese libraries. With a total of 2.4 million original Chinese books and more than one billion pages of information searchable, it provides readers with a full-text search that penetrates the book content.

¹⁸ <http://www.cnki.net/>, 2024.10.10. The English version can be changed from the top bar on the homepage.

By choosing the 9th item and opening it, a page in which several kinds of downloading markers can be found in the northwestern direction.



Figure 13 A dissertation example that can be downloaded

Laborious, manual, time-consuming, and stupid-seem efforts constructed 303 MA theses (around 5.1 million words) and 112 PhD dissertations (around 5.3 million words) into the ChAcE Corpus, which is general for its composition, representation, and purpose, contrasts with a specialized corpus; comparable because students in different educational level, i.e., Master and PhD levels may behave diversely in writing their theses that the collection of these varieties of the same language can be applied to make a comparison (Hunston, 2002, p. 15); a learner corpus that it is produced by students who learn English; diachronic by including texts from 2005 to 2014 that linguistics changes can also be considered of researching; and fixed on account of its invariable size though more original texts can be added manually and were not set as an information-autonomous-grabbing corpus. These theses and dissertations are all written by Chinese students from English major (a summarized name for English-related disciplines such as English-Chinese Translation, American-English Literature studies, English Linguistics, TESOL, Cultures in English-American countries, etc.) because theses from other majors in Chinese universities would never be written in English, which makes non-sense to searching for those texts and conduct English linguistics research.

Section \ Year	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	Count
Culture	3	1	2	3	1	2	4	7	8	6	37
Literature	1	2	1	1	1	6	2	9	5	5	33
Listening	1	-	3	1	1	1	2	3	3	5	20
Speaking	-	2	1	2	3	3	2	3	1	3	20
Reading	1	1	3	2	1	2	1	3	4	5	23
Writing	1	2	1	2	2	3	2	2	3	3	21
Phonology	-	3	-	4	1	2	5	2	5	3	25
Morphology	-	1	1	2	1	3	2	3	4	3	20
Syntax	1	1	2	2	1	2	3	4	4	2	22
Semantics	1	1	1	1	2	2	2	5	7	4	26
Pragmatics	3	2	2	3	1	4	3	2	3	7	30
Methodology	3	2	2	1	2	3	3	2	4	4	26
Count	15	18	19	24	17	33	31	45	51	50	303
Percentage	4.95%	5.94%	6.27%	7.92%	5.61%	10.89%	10.23%	14.85%	16.83%	16.50%	

Table 12 The downloading numbers of CMAC corpus from each disciplinary section and year
For the Master group (CMAC, Chinese MA Corpus), texts were collected and categorized into

sections according to their themes and years in which theses before 2010 were not in large quantities because the number of uploading theses on the website is not big enough on the one hand and graduation students were no more than those of later years on the other. If Culture and Literature are more related to literature studies parts, Listening, Speaking, Reading and Writing are classified into English language basic education, then Phonology, Morphology, Syntax, Semantics, and Pragmatics are traditional linguistics categories, and Methodology at last actually signifies any thesis focusing more on methodological innovation and discussion like teaching methods (pedagogical), learning strategies, course design, research methods, etc. Translation is not considered a standard of classification when specific objects of study are probed because any translation thesis can be grouped into the category mentioned above. Table 12 shows the basic downloading information of CMAC.

Thus, Culture and Literature theses account for almost a quarter (23.1%) of the CMAC, basic English language studies account for another quarter (27.75%), and half of the whole MA corpus is about linguistics research.¹⁹ (Statistics from Table 13) By counting downloading numbers of each year, over two-thirds (69.31%) are from the year 2010, while those from 2013 and 2014 take up to one-third (33.33%). Though thesis numbers per year are not equalized, linguistics features development through the year can be traced for at least 15 theses in the year 2015, accompanied by 280 thousand words in total.

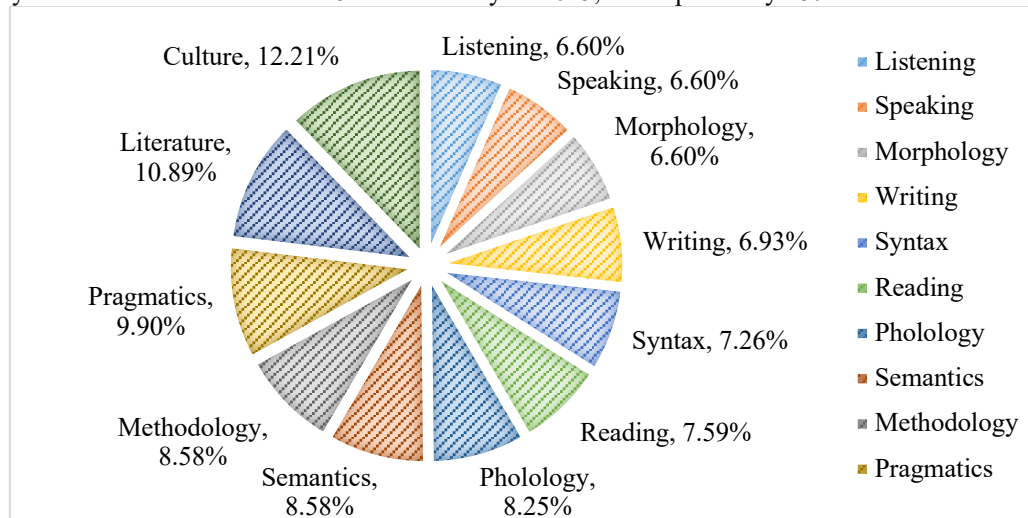


Table 13 The proportion of disciplinary sections of CMAC corpus from the least to the most

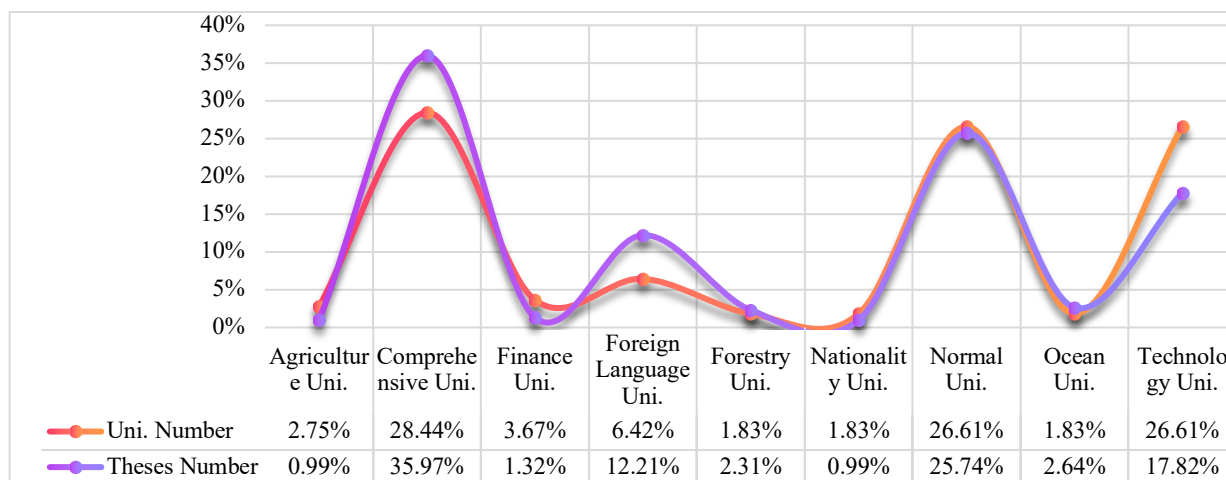


Table 14 The percentage of university categories and corresponding thesis numbers of CMAC corpus

Besides taking disciplines and years into consideration in downloading theses, geographical aspects (universities and districts) are also thought about for allocation. From Table 14, the number of Comprehensive, Technology, and Normal (for teachers) universities takes the largest proportion of the 109 universities in China that the CMAC comes from. It is no wonder that the first two kinds of universities have the highest percentage because more than half of Chinese universities belong to these

¹⁹ It does not mean that half of theses written by Chinese students of English major are about linguistics themes, it only signs the statistics reality of CMAC that was built for this research.

two categories, normal universities, by providing schools with a large number of English teachers every year, are thus seen as an important corpus source. Accordingly, the number of theses from these three types accounts for two-thirds of the whole corpus. Though only seven foreign language universities appear in the statistics, more than 10 percent of the theses are collected from these institutions: this is almost the same as the percentage of English-Major students graduating from foreign language universities in Chinese universities.

109 universities are distributed in six major geographical parts in China, half of which come from eastern and middle-southern districts such as the Yangtze River delta area (Shanghai, Zhejiang, Jiangsu), Shandong, Hubei, Hunan, Guangdong, etc., one third are from northern and northeastern provinces like Beijing, Hebei, Shanxi, Heilongjiang, Jilin, Liaoning, etc. This is consistent with university distribution because the population base in the eastern part of China is larger than that in the western part, and the economic development condition in the former is better than that of the latter. Richer districts attract more advanced students to better universities; thus, half of the theses are from eastern and northeastern areas, and nearly one-third are written in middle-southern and northeastern provinces. (See Table 15)

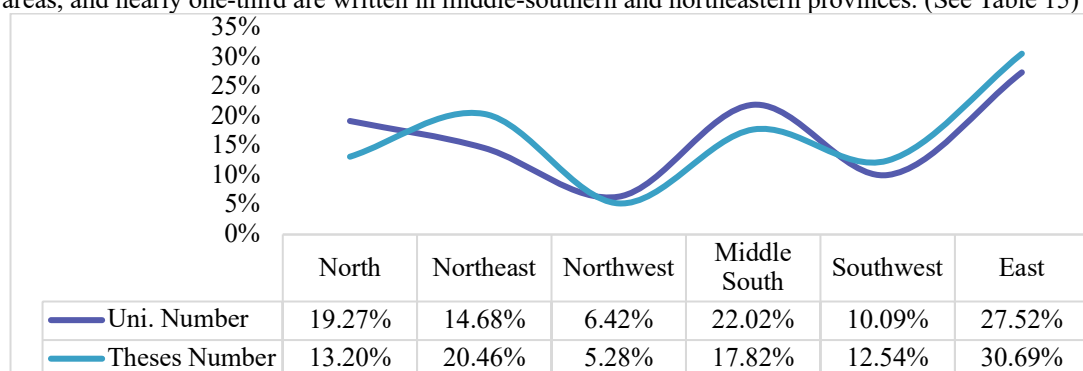


Table 15 The percentage of university and thesis numbers of CMAC corpus from each district

For the PhD group (CPhD, Chinese PhD Corpus), 112 dissertations were randomly picked, i.e., it was not organized after corpus allocation designs. From the perspective of year, they also range from the year 2005 to 2014, in which the percentage of dissertations in 2007-2009 and 2011 takes altogether three quarters after collating. (See Table 16) Seen from themes or contents, three big sections of disciplines are included. Fundamental English language studying takes the lowest proportion (23.21%), and pedagogical research has the highest proportion (41.96%). Again, translation-related dissertations that are divided into methodology, linguistics, or language studies are thus not listed individually in the section.

Section Year	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	Count
Listening	-	2	-	-	-	-	-	-	-	-	2
Speaking	2	-	-	2	-	-	-	-	4	-	8
Reading	-	-	2	2	-	-	1	-	-	1	6
Writing	-	-	-	-	10	-	-	-	-	-	10
Grammar	-	2	4	-	-	-	-	-	-	-	6
Phonology	-	-	-	-	-	-	-	-	-	3	3
Lexicology	-	-	1	1	2	-	4	-	-	-	8
Syntax	-	-	1	-	3	-	-	-	-	-	4
Semantics	-	-	-	1	-	-	1	-	-	1	3
Pragmatics	-	-	8	1	1	-	5	-	-	-	15
Teaching	-	-	-	8	-	-	2	-	-	1	11
Learning	-	2	1	2	2	2	4	4	-	2	19
Methodology	-	1	1	10	-	-	4	-	-	1	17
Count	2	7	18	27	18	2	21	4	4	9	112
Percentage	1.79%	6.25%	16.07%	24.11%	16.07%	1.79%	18.75%	3.57%	3.57%	8.04%	-

Table 16 The downloading numbers of CPhD corpus from each disciplinary section and year

To investigate each section in Table 17, Learning (including second language acquisition, EFL learning, computer-assisted learning, learning approach and strategies, etc.), Methodology (e.g., test performance research, incidental negative feedback, L1 translation role, etc.) and Pragmatics (discourse analysis, pragmatic cognition, appraisal and evaluative perspectives, etc.) researches account for almost half of the whole CPhD corpus, which marks the pragmatism of PhD studies in helping pervasive Chinese students more efficiently acquire English language and providing them with more intellectual

methods of teaching and employment.

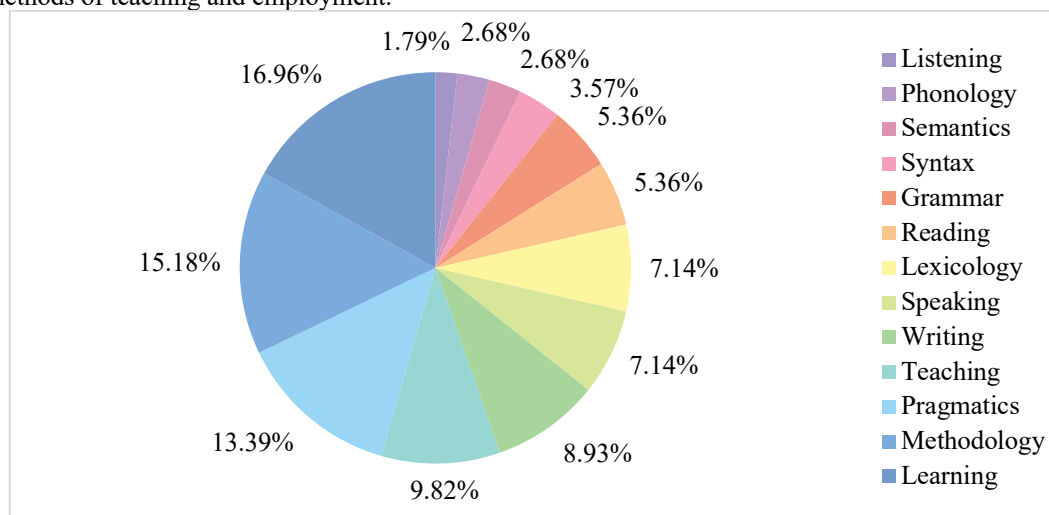


Table 17 The proportion of disciplinary sections of CPhD corpus from the least to the most

Being diversified from the broad geographical source of the CMAC group, PhD dissertations originate from only six universities, and more than 80% are written by graduates of Shanghai International Studies University (SISU). This does not mean the graduation ratio of PhD candidates from thousands of universities in China is only because more students in SISU are willing to upload their dissertations to websites like CNKI than those in others. (See Table 18)

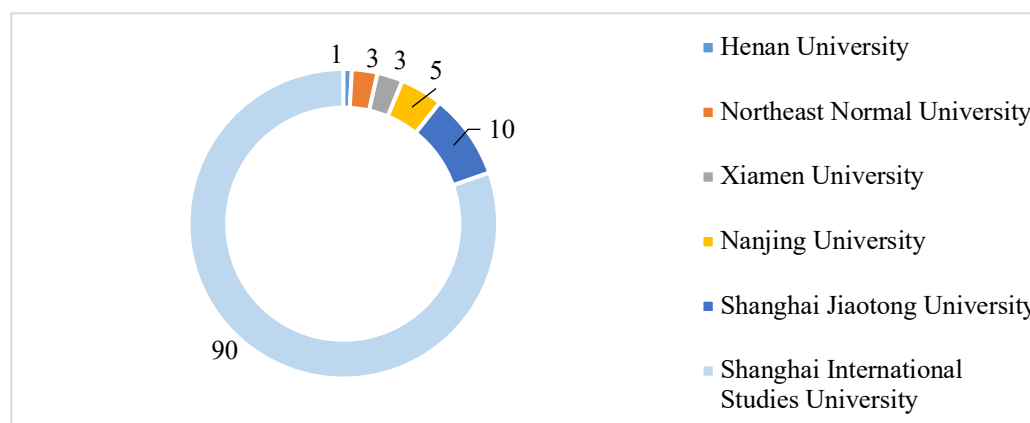


Table 18 The number distribution of dissertations from the CPhD corpus in each University

Data Conversion and Cleaning (Preliminary)

When articles are downloaded from *CNKI*, the format of these files is usually CAJ and should be opened by the software *CAJViewer 8.1*²⁰, which is not convenient for language queries. Thus, the whole corpus was converted to the format “.txt” by copying and pasting manually, and then each file was named as the following layout: “CMAC+YEAR+SECTION_number”²¹ (the number is increasing by the classification of section) for the CMAC group and “CPhD+YEAR_number” (the number is changing by year) for CPhD group, e.g., “CMAC13SE_9” means the 9th thesis in the section Semantics written in the year 2013 and “CPhD07_14” marks the 14th dissertation in the year 2017. For more accurate linguistics retrieval results in software like AntConc, detailed and redundant information of each file such as personal information (usually the first page which notes authors and supervisors’ names, graduation universities, years, etc.), titles, page numbers, footnotes, figures and tables, examples of corpus, appendix, references, and acknowledgments are all deleted because parts of them are not clauses written by the author, some are only phrases and numbers, some are excluded as to the main body of thesis themes, and acknowledgments are not considered significant in researching clause complexes in thesis writing.

²⁰ <https://cajviewer.oversea.cnki.net/en/download.html>. Till 12.12.2024, the newest version is 8.1.

²¹ Abbreviations for each section in naming each file in the CMAC group: PR-pragmatics, SE-semantics, PH-phonology, MO-morphology, SY-syntax, ME-methodology, LIS-listening, SP-speaking, RE-reading, WR-writing, LIT-literature, and CU-culture.

Data Information

After cleaning unnecessary parts of a thesis, the word number information is fixed, calculated, and then recorded by year (Table 19) and section (Table 20) in both groups that the average word number of CPhD group is evidently almost three times of that in CMAC group. Words percentage by year in CMAC is increasing because Master theses can be easier downloaded after the year 2009 while that in PhD is fluctuant because dissertations in recent three years are usually hard to download whenever searching is conducted: either due to the reason that the audit of dissertation contents takes a longer time for uploading, either the uploading ratio of dissertations is too much lower than the update rate of graduation. Another tendency is that the average words number of theses through temporal changes is decreasing in both groups, which might be affected by demand changes for acquiring a degree in universities in recent years.

CMAC				Year	CPhD			
Average Words No.	Thesis No.	Words Perc.	Words No.		Words No.	Words Perc.	Thesis No.	Average Words No.
18717.13	15	5.52%	280757	2005	103311	1.95%	2	51655.50
16138.56	18	5.71%	290494	2006	340334	6.42%	7	48619.14
17025.89	19	6.36%	323492	2007	819497	15.47%	18	45527.61
16312.13	24	7.70%	391491	2008	1284159	24.24%	27	47561.44
15835.59	17	5.30%	269205	2009	942739	17.79%	18	52374.39
16909.85	33	10.98%	558025	2010	59740	1.13%	2	29870.00
17501.77	31	10.67%	542555	2011	986649	18.62%	21	46983.29
16549.11	45	14.65%	744710	2012	176267	3.33%	4	44066.75
16874.20	51	16.93%	860584	2013	198744	3.75%	4	49686.00
16453.24	50	16.18%	822662	2014	387238	7.31%	9	43026.44
16778.80	303		5083975	Total	5298678		112	47309.63

Table 19 The (average) words number and their percentage in CMAC and PhD corpus by year

CMAC				Section	CPhD			
Ave. Words No.	Thesis No.	Words Perc.	Words No.		Words No.	Words Perc.	Thesis No.	Ave. Words No.
16453.84	37	11.97%	608792	Culture	-	-	-	-
16158.85	33	10.49%	533242	Literature	-	-	-	-
17745.00	20	6.98%	354900	Listening	97401	1.84%	2	48700.50
18192.10	20	7.16%	363842	Speaking	389292	7.35%	8	48661.50
16116.78	23	7.29%	370686	Reading	227777	4.30%	6	37962.83
16460.14	21	6.80%	345663	Writing	593240	11.20%	10	59324.00
18377.28	25	9.04%	459432	Philology	121505	2.29%	3	40501.67
-	-	-	-	Lexicology	366581	6.92%	8	45822.63
16303.45	20	6.41%	326069	Morphology	-	-	-	-
-	-	-	-	Grammar	302904	5.72%	6	50484.00
15527.41	22	6.72%	341603	Syntax	185667	3.50%	4	46416.75
15374.42	26	7.86%	399735	Semantics	176996	3.34%	3	58998.67
17994.53	30	10.62%	539836	Pragmatics	718424	13.56%	15	47894.93
16929.81	26	8.66%	440175	Methodology	764600	14.43%	17	44976.47
-	-	-	-	Teaching	531140	10.02%	11	48285.45
-	-	-	-	Learning	823151	15.54%	19	43323.74
16778.80	303		5083975	Total	5298678		112	47309.63

Table 20 The (average) words number and their percentage in CMAC and PhD corpus by section

For the word number comparison in each section, the word percentage of the linguistics part in CMAC takes up to 40%, while that in CPhD (35%) is not more than the proportion of the pedagogic part (40%). So we can speculate that when master students are still focusing more on investigations of linguistics phenomena, PhD candidates have been integrating their linguistics and applied linguistics theories and then applying them to didactics to foster methodological consciousness of their students, in which some candidates are English teachers who make the best of helping undergraduates or high school students solve learning dilemma when they are getting their degrees, some are assistants for their supervisors who are easier to collect data from their classes and then conduct experimental and comparison works. Interestingly, postgraduates would write more words for phonology, speaking, and pragmatics topics while fewer words for syntax and semantics parts; candidates use more words in describing writing, semantics, and grammar sections while less in phonology and reading-related texts. Changes in

semantics and phonology can be explained by data error because of the small size of the PhD group for one thing and disciplinary characteristics that phonology gets more difficult for research breakthrough as degree level promotes while semantics becomes easier for another. Syntax for CMAC and Reading for CPhD have the lowest average word numbers, which may be because a large quantity of examples (written by others) is deleted for calculation. However, whether the higher average word number in pragmatics and speaking in CMAC can be attributed to descriptive languages and that of writing and grammar in CPhD to interpretative clauses are inspiring points that will not be investigated in this research on account of length limitation.

Limitation

A problem should be stated briefly: most writers of the MA theses did not study further or did not upload their PhD dissertations onto websites, or the collector did not download the MA thesis and PhD dissertation from the same writer on purpose. Thus, almost all writers from the MA group are not the same as those from the PhD group. Another problem can be “gender” distinguishing difficulties that the gender of these writers is unknown by only judging from their names, which lies in Chinese language properties, i.e., the system of Chinese phonetic notation makes the Chinese character of each phone indecisive and is not in the same condition as languages like German that feminine names usually end with vowel *a*, *e* while masculine names are accustomed to end with consonant *r*, *d*, *t*, *g*, *k*, *n*, *m*, *o*, etc. Thus, the comparison of linguistic features in this corpus or with other corpora between females and males tends to be impossible. But one thing for sure is that most of the writers in this corpus are female due to discipline specialty and enrollment plan of universities enacted by the PRC Ministry of Education: female has inborn advantages over males in acquiring a language, and being a teacher is bidirectional more suitable for females, so females outnumbered males in English major in Chinese universities from Bachelor’s level. In addition, the percentage of theses and dissertations in each year and section in both groups is not equally and continuously distributed, and the two groups are thus not quite in the same condition, especially as to areas and universities, can they reasonably support temporal development of fixed language changes? Does the PhD group exactly represent the general average language usage of PhD candidates coming from all over China? Does it make any sense to compare linguistics items by choosing examples from the corpus?

To avoid awkward situations mentioned above and ensure the veracity and validity of statistics studies, gender differences in writing are firstly not included and regarded as one purpose in this research; two theses from CMAC and two dissertations from CPhD (decided by the minimum number from the year 2005 and 2010) in each year are chosen for temporal development investigations; writing habits of each student are not totally influenced by their education in universities, i.e., the factor of university diversity is not the biggest one on language usage differences, and before PhD-level studying, most of SISU candidates conducted their MA and Bachelor studies (when their writing habits might have already been formed and be stable) in other universities/districts, thus the PhD group in a sense can be representative; the last question is a significant research question that is answered in Sections Five and Six.

4.2 Research Methods

4.2.1 Qualitative & Quantitative Methods

4.2.1.1 Qualitative

Qualitative means an emphasis on the qualities of entities and on processes and meanings that are not experimentally examined or measured as to quantity, amount, intensity, or frequency. (Denzin & Lincoln, 2005, p. 10) The qualitative method refers to an in-depth study using face-to-face or observation techniques to collect data from people in natural settings. (McMilla & Schumacher, 2014, p. 5) Qualitative researchers usually describe and explain the truth, nature, and laws of things based on their attributes, relationships, and their inherent stipulations. They stress the socially constructed nature of reality, the intimate relationship between themselves and the studying object, and the situational constraints that shape inquiry; emphasize the value-laden nature of inquiry; and seek answers to how social experience is created and given meaning. (Denzin & Lincoln, 2005, p. 10)

From the analysis goal, qualitative research can be main cause analysis, process analysis, mechanism analysis, phenomenon analysis, factor analysis, structural analysis, function analysis, result analysis, etc. (Guo, C., 2015, p. 18) Four main stages in the qualitative analysis process could be: interpretation and coding, abstraction or data reduction, data display, and verification/conclusion drawing, in which a “code” usually refers to a word or phrase that indicates how the relevant data affects the research objective; and “abstraction” is the repeat summary of data. (ibid, p. 162-4) The main qualitative methods include ethnography, narrative analysis, discourse analysis, textual analysis, etc., and

the analysis of qualitative data is primarily to identify, examine, compare, contrast, discover, and explain meaningful patterns and themes in the data. (ibid, p. 161-2) Questions usually to be answered are: What are patterns and common themes in the collected data, and what implications do these patterns have for further research? Are there some deviations from these patterns? What factors could explain atypical cases with interesting stories, and what inspiration do they bring to further research? Is there any additional data that needs to be collected? Do the research questions need revision, and do the emerging patterns justify the conclusions? etc. (ibid)

4.2.1.2 Quantitative

Quantitative is a research paradigm in which objective data are gathered and analyzed numerically. (McMilla & Schumacher, 2014, p. 5) Quantitative methods emphasize objective measurements and the statistical, mathematical, or numerical analysis of data collected through polls, questionnaires, and surveys or pre-existing statistical data manipulated by computational techniques. (Babbie, 2010; Muijs, 2010) Quantitative research usually attempts to prove or disprove (i.e., falsify) certain precise hypotheses or to answer some specific, measurable phenomenological questions. (Guo, C., 2015, p. 18) It focuses on gathering numerical and unchanging data and generalizing it across groups of people (Babbie, 2010; Muijs, 2010) with an objective stance. The goal of quantitative research is to determine the relationship between one thing (an independent variable) and another (a dependent or outcome variable) within a population. Designs for quantitative research are either descriptive (subjects usually measured once) or experimental (subjects measured before and after treatment); a descriptive study establishes only associations between variables, while an experimental study establishes causality. (Babbie, 2010; Brians et al. 2011; McNabb, 2008; Singh, 2007)

Quantitative researchers often have clearly defined research questions to which objective answers are sought, and then try to recognize and isolate specific variables contained within the study framework, seek correlation, relationships, and causality, and attempt to control the environment in which the data is collected to avoid the risk of variables, other than the one being studied, accounting for the relationships identified. By using structured research instruments in gathering data and arranging them in tables, charts, figures, or other non-textual forms such as numbers and statistics, the research results are usually representative of the population based on larger sample sizes, and the whole project can be replicated or repeated to generalize wider concepts, predict future results, or investigate causal relationships because of its high reliability²² and validity²³.

The aspects of quantitative analysis to be investigated mainly are descriptive statistics (measurement of frequency, percentage, central tendency, and the degree of variation/dispersion); central tendency (mode, median, mean), including the measurement level of the variable and the distribution skewness of the data; and the degree of dispersion (range, interquartile range, variance, standard deviation)

4.2.1.3 Qualitative vs. Quantitative

By concluding descriptions above, qualitative and quantitative research methods are primarily different in the following aspects (McMilla & Schumacher, 2014, p. 20-1): assumptions about the world, the research purpose, research methods and process, prototypical studies, the researcher's role, and the importance of context in the study (Denzin & Lincoln, 2000, p. 11-12):

1. Assumptions about the world: Qualitative research is mainly based on constructionism by assuming that multiple realities are socially constructed through individual and collective perceptions or views of the same situation; while quantitative research is based on some form of positivism by assuming that there are stable, social facts with a single reality that is separated from the feelings and beliefs of individuals.
2. The research purpose: Qualitative research is concerned with understanding the social phenomenon from the participants' perspectives and thus asks the researcher's participation to some degree in the lives of those persons; while quantitative research seeks to establish relationships and explain causes of changes in measured outcomes.
3. Research methods and process: In qualitative studies, there is greater flexibility in both the strategies and the research process. Researchers typically use an emergent design and revise decisions about the

²² Reliability: the extent to which the results can be reproduced when the research is repeated under the same conditions.

²³ Validity: the extent to which the results really measure what they are supposed to measure.

data collection strategies during the study. In contrast, in quantitative studies, there is an established set of procedures and steps that guide the researcher to choose methods as part of a pre-established design that is determined before data collection.

4. Prototypical studies: Qualitative research considers subjectivity in data analysis and interpretation, and the proto-typical qualitative study is an ethnography, which helps readers understand the multiple perspectives of the social scene or system by the persons studied. Whereas quantitative research seeks to control for bias through design, a quantitative researcher usually employs experimental or correlational designs to reduce error, bias, and the influence of extraneous variables.

5. Researcher's role: As qualitative research is noted for "disciplined subjectivity" (Erickson, 1973) and "reflexivity" (Mason, 1996), which means critical self-examination of the researcher's role throughout the entire research process, qualitative scholars emphasize the importance of data collected by a skilled, prepared person in contrast to an instrument, and thus become immersed in the situation and the phenomenon being studied. However, quantitative researchers prefer to be detached from the study to avoid bias.

6. The importance of the context in the study: Qualitative researchers believe that human actions are strongly influenced by the settings in which they occur and cannot be explained without understanding the framework within which subjects convey their thoughts, feelings, and actions. Thus, they focus more on this context during data collection and analysis. Nevertheless, most quantitative researchers attempt to establish universal, context-free generalizations in their studies.

Qualitative methods are advantageous because they yield comprehensive and nuanced data that preserves the perspectives of participants and offers multiple contextual frameworks for interpreting the research phenomenon. As previously outlined by Anderson (2010), Denzin & Lincoln (2000), and Merriam (2009), among others, qualitative research allows researchers to: 1. gain an understanding of participants' perspectives on the research through cultural or situational immersion and direct interaction with participants (in their own language and terms); 2. respond to local situations, conditions, participants' needs, and changes that occur while the research is being conducted, and change the research focus accordingly in real time; 3. describe existing phenomena and current situations, and develop flexible methods for data collection, subsequent analysis, and interpretation of collected information; 4. improve the ability to describe raw and unstructured data; 5. produce results that help open up new ways of understanding; 6. gain a comprehensive understanding of the phenomenon being studied and obtain a more realistic view of the world of life that cannot be experienced in numerical data and statistical analysis. Qualitative research is, therefore, frequently employed to illustrate phenomena vividly or to facilitate cross-case comparisons and analyses of individuals or groups.

Qualitative research methods are not without limitations, as evidenced by the findings of Babbie (2010), Brians et al. (2011), McNabb (2008), and Singh (2007). Such limitations include: 1. The process of data collection and analysis is often time-consuming and/or expensive; 2. The quality and quantity of information obtained from different respondents can vary considerably, making it challenging to draw meaningful conclusions; 3. It is not possible to investigate causal relationships between different research phenomena; 4. The generation of large amounts of data makes it difficult to be randomly divided into manageable parts for analysis; 5. The original research objectives can be deviated from due to changes in the research environment; 6. Conclusions drawn from the same information can vary due to the characteristics of the researcher; 7. Extensive experience is necessary for the researcher to obtain the target information from the respondents; 8. The use of human subjects increases the possibility of ethical dilemmas, which can undermine the overall validity of the study; 9. Lack of consistency and reliability is another issue, as researchers can use different exploration techniques, and respondents can choose to tell specific stories and ignore others; and 10. To replicate a study will be very difficult.

Quantitative methods are regarded as objective techniques for investigating problems, accumulating data, and identifying the underlying causes of behavior by controlling and measuring variables. The application of quantitative methods allows researchers to achieve the following: 1. Adhere to well-established standards, ensuring the research can be replicated, analyzed, and compared with similar studies. 2. Conduct more extensive research involving more subjects and enhance the generalizability of the results. 3. Summarize a large number of information sources and make comparisons across categories and time. 4. Obtain data with greater objectivity and accuracy. 5. Avoid personal bias by maintaining a "distance" from the subjects involved and using recognized computing techniques. (cf. Anderson, 2010; Denzin & Lincoln, 2000; Merriam, 2009). Consequently, the outcomes of quantitative studies may be statistically significant yet often inconsequential from a human perspective.

However, an overreliance on quantitative methods may result in researchers (cf. Babbie, 2010; Brians et al., 2011; McNabb, 2008; Singh, 2007): 1. Collecting narrower, even superficial data sets; 2. Using static and rigid methods, thereby adopting an inflexible discovery process; 3. Formulating

standard questions that may lead to “structural bias” and misrepresentation; 4. Conducting research in unnatural, artificial environments to control the production of “laboratory results” rather than “real-world results”; 5. Obtaining results with less detail about the subjects’ behaviors, attitudes, and motivations, as well as the research context; 6. Obtaining limited numerical descriptions rather than detailed narratives; and 7. Obtaining presumed answers that do not necessarily reflect how people really feel about a topic, and in some cases, these answers may simply be the closest to preconceived hypotheses.

Qualitative studies are inductive in that a theory is usually derived from the research results, while quantitative research is deductive in that hypotheses are developed from known theory and then are proved (or disproved) in empirical investigation. (Litosseliti, 2010, p. 52) Due to the advantages and disadvantages described above, and in Schmied’s (1993, p. 85) words:

“A good quantitative study must be based on methodological and terminological clarity. Thus, the qualitative approach, which established the basic concepts and system elements, must precede any quantitative investigation, if the study wants to transcend simple language-statistical calculations. The quantitative paradigm, on the other hand, relativized the qualitative categories, in particular regarding the gradient nature of language. It is thus necessary to combine the quantitative and the qualitative approaches in any serious study of the variation of grammatical phenomena.”

Qualitative and quantitative research methods can be complementary and thus be combined for comprehensively conducting research in solving questions raised in Section 1.3. The need and practicability of integration have been developed into the Mixed Methods research in recent years that a researcher often

“collects and analyzes persuasively and rigorously both qualitative and quantitative data (based on research questions); mixes (or integrates or links) the two forms of data concurrently by combining them (or merging them), sequentially by having one build on the other, or embedding one within the other; gives priority to one or to both forms of data (in terms of what the research emphasizes); uses these procedures in a single study or in multiple phases of a program of study; frames these procedures within philosophical worldviews and theoretical lenses; and combines the procedures into specific research designs that direct the plan for conducting the study.” (Creswell & Clark, 2011, p. 6)

In this research, mixed designs are mainly Sequential, Conversion, Multilevel, Fully-integrated mixed (Teddlie & Tashakkori, 2009, p. 31-32), and Embedded (Creswell & Clark, 2017, p. 57) according to different taxonomy and their definitions. The detailed methods used in each part are demonstrated below:

Grammatical annotation: QUAL
Grammatical calculation: QUAN
Grammatical analysis: quan → QUAL
Lexical annotation: QUAL
Lexical calculation: QUAN
Lexical analysis: quan → QUAL
Recursion patterns annotation: QUAL
Recursion patterns analysis: QUAL(quan)
Four additions & related: Qual
Complexity of CC: [QUAN+qual] → QUAN
Verification & modification of the CC complexity measurement model: QUAL →← QUAN

Table 21 The detailed mixed methods research designs in this project

In which “QUAL” and “QUAN” mean prioritized or emphasized designs; “qual” and “quan” mean lesser priority designs; “→” means sequence that the latter follows the former; “()” means that the latter is embedded within a larger intervention design; “Qual” or “Quan” means the only design; “+” means concurrence; “[]” means a core mixed methods design; and “QUAL →← QUAN” means that the methods are implemented in a recursive process. (See Creswell & Clark, 2011)

4.2.2 Qualitative Methods

For qualitative analyzing methods used in this research, exploratory (heuristic), comparative, experiential summary, grounded theory, and case study are combined to explain how some results are received and how some ideas are formed. The main software used for qualitative research is SysFan, developed by Wu (2000), which enables researchers to organize texts, manually annotate grammatical

structures, record annotation results, and automatically calculate the numbers and percentages of each grammatical category.

4.2.2.1 Exploratory Research (Heuristic Analysis)

As an exploratory analysis method, heuristic analysis is irrelevant to discovering theories and proving hypotheses but is directly related to people's cognition (especially self-inquiry or introspection). *Heuristic* refers to the process of people's inner awakening to knowledge, i.e., to discover the essence and meaning of the world through people's feelings. The method indicates that the researcher plays a vital role in the research and that his/her real feelings are the main focus. In the heuristic analysis method, researchers explore and clarify phenomena's nature and meaning through their inner awakening. They are also encouraged to explore the creativity of ideas, put personal experience (monotonous) above numbers, and dig out tacit knowledge based on subjectivity and creativity between researchers and phenomena. As heuristic analysis is not a disorderly process, researchers are required to follow the disciplines of scientific research to ensure the authenticity and reliability of the research. After clause complexes were annotated, four kinds of additions and even categories in the theoretical framework (see Section 7.2) would be dug out using heuristic analysis.

As to the exploratory analysis, the model of the complexity of the clause complex in Section 7.5 is explored to determine which model is the best one to measure the complexity and whether the most suitable one can be applied in any text for better validity and reliability without any modification remains to be answered.

4.2.2.2 Comparative

Just as the word "comparative" implies, the method is to conduct a comparison. In this research, the most correlated application is to compare theses written by the Master group and the Doctor group of their clause complexes usage differences and similarities in grammatical (73 items), lexical (283 items) and recursion patterns choices, logico-semantic functions that are performed in construing their meanings and opinions in academic writing, explicit versus implicit clause complexes, errors versus should-be clause complexes found in theses, etc. In addition, by the comparison of existing literature reviews, the development of clause complex research can be made clear, margins can be found for further investigations in this research, and the significance and limitations of the research can be obvious to other scholars. The qualitative comparisons are mainly described in discussion chapters, and the quantitative comparisons are demonstrated by figures and charts with numerous numbers and percentages inevitably.

4.2.2.3 Experiential Summary

The most important experience of conducting this research comes from the annotation of clause complexes, and the most suitable is manual annotation in a recording-supporting software instead of automatic annotation by computational software after parsing their clauses, words/phrases, part-of-speech tagging, etc. Though the latter method can be further verified, software for combining parsing and clause complexing analysis is still absent until now. Research in the future can solve these difficulties and then improve encoding methods in new software with higher reliability and validity. Another experience is to distinguish one kind of clause complex relationship from the other in the whole logico-semantic system, writing and reading abilities of clause complexes can thus be enhanced.

4.2.2.4 Grounded Theory

Grounded theory emphasizes the generation of *theory* and the *data* in which that theory is *grounded* (Glaser, 1978). It "is a *detailed* grounding by systematically" and intensively "analyzing (deconstructing and reconstructing) data, often sentence by sentence, phrase by phrase or section by section of the field note, interview, or other document; by 'constant comparison,' data are extensively collected and coded," and then are combined for higher-level themes and meanings (Riazi, 2016, p. 129-130) or extracted of relevant categories and attributed according to the relationship between data and theories (Guo, C., 2015, p. 166), thus producing a well-constructed (with density, variability and a high degree of concepts integration) theory based on empirical data with a bottom-up method. (Strauss, 1987, p. 22-23)

"Grounded theory is based on a concept-indicator model, which directs the conceptual coding of a set of empirical indicators, i.e., actual data (at first provisionally but later with more certainty), such as behavioral actions and events, observed or described in documents, etc." (ibid, p. 25) It allows researchers to use a systematic procedure to generalize a phenomenon, provide a theoretical explanation for the general characteristics of the theme, produce crystallized patterns about the phenomenon (Riazi, 2016, p. 129-130), and at the same time make this explanation grounded in empirical observations of the data.

Two basic sub-processes are included for inductive analysis: unitizing and categorizing. Unitizing is the process of coding operations that separate information units one by one from the text; and categorization (called a “constant comparative method” by Glaser & Strauss, 1967) is to organize the information units obtained by unitization into different categories according to their similarity in meaning. (Rudestam & Newton, 2014, p. 242)

For coding data, three levels, i.e., open coding, axial coding, and selective coding, would be conducted (Riazi, 2016): 1. at the open coding level, researchers read the data line by line and assign meaningful concepts or codes to chunks of the data varying from a line to a whole paragraph; 2. for axial coding, researchers conduct intense analysis around one category at a time, in terms of the paradigm items (conditions, consequences, and so forth), and then attempt to make links between the codes around certain axes, which “results in cumulative knowledge about relationships between that category and other categories and even subcategories” (Strauss, 1987, p. 32); 3. selective coding, which pertains to coding systematically and concertedly for the core category, i.e., “the other codes become subservient to the key code under focus. The analyst delimits coding to only those codes that relate to the core codes in sufficiently significant ways as to be used in a parsimonious theory” (ibid, p. 33), and strives to produce a coherent, refined story and explanation of what is happening in the research site by using more abstract themes that can link categories or groups of categories together.

In this research, the integration and summary of literature reviews are carried out by the grounded theory method. But strictly speaking, if a well-constructed theory must be produced afterward, then just some steps of the whole process are applied, i.e., documents data systematically coding (opening, axial, and selective), unitizing and categorizing, constant comparing, crystalized themes and meanings generalizing, etc., in supporting for analysis. In the Four Additions (Section 7.2) discussion, though these categories are concluded from grammatical patterns that can be seen as already selective coding items (as to the system of clause complexing), the process of producing the Four Categories strictly follows the grounded theory method that axial and further selective coding can be conducted for modification from more theoretical (the analyst decides on analytic grounds what data to collect next and where to find them) (ibid, p. 38) and even selective (the calculated decision to sample a specific locale according to a preconceived but reasonable initial set of dimensions (such as time, space, identity) which are worked out in advance for a study) (ibid, p. 39) samplings.

4.2.2.5 Case Study

In Thomas & Yin’s (2018, p. 8) words: “Every research method can be used for all three purposes: exploratory, descriptive, and explanatory studies,” the categorization of case studies can also be these three accordingly, i.e., exploratory case studies, descriptive case studies, or explanatory case studies. Exploratory and descriptive case studies primarily give full exploration and description to cases to facilitate readers’ thorough understanding that the former is often used to lay the foundation for further, probably “more quantitative studies by defining questions and hypotheses” when the examined case is rarely known; and the latter intends to demonstrate a particular phenomenon’s details and contextualization (Heigham & Croker, 2009, p. 71). The explanatory case study explains the cause-effect relationships of a case by focusing on quantitative analysis (testing hypotheses related to the case), is usually a long-term or longitudinal case study (ibid), and consists of (a) an accurate rendition of the facts of the case, (b) some consideration of alternative explanations of these facts, and (c) a conclusion based on the single explanation that appears most congruent with the facts. (Yin, 1981, p. 61) From the beginning of this research, an exploratory study has been made on clause complex usage conditions in each Master thesis and PhD dissertation to conduct preliminary comparisons and make assumptions of testing hypotheses on results obtained. A descriptive and explanatory study would be applied to specific clause complexes in texts to demonstrate their uniqueness and peculiarity; clause complex recursion patterns in some texts are described in detail and explain their most possible reasons.

4.2.2.6 Software: SysFan

“Linguistic resources are stratified into three strata: semantics, lexicogrammar and phonology/graphology, which are related by means of realization. The lower stratum realizes, and is contextualized by, its immediate higher one.” (Wu, 2000, p. 114) “Analysing an example lexicogrammatically means locating it within the overall lexicogrammatical system.” (Matthiessen, 1995, p. 116) 错误!未找到引用源。 shows the “level” of analysis (stratum/rank/axis), increase in labor intensity, and decrease in potential for automation (Matthiessen, 2006, p. 109) and also the location of *SysFan*, *SysConc* and other tools for text analysis along stratification and rank. (Wu, 2000, p. 112)

As we move higher up, we are adding more value to the text: word-class tagging with taggers, grammatical annotation with parsers, and multi-dimensional and multi-functional analyses with manual

analyzers. However, the higher we move up, the less the analysis can be automated. As linguistic parsing often involves very rich information (much richer than what current NLP systems can cope with and certainly much richer than concordance and pattern matching), and may include many steps of analysis, fully automatic analysis of large volumes of texts thus has not yet become a reality. (Wu, 2009, p. 134-5)

Considering the fact that current systems are constrained with respect to the complexity of analysis they can cope with, a tool for supporting manual higher-level systemic-functional analysis, i.e., SysFan, was developed in the hope that the result of the manual analysis will provide fundamental insights for moving towards automatic systems in the future. (Wu, 2000, p. 113-4)

As demonstrated in Figure 14, SysFan is located at the other end of the scale and is one of the tools that operate at the highest rank within the lexicogrammatical stratum for doing systemic functional analysis of texts. It is theoretically comprehensive and can be used to process large volumes of texts to produce THEME, MOOD, and TRANSITIVITY analyses of clauses and other systemic and functional analyses of other grammatical units such as clause complexes manually, semi-automatically, or fully automatically depending on the amount of human intelligence involved, with the ultimate goal to achieve fully automatic analysis of text at all levels. (Wu, 2009, p. 134)

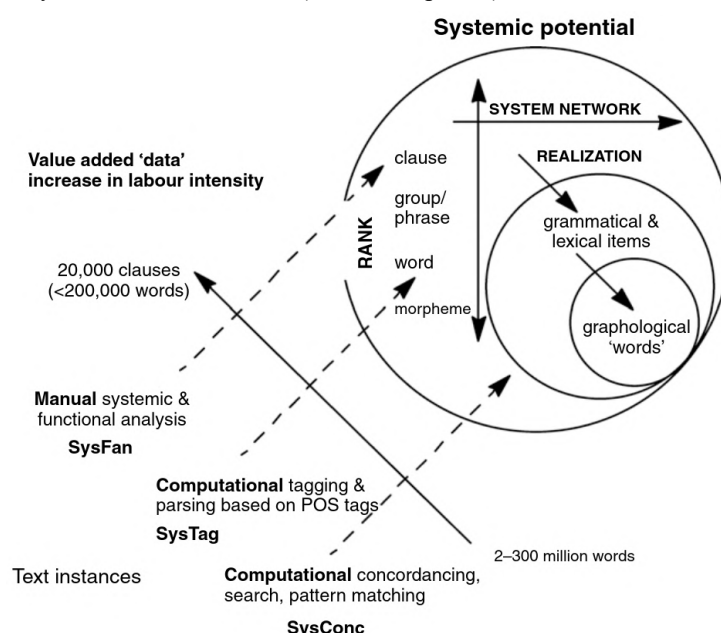


Figure 14 The increase of data annotation labor intensity and the location of the software SysFan for text analysis along with stratification and rank

CC Analysis with SysFan:

It has been described in Chapter 3 that the clause complex is equivalent to the sentence except that “a sentence is a constituent of writing while a clause complex is a constituent of grammar,” which can be used to account in full for the functional organization of the former. A clause complex is a grammatical unit over the clause; not placed above it but next to it on the rank scale because of their logical relationship instead of constituent one (Halliday, 1985a, p. 193; Eggins, 1994, p. 129) (cf. Wu, 2000, p. 121)

The logical relations between clauses can be interpreted as functional components of the linguistic system: the system of INTERDEPENDENCY (“parataxis” and “hypotaxis”) and the system of LOGICO-SEMANTIC RELATION (“projection” and “expansion”). Parataxis is the relation between two clauses with equal status: one initiating and the other continuing, thus is represented conventionally by numerical notations such as [1 2 3...]; whereas hypotaxis is the relation between two clauses with unequal status: one clause is dependent on the other, the is represented by the Greek letter notation as [α β γ ...]. Expansion means that a clause expands another clause by elaborating [=], extending [+], or enhancing [x] it, and projection means that a clause is projected through the other clause as a locution [‘] or an idea [‘] (see Halliday, 1994a, p. 218-220; Matthiessen, 1995, p. 130). These symbols combine, providing a convenient notation for representing the logical structure of the clause complex. (ibid, p. 122)

As SysFan aims to reduce the task of the linguist and make analysis more efficient and more enjoyable, it allows the user to analyze texts lexicogrammatically by making selections from menus or networks. (ibid, p. 115) The analysis in SysFan starts with the creation of text archives: any kind of computer-readable text can be used as input to the system. Once a text is added to the archives, it is

automatically assigned a unique Text ID (such as CPhD17_0002) and a creation date. Then general information about the text, such as Text Title, Text Source, the Author of the text, and their own notes or commentary, can be input into columns with word numbers of the text automatically counted. (See Figure 15)

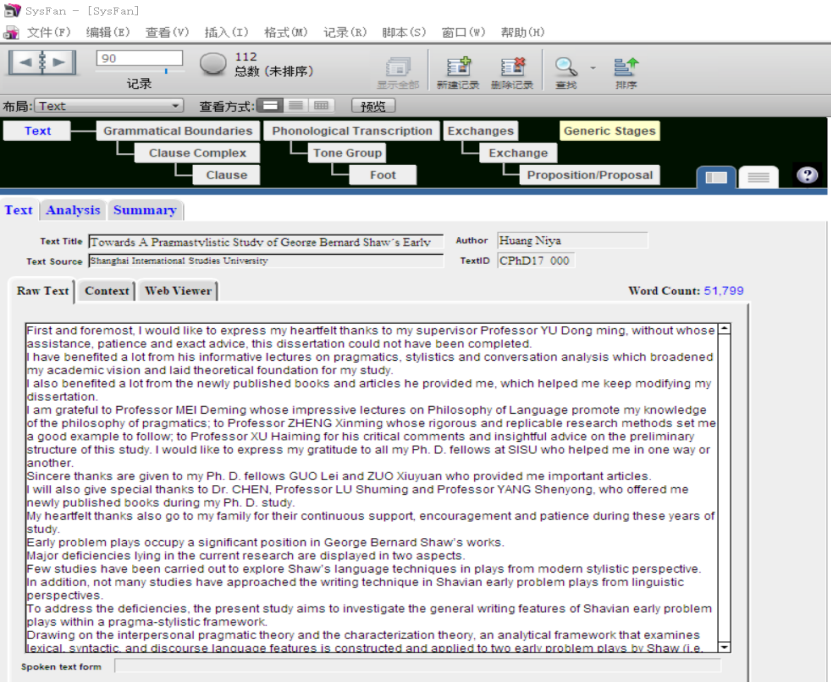


Figure 15 The text import interface of SysFan with related information filling in columns

For lexicogrammatical analysis, the text is chunked into clause complexes (though they can be further chunked into smaller units such as clauses, group/ phrase complexes, etc.): the user can select a clause complex in the text and click the button New (which means “creating a new record for the clause complex”) until there are no more clause complexes left out in the text; alternatively, researchers can manually insert a clause complex divider (e.g. “|||”) between clause complexes, and then create all the clause complexes in one go by clicking the button Create All Records. If researchers want to step into the analysis at a lower level, then a click on the icon next to Complex ID will take us to the analysis of the clause complex (ibid, p. 118-9) (See Figure 16)

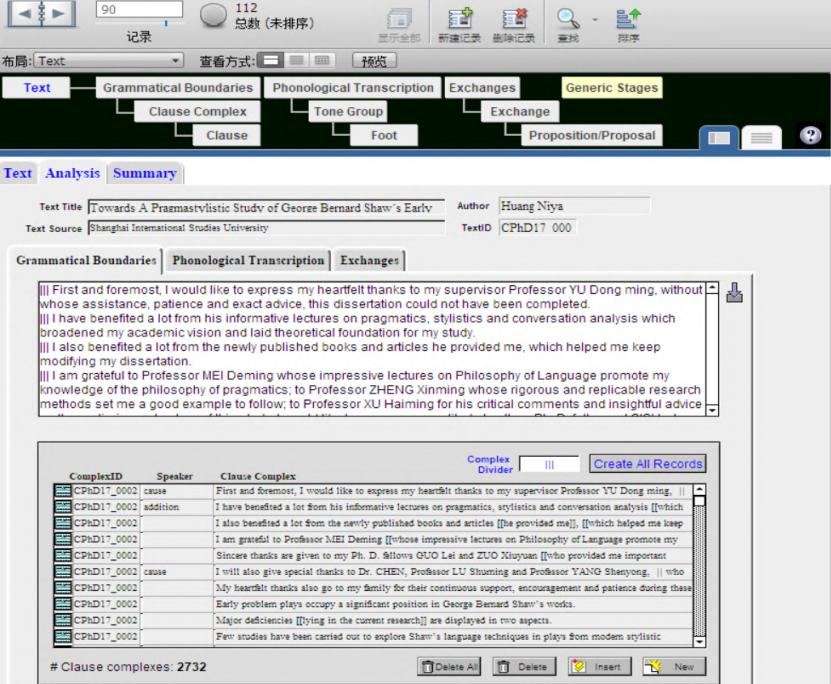
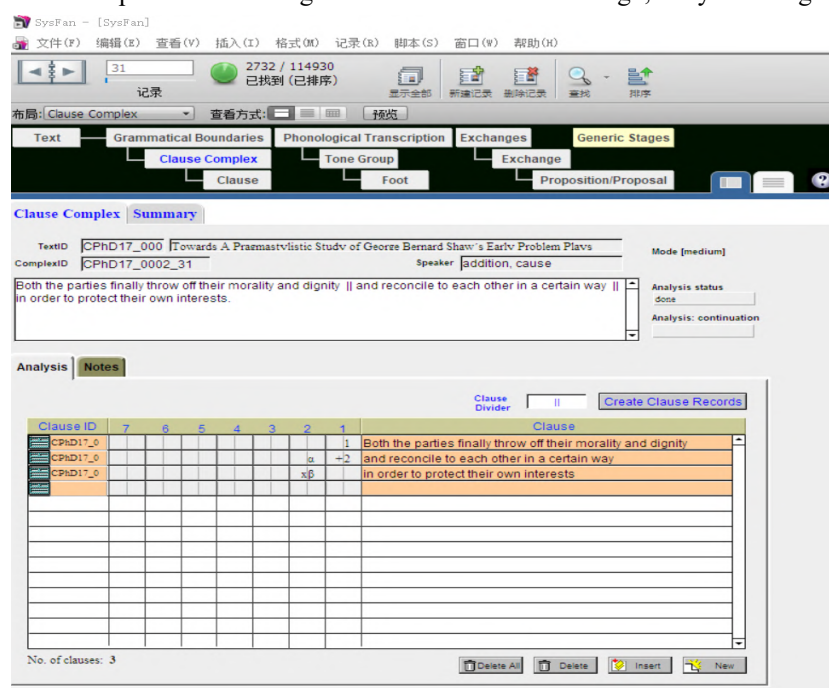


Figure 16 The chunk interface of texts into clause complexes in SysFan

If the user deletes or adds some clause complexes at any time, the following clause complex IDs will be updated to reflect this change as long as the deleted or added clause complex is not the last one in the text,

such that there is no discontinuity in the IDs. (ibid, p. 120) SysFan enables both the systemic and the structural analyses of the clause complex. Once clause complexes have been chunked out of a text, they are stored in the clause complex database and can be pulled out at any time for analysis. To chunk the clause complex into ranking clauses, researchers can fulfill this either by inserting the clause divider (e.g., “||”) in the “clause complex” and creating all the clause records in one go, or by selecting one clause after



another and creating clause records one by one until all clause complexes have been fully chunked.

Figure 17 The clause complex annotation interface in SysFan

In analyzing the clause complex (as shown in Figure 17), a clause in the clause complex can be assigned a univariate function by clicking the right cell to the left of the clause in the column that represents the appropriate level of nesting, and selecting from a menu of options: a paratactic or hypotactic value [1 2 3 ...; $\alpha \beta \gamma$...]; then clicking the left cell in the same column, and selecting a value from another menu of options: a projecting or expanding one [= + x; “ ‘]. (ibid, p. 123)

During the process of annotation, some comments of related clause complex relationships can be written down in the column Notes with the number of clauses and also the number of words in the clause complex shown right as in Figure 18. The specific words of comments in The Notes part can be searched if needed and necessary afterward.

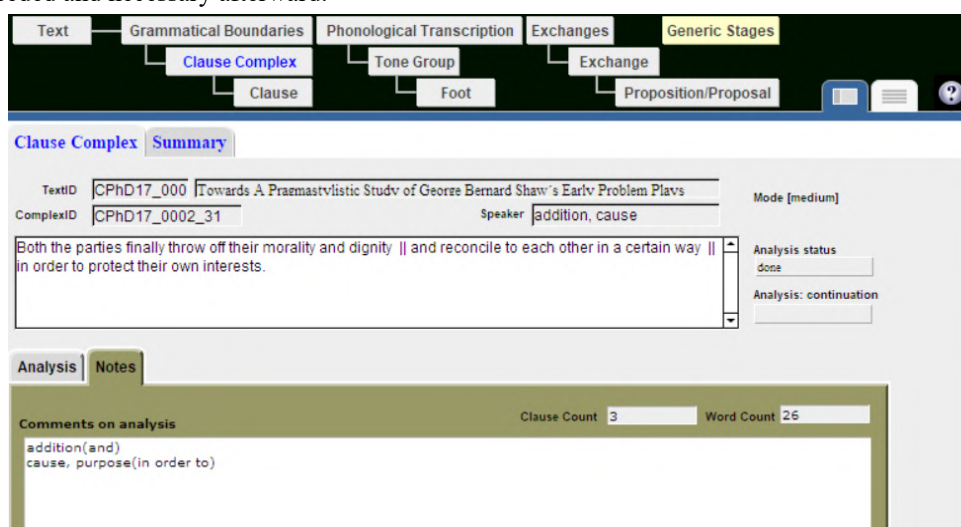
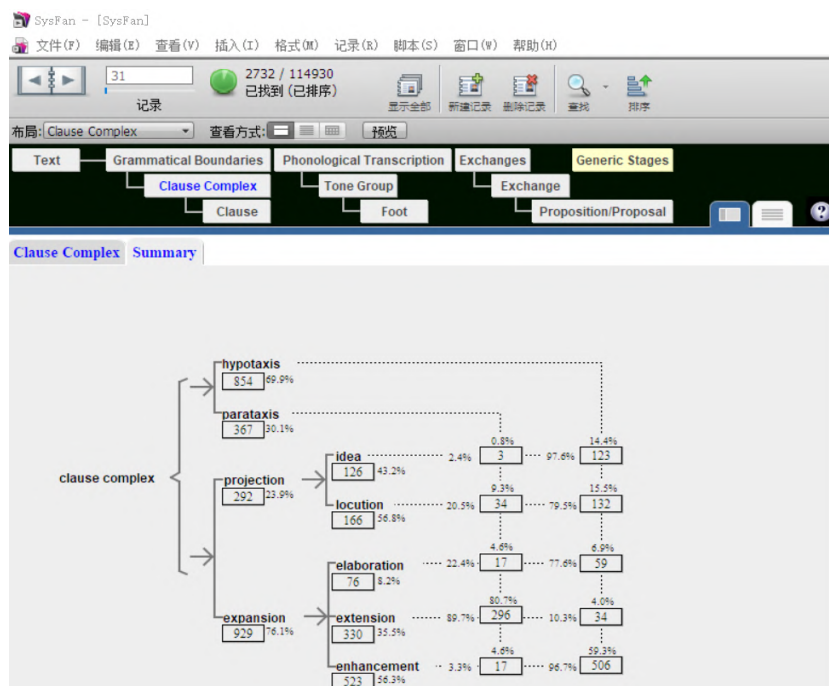


Figure 18 The interface of clause complex annotation notes in SysFan

As all analyzing results are stored in a database, they can be accessed through query languages like analyzing features or specific lexical items, and corresponding reports can be automatically or manually generated such as analyzing matrixes (see Figure 19), occurrences of specific categories and their related

percentages in a text or a group of texts, the number and categories of recursion patterns on each level



from 1 to 6, the number of words per clause complexes in each text, etc. (ibid, p. 15-6)

Figure 19 The interface of analyzing the matrix of clause complex categories in one text in SysFan

The statistical information, which dynamically reflects the clause complex analysis and changes through analyzing modification, is automatically obtained in SysFan. On the one hand, this analyzing result gives a whole picture of the numbers and percentages of clause complex categories; on the other, this provides an effective method of checking the reliability, validity, and consistency of the analysis. (ibid, p. 124)

One thing to note is that if a clause complex is embedded in a clause or words/phrases, i.e., functioning at group/phrase rank and marked as [[]], it will be annotated and recorded with an ID and then counted according to various statistical queries.

4.2.3 Quantitative Methods

For quantitative analyzing methods used in this research, descriptive, corpus linguistics basics, and mathematical models are necessary for comparing the differences and similarities of two groups selected, seeking for level development, bridging the gap between theory constructing and practical measurement, predicting structures usage habits of students' writing, and raising the reliability and validity of this research for replication and repetition. The main software for quantitative research is *JASP*²⁴ for statistics analysis and results generation.

4.2.3.1 Descriptive Research

For three main purposes of research, descriptive studies often appear following creative exploration and serve to organize the findings for better explorations and then test or validate those explanations (Krathwohl, 1993). It illuminates knowledge that we might not otherwise notice or even encounter. With the goal of describing a phenomenon and/or behavior of a sample population with its characteristics, descriptive research is concerned more with "what" rather than "how" or "why" something has happened. (Gall, Gall, & Borg, 2007) It "aims at casting light on current issues or problems through a process of data collection that enables them to describe the situation more completely than was possible without employing this method." (Fox & Bayat, 2007, p. 45)

Descriptive research can be either qualitative or quantitative. It involves gathering data that describe phenomena and then organizing, tabulating, depicting, and describing the data collection (Glass & Hopkins, 1984). When in-depth, narrative descriptions of a small number of cases are involved,

²⁴ JASP: Jeffreys's Amazing Statistics Program, a free and open-source program for statistical analysis supported by the University of Amsterdam. It offers standard analysis procedures in both classical and Bayesian form.

description is used as a tool to organize data into patterns that emerge during analysis, such as describing categories of clause complex recursion patterns that are annotated in each group in this project. It can also involve collections of quantitative information that can be tabulated along a continuum in numerical form, such as the length of a clause complex in each text chosen from the corpus. Because the human mind cannot extract the full import of a large mass of raw data, descriptive statistics are significant in reducing the data to a manageable form with visual aids such as graphs and charts for researchers and readers to understand the data distribution. As descriptive research can also be effective in getting information for developing hypotheses and proposing associations, data collection, and analysis techniques that yield reports concerning the measures of central tendency (mean, median, mode, and deviance from the mean), variation, percentage, and correlation between variables in descriptive statistics are utilized. The combination of its characteristics' summary and correlational statistics, along with its focus on specific types of research questions, methods, and outcomes, distinguish descriptive research from other research types.

Descriptive research can be either longitudinal or latitudinal, thus, how clause complexes are construed and how grammatical and lexical categories are chosen in each group (MA and PhD) and between two groups will be described in detail.

Descriptive research can be either diachronic or synchronic. Though the writing year of texts chosen from ChAcE and CNKI corpus ranges from 2005 to 2020 for both the Master's group and the Doctor's group, this project does not pay much attention to diachronic description because of the inconsistency of participants and also the research purposes.

4.2.3.2 Corpus Linguistics

As described in Section 3.3.1, the term *corpus* refers to a large collection of texts that represents a particular variant of a language or a sample of usages and is often stored in the computer with machine-readable forms. Corpus linguists emphasize text samples used in the real world in the hope of summarizing a series of abstract rules that govern natural languages and various rules of language in actual use, verifying existing linguistic theories, and proposing new ones. (Guo, 2015, p. 63)

Corpus analysis can be qualitative and quantitative as the former uses data as a basis for identifying and describing language usage to provide "real-life" examples of particular phenomena, while the latter classifies linguistic features that are identified in the data, counts them, and assigns frequencies, and even constructs more complex statistical models to explain observed phenomena. (McEnery & Wilson, 2001, p. 76) The qualitative approach to corpus analysis offers a rich and detailed perspective on the data that enables fine distinctions to be made since rare phenomena should receive the same attention as the more frequent phenomena, and classifying all the data into a finite number of categories is sometimes impossible. In addition, the ambiguity of language use, which is inherent in human language - by accident or deliberately - can be fully recognized: qualitative analysis does not force a potentially misleading interpretation. However, qualitative findings cannot be extended to wider populations as some specific ones cannot be considered statistically significant or representative or as just chance. The quantitative approach enables researchers to get a precise picture of the frequency and rarity of particular phenomena with relative normality or abnormality, though it may entail a certain part of idealized data and, at the same time, sideline rare occurrences. (ibid, p. 76-77)

The methods and perspectives of corpus linguistics thus offer us the foundation for analyzing a large number of texts and annotating specific language structures or theoretical categories according to research purposes, just as what Wallis & Nelson (2001) have put forward: the "3A perspective" (Annotation - Abstraction - Analysis). Though annotation can usually be undertaken by hand, automatically, or with some assistance from a computer program, and can tag features relating to lexis (lemmas, parts-of-speech (POS) tags, or semantic characteristics), syntax (parsing), and/or discourse features (co-reference relations, functions, or stylistic characteristics), etc., (Hyland & Shaw, 2016, p. 212) annotation of texts in this project can be found in Section 3.1 and SysFan part before that only lexis (semantically) and clause complexes are annotated manually with the basic corresponding corpus-building information described in Section 4.1. Abstraction consists of the translation (mapping) of grammatical terms in the system to terms in a theoretically motivated model or dataset. As statistics can reveal whether the finite samples available to us are in any probable sense representative of the language as a whole when it is applied directly to data-intensive linguistics, it has been a general tool for handling finite samples of potentially infinite (or at least unmanageably large) datasets in corpus linguistics. (Brew & Moens, 2004, p. 14) Analysis then includes statistical probes, manipulations, and generalization from the dataset, evaluations, optimizations of modeling, knowledge discovery, theory construction, etc.

4.2.3.3 Modelling

Modeling is not the most important method in this project, though it will be applied in the discussion part, and detailed questions and descriptions will be demonstrated in Section 7.5 when the measurement of the Complexity of Clause Complex in a text is presumed to be conducted for comparison. From the probability distribution of statistics that are yielded from the corpus, inductive reasoning and hypotheses testing are completed through *JASP*, and then the numerical analysis method is applied to optimize a functional integration for research purposes.

4.2.3.4 JASP & Bayesian Independent T-test

JASP Introduction

With the minimalism principle, *JASP* can afford every information that is needed with more in optional selection columns. The software can be downloaded and used for free from its website²⁵, which can not only be an alternative to *SPSS* but can also provide complex calculations such as Bayesian analyses. Data can be uploaded and stored in the computer as comma-separated values (.csv) files.

Descriptive Statistics

After data are split into groups with corresponding variables, boxplot, and violin elements will be demonstrated, though these two are almost the same in their functions for continuous variables. As data are not randomly scattered, scatter plots are not included. Also, distribution or correlation plots are excluded because of data characteristics, which are shown in Figure 20.

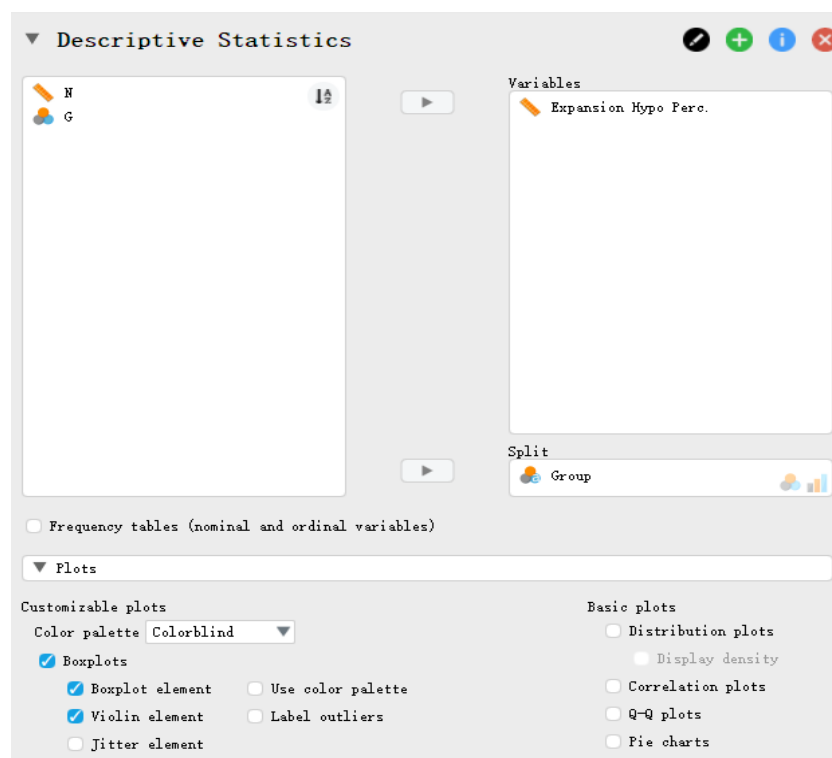


Figure 20 The plots choosing interface of descriptive statistics in JASP

As data distribution can generally be estimated from boxplots, various parameters are calculated to see differences between these two groups. In detail, Central Tendency (mean, median, mode²⁶, sum), Dispersion (standard error of the mean²⁷, standard deviation²⁸, variance, range, minimum value, and maximum value), and Distribution (See) are all considered as significant in descriptive statistics that skewness tells the distortion or asymmetry that deviates from the symmetrical bell curve, or normal distribution, in the distribution of the data; kurtosis refers to measure of the combined weight of a distribution's tails relative to the center of the distribution; Shapiro-Wilk test is a test of normality that

²⁵ <https://jasp-stats.org/>

²⁶ If more than one mode exists, only the first one will be reported.

²⁷ It equals the standard deviation by dividing the square root of the number of samples.

²⁸ 68% of the data lie within the mean plus the standard deviation and the mean minus the standard deviation; 95% of the data lie within the mean plus two times the standard deviation and the mean minus two times the standard deviation.

correlation coefficient gets closer to 1 would indicate that the sample has a higher matching degree of normality distribution.

The screenshot shows the 'Statistics' panel in JASP. It is divided into several sections with checkboxes for different statistical tests and measures:

- Percentile Values:**
 - ☐ Quartiles
 - ☐ Cut points for: 4 equal groups
 - ☐ Percentiles:
- Dispersion:**
 - ☒ S. E. mean
 - ☒ Std. deviation
 - ☐ MAD
 - ☐ MAD Robust
 - ☐ IQR
 - ☒ Variance
 - ☒ Range
 - ☒ Minimum
 - ☒ Maximum
- Central Tendency:**
 - ☒ Mean
 - ☒ Median
 - ☒ Mode
 - ☐ Sum
- Distribution:**
 - ☒ Skewness
 - ☒ Kurtosis
 - ☒ Shapiro-Wilk test

Figure 21 The testing parameters of descriptive statistics in JASP in this project

The positive value of skewness means a right-side, and the negative a left-side skewness with a bigger value means longer tails seen from the coordinate axis. When the ratio of skewness to standard error of skewness is less than -2 or more than +2, then the normal distribution of the sample can be neglected. When the value of kurtosis is 0, the sample is near to normal distribution; a positive value means leptokurtic distribution, and a negative value means platykurtic distribution. Again, if the ratio of kurtosis to the standard error of kurtosis is less than -2 or more than +2, the normal distribution of the sample can also be neglected. As to the P-value of Shapiro-Wilk, the data is abnormally distributed when it is less than 0.05.

Independent Samples T-test

The screenshot shows the 'Independent Samples T-Test' panel in JASP. It includes several sections for configuring the test:

- Variables:** Expansion Hypo Perc.
- Grouping Variable:** Group
- Tests:**
 - ☒ Student
 - ☒ Welch
 - ☒ Mann-Whitney
- Alt. Hypothesis:**
 - ☐ Group 1 $\neq$ Group 2
 - ☒ Group 1 > Group 2
 - ☐ Group 1 < Group 2
- Assumption Checks:**
 - ☐ Normality
 - ☐ Equality of variances
- Additional Statistics:**
 - ☐ Location parameter
 - ☐ Confidence interval 95.0 %
 - ☒ Effect Size
 - ☒ Cohen's d
 - ☐ Glass' delta
 - ☐ Hedges' g
 - ☐ Confidence interval 95 %
 - ☐ Descriptives
 - ☐ Descriptives plots
 - Confidence interval 95.0 %
 - ☒ Vovk-Sellke maximum p-ratio
- Missing Values:**
 - ☒ Exclude cases per dependent variable
 - ☐ Exclude cases listwise

Figure 22 The interface of the Independent Samples T-test in JASP

The independent samples t-test allows researchers to estimate the effect size (either measures the sizes of association between variables or the sizes of differences between group means and is independent of sample size) and test the null hypothesis that the population means of two independent groups are equal on the assumptions that 1. the dependent variable is continuous; 2. the observations in both groups are a

random sample for population; 3. the dependent variable is normally distributed in both populations; and 4. the population variances in the two groups are homogeneous, it has been used as the main test to check the grammatical variances between Master's and Doctor's groups in this project. (See Figure 22)

When the variances of the two populations are equal, the Student's test is referred to; when they are unequal, Welch's index is chosen. The t-value of these two tests is calculated by the following equation:

$$t = \frac{m_1 - m_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

in which m_1 and m_2 are the means of each group, n_1 and n_2 are the sample sizes, and S_1^2 and S_2^2 are the variances. "df" is the degree of freedom that for the student's t-test, $df = n_1 + n_2 - 2$, and for the Welch's,

$$df = \frac{(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2})^2}{\frac{(\frac{S_1^2}{n_1})^2}{n_1 + 1} + \frac{(\frac{S_2^2}{n_2})^2}{n_2 + 1}} - 2$$

SE Difference is the standard error of the difference in means. As the Mann-Whitney U test has no requirement for the distribution of two samples, if the variable in one sample is abnormally distributed, the comparison of variances between two groups can be demonstrated by Mann-Whitney U's index. Usually, the default selection (Group 1 $\neq$ Group 2) of the alternative hypothesis is that the population means are equal (two-sided), "Group 1 > Group 2" is a one-sided alternative hypothesis that the population mean of Group 1 is larger than that of Group 2; while "Group 1 < Group 2" means that the population mean of Group 1 is smaller than that of Group 2. For the Student t-test and Welch t-test, the effect size in this project is presented by Cohen's d index (the pooled standard deviation standardizes the mean difference) as Cohen (1988, p. 40) suggested that $d=0.2$ can be considered as a "small" effect size, 0.5 represents a "medium" effect size and 0.8 a "large" effect size. If the difference between the two groups' means is less than 0.2 standard deviations, then the difference is negligible, even if it is statistically significant. For the Mann-Whitney U test, the effect size is given by the rank biserial correlation (Cureton, 1956), and it can be calculated through the U value that

$$r = \frac{2U_1}{n_1 n_2} - 1$$

in which U_1 relates to the U value from the larger or more favorable group, and n_1 , and n_2 are the number of two samples.

The calculation of the Mann-Whitney U test value is as follows:

$$U = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - \sum_{x_i} Rank(x_i)$$

In which n_1 and n_2 refer to the number of two samples. The test also assesses whether the Hodges-Lehmann estimate of the difference in central tendency between the two samples differs from zero, i.e., the location parameter in the Mann-Whitney U test is given by this estimate, which is the median of all possible differences between two observed samples.

The Vovk-Sellke Maximum p -Ratio is obtained by choosing the shape α of the distribution under H_1 so that the obtained p -value is *maximally diagnostic* (Vovk, 1993; Sellke et al., 2001). The bound $1/(-e p \log(p))$ is derived from the shape of the p -value distribution that it is uniform (0, 1) under the null hypothesis (H_0) while it decreases in p under the alternative (H_1), e.g., a beta (α , 1) distribution ($0 < \alpha < 1$). (See Figure 23)

The value thus is the ratio of the densities at point p under H_0 and H_1 , e.g., if the two-sided p -value equals .05 ("reject the null"), the Vovk-Sellke MPR equals 2.46 by indicating that this p -value is at most 2.46 times more likely to occur under H_1 than under H_0 .

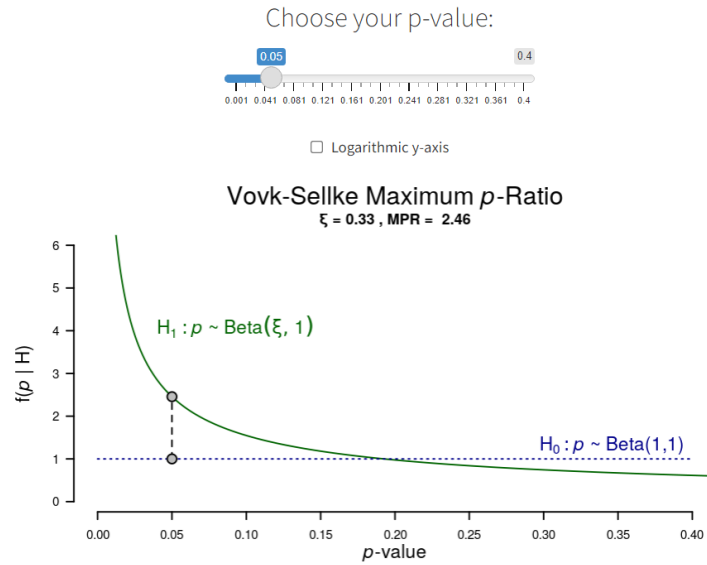


Figure 23 The Vovk-Sellke maximum p-ratio calculation application²⁹

Bayesian Independent Samples T-test

Because traditional independent Samples T-test can only tell that two groups are the same or have some differences while the Bayesian Independent Samples T-test can also tell researchers how big differences these two groups have, thus it is also considered an important factor in comparing grammatical choice variances conducted in this research. In Bayesian statistics, $P(d|h)$ refers to the likelihood of the observed data d given hypothesis h , the joint probability of the hypothesis, and the data is written as $P(d, h)$, and it can be calculated by multiplying the prior $P(h)$ by the likelihood $P(d|h)$, i.e., $P(d, h) = P(d|h)P(h)$. As dividing the joint probability of the hypothesis and the data $P(d, h)$ by the marginal probability of the data $P(d)$ gives out the posterior probability of the hypothesis given that the data have been observed, i.e.,

$$P(h|d) = \frac{P(d, h)}{P(d)}$$

to substitute $P(d, h)$ with $P(d|h)P(h)$, we get the formula for the posterior probability:

$$P(h|d) = \frac{P(d|h)P(h)}{P(d)}$$

This formula is the Bayes' rule. When a null hypothesis h_0 and an alternative hypothesis h_1 are compared, the posterior probability of the null hypothesis using Bayes' rule is:

$$P(h_0|d) = \frac{P(d|h_0)P(h_0)}{P(d)}$$

which shows the degree of belief that should have in the null hypothesis after observing the data d , and the same for the alternative hypothesis:

$$P(h_1|d) = \frac{P(d|h_1)P(h_1)}{P(d)}$$

Thus, equations can be written as:

$$\frac{P(h_1|d)}{P(h_0|d)} = \frac{P(d|h_1)}{P(d|h_0)} \times \frac{P(h_1)}{P(h_0)}$$

On the left side of the equation above, the posterior odds $\frac{P(h_1|d)}{P(h_0|d)}$ means the relative plausibility of the null hypothesis and the alternative hypothesis after seeing the data; on the right side, the prior odds $\frac{P(h_1)}{P(h_0)}$ indicates plausibility before seeing the data; and in the middle, the **Bayes factor** $\frac{P(d|h_1)}{P(d|h_0)}$ quantifies the degree/strength of evidence provided by the data and is the main index that is reported by researchers in the Bayesian statistics. By convention, equal consideration would be given to the null and the

²⁹ See <https://www.shinyapps.org/apps/vs-mpr/>

alternative hypothesis, which means that the prior odds $\frac{P(h_1)}{P(h_0)}$ equals 1, then the posterior odds $\frac{P(h_1|d)}{P(h_0|d)}$ becomes the same as the Bayes factor $\frac{P(d|h_1)}{P(d|h_0)}$.

For the interpretation of the Bayes factor as shown in Figure 24, “1-3” means negligible evidence, and “3-20” means positive evidence, which are common results that will be demonstrated in the analysis sections.

Bayes factor	Interpretation
1 - 3	Negligible evidence
3 - 20	Positive evidence
20 - 150	Strong evidence
>150	Very strong evidence

Figure 24 The interpretation of the Bayes factor (Jeffreys, 1961; Kass & Raftery, 1995)

Figure 25 The interface of the Bayesian Independent Samples T-test in JASP

In JASP (See Figure 25), $BF_{10} = \frac{P(data|H_1)}{P(data|H_0)}$, BF_{+0} thus indicates the comparison of H_+ against H_0 , i.e., the probability of accepting H_+ vs. H_0 (the value of which usually should >1) as compared to BF_{0+} (the probability of accepting H_0 vs. H_+). In this project, if H_0 (the original hypothesis) means that $D=M$, then H_+ (the alternative hypothesis) indicates that $D \neq M$; if H_0 (the original hypothesis) means that $D < M$, then H_+ (the alternative hypothesis) indicates that $D > M$, which will be particularly demonstrated in Section 5 & 6. (Rouder et al., 2009) Again, if the sample is non-normally distributed, the Mann-Whitney U test would be conducted. “Prior and posterior” (See Figure 26) test the prior (dashed line) and posterior (solid line) distribution of the effect size under the alternative hypothesis with the grey circles representing the height of the prior and the posterior density at the effect size $\delta=0$. The horizontal solid line shows the width of the 95% credible interval of the posterior distribution with its median, the detailed Bayes factor is clearly computed, and a probability wheel describing the odds of the data under the null vs. hypothesis is also displayed.

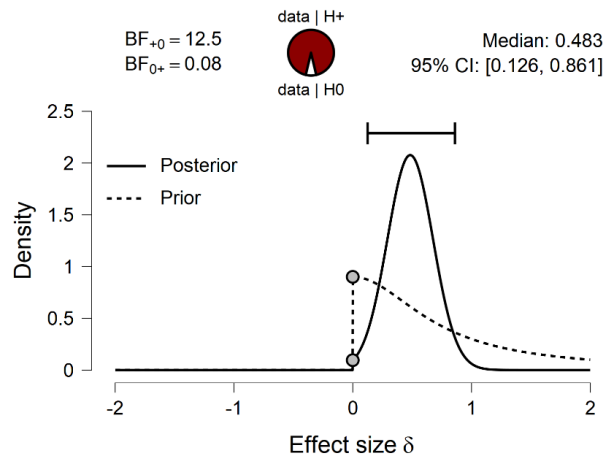


Figure 26 The inferential plot interface of prior and posterior in JASP (words per clauses of Masters (H0) and PhD (H+) groups as an example)

Bayes factor robustness check (See Figure 27) displays the Bayes factor as a function of the width of the Cauchy prior (the scale ranges from 0 to 1.5) on effect size that circles in different colors represent the Bayes factor computed with different priors.

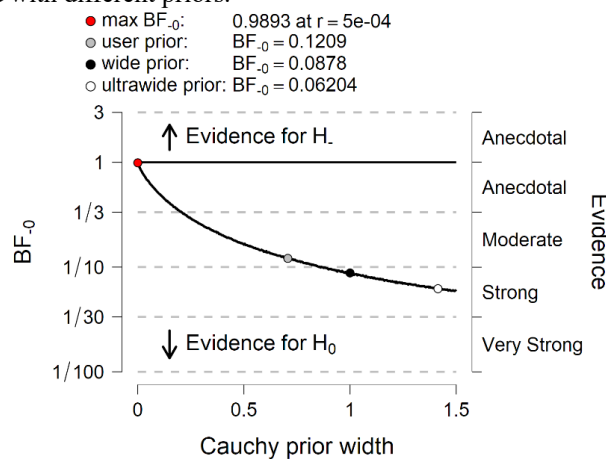


Figure 27 The inferential plot interface of the Bayes factor robustness check in JASP (clause nexus percentage on Level 1 of Masters (H0) and PhD (H+) groups as an example)

The sequential analysis (See Figure 28) displays the development of the Bayes Factor as a function of the number of observations (n) by using the user-defined prior (as well as the wide and the ultrawide prior distribution) that, with the decisiveness of the evidence from Jeffreys' (1961) categories showed on the left side.

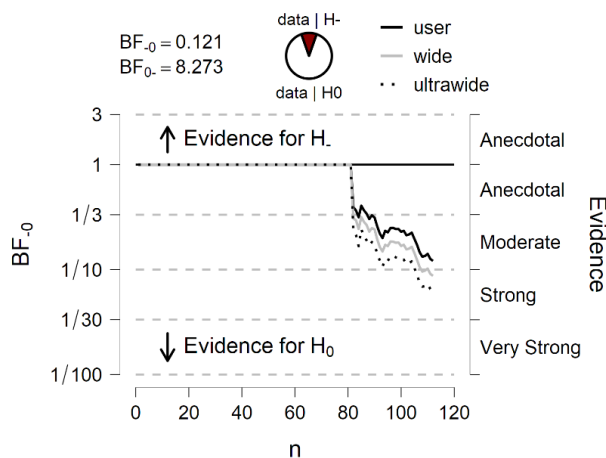


Figure 28 The inferential plot interface of the sequential analysis in JASP (clause nexus percentage on Level 1 of Masters (H0) and PhD (H+) groups as an example)

As the three plots described above are of equal importance in explaining how different Master and PhD groups are for clause complex choices in thesis writing, they will be listed one by one after each

Bayesian independent samples t-test is conducted. In the output table, “error%” is the error of the Gaussian quadrature integration routine used for the computation of the Bayes factor. “W” signifies the Wilcoxon rank sum test result (van Doorn et al., 2020) and corresponds to the sum of the ranks of the first sample with the minimum value subtracted or not. “Rhat” is a measure of MCMC convergence (Morey et al., 2011) for the Wilcoxon test, a ratio by comparing the between- and within-chain variance of the MCMC estimates for the delta parameter. If the value is less than or equal to 1, then the chain variance is convergent.

4.2.4 Summary

To make a conclusion about the research methods that are applied in this project, both qualitative and quantitative methods are adopted because of their advantages and also disadvantages. The former is more concerned with logo-genesis, while the latter is phylogenesis. They are not merely combined but also intersected within diverse content that a mixed-methods research is actually conducted for better explain analyzing process. The qualitative methods include exploratory (heuristic) analysis, comparative research, experiential summary, grounded theory, and case study. The main software for corpus annotation and recording is SysFan. Quantitative methods contain descriptive research, corpus linguistics, modeling, and statistical analysis. Though AntConc (with SysFan) is also used for lexical markers calculating, it is not described in this part as its secondary status (will be simply presented in Section 5&6). JASP is important because of its descriptive results of the data, its independent samples t-test, and Bayesian independent samples t-test for data comparison in this project.

4.3 Chosen Data: CMAC & CPhD

4.3.1 Corpus Design

In corpus designing, the research purpose determines the target population (e.g., language(s), language variety, genre, register, etc.). Thus, the target population in this research is English theses written by Chinese students for comparison. The representativeness, i.e., “the extent to which a sample includes the full range of variability in a population” (Biber, 1993p. 243), comes from two aspects: 1. balance or the range of genres and registers included in the corpus (broad range of genres is essential for general corpora while closure (saturation) at lexical level is essential for specialized corpora); and 2. sampling techniques or how the text excerpts for each genre are selected e.g., simple random sampling and stratified random sampling (proportionality emphasized). For CMAC (Chinese Masters’ corpus) and CPhD (Chinese PhDs’ dissertations), though thesis writing genres are enough for lexical level analysis, still eight directions such as translation, linguistics, cultural studies, etc. are classified (See Section 4.1) for grammatical level comparison, and though these specific directions are not the main reasons that lead to differences, this is where sampling stratification is realized in with the range of years together. So, the balanced instead of pyramidal or opportunistic designing method is adopted in corpus construction. Though language would change over time and diachronic corpora are used to monitor the dynamic of language development, synchronic corpora that are used for static language modeling change over time are still representative of the period chosen while designing the corpus. Thus, the synchronic instead of the diachronic method remains the focus of corpus research in this project, except that the time of the sampling lasts from 2005 to 2020. For the language of the corpus, of course, living languages that are used to examine naturally occurring linguistic feature distributions are the research object, and it is monolingual (not multilingual) as it only contains the English language; original (not translations) texts that are produced by English language learners instead of native speakers; and written (not spoken) materials that are downloaded from websites and stored as electronic text in .txt. or .doc format. Because the corpus would not be broadened automatically or manually, it is a fixed-, predefined-size corpus that relates to research purpose and is thus not too exhaustive for collecting and annotation. The size could not be too large or too small because using a large corpus enables researchers to make relatively robust generalizations about the occurrences of features in a given genre, and using a small corpus allows the analyst to delve deeper into the complexities of certain linguistic features. In a smaller corpus, for example, the researcher can code the data comprehensively in order to identify features that cannot all be identified by surface formal features alone. (Hyland & Shaw, 2016, p. 386) The scope of CMAC and CPhD is a medium for mark-up analysis in full texts (not text chunks) so as to improve corpus representativeness and discover design lapses and errors at any moment.

4.3.2 Corpora General

4.3.2.1 General Information

The relationship of clause complex is more semantic that it is necessary to be annotated within clause complexes. To eliminate the effects of thesis structures on the choice of clause complex, i.e., more elaborations might be found in the research methods part, and projection would be preferred in the literature review part, the whole thesis instead of text chunks or paragraphs or separated clause complexes is chosen as a unit in corpus building. To make statistics results more meaningful and convincing, at least 30 participants in each testing group should be included, and considering the year from 2005 (the first year of texts collected in ChAcE) till the year for annotation, i.e., 2020, the year range can be 16 years. This comes out of 32 PhD dissertations (one or two texts in each year from 2005 to 2014 in ChAcE and from 2015 to 2020 in CNKI). As seen from Section 4.1, the number of words, clauses & clause complexes (sentences) (See Appendix 4.3.2.1 for general data information of the CMAC & CPhD corpus) in a PhD dissertation is almost 2.5 times that in an MA thesis, altogether 80 MA theses are chosen (50 texts from ChAcE and 30 from CNKI separately). The detailed number of texts chosen each year and from different directions can also be referred to in Appendix 4.3.2.1.

4.3.2.2 Corpora Importing & Annotating

After choosing texts, the first thing to make sure is that irrelevant information in each text (.txt format) has been deleted. Then these texts are transformed into .doc format and stored in “Word” for annotating each clause complex: select all, and replace “.” with “.^p||”, thus all sentences are automatically organized into separate clause complexes with “Enter” at the end and blanks at the beginning. As many “.” such as those in numbers, symbols, tables, abbreviations, etc., are unnecessarily replaced, manually checking and cleaning are needed though consuming time and energy. The importing process of corpora is to open SysFan, click the “text” and “new record” menu first; paste cleaned, original texts in the “text-raw text” window, and clause complexes separated data in blanks of “grammatical boundaries - analysis” window then; and repeat all these steps to ensure that all 112 texts have been correctly imported. The annotation of clause complexes can be started by clicking the “complex ID” menu and choosing numbers or alphabets with symbols as described in Section 4.2.2.6. Because more than 0.1 million clause complexes are imported in SysFan, which is beyond the capability of manual annotation, half of each text (either odds or evens or the former half part) would be actually marked. “||” (double vertical lines) is used for departing clause nexuses, “[]” (double square brackets) for departing embedded clauses, “[]” (square brackets) for embedded phrases, “<< >>” (double angle brackets) for inserted clauses, and “< >” (angel brackets) for inserted phrases. The relationship of clause complexes lies in clauses and clauses, thus, embedding and insertion are not taken into consideration and even for calculation, except that clause complexes are embedded in a clause. By clicking “create clause records” and menus including numbers (1, 2, 3...), alphabets (α , β , γ ...), symbols ($=$, $+$, $\times$), and quotation marks (‘, “) in each level before “clause nexus” one by one, annotation is fulfilled, which takes more than six months by the researcher. In addition, keywords denoting corresponding clause complex relationships and grammatical categories are recorded in “note” windows for further calculation and searching.

4.3.3 Corpora Details

4.3.3.1 The Number of Clauses

The information on the clauses’ number in each group is demonstrated in Table 4.3.3.1-1. (see Appendix 4.3) From the mean (1765.69 in group D and 708.34 in group M), it is obvious that the average number of clauses found in a Doctoral dissertation is 2.5 times that in a Master’s thesis, which provides the choice of thesis number in each group. A larger standard deviation and smaller number of the sampling bring out a larger standard error of mean, and a larger standard deviation with a wider range in group D shows that the doctoral dissertation has a bigger variance in the number of clauses, which can also be proved by the boxplot below (Figure 29) that the highest in group D is larger than 2500 with the smallest around 1000 while in group M the highest about 1600 with the smallest number 400. The larger number of skewness means that group M has a longer tail on the right side (as both values are positive). When the value of skewness divides that of the standard error of skewness, group D is less than 1, while group M is too much larger than +2, a confirmation can then be given that the data in group M is abnormally distributed. The same conclusion can be received from the division of the kurtosis value by the standard error of the kurtosis value that D is with -2 to 0 while M is almost 10 ($> +2$). A negative value in kurtosis means a platykurtic distribution of the data, while a positive value means a leptokurtic distribution. For the Shapiro-Wilk test value, group D gets closer to 1 (0.973) than group M (0.852), one thing for sure is that

the data in group D accords with normal distribution. It couldn't be confirmed that group M is non-normally distributed until the P-value of Shapiro-Wilk is judged as less than 0.05 (which is actually $< .001$ in group M). From the sum of the clause numbers, it is almost the same in the two groups (32 texts in group D and 80 texts in group M). Thus, further statistical analysis can be conducted on this basis. One explanation of the peculiarities of clause number in group M originates from students' willingness to write theses, and they may have uneven English application abilities, though they belong to the same educational level. But, do more clauses bring out more clause complexes, and are these texts more complex in choosing clause complexes remain unknown on the surface?

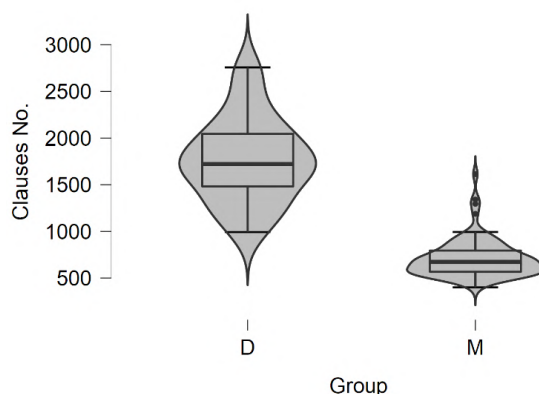


Figure 29 The distribution boxplot of the number of clauses in Master and PhD sub-corpora

4.3.3.2 Words/Clauses

For the average words used in each clause in two groups, basic information can be interpreted from Table 4.3.3.2-1 (see Appendix 4.3) that the mean of group D (25.678) is totally larger than that of group M (23.102), which means that at least two more words can be found in Doctors' dissertations in each clause(sentence) and thus they write longer clauses in theses than Master students with both larger minimum and maximum value (18.960 vs. 18.006, and 32.534 vs. 30.938). The broader range also indicates this condition. Though the two groups have almost the same standard deviation and variance, group D has nearly no skewness, while group M is skewed. With the division of skewness and standard error of skewness, group D is proved to be normally distributed, while group M is abnormal. The kurtosis of group D is only 0.028, which indicates normal distribution as it gets close to 0; while the kurtosis in group M is negative. The P-value of Shapiro-Wilk in group M ($< .001$) confirmed the abnormal distribution of its data and can also be seen in Figure 30.

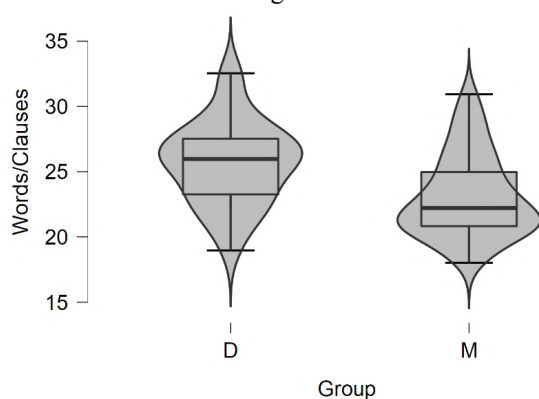


Figure 30 The distribution boxplot of the number of words in clauses in Master and PhD sub-corpora

As the results in Table 4.3.3.2-1 show that the two groups have the same variance with means being different from their medians, and the distribution of group M is skewed, the Mann-Whitney U test is thus suitable for checking data differences in these two groups. By the equation given in Section 4.2.3.4, the statistic of the Mann-Whitney U test is shown in Table 22. When the p-value is less than .001 for two-tail test, the value of VS-MPR is 382.340, which means that the maximum possible odds in favor of H_1 (the alternative hypothesis that group D is greater than group M) over H_0 (the original hypothesis that group M is greater than group D) is very large. The location parameter value of 2.864 indicates that the average number of words in clauses in group D is 2.8 more than that in group M. The effect size is 0.45 based on the equation in Section 4.2.3.4, and this value is medium. A conclusion can thus be reached that around three more words would be used in writing clauses in dissertations by PhD candidates than MA students.

Independent Samples T-Test								
	Test	Statistic	df	p	VS-MPR*	Location Parameter	SE Difference	Effect Size
Words/Clauses	Mann-Whitney	1856.000		< .001	382.340	2.864		0.450
<i>Note.</i> For all tests, the alternative hypothesis specifies that group <i>D</i> is greater than group <i>M</i> .								
<i>Note.</i> For the Mann-Whitney test, effect size is given by the rank biserial correlation.								
<i>Note.</i> For the Mann-Whitney test, location parameter is given by the Hodges-Lehmann estimate.								
* Vovk-Sellke Maximum <i>p</i> -Ratio: Based on a two-sided <i>p</i> -value, the maximum possible odds in favor of H_1 over H_0 equals $1/(-e p \log(p))$ for $p \leq .37$ (Sellke, Bayarri, & Berger, 2001).								

Table 22 The Independent Samples T-test (Mann-Whitney U test) for the number of words in clauses in Master and PhD sub-corpora

To better demonstrate the differences between the two groups, the Bayesian independent samples T-test is then performed with results shown in Table 23.

Bayesian Mann-Whitney U Test

	BF ₊₀	W	Rhat
Words/Clauses	12.500	1856.000	1.007

Note. For all tests, the alternative hypothesis specifies that group *D* is greater than group *M*.

Note. Result based on data augmentation algorithm with 5 chains of 1000 iterations.

Table 23 The Bayesian Mann-Whitney U test for the number of words in clauses in Master and PhD sub-corpora

As the data in group M has been proved to be abnormally distributed, the Bayesian Mann-Whitney U test is conducted with the Bayes Factor indicating that the probability of accepting the alternative hypothesis is 12.5 times that of accepting the original hypothesis, which gives positive evidence that group D is greater than group M.

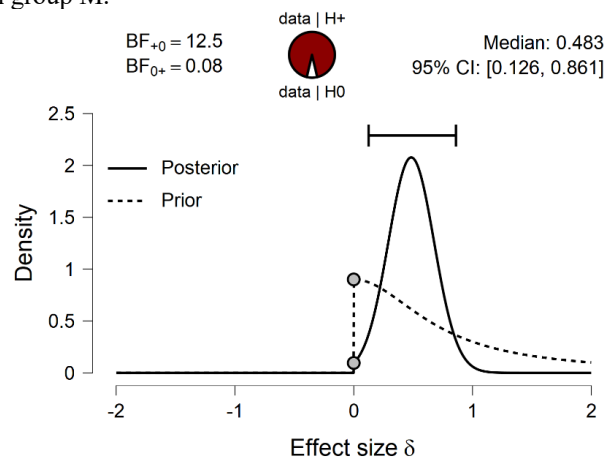


Figure 31 The inferential plot of the number of words in clauses in Master and PhD sub-corpora. Again, “W” is the Wilcoxon rank sum test result, and “Rhat” is a measure of MCMC convergence for the Wilcoxon test that the chain variance is convergent for this parameter when it nearly equals 1. The inferential plot of word numbers per clause by prior and posterior test in Figure 31 shows the tendency of difference density through effect size.

One explanation for PhD candidates using more words in a clause complex in writing theses is from contents that more complex concepts are discussed, more adjectives or adverbs are necessary for describing, and more verbs are chosen for events acting. Another is grammatical that more “additive: and” relationships with the marker *and* are applied in connecting phrases or even clauses, which might be deleted unconsciously in MA theses as an implicit phenomenon; and phrases might be more preferable by PhD candidates in transmitting the same clause complex functions as what simple words can do. But, as the research does not pay attention to the relations below clauses, the former explanation cannot be tested or proved yet. The latter, with enough data being analyzed qualitatively or quantitatively, can be answered gradually.

4.3.3.3 Clause Complexes/Sentences

Chosen CC (the CC here indicates clause complexes plus clause simplexes. The aim is to calculate how many sentences are actually chosen for annotation and what is the probability of complex sentences compared with simplex sentences in chosen sentences.)

The words of clause complexes found in sentences chosen for annotation are altogether 15.8 thousand in Doctors' dissertations and 16 thousand in Masters' theses. In each text, the former has a larger range of clause complexes, from 223 to 760, than the latter from 105 to 389. The mean of chosen clause complexes in group D is 2.5 times that in group M, as presented in Table 4.3.3.3-1 (See Appendix 4.3), thus explains the decision of sampling numbers of texts in each group that when group D has 32 texts, group M then has 80 texts. The standard error of the mean comes from the square root of the valid sampling number in each group divided by the standard deviation that larger value indicates greater dispersion of the data. The skewness of Group D is only 0.091, which means that the data in Group D is normally distributed; while that of Group M is 1.353, which is far away from zero thus, the data in Group M is abnormally distributed with the right tail to some degree long, i.e., longer texts in group M result to more chosen clause complexes. The skewness can also be judged from the division of skewness by the standard error of skewness that it is within -2 to +2 in group D (the value is 0.22) while larger than +2 in group M (5.03), then the same conclusion can be reached. By kurtosis, -0.496 of group D represents a platykurtic data distribution, while 1.9 of group M signifies a leptokurtic data distribution. When the value of kurtosis is divided by that of the standard error of kurtosis and is compared with -2/+2, group D is normally distributed with its value 0.613 (within -2 to +2), while group M is not with its value 3.57 (more than +2). The Shapiro-Wilk value in group D is closer to 1 than that in group M, and the P-value of Shapiro-Wilk is <.001(<.05). All of these parameters give evidence to the data distribution in these two groups that group D is normal but group M is non-normal distributed. The comparison can be more obvious when seen from the boxplots in Figure 32.

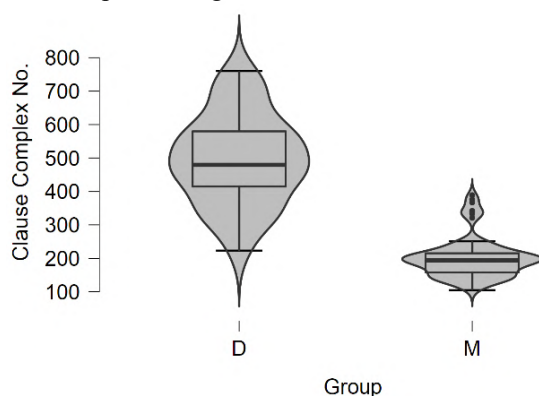


Figure 32 The boxplots of the number of clause complexes chosen for annotation in Master and PhD sub-corpora

4.3.3.4 Chosen CC/Sentences (1/2 clauses) Percentage

As sentences chosen for analysis are half of the sentences written in each text, they are traditionally called "sentence" formally but in SFG contain clause complexes and also clause simplexes (One clause is the whole sentence), in which the former structures are the research focus and for annotation. Hence, the proportion of actually chosen clause complexes as to sentences (half sentences written in each text) is calculated and compared between two groups. The results in Table 4.3.3.4-1 (see Appendix 4.3) demonstrate that the mean of clause complex/sentences in group D (56.045%) is slightly lower than that in group M (57.064%) with a maximum of 66.57% and minimum of 42.06% in group D and a range from 39.18% to 79.35% in group M. The broader range in the latter group (40.17%, can also be judged from a bigger standard deviation value in group M) comes from particular cases in which some students especially prefer to use clause complexes in expressing opinions. The skewness in group D is negative, which means that more values are lower than the mean; while in group M, it is positive, i.e., more values are beyond the line of mean value. The division of skewness to the standard error of skewness in both groups is within -2 to +2 (-0.534 in group D and 0.959 in group M), which indicates that these two groups are seemingly normally distributed. The data distributions in both groups are leptokurtic as the kurtosis values are negative, and they are both seemingly normally distributed when calculating the value of kurtosis divided by the standard error of kurtosis that both are within -2 to +2 (-0.23 in group D and -0.73 in group M). The Shapiro-Wilk value in both groups is perfectly close to 1, and the P-value is bigger than 0.05, a conclusion can be reached that the data are normally distributed in both group D and group M. But

the boxplots in Figure 33 show that the percentage also has a kurtosis within 60% and 70% except for 50% to 60%. Though the data distribution is imperfect because of the sampling size, it can be changed to normally distributed when enough samples are collected for analysis. In this research, this kind of condition is solved by regarding them still as un-normally distributed data, and the suitable test for comparison is the Mann-Whitney U test.

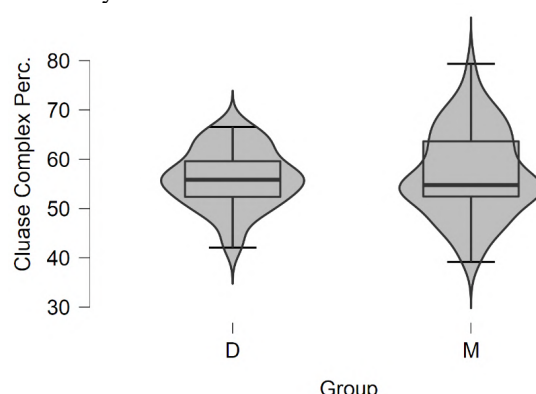


Figure 33 The boxplots of the percentage of clause complexes in chosen sentences in Master and PhD sub-corpora

Table 24 shows the results of the Mann-Whitney U-test comparing the clause complex percentage in chosen sentences in two groups. When the p-value is 0.05, the Vovk-Selke Maximum p-ratio is 2.46, which means that the maximum diagnosticity of a two-sided p-value is only 2.46 times more likely under the best H_1 than under H_0 . While the VS-MPR value here is only 1 and the corresponding p-value is 0.382, i.e., the probability of rejecting the null hypothesis is 76.4% (two-sided), thus it indicates that the probability of accepting the alternative hypothesis (group D is less than group M) is 76.4%. The location parameter value -0.49 means that the percentage of clause complex per chosen sentence in group D is 0.49% less than group M, and the effect size is only -0.037; it suggests that this test has extremely small effect size and the difference between group D and group M is actually very small, to be specific.

Independent Samples T-Test								
	Test	Statistic	df	p	VS-MPR*	Location Parameter	SE Difference	Effect Size
Clause Complex Perc.	Mann-Whitney	1233.000		0.382	1.000	-0.490		-0.037
<i>Note.</i> For the test, the alternative hypothesis specifies that group D is less than group M.								
<i>Note.</i> For the Mann-Whitney test, effect size is given by the rank biserial correlation.								
<i>Note.</i> For the Mann-Whitney test, location parameter is given by the Hodges-Lehmann estimate.								
* Vovk-Sellke Maximum p -Ratio: Based on a two-sided p -value, the maximum possible odds in favor of H_1 over H_0 equals $1/(-e p \log(p))$ for $p \leq .37$ (Sellke, Bayarri, & Berger, 2001).								

Table 24 The Mann-Whitney U test for comparing the clause complex per chosen sentences percentage in Master and PhD sub-corpora

As the independent samples t-test cannot best demonstrate the variance between two groups, the Bayesian independent samples t-test is then conducted. From Table 25, the results of the Bayesian Mann-Whitney U test (as data in two groups are regarded as abnormally distributed) show that the probability of accepting the null hypothesis (group D > group M) is 3.338 times accepting the alternative hypothesis (group D < group M). W is the result of the Wilcoxon rank sum test, and Rhat is the MCMC convergence that the chain variance is slightly dis-convergent as the value is 1.056.

Bayesian Mann-Whitney U Test

	BF ₀₋	W	Rhat
Clause Complex Perc.	3.338	1233.000	1.056

Note. For the test, the alternative hypothesis specifies that the location of group D is smaller than the location of group M.

Note. Result based on data augmentation algorithm with 5 chains of 1000 iterations.

Table 25 The Bayesian Mann-Whitney U test for comparing the clause complex per chosen sentences percentage in Master and PhD sub-corpora

The same results can be inferred from the inferential plots of prior and posterior tests in Figure 34 that the

probability of accepting the null hypothesis is 3.34 times of rejecting it, with a quite small effect size.

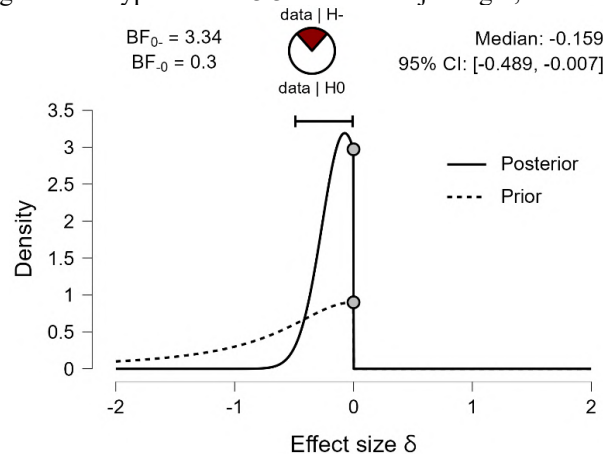


Figure 34 The inferential plots comparing the clause complex per chosen sentences percentage in Master and PhD sub-corpora with the null hypothesis group D > group M

But does Group D really have a higher percentage than Group M? When the alternative hypothesis is set up as Group D > Group M in the Bayesian Mann-Whitney U test, a different result can be reached, just as Figure 35 demonstrates that the probability of accepting the null hypothesis now is 6.212 times of rejecting it. So, the conclusion is that the percentage of clause complex per chosen sentence in group D is less than that in group M. Interestingly, when the alternative hypothesis is chosen as group D $\neq$ group M in the test interface, another result is obtained that the probability of accepting the null hypothesis is 4.34 times of rejecting it. (See Figure 36) Thus, the percentage of clause complex per chosen sentence in group D is the same as that in group M?

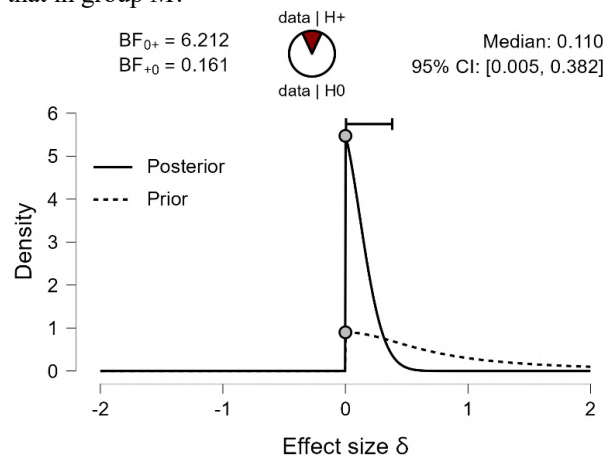


Figure 35 The inferential plots comparing the clause complex per chosen sentence percentage in Master and PhD sub-corpora with the null hypothesis group D < group M

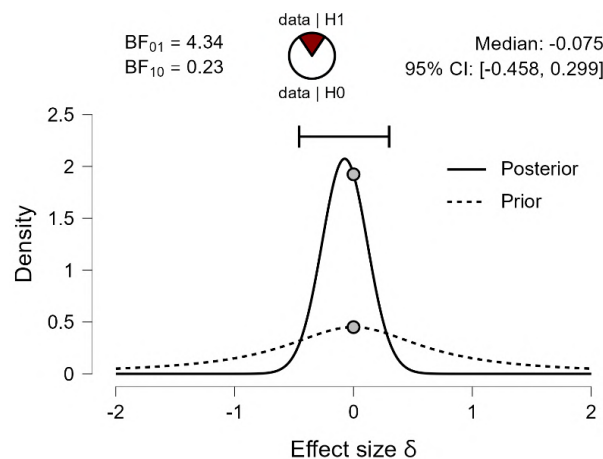


Figure 36 The inferential plots comparing the clause complex per chosen sentence percentage in Master and PhD sub-corpora with the null hypothesis group D = group M

After comparing the probabilities (group D>group M is 3.34, group D<group M is 6.21, group D=group M is 4.34) received from the Bayesian Mann-Whitney U tests above, a conclusion can be drawn that a slight difference can be found on the parameter (clause complex per chosen sentences percentage) with its effect size value equals to only 0.037. Although in the sampling, group D is to some degree lower than group M. There are reasons to believe that no difference can be discovered when the sampling is big enough. As the sample size of group M is 2.5 times that of group D, it explains that more peculiarities, i.e., students of various writing habits and language application abilities though in the same Master level, may bring to a broader range of the data, which then leads to a false appearance that Master students prefer to use more clause complexes in sentences written in their thesis than Doctoral candidates.

4.3.3.5 Clause Nexus/CC

The results of the percentage of clause complexes found in chosen sentences in each text in two groups that almost no variance can be found, as demonstrated above, provide the foundation for further comparison of whether the percentage of clause nexuses written in each clause complex within the two groups is the same or not. From the descriptive statistics in Table 4.3.3.5-1 (see Appendix 4.3), group M (0.772) has a broader range again and a bigger mean than group D (0.435) that in each clause complex, about 1.5 clause nexuses are written, i.e., 3 clause nexuses appeared in 2 clause complexes. So, structures of sentences in theses from this corpus can be the free combination of two-clause simplexes, one-clause complex (one-clause nexus), and one-clause complexes (two-clause nexuses) with a high probability. The standard deviation of group M (0.172) is bigger than that of group D (0.107), which means that the former has a larger data difference because of its bigger sample size compared with the latter. Both groups have a kind of skewness to the right side, as their positive values have shown.

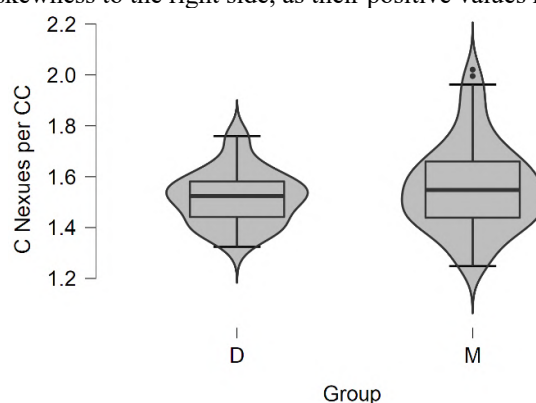


Figure 37 The boxplots of clause nexuses per clause complex in Master and PhD sub-corpora

When the division of the skewness to the standard error of skewness is conducted, and their results are compared to $-2/+2$, group D is 0.85, and group M is, however, 2.25, which denotes normally and abnormally distributed data separately. As to the kurtosis, the value of group D is almost zero (normally distributed), and that of group M is 0.248 (not very leptokurtic); and the division of the kurtosis to the standard error of kurtosis in group D is 0.109, and that in group M is 0.466, both of which are within the range of -2 to $+2$. The Shapiro-Wilk value in both groups is nearly 1, while the P-value of Shapiro-Wilk in group M is 0.05, which brings the dispute to the judgment of data distribution of group M. Thus, the boxplots in Figure 37 are necessary for presenting whether the data are normally or abnormally distributed in group M. Except for two cases, the data seems to be perfect in group M.

Independent Samples T-Test						
	W	df	p	VS-MPR*	Hodges-Lehmann Estimate	Rank-Biserial Correlation
C Nexuses per CC	1115.000		0.145	1.315	-0.030	-0.129
<i>Note.</i> For the Mann-Whitney test, effect size is given by the rank biserial correlation.						
<i>Note.</i> For all tests, the alternative hypothesis specifies that group D is less than group M.						
<i>Note.</i> Mann-Whitney U test.						
* Vovk-Sellke Maximum p -Ratio: Based on a two-sided p -value, the maximum possible odds in favor of H_1 over H_0 equals $1/(-e p \log(p))$ for $p \leq .37$ (Sellke, Bayarri, & Berger, 2001).						

Table 26 The Mann-Whitney U test for clause nexuses per clause complex in Master and PhD sub-corpora Again, the Mann-Whitney U test is consulted for comparison between these two groups that the probability of rejecting the null hypothesis (group D > group M) is 0.29 (two times the p -value because of

two tails of data distribution) and when the p-value is 0.145, the VS-MPR is 1.315 that the alternative hypothesis is 1.315 times the null hypothesis for acceptance. (See Table 26)

The Hodges-Lehmann Estimate is the value of the location parameter, and the difference between groups D and M is only 0.03. As the effect size is given by the rank biserial correlation, its value is -0.129, which suggests a low effect size. One conclusion can be that the usage of clause nexuses per clause complex in group D is a little bit less than that in group M, though the variance is not quite obvious. The more evident comparing results can be demonstrated in the Bayesian Mann-Whitney U test (Table 27) and its referential plots (Figure 38) of clause nexuses per clause complex that the probability of accepting the null hypothesis is 8.8 times of accepting the alternative hypothesis, which means that group D actually has fewer clause nexuses per clause complex than group M, though the effect size of this test is very small.

Bayesian Independent Samples T-Test

Bayesian Mann-Whitney U Test

	BF ₀₊	W	Rhat
C Nexuses per CC	8.800	1115.000	1.047

Note. For all tests, the alternative hypothesis specifies that the location of group *D* is greater than the location of group *M*.

Note. Result based on data augmentation algorithm with 5 chains of 1000 iterations.

Table 27 The Bayesian Mann-Whitney U test for the clause nexuses per clause complex in Master and PhD sub-corpora

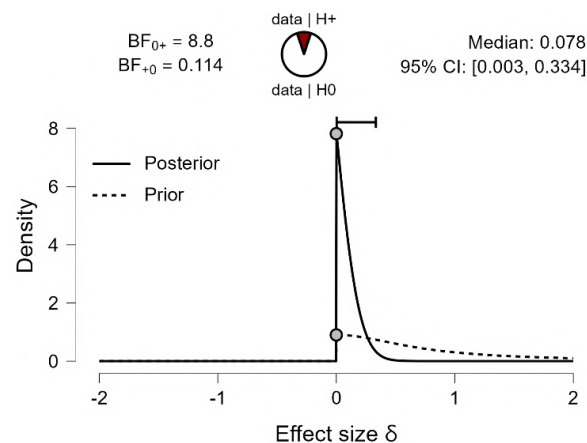


Figure 38 The inferential plots of the clause nexuses per clause complex in Master and PhD sub-corpora

To give a conclusion here, two particular cases in group M with a large number of clause nexuses per clause complex result in differences in this parameter. In predicting the population, this variance can actually be eliminated and ignored when more samples are collected for testing.

4.3.3.6 Clause Nexuses (1-7) Information

4.3.3.6.1 General Distribution

Before moving on to the next parameter for comparison, an example (E.g., 4.3.1) for clause complex annotation is chosen, and the language recursion and depth of grammar are clearly demonstrated in the SysFan interface (Figure 39):

Clause ID	7	6	5	4	3	2	1	Clause
CMAC15_							1	Amir blames himself badly,
CMAC15_					1	1	x2	for he betrayed Hassan
CMAC15_					+2			and planned to frame Hassan,
CMAC15_					α	x2		so in Clause 7 he said
CMAC15_				α	" β "			he "was a liar, a cheat, and a thief",
CMAC15_			α	= β				which intends
CMAC15_		1	" β "					that he keeps the dirty secret,
CMAC15_		+2						deceives his father
CMAC15_		+3						and conceals the truth

Figure 39 The clause complex recursion annotation interface of E.g., 4.3.1 in SysFan

E.g., 4.3.1 ||| Amir blames himself badly, || for he betrayed Hassan || and planned to frame Hassan, || so in Clause 7 he said || he "was a liar, a cheat, and a thief," || which intends || that he keeps the dirty secret, ||

deceives his father || and conceals the truth. || ---- CMAC15_05_590

The linear organization of clauses cannot reveal in which level the clause is and which clauses are actually on the same level, though they might not be in the same clause nexus. But this can be fulfilled in SysFan that in the example, the paratactic enhancement: cause, purpose relationship marked by *for* is on the 1st level; the paratactic enhancement: cause, reason relationship marked by *so* is on the 2nd level; the paratactic extending: addition, additive relation signed by “and” and the hypotactic locution: verbal: proposition relation marked by *say* are on the 3rd level; the hypotactic elaboration: clarification relation noted by *which* are on the 4th; the hypotactic idea: mental, proposition relation noted by *intend* is on the 5th; and two paratactic extending: addition, additive relations marked by an implicit one and the word *and* are on the 6th. Usually, clause complexes written in theses by Master and Doctoral students can include clauses located on levels from 1 to 3, not so much on level 4, and seldom on levels 5 and even 6, not to mention level 7. The detailed level distribution of clauses in each text can be referred to in Appendix 4.3.3.6.1. The total number of clause nexuses in group D is 24,122, and in group M is 25,326. These clause nexuses are normally distributed with a platy-kurtosis (-0.288) in group D as judged from its skewness (0.053, close to zero), the division of skewness to the standard error of skewness (0.128, within -2 to +2), the division of kurtosis to the standard error of kurtosis (-0.36, within -2 to +2), the Shapiro-Wilk value (0.986, closes to 1), and the P-value of Shapiro-Wilk (0.946 >0.05); while are abnormal, leptokurtic distributed in group M with its skewness larger than 0 (1.166, right-sided), the division of skewness to the standard error of skewness bigger than +2 (4.33), the kurtosis (1.203), the division of kurtosis to the standard error of kurtosis bigger than +2 (2.26), and the P-value of Shapiro-Wilk <0.001. (See Table 4.3.3.6.1-1 in Appendix 4.3) In fact, the data distribution in group M has a strange shape, as showed in Figure 40 that 5 cases particularly have a big number of clause nexuses ranging from 500 to 800, and two kurtoses can actually be found within 200 to 400 in each text.

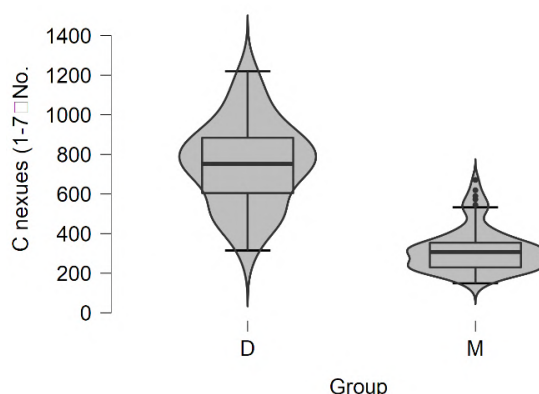
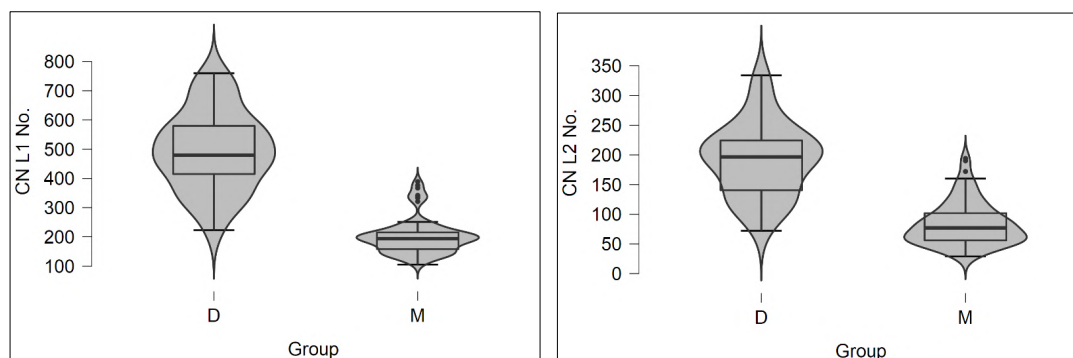


Figure 40 The boxplots of the clause nexuses in Master and PhD sub-corpora

On levels 1-3, the basic information of the clause nexus numbers between two groups is represented in Table 4.3.3.6.1-2 (See Appendix 4.3). The mean in group D is almost 2.5 times that in group M on Level 1 (495 in group D and 201 in group M) while just 2 times on Level 3 (51 in group D and 25 in group M). Comparing the sum of clause nexuses on each level in two groups, the number of group D is always less than that of group M, in which the variance on Level 2 (6,161 in group D while 6,697 in group M) is the biggest. Seen from the value of the skewness (with 0), the division of the skewness by the standard error of skewness (within -2/+2), the kurtosis (with 0), the division of the kurtosis by the standard error of the kurtosis (within -2/+2), and the P-value of Shapiro-Wilk (>0.05), Group D is normally distributed with platy kurtosis to the left side while group M is abnormally distributed with leptokurtosis to the right side on each level. From Figure 41, it is obvious that the number of clause nexuses in group D is normally distributed from Level 1 to Level 3, and the same condition can be found in group M when 6 cases are excluded on Level 1, 3 on Level 2 and 3 on Level 3.



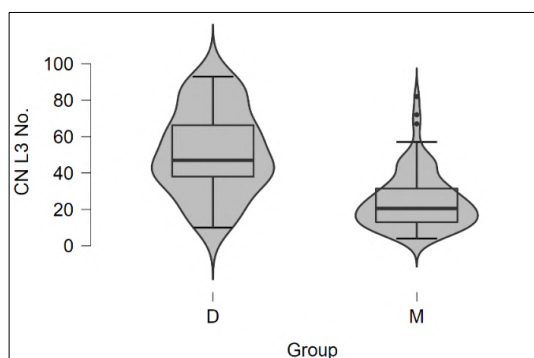


Figure 41 The boxplots of clause nexuses on levels 1-3 in Master and PhD sub-corpora

On levels 4-7, the basic information of the clause nexus numbers between the two groups is shown in Table 4.3.3.6.1-3 (see Appendix 4.3) as not all texts have clause nexus written on Levels 4 to 7, parts of texts are thus calculated in the Missing column. The sum in group D is less than in group M because of more clauses annotated in the latter group, and more peculiarities are found from a larger sample size. From around 25 thousand clause nexuses, only 30 and 32 are on Level 4 in each group, 11 (D) and 14 (M) on Level 5, 2 (D) and 6 (M) on Level 6, and 1 in each on Level 7. So, the recursion of clauses can be a tough challenge for both Master's and Doctoral students because of their language inability in academic writing or simply logical thinking in organizing or just unconsciousness of clause recursion. Being different from levels 1-3, all clause nexuses are abnormally distributed in each group from Level 4-7 as uniqueness.

To eliminate the effects of the numbers on the comparison of clause nexuses between Master theses and PhD dissertations, the parameter of clause nexuses (CNs) on each level per clause nexuses found in each text is regarded as an important indication of choosing difference in theses writing. As the choice of clause nexuses on levels 5-7 is rare, only the percentages of clause nexus choice on levels 1-4 are thus compared within two groups. From Table 4.3.3.6.1-4 (see Appendix 4.3), the mean of the CNs percentage in each text of group D is a little bit larger than that of group M on Level 1 (66.17 versus 64.88) while smaller on Level 3 (6.64 vs. 7.33), and almost the same on Level 2 (25.34 vs. 25.83) and Level 4 (1.52 vs. 1.55), which means that some Master students choose to use clause complexes with more recursion clauses in conveying their ideas. The broader ranges in group M on each level can only be explained by a larger sample size, and more students bring more possible choices to the data. In all 7 levels, around 2/3 of clause nexuses are on level 1, 1/4 are on level 2, and the remaining 10% are for levels 3-7, which denotes that students would usually stop recursing after two clause nexuses because these have already been enough for thoughts transmitting. The negative values of skewness mean left-sided skewness of data distribution, and the negative values of kurtosis mean leptokurtic distribution. As these values are within $-1/+1$, data are normally distributed. The same conclusion can be received from the values of skewness divided by the standard error of skewness and that of the kurtosis divided by the standard error of kurtosis when they are within $-2/+2$. One exception is the skewness divided by the standard error of skewness on Level 4 in group M. As the P-value of Shapiro-Wilk in this column is less than 0.05, the clause nexus percentage on Level 4 in group M is abnormally distributed. Thus, the Mann-Whitney U test would be used for comparison in two groups on Level 4, while the Welch t-test would be used for levels 1-3 in two groups as they have different variances.

4.3.3.6.2 Clause Nexuses on Level 1

On Level 1, more texts have a clause nexus percentage ranging from 65% to 70% in Group D and from 60% to 65% in Group M. (See Figure 42)

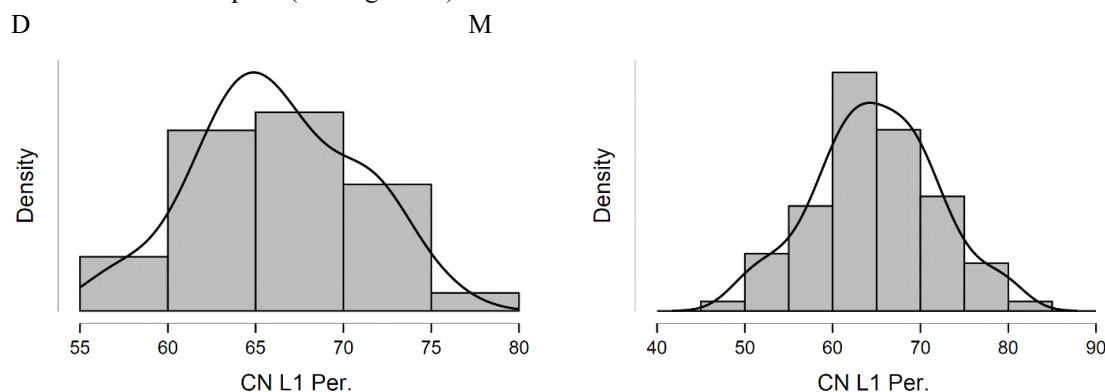


Figure 42 The distribution plots of clause nexuses percentage on Level 1 in Master and PhD sub-corpora

In the Welch t-test (Table 28), as the t-value (1.153), location parameter (1.294) and SE difference (1.123) are all positive, the percentage of CNs on Level 1 of group D is a little bigger than that of group M, and the p-value>0.05, the alternative hypothesis is accepted than the null one and the probability of accepting the former is 1.409 times of rejecting it (see VS-MPR). But with a low effect size (0.22), the independent samples t-test (Welch) primarily proves the small difference between group D and group M.

Independent Samples T-Test								
	Test	Statistic	df	p	VS-MPR*	Location Parameter	SE Difference	Effect Size
CNs L1 Per.	Welch	1.153	85.354	0.126	1.409	1.294	1.123	0.220
<i>Note.</i> For the test, the alternative hypothesis specifies that group <i>D</i> is greater than group <i>M</i> .								
<i>Note.</i> For the Welch t-test, effect size is given by Cohen's d.								
<i>Note.</i> For the Welch t-test, location parameter is given by mean difference.								
* Vovk-Sellke Maximum <i>p</i> -Ratio: Based on a two-sided <i>p</i> -value, the maximum possible odds in favor of H_1 over H_0 equals $1/(-e p \log(p))$ for $p \leq .37$ (Sellke, Bayarri, & Berger, 2001).								

Table 28 The Welch's test of the clause nexus percentage on Level 1 in Master and PhD sub-corpora

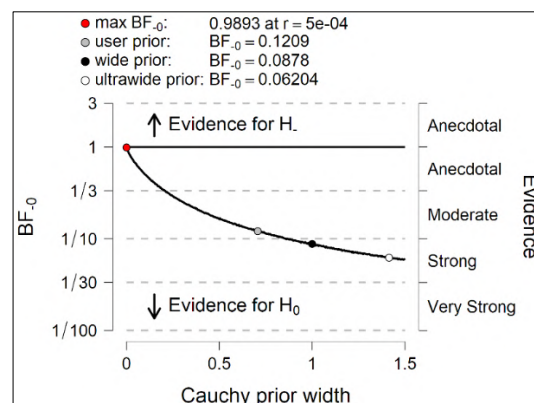
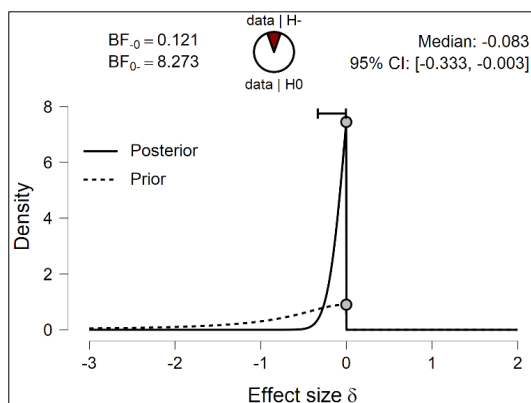
In the Bayesian independent samples t-test (Table 29 & Figure 43), the probability of accepting the alternative hypothesis is only 0.121 times accepting the null hypothesis, with group D being larger than group M. Thus, the clause nexus percentage on Level 1 in Group D is larger than that in Group M, as confirmed in Welch's test, with quite a small effect size (see the prior and posterior inferential plots).

Bayesian Independent Samples T-Test		
	BF ₀	error %
CNs L1 Per.	0.121	~0.004
<i>Note.</i> For all tests, the alternative hypothesis specifies that group <i>D</i> is less than group <i>M</i> .		

Table 29 The Bayesian independent t-test of the clause nexus percentage on Level 1 in Master and PhD sub-corpora

Prior and Posterior

Bayes Factor Robustness Check



Sequential Analysis

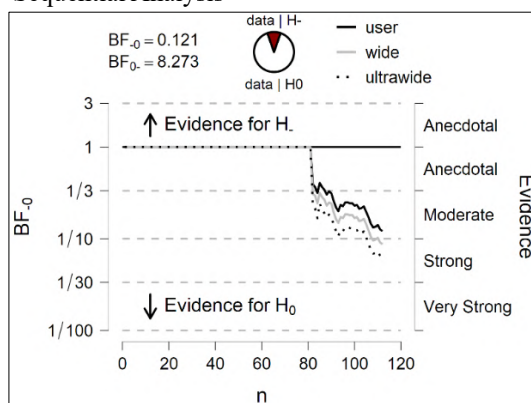


Figure 43 The Bayesian Independent samples t-test inferential plots of the clause nexus percentage on Level 1 in Master and PhD sub-corpora

A bigger sample size causes more evidence to support the null hypothesis, as can be predicted from the sequential analysis, and the Bayes factor robustness checks that when the sample size and the Cauchy

prior width increase, the tendency for the null hypothesis falls to strong evidence.

4.3.3.6.3 Clause Nexuses on Level 2

On Level 2, the clause nexus percentage concentrates on 22%-23% and 25%-27% in group D while 25% to 28% in group M. (See Figure 44)

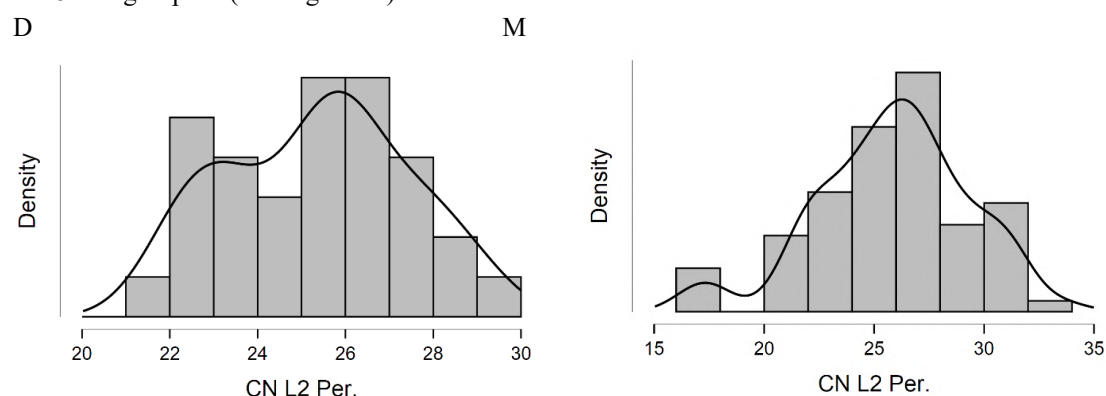


Figure 44 The distribution plots of clause nexuses percentage on Level 2 in Master and PhD sub-corpora

Again, the Welch's t-test is referred to (see Table 30) that the negative values of the statistic (-0.914) and the location parameter (-0.489) indicate a slight difference between group D when compared with group M, and SE difference can tell that this difference is no larger than 1 percent considering the clause nexus percentage on Level 2. The p-value (0.181>0.05) shows that the probability of neglecting the null hypothesis cannot be rejected, and with the VS-MPR value equal to 1.188, the probability of accepting the alternative hypothesis is only 1.188 times of rejecting it. So, the clause nexus percentage on Level 2 in Group D is less than that in Group M, and this variance is not big at all, with the effect size value only -0.17.

Independent Samples T-Test								
	Test	Statistic	df	p	VS-MPR*	Location Parameter	SE Difference	Effect Size
CNs L2 Per.	Welch	-0.914	94.274	0.181	1.188	-0.489	0.534	-0.170
<i>Note.</i> For the test, the alternative hypothesis specifies that group <i>D</i> is less than group <i>M</i> .								
<i>Note.</i> For the Welch t-test, effect size is given by Cohen's d.								
<i>Note.</i> For the Welch t-test, location parameter is given by mean difference.								
* Vovk-Sellke Maximum <i>p</i> -Ratio: Based on a two-sided <i>p</i> -value, the maximum possible odds in favor of H_1 over H_0 equals $1/(-e p \log(p))$ for $p \leq .37$ (Sellke, Bayarri, & Berger, 2001).								

Table 30 The Welch's test of the clause nexus percentage on Level 2 in Master and PhD sub-corpora

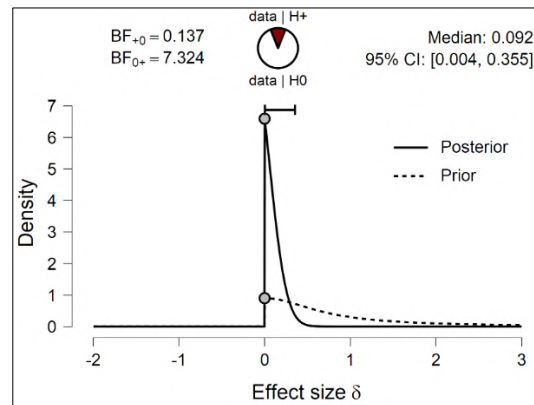
When the results of the independent samples t-test are not obvious enough to demonstrate where the exact variance lies, the Bayesian test then becomes useful and necessary. In Table 31, it is apparent that the probability of accepting the alternative hypothesis (group D>group M) is only 0.137 times of accepting the null hypothesis (group D<group M). So, the clause nexus percentage on Level 2 in Group D is less than that in Group M.

Bayesian Independent Samples T-Test		
	BF ₀₊	error %
CNs L2 Per.	0.137	~ 0.005
<i>Note.</i> For all tests, the alternative hypothesis specifies that group <i>D</i> is greater than group <i>M</i> .		

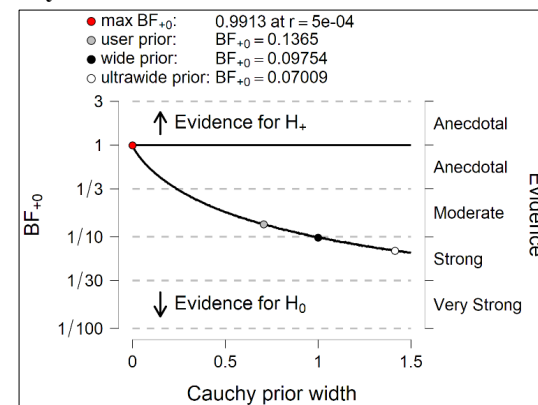
Table 31 The Bayesian independent t-test of the clause nexus percentage on Level 2 in Master and PhD sub-corpora

More details are presented in the inferential plots (Figure 45) that BF₀₊ is 7.324 times BF₀ with low effect size in the prior and posterior plot, and when the Cauchy prior width and the sample size increase, the evidence for supporting the null hypothesis (group D<group M) becomes strong. The statement that the clause nexus percentage on Level 2 of Group D is less than that of Group M has thus been confirmed.

Prior and Posterior



Bayes Factor Robustness Check



Sequential Analysis

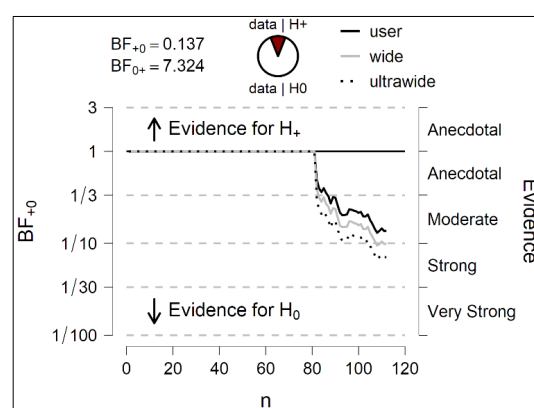


Figure 45 The Bayesian Independent samples t-test inferential plots of the clause nexus percentage on Level 2 in Master and PhD sub-corpora

4.3.3.6.4 Clause Nexuses on Level 3

On Level 3, the clause nexus percentage concentrates on 6%-8% in both group D and group M. (See Figure 46)

D

M

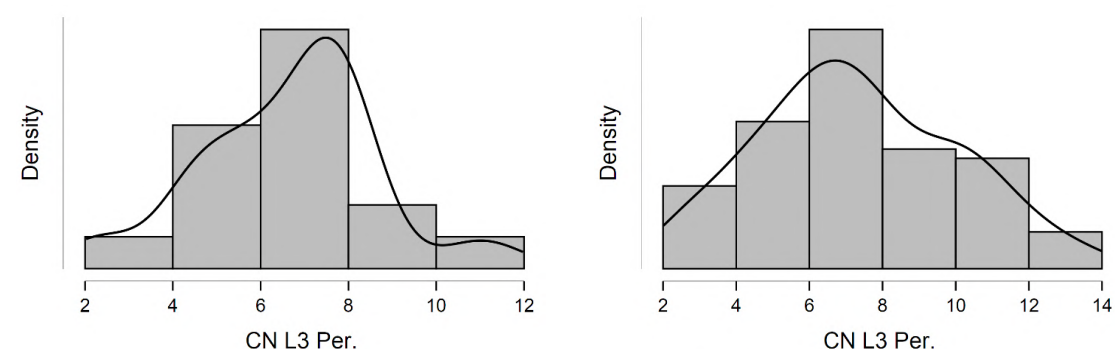


Figure 46 The distribution plots of clause nexuses percentage on Level 3 in Master and PhD sub-corpora

The effect size of Welch's t-test (Table 32) is also very small, with a value of -0.289. The negative values of statistic (-1.476) and the location parameter (-0.69) also indicate that group D is less than group M, and the p-value gets very close to 0.05, the null hypothesis is also possible. SE difference notes that the percentage difference between these two groups is no larger than 1%. VS-PMR value indicates that the probability of accepting the alternative hypothesis is 1.941 times of neglecting it.

Independent Samples T-Test								
	Test	Statistic	df	p	VS-MPR*	Location Parameter	SE Difference	Effect Size
CNs L3 Per.	Welch	-1.476	76.542	0.072	1.941	-0.690	0.468	-0.289
<i>Note.</i> For the test, the alternative hypothesis specifies that group <i>D</i> is less than group <i>M</i> .								
<i>Note.</i> For the Welch t-test, effect size is given by Cohen's d.								
<i>Note.</i> For the Welch t-test, location parameter is given by mean difference.								
* Vovk-Sellke Maximum <i>p</i> -Ratio: Based on a two-sided <i>p</i> -value, the maximum possible odds in favor of H_1 over H_0 equals $1/(-e p \log(p))$ for $p \leq .37$ (Sellke, Bayarri, & Berger, 2001).								

Table 32 The Welch's test of the clause nexus percentage on Level 3 in Master and PhD sub-corpora

For further comparison, the Bayesian independent samples t-test is conducted with the result BF_{+0} (the alternative hypothesis specifies that group *D* is greater than group *M*) equals 0.104 (See Figure 47), which gives preference to the null hypothesis. But the effect size is smaller as compared to that in the former two levels. When the sample size increases, the evidence for H_0 , as shown in the sequential analysis plots, falls quickly to strong. Thus, it is suggested that the probability of accepting the null hypothesis (group $D < \text{group } M$) is 9.655 times of accepting the alternative hypothesis (group $D > \text{group } M$).

Prior and Posterior

Sequential Analysis

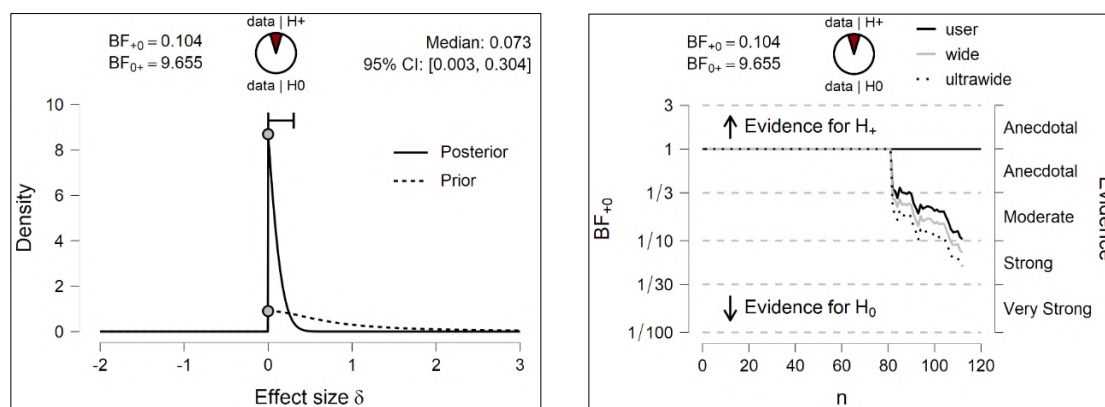


Figure 47 The Bayesian Independent samples t-test inferential plots of the clause nexus percentage on Level 3 in Master and PhD sub-corpora

4.3.3.6.5 Clause Nexuses on Level 4

On Level 4, though the percentage is quite small in both groups and the most frequent is around 1%, hundreds of clause nexuses can be found in each group, which still occupies an important status and thus needs attention and discussion. (See Figure 48)

D

M

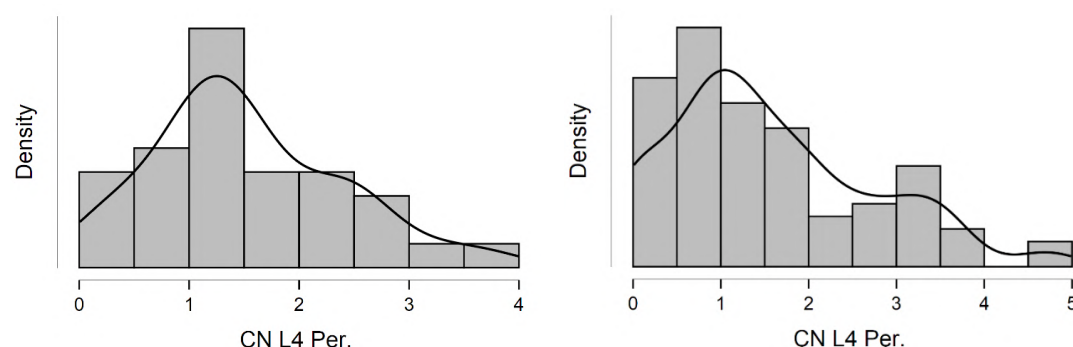


Figure 48 The distribution plots of clause nexuses percentage on Level 4 in Master and PhD sub-corpora

It seems that group D has a lower clause nexus percentage on Level 4 compared to group M, but when the independent samples t-test (Table 33) results are checked, the t-statistic, mean difference, and SE difference values are extremely small, with p-value $0.558 > 0.05$, and the effect size measured by the Cohen's d -0.029 (almost equals to 0), a conclusion is very hard to reach.

Independent Samples T-Test						
	t	df	p	Mean Difference	SE Difference	Cohen's d
CNs L4 Per.	-0.146	76.317	0.558	-0.029	0.200	-0.029
<i>Note.</i> For all tests, the alternative hypothesis specifies that group <i>D</i> is greater than group <i>M</i> .						
<i>Note.</i> Welch's t-test.						

Table 33 The Welch's t-test comparing the clause nexus percentage on Level 4 in Master and PhD sub-corpora

Then, in the Bayesian independent samples t-test, the method is the Mann-Whitney U-test, as the data distribution in group M is not normally distributed. Results are also intriguing because the probabilities of accepting the null hypothesis under different conditions are quite similar, which brings uncertainties in comparing these two groups. When the null hypothesis is that group D is larger than group M, the probability of accepting it is 6.602 times of neglecting it, with the Rhat is 1.063. (See Table 34)

Bayesian Mann-Whitney U Test			
	BF ₀₋	W	Rhat
CNs L4 Per.	6.602	1334.000	1.063
<i>Note.</i> For all tests, the alternative hypothesis specifies that the location of group <i>D</i> is smaller than the location of group <i>M</i> .			
<i>Note.</i> Result based on data augmentation algorithm with 5 chains of 1000 iterations.			

Table 34 The Bayesian Mann-Whitney U Test for comparing the clause nexus percentage on Level 4 in Master and PhD sub-corpora

The same conclusion can be found in the inferential plots (Figure 49, left 1) with a small effect size. But when the null hypothesis is that group D is smaller than group M, the probability of accepting it is 3.057 times of neglecting it (Figure 49, right 1), with the effect size larger than the former hypothesis; and when the null hypothesis is that group D equals to group M, the probability of accepting the null hypothesis is 4.179 times of rejecting it (Figure 49, left 2), with the effect size almost equals to zero. By comparing the probability of acceptance, it can be said that the clause nexus percentage on Level 4 of Group D is slightly bigger than that of Group M, though the variance is extremely small.

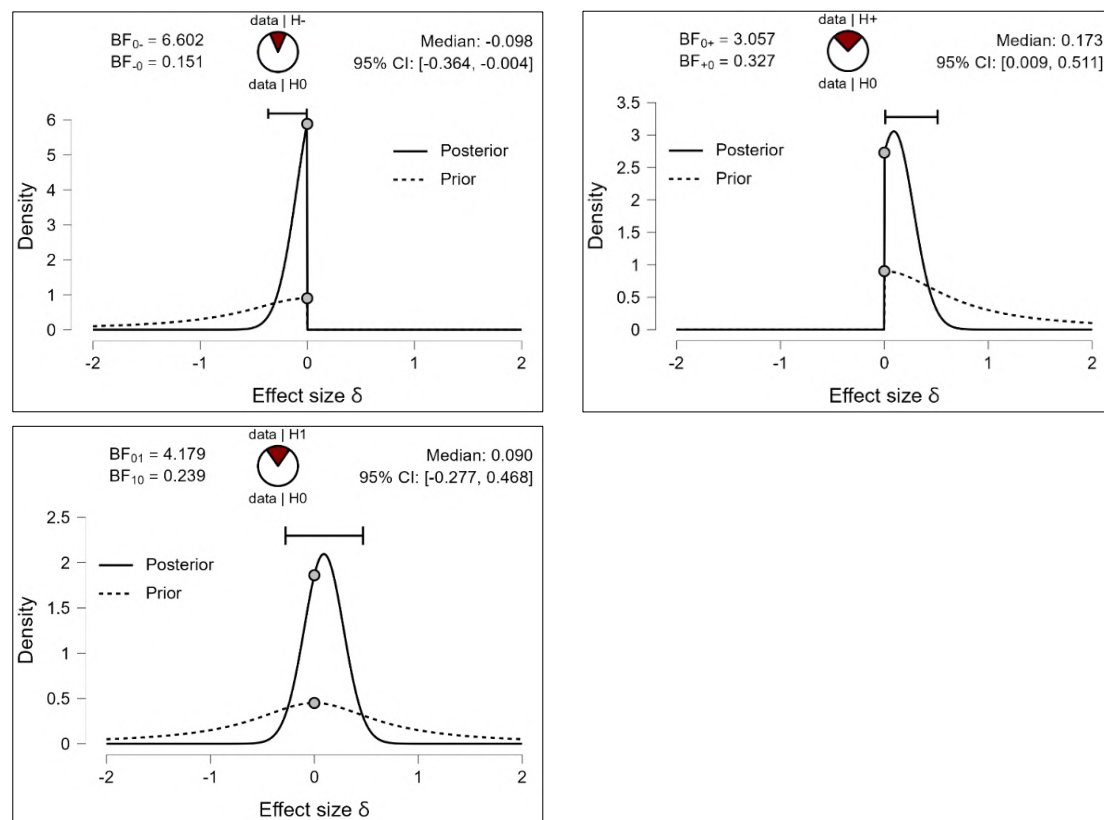


Figure 49 The inferential plots (prior and posterior) of the clause nexus percentage on Level 4 in Master and PhD sub-corpora

To summarize, Doctoral students would have a higher percentage of clause nexuses in each text on Level 1 and Level 4 and a lower percentage on Level 2 and Level 3 than Master students in thesis writing.

Though these differences are not obvious enough to be a certainty, and they may have been affected by the particular cases collected in group M, at least two points can be affirmed: more than 80 percent of clause nexuses written by both doctoral and master students are on Level 1 and Level 2, and less than 20 percent on are level 3 to level 7, which demonstrates students' inability and reluctance to write more complex structures with recursion patterns in their theses; and the frequencies and categories of recursion patterns have more universality on Level 1-3 and individuality on Level 4-7, which will be discussed detailly in Section 7.4.

Considering the taxis of clause nexus, what is the proportion of parataxis versus hypotaxis on each level, what is the most common grammatical category on each level, and on which level does each grammatical category have the highest frequency are worthwhile but time and energy-consuming questions as these parameters need manual statistical analysis, so they will not be answered in this thesis due to length limitation.

4.3.3.7 Parataxis/Clause Nexuses Percentage

At least the proportion of parataxis versus hypotaxis in all clause nexuses in both groups is explored. For the number of paratactic clause complexes, around 265 can be found in each doctoral dissertation, 123 can be found in each Master thesis, and about 8500 in total in group D and 9900 in group M can be analyzed. The distributions of data in two groups are not normal at all, as suggested by Figure 50. For the proportion of paratactic clause complex in two groups (see Table 4.3.3.7-1 in Appendix 4.3), the mean of group D is 34.89%, which is lower than that of group M (39.01%), and the higher variance of group M (56.96%) is not strange as it has a larger sample size. The negative skewness of group D means a left-tailed skew of data distribution. The division of skewness to the standard error of skewness in both groups is within $-2/+2$ (-0.68 in group D and 1.17 in group M), which denotes that the data are normally distributed in both two groups. The negative value of kurtosis in group M means platykurtic data distribution. Again, the division of kurtosis to the standard error of kurtosis in both groups is also within $-2/+2$ (0.25 in group D and -0.88 in group M), and the p-value of Shapiro-Wilk in both groups is larger than 0.05 (0.757 in group D and 0.141 in group M).

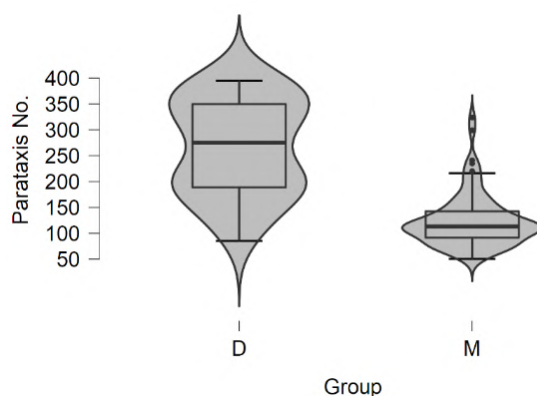


Figure 50 The boxplots of the amount of parataxis in Master and PhD sub-corpora

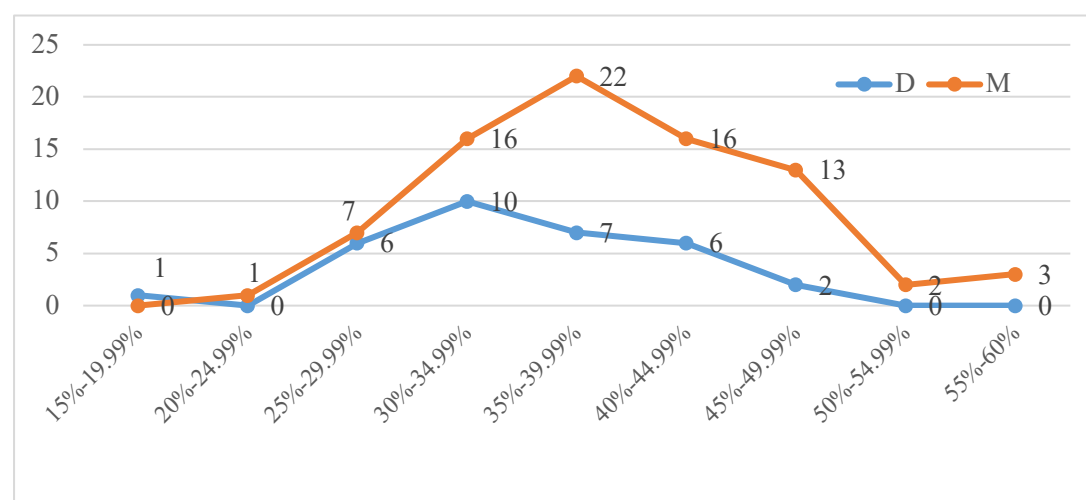


Table 35 The number of texts on each paratactic clause complex percentage stage in Master and PhD sub-corpora

All these parameters indicate that the data are normally distributed in both group D and group M. Another comparative parameter is the number of texts on each percentage stage of the paratactic clause complex (see Table 35), in which most numbers of group D concentrate on 25% to 45%, and the highest number lies in the range of 30%-34.99%; most numbers of group M concentrate on 25% to 49.99%, and the highest locates in the range of 35%-39.99%. Thus, the descriptive statistics of the paratactic versus hypotactic clause complex proportion preliminarily suggested that group D is inferior to group M.

To confirm the statements above, the boxplot also refers to the fact that data in two groups are indeed normally distributed, and the paratactic per clause complex percentage of group D is less than group M. (See Figure 51)

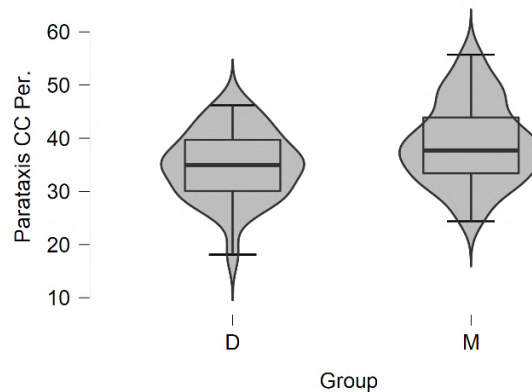


Figure 51 The boxplots comparing the paratactic clause complex percentage in Master and PhD sub-corpora

Because the variances in the two groups are different, the Welch's t-test (Table 36) is firstly considered that the negative values of the t (-2.930) and the mean difference (-4.117), and the positive value of the standard error difference of variance (1.405) all indicate how much the difference between two groups is. The p -value is 0.002 (<0.05), which gives a significant sign that the null hypothesis (group D > group M) is rejected, i.e., the alternative hypothesis that group D has a lower paratactic clause complex percentage than group M is accepted. The same conclusion can be drawn from the VS-MPR value that it is too much larger than the value 2.46 (when $p < .05$), so the probability of rejecting the alternative hypothesis is extremely small. This test has a medium-size effect as its Cohen's d is -0.590 (negative only means the value of group D is less than that of group M).

Independent Samples T-Test

	t	df	p	VS-MPR*	Mean Difference	SE Difference	Cohen's d
Parataxis CC Per.	-2.930	67.438	0.002	26.255	-4.117	1.405	-0.590

Note. For all tests, the alternative hypothesis specifies that group D is less than group M.

* Vovk-Sellke Maximum p -Ratio: Based on a two-sided p -value, the maximum possible odds in favor of H_1 over H_0 equals $1/(-e p \log(p))$ for $p \leq .37$ (Sellke, Bayarri, & Berger, 2001).

Note. Welch's t-test.

Table 36 The Welch's t-test of the paratactic clause complex percentage in Master and PhD sub-corpora

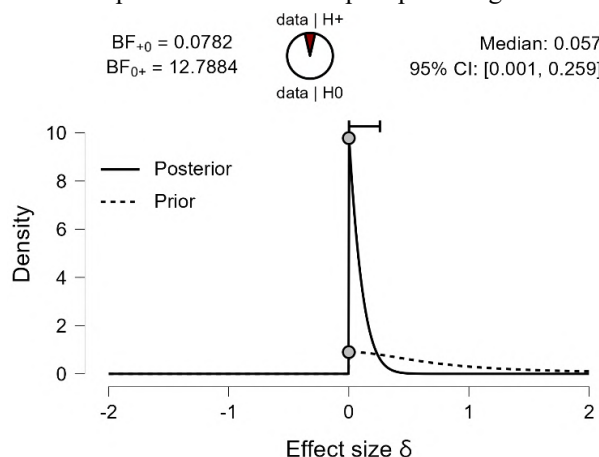


Figure 52 The inferential plots (prior and posterior) of the paratactic clause complex percentage in Master and PhD sub-corpora

In the Bayesian Mann-Whitney U-test, when the alternative hypothesis specifies that the location of group D is greater than the location of group M, the Wilcoxon test value is 886.5, and the Rhat is 1.143, which denotes that the between- and within-chain variance of the MCMC estimates for the delta parameter is not convergent. The BF_{0+} is 12.79 (see Figure 52), suggesting that the probability of accepting the null hypothesis (group D < group M) is 12.8 times of accepting the alternative hypothesis, with a small effect size. The reasons for this result, i.e., what paratactic grammatical structures are preferred by Master students compared to doctoral students, will be analyzed and explained in detail in Sections 5 & 6.

4.3.3.8 Hypotaxis/Clause Nexuses Percentage

There is no doubt that the hypotactic clause complex percentage of group D is larger than that of group M as it is opposite from the paratactic testing result. However, general information and the distribution of the data are still necessary for demonstration. Altogether, 15,623 and 15,450 hypotactic clause complexes can be found in groups D and M separately; the mean of the former is 488.22, and of the latter is 193.13. As the sample size of group D is $\frac{1}{4}$ that of group M, the range and the standard deviation of the former is thus bigger than that of the latter. The data of group M is skewed to the right side (skewness value is 1.098) and is leptokurtic as its kurtosis value is 1.266; while that of group D is platykurtic and a little bit skewed to the right side. Comparing the division of the skewness to the standard error of skewness and the kurtosis to the standard error of kurtosis in both groups, data in group D is normally distributed (values are 0.829 and 0.491 separately, and both are within $-2/+2$) while in group M is abnormally distributed (values are 4.082 and 2.380 separately and both are bigger than $+2$). The saying can also be approved by the P-value of Shapiro-Wilk that it is <0.001 in group M. (See Table 4.3.3.8-1 in Appendix 4.3) The number of texts on each percentage stage of the hypotactic clause complex would be the mirror image of that in Table 35 above, in which the percentage stage would be reduced by 1, ranging from 40% to 85%. Most numbers of group D concentrate on 55% to 75%, and the highest number lies in the range of 65%-69.99%, and most numbers of group M concentrated on 50% to 75% and the highest located in the range of 60%-64.99%.

For the hypotactic clause complex percentage, the boxplots of two groups are the reversed shape of that for the paratactic ones. In group D, the proportion ranges from 53.801% to 81.876% with a mean of 65.112% and the mode 53.801%; and in group M, the proportion ranges from 44.279% to 75.610% with a mean of 60.995% and the mode 44.279%. Thus, the mean of Group D is 4% larger than that of Group M. The independent Welch's t-test and the Bayesian Mann-Whitney U-test would provide the same values for hypotaxis and the opposite acceptance for the hypothesis, that is, all parameters can confirm the conclusion that group D is greater than group M in choosing hypotactic clause complexes in writing theses.

4.3.4 Summary

This section declares the principles of designing the corpus for this research that representativeness and intuition are both important for a balanced sampling; simply describes the general organization of the corpus for both lexical and grammatical investigations; detailly compares the Master theses group and the doctoral dissertations group from the number of clauses, the average words used in each sentence, the number of chosen clause complexes, the proportion of found clause complexes in chosen sentences in each text, the proportion of clause nexuses per clause complexes in each text, the descriptive and comparative information of clause nexuses written on level 1-7, the proportion of paratactic clause complexes in each text, and the proportion of hypotactic clause complexes in each text. Similarities and differences between two groups found in these parameters lay the foundation for further research that the same proportion of clause complexes in chosen sentences in each text makes more comparison within clause complex stratification possible and necessary, and the testing results of hypotaxis can be predicted from that of parataxis as they are complementary in clause complexes. Though paratactic and hypotactic relations on each level in two groups are worthwhile for comparing, they are excluded from this research because of length limitation.

4.4 Analysis Outline

To put more words about this research on the methods section, the knowledge of three intersected fields, i.e., Systemic Functional grammar, corpus linguistics and academic writing lay the foundation for practical and comparative activities that will be illustrated in later sections: theories and theoretical framework of clause complex in SFG is adopted and modified according to specific use; corpus linguistics is seen as a perspective that methods of large quantity of texts processing, annotating, patterns and system investigating, statistics calculating, and software applying are borrowed correspondingly; and academic writing is regarded as a genre that for the purpose of not only writing, framing problems, understanding issues, forming personal identities, and progressing to graduation for students but also collaborating and competing with academic researchers, revealing and creating knowledge, and defining academic allegiances, (Hyland, 2009, p. 1-2) how language is constructed for performing these functions and how the academic discourse reveal these performances are main directions to be explored in this research. The integration details are demonstrated in Figure 53:

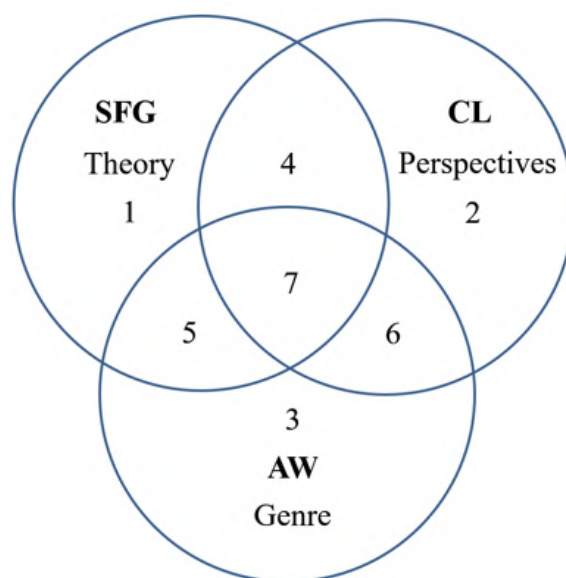


Figure 53 The integration of SFG, Corpus Linguistics, and Academic Writing theories and methods for conducting this research

Data or chosen corpus for the research are texts divided into two groups: doctoral group and Master's group with the number of texts, words, chosen sentences, clause complexes, clause nexuses for analyzing and hypotaxis vs. parataxis are listed in Table 37 that the same percentage of clause complexes per chosen sentences provides the probability for further comparisons within two groups.

	Texts	Words	Chosen Sentences	Clause Complexes	Clause Nexuses	Hypotaxes	Parataxes
D	32	1440311	28257	15843	24122	15623	8499
M	80	1282881	28380	16084	25326	15450	9876

Table 37 The general information of the chosen Master and PhD sub-corpora for the research

For research methods applied in each part on the basis of diverse theoretical background, they would differ accordingly with certain software for processing and fulfilling as shown in Table 38 (developed from Table 21) that grammatical and lexical annotation are conducted in SysFan manually and qualitatively based on SFG framework; grammatical and lexical calculation are completed quantitatively by JASP based on corpus linguistics theories; lexical analysis is realized in SysFan and AntConc quantitatively and qualitatively on genre and discourse analysis methods in academic writing; annotations for recursion patterns are fulfilled by SysFan manually and qualitatively on SFG framework and corpus linguistics methods; grammatical analysis borrowed SFG framework and academic writing theories for both quantitative and qualitative analysis in SysFan and AntConc; the preliminary theory developed from "four additions" are qualitatively completed on SFG and academic writing theories; the analysis for recursion patterns is mainly conducted by JASP and SysFan qualitatively with also quantitative calculations on intersections of academic writing and corpus linguistics ways; and the

complexity of clause complex requires the synthesise of all the three directions with the circulation of quantitative plus qualitative methods to mainly quantitative method.

Section	Contents	Method	Software
1. SFG	Grammatical annotation	QUAL	SysFan
	Lexical annotation	QUAL	
2. CL	Grammatical calculation	QUAN	JASP
	Lexical calculation	QUAN	
3. AW	Lexical analysis	quan $\rightarrow$ QUAL	SysFan, AntConc
4. SFG x CL	Recursion patterns annotation	QUAL	SysFan
5. SFG x AW	Grammatical analysis	quan $\rightarrow$ QUAL	SysFan, AntConc
	Four additions & related	Qual	
6. AW x CL	Recursion patterns analysis	QUAL(quan)	JASP, SysFan
7. SFG x AW x CL	Complexity of CC	[QUAN+qual] $\rightarrow$ QUAN	

Table 38 The specific methods with software for diverse research activities based on corresponding theories in this research

Chapter 5 Expansion

This chapter will specifically examine the similarities and differences between master's and doctoral students in the grammatical category of expansion clause complexes and their lexical choices in English academic writing. Through quantitative analysis of the theoretical system, Section 5.1 will demonstrate the use of expansion/hypotactic expansion/paratactic expansion, as well as the three major parts of expansion in master's and doctoral theses; Section 5.2 will focus on elaboration and investigate the selection of paratactic elaboration and hypotactic elaboration; Section 5.4 is about the choice of extension, examining the comparison of paratactic extension and hypotactic extension in master's and doctoral theses; Section 5.6 is about the choice of enhancement, which is also divided into parataxis and hypotaxis for discussion. Through the detailed categorical choice of elaboration in Section 5.3, namely exposition, exemplification, clarification, and their specific realizations; the categorical choice of extension in Section 5.5, namely addition, variation, alternation, and their main connecting lexis in clause complexes; and the categorical choice of enhancement in Section 5.7, namely temporal, spatial, manner, causal, concessive and the comparative qualitative analysis of marking lexis of each category in the examples, questions regarding which lexical meaning choices lead to the corresponding grammatical tendencies or how the application of grammatical categories are realized by relevant lexis in specific clause complexes will be answered, thereby confirming the statement that lexis and grammar are at two ends of a cline.

5.1 Expansion Comparison

In this part, the comparison between data in the doctoral group and Master's group focuses mainly on expansion, in which general conditions of expansion, paratactic expansion proportion, hypotactic expansion proportion, and the composition of expansion, i.e., the proportion of elaboration, extension, and enhancement will all be compared to see how different students are in choosing these grammatical structures in writing theses, transmitting ideas and performing academic functions.

5.1.1 General Distribution of Expansion

The total number of expansion clause nexuses identified in each group is considerable, with 19,571 in group D and 20,633 in group M. The mean for group D (611) is approximately 2.4 times that of group M (258). In some theses, the number of expansions is limited to 249 (in group D) and 120 (in group M). As illustrated in Table 5.1.1-1 (see Appendix 5.1) and Figure 54, the data in group D are typically platykurtic (kurtosis value of -0.573) with minimal skewness (skewness value of 0.029). In contrast, the data in group M exhibit an abnormal leptokurtic distribution (kurtosis value of 1.331) with a pronounced skew towards the right tail (skewness value of 1.193). With the exception of four specific instances within Group M, the distribution remains bimodal.

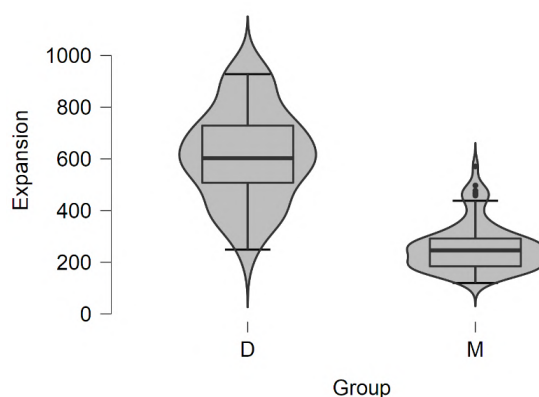


Figure 54 The boxplots of expansion distribution in Master and PhD sub-corpora

5.1.1.1 Expansion/Clause Nexuses

The proportion of expansion per clause nexus in each text in the two groups (See Section 4.3.3.3) is meaningful and comparable, given that the percentage of clause nexuses per chosen sentences in each text in the two groups is almost the same and the number of clause nexuses in the two groups is not significantly different.

A review of the mean for the two groups (Table 5.1.1.1-1 in Appendix 5.1) reveals that approximately 81% of clause nexuses are expansions rather than projection structures, with minimal variance. The large sample size increases the probability of expansion proportion, with the highest value observed in group M at 92% and the lowest at 64%. Group D displays a platykurtic distribution, as indicated by a kurtosis value of -0.869 and a right-sided skew, as reflected by a skewness value of 0.176. In contrast, group M exhibits a slightly leptokurtic distribution, as indicated by a kurtosis value of 0.26, and a left-sided skew, as suggested by a skewness value of -0.583. When the division of skewness to the standard error of skewness is considered, the value in group D is 0.425 (within -2/+2), while in group M it is -2.167 (smaller than -2). This indicates that the data distribution in group D is normal, while in group M, it is abnormal. Similarly, when the division of kurtosis to the standard error of kurtosis is considered, the value in group D is -1.074 (within -2/+2), and in group M is 0.489 (also within -2/+2). The Shapiro-Wilk value in both groups is approximately 1, and the P-value is larger than 0.05, a contradictory conclusion is reached here that the data in both groups are normally distributed. To provide further clarification, the boxplots (Figure 55) are referenced, demonstrating that except for one case in group M, the remaining data are normally distributed in both groups.

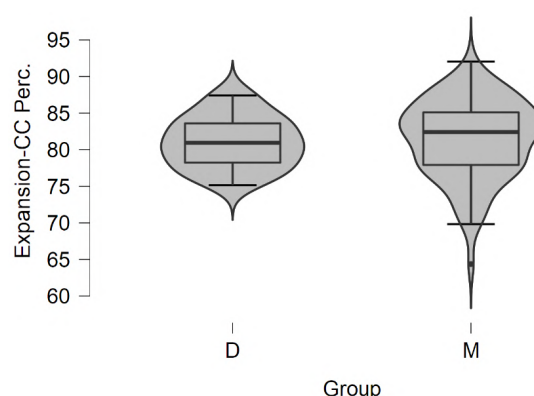


Figure 55 The boxplots of expansion per clause nexuses proportion in Master and PhD sub-corpora

Thus, for the independent samples t-test, the Mann-Whitney U test (Table 39) is performed with its results are discussed: the p-value is 0.224, which means that there is still a probability of accepting the original hypothesis (group D > group M), and the VS-MPR value (1.098) denotes that the possibility of accepting the alternative hypothesis (group D < group M) is 1.098 times of neglecting it. The Hodges-Lehmann Estimate (-0.746) suggests that the variance between the two groups is extremely small (0.746%), and group D is smaller than group M. But the Rank-Biserial Correlation is only -0.093, the effect size of this test is almost zero. So, the difference between the two groups is hard to tell by only conducting one test.

Independent Samples T-Test

Independent Samples T-Test						
	W	df	p	VS-MPR*	Hodges-Lehmann Estimate	Rank-Biserial Correlation
Expansion-CC Perc.	1161.500	0.224	1.098	-0.746		-0.093

Note. For the Mann-Whitney test, effect size is given by the rank biserial correlation.

Note. For all tests, the alternative hypothesis specifies that group D is less than group M.

Note. Mann-Whitney U test.

* Vovk-Sellke Maximum p -Ratio: Based on a two-sided p -value, the maximum possible odds in favor of H_1 over H_0 equals $1/(-e p \log(p))$ for $p \leq .37$ (Sellke, Bayarri, & Berger, 2001).

Table 39 The Mann-Whitney U test for comparing the expansion per clause nexuses proportion in Master and PhD sub-corpora

Bayesian Mann-Whitney U Test

	BF ₀₊	W	Rhat
Expansion-CC Perc.	6.654	1161.500	1.076

Note. For all tests, the alternative hypothesis specifies that the location of group D is greater than the location of group M.

Note. Result based on data augmentation algorithm with 5 chains of 1000 iterations.

Table 40 The Bayesian Mann-Whitney U test for comparing the expansion per clause nexuses proportion in Master and PhD sub-corpora

In the Bayesian Mann-Whitney U test, when the alternative hypothesis specifies that the location of

group D is greater than that of group M, the probability of accepting the original hypothesis is 6.654 times neglecting it (See Table 40) with a small effect size (See Figure 56, left 1), and the Rhat is 1.076, which means that the between- and within-chain variance of the MCMC estimates for the delta parameter is not convergent, i.e., a difference indeed exists between group D and group M. But can a conclusion here be reached that the expansion per clause nexuses proportion in group D is less than group M?

When the alternative hypothesis is set as group D < group M, then the probability of accepting the original hypothesis is 2.924 times (See Figure 56, right 1) of rejecting it with a medium effect size; when the original hypothesis signifies that group D = group M, the possibility of accepting it is 4.063 times (See Figure 56, left 2) of rejecting it with a larger effect size; and when the alternative hypothesis is group D > group M, the probability of accepting the original hypothesis is 6.65 times of neglecting it (See Figure 56, left 1). So, by comparing these probabilities, a tendency to regard the two groups as the same is chosen, though the parameter of group D is slightly smaller than that of group M.

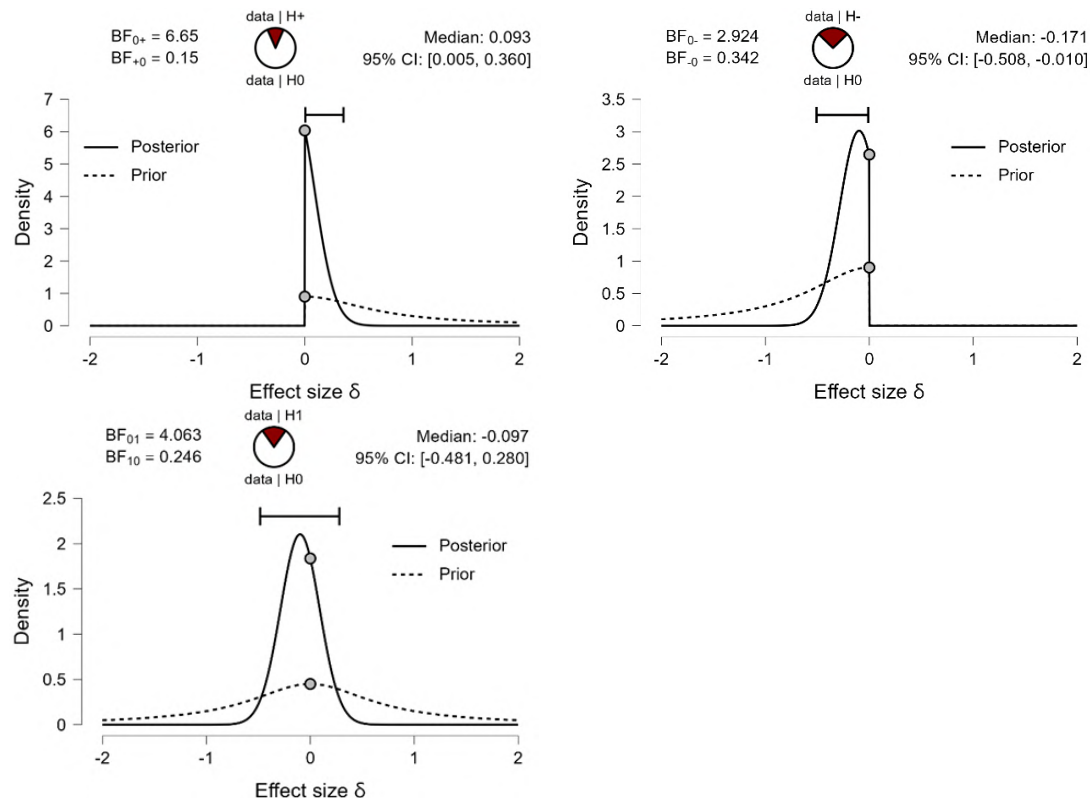


Figure 56 The inferential plots comparing the expansion per clause nexuses proportion in Master and PhD sub-corpora with different alternative hypotheses

5.1.2 The Distribution of Paratactic Expansion

Again, the same proportion of expansion per clause nexuses provides possibilities for further comparison within expansion structures.

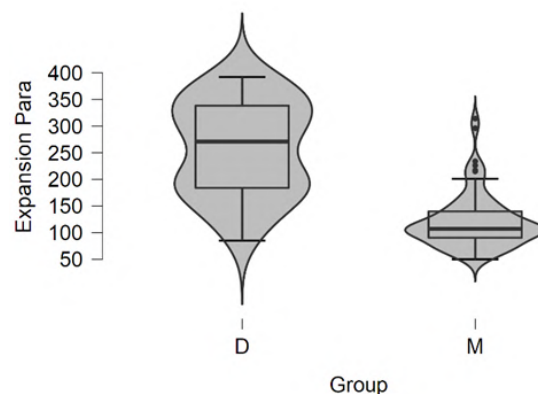


Figure 57 The boxplots of the number of paratactic expansions in Master and PhD sub-corpora

For paratactic expansion, the total number found in group D is 8230, and in group M is 9633 (Table 5.1.2-1 in Appendix 5.1), in which the minimum is 85 in the former and 50 in the latter; and the maximum can be 392 and 314 in each group. Data in group M is abnormal, right-sided skewed distributed as judged from the P-value of Shapiro-Wilk and also Figure 57, and in group D is normally, platykurtic distributed but with bimodal as seen from Figure 57.

5.1.2.1 Expansion Parataxis/Clause Nexuses

When the proportion of paratactic expansion is compared, the denominator can be clause nexuses, parataxis, and expansion structures, in which the former investigates the ratio of paratactic expansion in the whole clause complex system; the second compares the proportion of paratactic expansion versus paratactic projection structures; and the third compares paratactic expansion versus hypotactic expansion that they usually have the same testing results but reversed or complementary ratio.

In all clause nexuses, the mean ratio of paratactic expansion is around 34% in group D and 38% in group M (See Table 5.1.2.1-1 in Appendix 5.1), with a range from 18% to 45% in the former group and 23% to 54% in the latter. Data found in each text are left-sided skew with a skewness value of -0.116 and platykurtic distributed as its kurtosis is -0.135 in group D; and in group M are right-sided skew (the skewness value is 0.286) and also platykurtic distributed with its kurtosis value is -0.493. Seeing the division of skewness to the standard error of skewness, and kurtosis to the standard error of kurtosis, the values are -0.280 and -0.167 in group D relatively, and 1.063 and -0.927 in group M correspondingly, which are both within -2 to +2. The Shapiro-Wilk value in each group is near 1, and the P-value of Shapiro-Wilk are all bigger than 0.05 (0.834 in group D and 0.149 in group M), the data distribution of paratactic expansions per clause nexuses in each text is thus normally in both groups. This statement can also be confirmed by the boxplots shown in Figure 58 below:

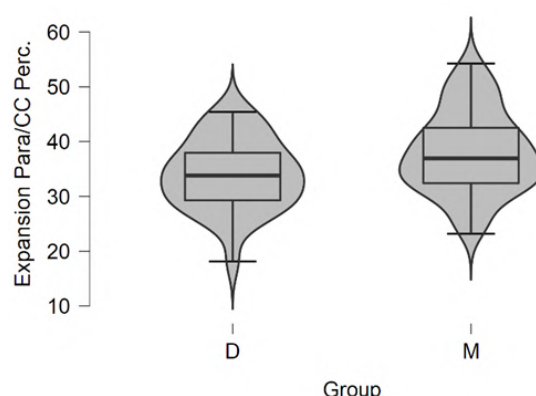


Figure 58 The boxplots comparing the proportion of paratactic expansion per clause nexuses in Master and PhD sub-corpora

As the variance in the two groups is different, the Welch's t-test (Table 41) is thus referred to that both values are negative (the statistic value is -2.88, and the mean difference is -4.051), which means that group D has a lower ratio of paratactic expansion per clause nexuses than group M. The standard error of difference (the value is 1.407) can also confirm this statement. The p-value is 0.003 (which is smaller than 0.05), the original hypothesis that group D is larger than group M is thus highly rejected. The VS-MPR value suggests that the probability of accepting the alternative hypothesis is 23.172 times of rejecting it. Cohen's d value (-0.583) means a medium effect size of the test. So, the conclusion here is that group D has a lower ratio of paratactic expansion per clause nexuses than group M, with a mean difference of 4.051%.

Independent Samples T-Test

	t	df	p	VS-MPR *	Mean Difference	SE Difference	Cohen's d
Expansion Para/CC Perc.	-2.880	66.026	0.003	23.172	-4.051	1.407	-0.583

Note. For all tests, the alternative hypothesis specifies that group D is less than group M.

* Vovk-Sellke Maximum p -Ratio: Based on a two-sided p -value, the maximum possible odds in favor of H_1 over H_0 equals $1/(-e p \log(p))$ for $p \leq .37$ (Sellke, Bayarri, & Berger, 2001).

Note. Welch's t-test.

Table 41 The Welch's t-test for comparing the paratactic expansion per clause nexuses proportion in Master and PhD sub-corpora

The same conclusion can also be demonstrated by the Bayesian Mann-Whitney U-test (see Figure 59) as

it is 13.18 times more probable than accepting the alternative hypothesis (group D > group M) with a small effect size and a Rhat value larger than 1. This indicates that the between- and within-chain variance of the MCMC estimates for the delta parameter is not convergent, thereby substantiating the existence of a difference between the two groups.

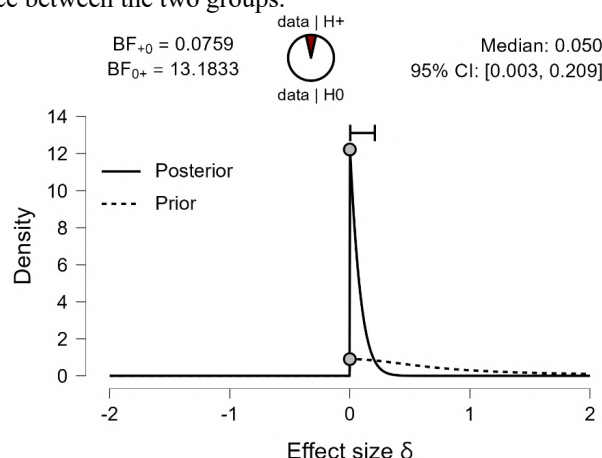


Figure 59 The inferential plots comparing the paratactic expansion per clause nexus proportion in Master and PhD sub-corpora (with the original hypothesis group D < group M)

The reasons for the higher percentage of paratactic expansion in clause nexuses found in group M will be illustrated in later sections in detail.

5.1.2.2 Expansion Parataxis/Parataxis

An amazing result is that in parataxis, about 97% are paratactic expansions, and in some texts, the percentage can even reach 100% in both groups, with a larger range (22%) in group D compared to group M (16%). Data are left-sided skew (the skewness value is -2.924 in group D and -1.864 in group M) and leptokurtic (the kurtosis is 10.338 in group D and 2.951 in group M) distributed in both groups as demonstrated in Table 5.1.2.2-1 (see Appendix 5.1). The values of the skewness divided by the standard error of skewness (-7.063 in group D and -6.929 in group M) being smaller than -2, the values of the kurtosis divided by the standard error of kurtosis (12.779 in group D and 5.547 in group M) being larger than +2, the low Shapiro-Wilk values (far away from 1 with 0.627 in group D and 0.706 in group M) and the P-value of Shapiro-Wilk being less than 0.05 all indicate that data are abnormally distributed in both groups. This can also be confirmed by the boxplots shown in Figure 60 below.

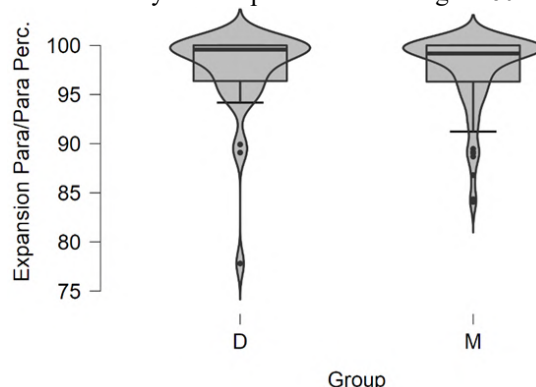


Figure 60 The boxplots comparing the paratactic expansion in parataxis in Master and PhD sub-corpora

Independent Samples T-Test

	W	df	p	VS-MPR *	Hodges-Lehmann Estimate	Rank-Biserial Correlation
Expansion Para/Para Perc.	1272.500	0.962	1.000	-3.511e-6	-0.006	

Note. For the Mann-Whitney test, effect size is given by the rank biserial correlation.

Note. Mann-Whitney U test.

* Vovk-Sellke Maximum p -Ratio: Based on a two-sided p -value, the maximum possible odds in favor of H_1 over H_0 equals $1/(-e p \log(p))$ for $p \leq .37$ (Sellke, Bayarri, & Berger, 2001).

Table 42 The Mann-Whitney U test for comparing the paratactic expansion in parataxis in Master and PhD sub-corpora (with the null hypothesis group D = group M)

Actually, data are normally distributed except for three particular cases in group D and six cases in group M. For comparing two groups, the Mann-Whitney U test is thus suitable, with results shown in Table 42. When the null hypothesis is that the percentage of paratactic expansion in the parataxis of group D equals that of group M, the p-value is 0.962 with VS-MPR equal to 1, the alternative hypothesis is thus totally rejected. But the Rank-Biserial Correlation value, which indicates the effect size is only -0.006, the U-test results are doubtful and need more tests to confirm hereby.

The Bayesian Mann-Whitney U-test results are also intriguing that when the alternative hypothesis specifies that the location of group D is larger than the location of group M, the BF_{0-} value is 5.154 (See Figure 61, left 1), which indicates that the probability of accepting the null hypothesis (group D < group M) is 5.154 times of rejecting it with the R_{hat} value is 1.017 (larger than 1, suggesting a difference in two groups); when the alternative hypothesis is that group D has a larger proportion than group M, the BF_{0+} value is 4.289, i.e., the probability of accepting the original hypothesis (group D > group M) is 4.289 times of rejecting it with a medium effect size (see Figure 61, right 1); and when the original hypothesis is that group D has the same proportion as group M, the possibility of accepting this saying is 4.682 (BF_{01}) times of rejecting it (see Figure 61, left 2) with a large effect size.

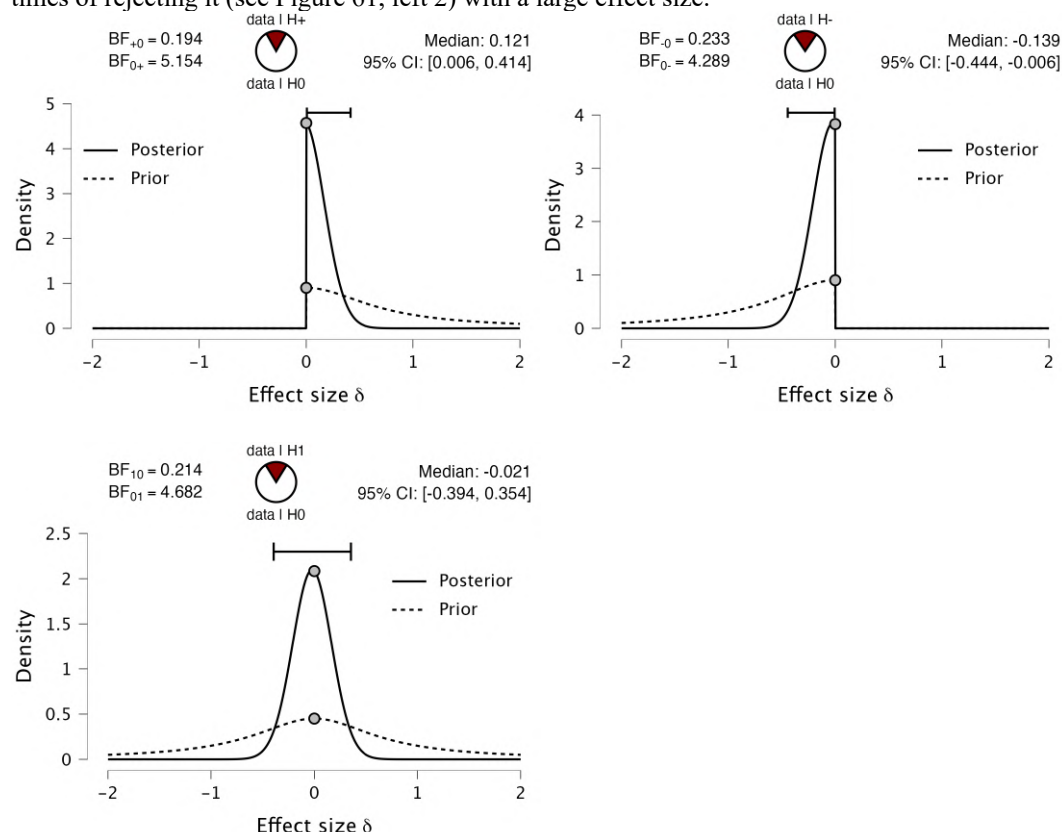


Figure 61 The inferential plots comparing the paratactic expansion per parataxis in Master and PhD sub-corpora (with different original hypotheses)

As a result, the proportion of paratactic expansion in all parataxis in each text in group D tends to be the same as that in group M, though the former is slightly smaller than the latter. One reason for this comes from more clause nexuses found in group D as the proportion of expansion per clause nexuses and the ratio of paratactic expansion per expansion are almost the same in the two groups; the other reason needs further exploration in paratactic expansion that perhaps paratactic elaboration or paratactic enhancement has a larger percentage in group M compared with group D.

5.1.2.3 Expansion Parataxis/Expansion

For the proportion of paratactic expansion in expansion structures, the minimum in group D is 23% and in group M is 28%, and the maximum in group D is 57% and in group M is 66%. (See Table 5.1.2.3-1 in Appendix 5.1)

The mean of group D (41.82%) is also smaller than that in group M (46.69%) except for the variance (51.62% in group D and 71.27% in group M). The data in group D is left-sided skew (the skewness value is -0.205) and slightly leptokurtic (the kurtosis value is 0.329) distributed, while in group M is platykurtic (the kurtosis value is -0.259) with almost no skew (the skewness value is 0.085). The division of skewness to the standard error of skewness in group D is -0.495, and in group M is 0.316; and

the division of kurtosis to the standard error of kurtosis in group D is 0.407, and in group M is -0.487: all values are within -2 to +2. The Shapiro-Wilk values in both groups are nearly 1, and the P-values of Shapiro-Wilk are both larger than 0.05. So, data in the two groups are both normally distributed, which can also be demonstrated in Figure 62.

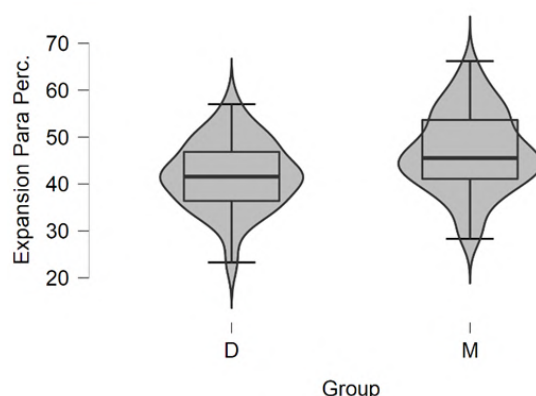


Figure 62 The boxplots comparing the proportion of paratactic expansion in expansion in Master and PhD sub-corpora

With different variance values but the same normal distribution, the Welch's t-test is performed. From Table 43, the values of t (-3.080) and mean difference (-4.873) are all negative, which means that group D has a lower percentage of paratactic expansion in expansions on average compared with that in group M and the SE difference is 1.582, which also presents differentiation in some degree. The p -value is 0.002 (smaller than 0.05), the original hypothesis that group D is larger than group M is hereby totally rejected with the VS-MPR value (37.629) indicating that the probability of accepting the alternative hypothesis (group D < group M) is 37.629 times of rejecting it in this test. The effect size signed by Cohen's d value is medium (-0.622), which suggests a useful test.

Independent Samples T-Test

	t	df	p	VS-MPR*	Mean Difference	SE Difference	Cohen's d
Expansion Para Perc.	-3.080	66.709	0.002	37.629	-4.873	1.582	-0.622

Note. For all tests, the alternative hypothesis specifies that group D is less than group M.

* Vovk-Sellke Maximum p -Ratio: Based on a two-sided p -value, the maximum possible odds in favor of H_1 over H_0 equals $1/(-e p \log(p))$ for $p \leq .37$ (Sellke, Bayarri, & Berger, 2001).

Note. Welch's t-test.

Table 43 The Welch's t-test for comparing the proportion of paratactic expansion in expansions in Master and PhD sub-corpora

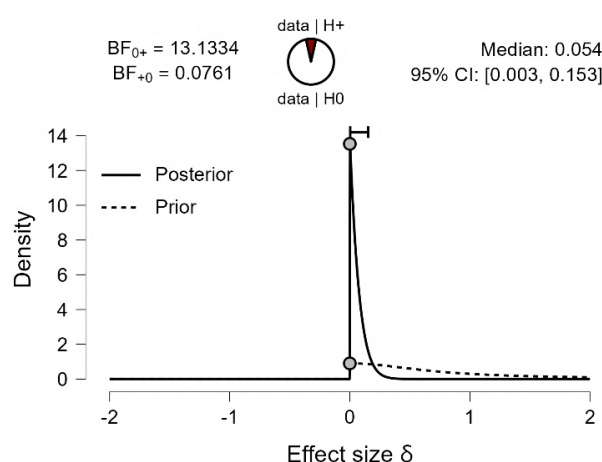


Figure 63 The inferential plots comparing the proportion of paratactic expansion in expansions in Master and PhD sub-corpora (with the original hypothesis group D < group M)

This obvious difference between the two groups can also be confirmed by the Bayesian Mann-Whitney U-test results when the probability of accepting the original hypothesis (group D < group M) is 13.13 times rejecting it, with a small effect size though, as suggested by Figure 63. The R_{hat} value is 1.053; thus, the between- and within-chain variance of the MCMC estimates for the delta parameter is

not convergent, i.e., a difference indeed exists within the two groups.

5.1.3 The Distribution of Hypotactic Expansion

The number of hypotactic expansions found in group D (11,341) is larger than that in group M (11,000), and the mean of group D (354) is around 2.5 times that of group M (137). More detailed statistics can be found in Table 5.1.3-1 (see Appendix 5.1), in which data in group M is abnormally distributed as the P-value of Shapiro-Wilk is less than 0.001. Data in group D seems to be normally distributed as suggested by the division of skewness to the standard error of skewness (0.911, within $-2/+2$), the division of kurtosis to the standard error of kurtosis (-0.277, within $-2/+2$), and the Shapiro-Wilk value 0.970 (nearly to 1). But Figure 64 suggests that data distribution in group D looks like a calabash with one particular case and in group M, a right-sided skewness with four peculiar cases.

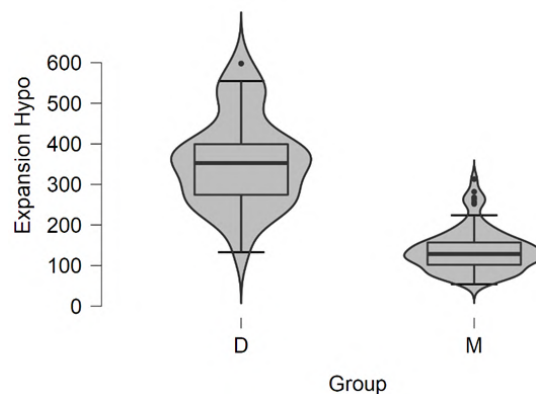


Figure 64 The boxplots comparing the number of hypotactic expansions in Master and PhD sub-corpora

5.1.3.1 Expansion Hypotaxis/Clause Nexuses

Likewise, when the proportion of hypotactic expansion is compared, the denominator can be clause nexuses that investigate the ratio of hypotactic expansion in the whole clause complex system; parataxis that compares the proportion of hypotactic expansion versus hypotactic projection structures; and again, expansion structures that concern paratactic expansion versus hypotactic expansion with same testing results but reversed or complementary ratio.

On the proportion of hypotactic expansion in clause complex system, roughly 34% to 60% of clause nexuses are hypotactic in group D and 25% to 64% in group M, the mean of the former is 47%, and of the latter is 43%. (See Table 5.1.3.1-1 in Appendix 5.1) Data in the two groups almost have no skewness (the value in group D is 0.096 and in group M is -0.017) and are a little bit leptokurtic as values of kurtosis are 0.273 in group D and 0.355 in group M. The division of skewness to the standard error of skewness values is 0.232 in group D and -0.063 in group M (both are within $-2/+2$), and the division of kurtosis to the standard error of kurtosis values is 0.337 in group D and 0.667 of group M (both are also within $-2/+2$). The Shapiro-Wilk values are very close to 1 (0.987 in group D and 0.991 in group M), and the P-value of Shapiro-Wilk in both groups are larger than 0.05 (0.963 in group D and 0.864 in group M), so data are normally distributed for both two groups seeing from these statistics. When the boxplots (Figure 65) are referred to, it is obvious that except for one particular case in group M, other data are indeed normally distributed.

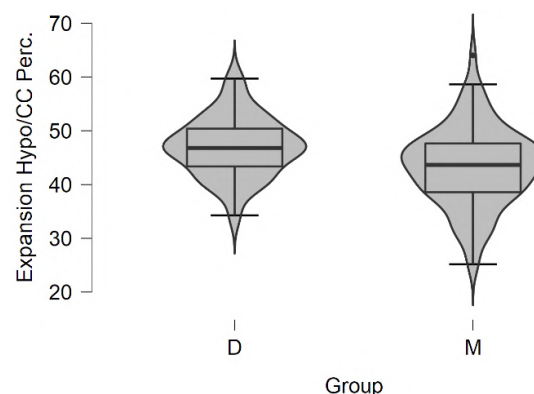


Figure 65 The boxplots comparing the proportion of hypotactic expansion in clause nexuses in Master and PhD sub-corpora

For preciseness, both Welch's t-test and Mann-Whitney U test results (See Table 44) are listed here for comparison. In the former test, the positive values of statistic (2.890) and location parameter (3.690) mean that the proportion in group D is 2.89% and 3.69% larger than group M as to hypotactic expansion in different calculation methods, and the SE difference (1.277) also confirms this diversity to some degree. As the p-value is 0.003 (which is smaller than 0.05), the original hypothesis that group D is less than group M is totally rejected with VS-MPR, meaning that the probability of accepting the alternative hypothesis is 24.494 times of rejecting it. Medium effect size (0.567) proves this test to be efficient. The same conclusion can be received from the Mann-Whitney U-test results as the p-value is 0.004 and the VS-MPR is 16.981 with a small effect size (0.323). So, group D has a higher proportion of hypotactic expansion in clause nexuses compared to group M.

Independent Samples T-Test

	Test	Statistic	df	p	VS-MP R*	Location Parameter	SE Difference	Effect Size
Expansion Hypo/CC Perc.	Welch	2.890	75.753	0.003	24.494	3.690	1.277	0.567
	Mann-Whi tney	1693.500		0.004	16.981	3.455		0.323

Note. For all tests, the alternative hypothesis specifies that group D is greater than group M.

Note. For the Welch t-test, effect size is given by Cohen's d. For the Mann-Whitney test, effect size is given by the rank biserial correlation.

Note. For the Welch t-test, location parameter is given by mean difference. For the Mann-Whitney test, location parameter is given by the Hodges-Lehmann estimate.

Table 44 The independent samples t-test for comparing the proportion of hypotactic expansion in clause nexuses in Master and PhD sub-corpora

In the Bayesian Mann-Whitney U-test, when the original hypothesis is that group D is larger than group M in the data proportion location, the possibility of accepting this original hypothesis is 15.44 times of accepting the alternative one (see Figure 66), with the Rhat value is 1.023 and a small effect size. Thus, group D has a higher proportion of hypotactic expansion in clause nexuses than group M with sufficient proof.

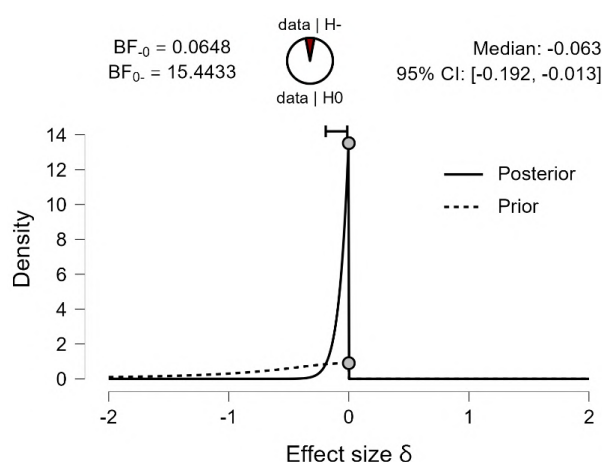


Figure 66 The inferential plots comparing the proportion of hypotactic expansion in clause nexuses in Master and PhD sub-corpora (with the original hypothesis group D>group M)

5.1.3.2 Expansion Hypotaxis/Hypotaxis

Hypotaxis includes hypotactic expansion and hypotactic projection, in which at least 47% in group M and 62% in group D are hypotactic expansions, and to the maximum, 83% in group D and even 90% in group M can be found. (See Table 5.1.3.2-1 in Appendix 5.1) The mean in group D (72.30%) is only 1% larger than that of group M (71.15%), but the variance in the latter group is higher (71.16) with a broader range (42.97%) compared to the former group. Data in group D is right-sided skew (the skewness value is 0.214), leptokurtic (the kurtosis value is -0.498) and normally distributed as the division of skewness to the standard error of skewness is 0.517 (within -2/+2), the division of the kurtosis to the standard error of kurtosis is -0.616 (within -2/+2), the Shapiro-Wilk value is 0.982 (close to 1) and the P-value of Shapiro-Wilk (0.86) is larger than 0.05. But the data in group M is left-sided skew (the skewness value is -0.806), leptokurtic (the kurtosis value is 0.699), and abnormally distributed as values of two divisions are -3 (smaller than -2) and 1.3 for each and the P-value of Shapiro-Wilk is 0.002 (smaller than 0.05). The

boxplots (Figure 67) can bring out more information directly on data distribution that five peculiar cases make the whole data normally distributed in group M.

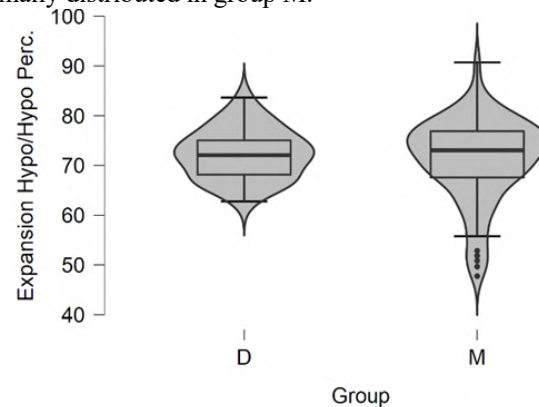


Figure 67 The boxplots comparing the proportion of hypotactic expansion in hypotaxis in Master and PhD sub-corpora

On account of this, the Mann-Whitney U-test (Table 45) is firstly conducted with the original hypothesis group D equal to group M. As the P-value is 0.982, the original hypothesis is thus accepted, and the Hodges-Lehmann Estimate is -0.04, the difference between the two groups is very small. But this test has a very low effect size due to the rank-biserial correlation value of only -0.003.

Independent Samples T-Test

	W	df p	VS-MPR *	Hodges-Lehmann Estimate	Rank-Biserial Correlation
Expansion Hypo/Hypo Perc.	1276.000	0.982	1.000	-0.040	-0.003

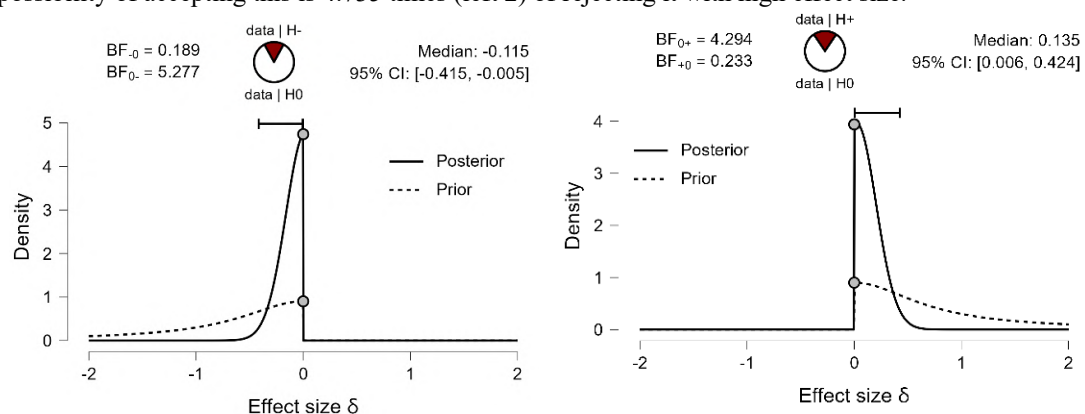
Note. For the Mann-Whitney test, effect size is given by the rank biserial correlation.

Note. Mann-Whitney U test.

* Vovk-Sellke Maximum p -Ratio: Based on a two-sided p -value, the maximum possible odds in favor of H_1 over H_0 equals $1/(-e p \log(p))$ for $p \leq .37$ (Sellke, Bayarri, & Berger, 2001).

Table 45 The Mann-Whitney U test for comparing the proportion of hypotactic expansion in hypotaxis in Master and PhD sub-corpora

When the independent samples t-test cannot clearly demonstrate the differentiation between two groups, the Bayesian Mann-Whitney U test has to be conducted. From Figure 68, when the alternative hypothesis specifies that the location of group D is smaller than the location of group M, the probability of rejecting it, i.e., accepting the null hypothesis is 5.277 times of accepting the alternative with medium effect size as demonstrated in left 1 picture, and the Rhat value is 1.012; when the alternative hypothesis is group $D > \text{group M}$, the possibility of accepting the original hypothesis (group $D < \text{group M}$) is 4.294 times (right 1) of rejecting it with medium effect size; and when the original hypothesis is that group $D = \text{group M}$, the possibility of accepting this is 4.735 times (left 2) of rejecting it with high effect size.



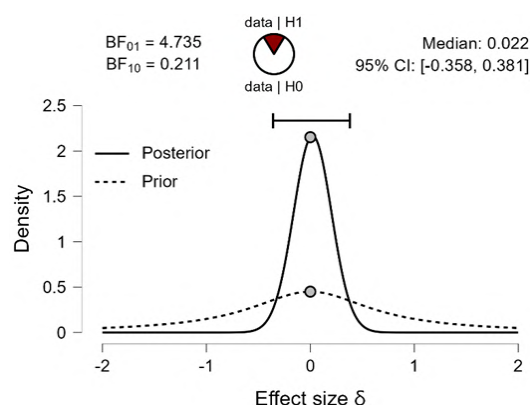


Figure 68 The inferential plots comparing the proportion of hypotactic expansion in hypotaxis in Master and PhD sub-corpora with diverse original hypotheses

Thus, a conclusion can be reached that group D has a larger proportion of hypotactic expansion in hypotaxis compared to group M, though the difference is quite small.

5.1.3.3 Expansion Hypotaxis/Expansion

The testing results in Section 5.1.2.3 can confirm that in expansion structures, group D has a higher proportion of hypotactic expansion than group M, and data in these two groups are normally distributed. Hence, only the descriptive statistics will be demonstrated here for detailed information. In Table 5.1.3.3-1 (see Appendix 5.1), the minimum and maximum values of group D (43% and 76%) are both larger than those of group M (33% and 71%) and so does the mean (58% in group D compared to 53% in group M).

To make a conclusion for paratactic and hypotactic expansions, the proportion of paratactic expansion per clause nexuses and paratactic expansion per expansions of group D are obviously smaller than those of group M; and correspondingly, the proportion of hypotactic expansion per clause nexuses and hypotactic expansion per expansions of group D are larger than those of group M, a preliminary statement can thus be that doctoral students prefer to choose more complex expansion structures in writing their theses compared to Master students if hypotaxis is consumed to be harder than parataxis for students to learn, organize, convey ideas and even transmit knowledge. But what grammatical structures make the higher proportion of hypotactic expansions in group D and why the related lexis is chosen instead of another lexis that probably indicates parataxis is to be answered in the next sections.

5.1.4 Expansion Composition

From Sections 3.1.2 & 3.1.3, the expansion in clause complex system usually includes elaboration, extension, and enhancement structures that perform diverse functions according to discourse contexts and are realized by kinds of lexis that can be selected in performing these functions. So, in comparing expansion choices, the proportion of elaboration, extension, and enhancement structures are inevitable for discussion.

5.1.4.1 Elaboration (Descriptive Statistics; Elaboration/Expansion)

The choice of elaboration in some texts can only be 24 (in group D) and 7 (in group M), while it can be 177 (in group D) and 98 (in group M).

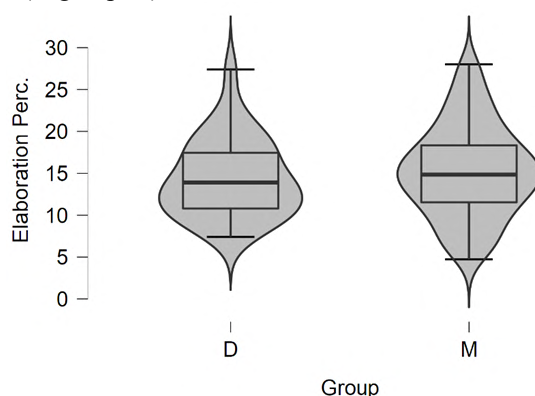


Figure 69 The boxplots of the number of elaborations in Master and PhD sub-corpora

Detailed information from Table 5.1.4.1-1 (see Appendix 5.1) shows that data are leptokurtic (the kurtosis values are 0.669 in group D and 1.025 in group M) distributed with a right-sided skewness (values are 0.768 in group D and 0.867 in group M) in both groups, which means that the number of elaborations concentrates in scope but with some extremely high conditions. A clearer data distribution of the number of elaborations in each group is suggested by Figure 69 that one particular case in group D and four particular cases make the elaboration means higher.

Elaboration/Expansion

In expansion, only 14% (group D) or 15% (group M) structures are elaborations and the highest percentage is only 27% (group D) or 28% (group M). (See Table 5.1.4.1-2 in Appendix 5.1) This means that the elaboration proportion cannot be higher than 1/3 in every text written by these students from the corpus compared to the other two grammatical categories. From the skewness and kurtosis values, data in group D is right-sided skew (0.781) and leptokurtic (0.44) distributed; and in group M is also right-sided skew (0.264) but platykurtic (-0.34) distributed. From the division of skewness to the standard error of skewness (1.886 in group D and 0.981 in group M, within -2/+2), the division of kurtosis to the standard error of kurtosis (0.544 in group D and -0.639 in group M, within -2/+2), the Shapiro-Wilk (0.954 in group D and 0.984 in group M, close to 1) and the P-value of Shapiro-Wilk (0.189 in group D and 0.403 in group M, larger than 0.05), data in two groups are both normally distributed. This can also be demonstrated in Figure 70.

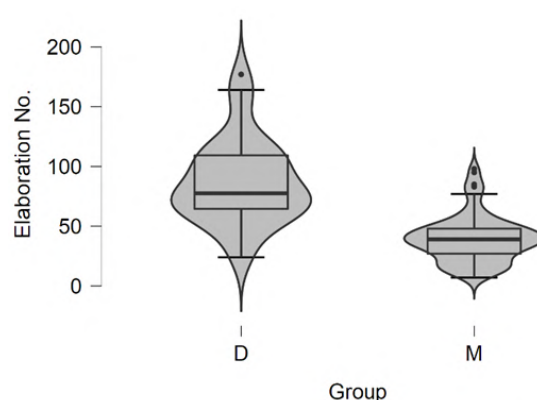


Figure 70 The boxplots comparing the proportion of elaboration per expansion in Master and PhD sub-corpora

With diverse variance values, Welch's t-test results are considered useful, in which the negative values of t (-0.858) and mean difference (-0.884) indicate that group D has a slightly lower percentage of elaboration compared to group M. The SE difference is 1.030, i.e., the difference between the two groups is only 1.03%, judged from this parameter. As the p -value is 0.197, possibilities still exist in supporting the original hypothesis; and the VS-MPR is 1.15, which indicates that the probability of accepting the alternative hypothesis (group $D < \text{group M}$) is only 1.15 times rejecting it with a small effect size (the Cohen's d value is -0.173). (See Table 46)

Independent Samples T-Test

	t	df	p	VS-MPR*	Mean Difference	SE Difference	Cohen's d
Elaboration Perc.	-0.858	66.745	0.197	1.150	-0.884	1.030	-0.173

Note. For all tests, the alternative hypothesis specifies that group D is less than group M .

* Vovk-Sellke Maximum p -Ratio: Based on a two-sided p -value, the maximum possible odds in favor of H_1 over H_0 equals $1/(-e p \log(p))$ for $p \leq .37$ (Sellke, Bayarri, & Berger, 2001).

Note. Welch's t-test.

Table 46 The Welch's t-test for comparing the proportion of elaboration in expansion in Master and PhD sub-corpora

Because the test above cannot present a certain result, the Bayesian Mann-Whitney U test is referred to. When the alternative hypothesis specifies that the location of group D is greater than the location of group M , the probability of accepting the original hypothesis (group $D < \text{group M}$) is 6.477 times (Figure 71, left 1) of rejecting it with medium effect size and the R_{hat} value is 1.065, which indeed suggests a diversity in two groups; when the alternative hypothesis is that group D has lower proportion than group M , the possibility of accepting the original hypothesis (group $D > \text{group M}$) is 3.077 times (Figure 71, right 1) of rejecting it with medium effect size; and when the original hypothesis is that group D equals to group M , the probability of accepting this then is 4.17 times (Figure 71, left 2) of rejecting it with large effect size. So, group D is considered to have a lower proportion of elaboration in expansion than group

M after comparing these Bayesian factors. Though the difference between the two groups is not quite obvious.

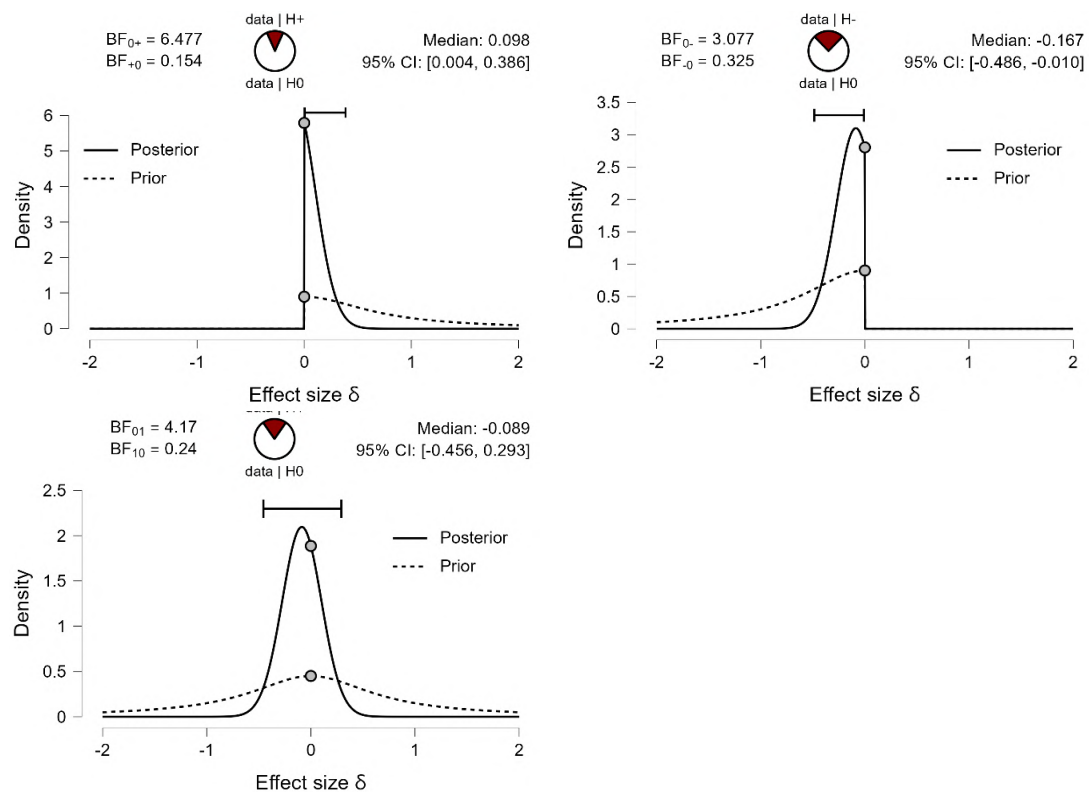


Figure 71 The inferential plots comparing the proportion of elaboration in expansion in Master and PhD sub-corpora (with diverse original hypotheses)

The conclusion here can be affected by the proportion of extension and enhancement in expansion, which will be approved in the next two sections.

5.1.4.2 Extension (Descriptive Statistics; Extension/Expansion)

The total number of extensions in group D is 7768, and in group M is 8594; the mean of group D (242) is 2.26 times of group M (107). (See Table 5.1.4.2-1 in Appendix 5.1) In some texts, only 71 (in group D) or 50 (in group M) extensions can be found, while in some texts, as many as 400 (in group D) or 263 (in group M) extensions are used. Data seems to be normally and leptokurtic distributed in group D with the kurtosis value of -1.107 and the P-value of Shapiro-Wilk 0.197; and is abnormally, right-sided skew, and leptokurtic distributed in group M when seeing from the skewness value (1.344), the kurtosis value (2.403) and the P-value of Shapiro-Wilk (<0.001). However, boxplots in Figure 72 seem to disapprove of these statements that data in group D are normally and bimodal distributed, and in group M is also normally distributed except for two cases with a high number of extensions.

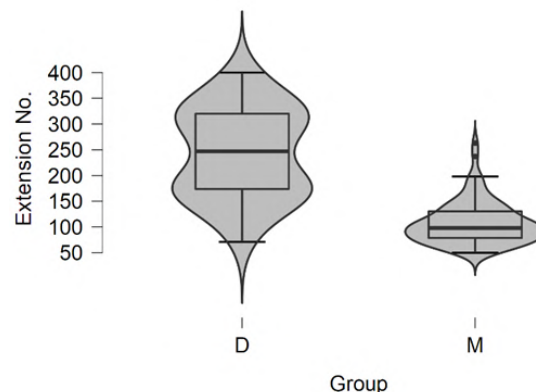


Figure 72 The boxplots of the number of extensions in Master and PhD sub-corpora

Extension/Expansion

In some texts, extension accounts for 19% (in group D) or 24% (in group M) in all expansion structures and can account for 56% (in group D) or 64% (in group M) in some texts, i.e., group M has a broader range of extension proportion; the mean of group D is 39.39% and of group M is 42%, the general statistics seem to indicate that group D has a lower proportion of extension than group M. (See Table 5.1.4.2-2 in Appendix 5.1) Data in group D is leptokurtic (the kurtosis value is 0.484), normally distributed seeing from the division of skewness to the standard error of skewness (-0.012, within -2/+2), the division of the kurtosis to the standard error of the kurtosis (0.598, within -2/+2), the Shapiro-Wilk (0.980, close to 1) and the P-value of the Shapiro-Wilk (0.814, larger than 0.05); and in group M is right-sided skew (the skewness value is 0.312), a little bit leptokurtic (the kurtosis value is 0.231), and also normally distributed as two division values are 1.160, 0.434 (within -2/+2) for each parameter, the Shapiro-Wilk value is 0.990 (close to 1) and the P-value of Shapiro-Wilk is 0.767 (larger than 0.05). To be clearer, the boxplots (Figure 73) are put here for reference. Except for one case of low percentage in group D and one of high percentage in group M, other data are indeed normally distributed in two groups.

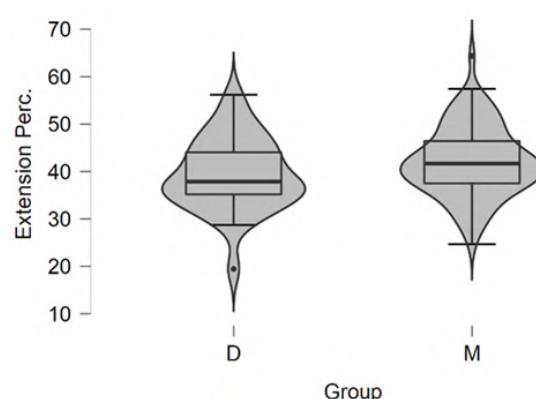


Figure 73 The boxplots comparing the proportion of extension in expansion in Master and PhD sub-corpora

In this condition, both Welch's t-test and the Mann-Whitney U-test results are listed in Table 47 for further discussion. The negative values of statistic (-1.660) and location parameter (-2.647) in Welch's indicate how much the difference is between the two groups (group D compared to group M). Thus, group D is 1.66% or 2.647% lower than group M for this parameter from a diverse perspective. The same condition goes to the SE difference value. As the P-value is 0.051, which gets very close to 0.05, the original hypothesis that group D is larger than group M is highly rejected, and the probability of accepting the alternative hypothesis is 2.415 times rejecting it, as seen from the VS-MPR value. The effect size is small to medium when the value is -0.350. So, for these two tests, results suggest that group D has a lower proportion of extension used in expansion structures as compared to group M.

Independent Samples T-Test

	Test	Statistic	df	p	VS-MP R*	Location Parameter	SE Difference	Effect Size
Extension Perc.	Welch	-1.660	55.543	0.051	2.415	-2.647	1.594	-0.350
	Mann-W hitney	1017.000			0.045	2.619	-2.624	-0.205

Note. For all tests, the alternative hypothesis specifies that group D is less than group M.

Note. For the Welch t-test, effect size is given by Cohen's d. For the Mann-Whitney test, effect size is given by the rank biserial correlation.

Note. For the Welch t-test, location parameter is given by mean difference. For the Mann-Whitney test, location parameter is given by the Hodges-Lehmann estimate.

* Vovk-Sellke Maximum p -Ratio: Based on a two-sided p -value, the maximum possible odds in favor of H_1 over H_0 equals $1/(-e p \log(p))$ for $p \leq .37$ (Sellke, Bayarri, & Berger, 2001).

Table 47 The independent samples t-test for comparing the proportion of extension in expansion in Master and PhD sub-corpora

The result of the Bayesian Mann-Whitney U test also confirms the conclusion above with the Bayesian factor is 10.39 (see Figure 74), which means that the probability of accepting the original hypothesis (group D < group M) is 10.39 times accepting the alternative hypothesis (group D > group M), with a small effect size and the Rhat value 1.061 meaning that this difference truly exists within two groups.

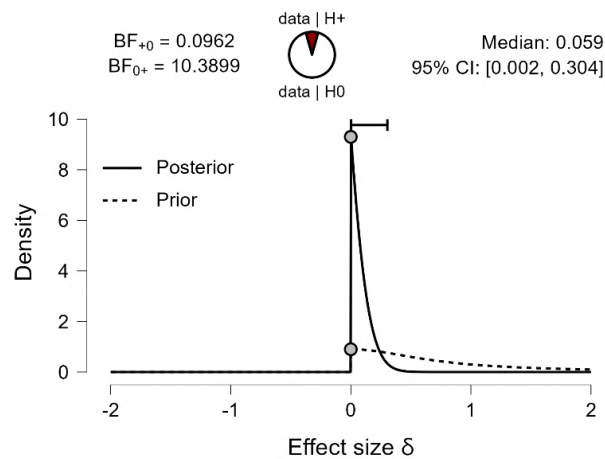


Figure 74 The inferential plots comparing the proportion of extension in expansion in Master and PhD sub-corpora (with the original hypothesis group D < group M)

So, unlike conditions in elaboration proportion, group D has a significantly lower proportion of extension in expansion structures than group M in the writing thesis. The exact reasons are to be investigated, but at least one thing is for sure, i.e., the proportion of enhancement in group D is certainly larger than that in group M (see Section 5.1.4.3). If a higher percentage of enhancement means higher writing abilities and more complex grammatical structures, then doctoral students can be regarded as, to some degree, one level in advance in writing as compared to Master students.

5.1.4.3 Enhancement (Descriptive Statistics; Enhancement/Expansion)

The total, general information of the number of enhancements used in each group is demonstrated in Table 5.1.4.3-1 (see Appendix 5.1), in which the total number of group D is 9017 and in group M is 8875. The number in each text can range from 105 to 522 in group D and from 50 to 280 in group M, with a mean of 281 in the former group and 110 in the latter. Data in group D is right-sided skew (the skewness value is 0.460), a little bit leptokurtic (the kurtosis value is 0.233), and normally distributed (the division of the skewness to the standard error of skewness is 1.11, the division of the kurtosis to the standard error of kurtosis is 0.288, the Shapiro-Wilk value is 0.973 and the P-value of Shapiro-Wilk is 0.589) while in group M is totally right-sided skew (the skewness value is 1.454), leptokurtic (the kurtosis value is 2.331) and abnormally distributed with two division values are 5.41, 4.38 separately (larger than +2), and the P-value of Shapiro-Wilk is less than 0.001. More information can be found in the boxplots (Figure 75) that one case in group D and four cases in group M raise the average number of enhancements, as demonstrated in the table described above.

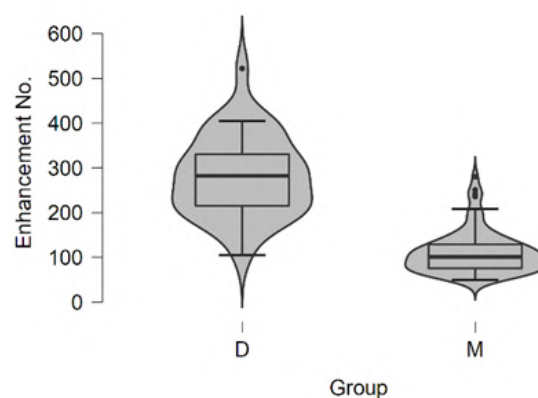


Figure 75 The boxplots of the number of enhancements in Master and PhD sub-corpora

Enhancement/Expansion

It can be inferred from the preceding two sections that Group D has a higher proportion of enhancement in expansion than Group M. However, the magnitude of this difference and the general conditions of these two groups remain open for discussion. Consequently, the same analytical process is employed as previously described in the Methods section and conducted in the preceding sections. The proportion of enhancement in some texts in Group D ranges from 33% to 59%, while in Group M, it ranges from 25% to 64%. The mean value for group D is 46%, while for group M, it is approximately 43%. (Please refer to

Table 5.1.4.3-2 in Appendix 5.1.) The data in group D exhibits minimal skewness (0.038) and is platykurtic (kurtosis value is -0.387), indicating a normal distribution as evidenced by the statistical analysis (skewness to the standard error of skewness ratio is 0.092, kurtosis to the standard error of kurtosis ratio is -0.478, the Shapiro-Wilk value is 0.994 and the P-value of Shapiro-Wilk is 1). Group M exhibits right-sided skew (skewness value: 0.359) and leptokurtic (kurtosis value: 1.139) distribution. The normality of the data is undecided, as evidenced by the two divisions of the Shapiro-Wilk value, which are 1.334 and 2.14, respectively. The P-value of the Shapiro-Wilk test is 0.319, which does not provide sufficient evidence to reject the null hypothesis of normality. It is not possible to reach a definitive conclusion until the boxplots (Figure 76) are consulted.

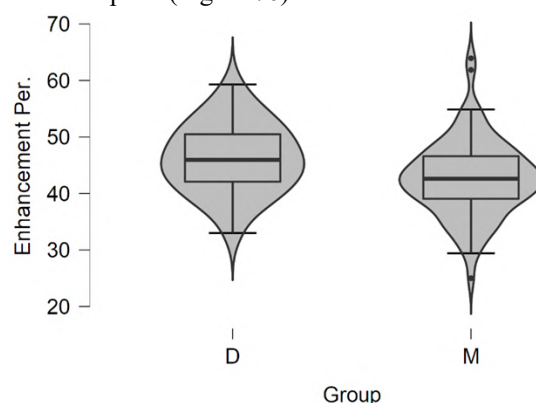


Figure 76 The boxplots comparing the proportion of enhancement in expansion in Master and PhD sub-corpora

From the Mann-Whitney U test result (Table 48), the original hypothesis is totally rejected as the p-value is 0.005, with the VS-MPR value 14.007, i.e., the probability of accepting the alternative hypothesis is 14 times of rejecting it, so the proportion of enhancement in expansion of group D is larger than that of group M; the Hodges-Lehmann Estimate denotes that the difference can be 3.637% within two groups with a small to medium effect size as the rank-biserial correlation value is 0.313. The data in group D is normally distributed, while that in group M is non-normally distributed, with two cases of high percentage and one of low percentage.

Independent Samples T-Test

	W	df	p	VS-MPR*	Hodges-Lehmann Estimate	Rank-Biserial Correlation
Enhancement Per.	1681.000		0.005	14.007	3.637	0.313

Note. For the Mann-Whitney test, effect size is given by the rank biserial correlation.

Note. For all tests, the alternative hypothesis specifies that group D is greater than group M.

Note. Mann-Whitney U test.

* Vovk-Sellke Maximum p -Ratio: Based on a two-sided p -value, the maximum possible odds in favor of H_1 over H_0 equals $1/(-e p \log(p))$ for $p \leq .37$ (Sellke, Bayarri, & Berger, 2001).

Table 48 The Mann-Whitney U test for comparing the proportion of enhancement in expansion in Master and PhD sub-corpora

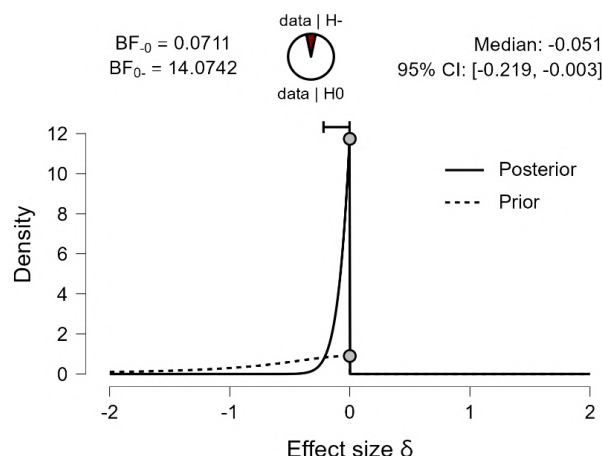


Figure 77 The inferential plots comparing the proportion of enhancement in expansions in Master and PhD sub-corpora (with the original hypothesis group D>group M)

The Bayesian Mann-Whitney U test also confirms this result as the Bayesian factor is 14.07 (See Figure 77), i.e., the possibility of accepting the original hypothesis (group D>group M) is 14.07 times rejecting it with a Rhat value 1.040 (the parameter in two groups is diverse) and a small effect size.

5.1.5 Summary of Expansion Comparison

To sum up, for expansion clause complexes, doctoral students has almost the same proportion of expansion per clause nexuses as Master students though the former is a little bit smaller than the latter one; for paratactic expansion, the proportion of paratactic expansion per clause nexuses and per expansion in doctoral students' theses is firmly smaller than the latter, and the proportion of paratactic expansion in parataxis is almost the same with the doctors' slightly bigger than Masters'; for hypotactic expansion, the proportion of hypotactic expansion per clause nexuses and expansion in doctoral students' is thus larger than that of the Master students', and slightly larger than the latter group for the proportion of hypotactic expansion in hypotaxis; for expansion constituents, small difference can be found for the proportion of elaboration though doctoral students has a lower one, and Master students has a higher proportion of extension than doctors' while smaller proportion of enhancements. Should hypotactic structures and enhancements be considered more challenging to incorporate into a thesis, doctoral students are likely to encounter greater complexity, given the increased likelihood of including these elements in their work.

5.2 Elaboration (Quantitatively Comparison)

As has been described in Section 3.1.3.1, elaboration is the relationship of one clause that develops the prior one by providing clarifications or descriptively commenting on the information introduced in the prior clause in a clause complex (Halliday & Matthiessen, 2004). It is in traditional grammar presented as non-defining relative clauses that the secondary clause usually provides background information, characterization, or interpretation for the former one (Nagar & Fine, 2013). From structures, elaboration can be hypotactic or paratactic or can be exposition, exemplification, and clarification from functions they perform in clause complexes. Because calculations for functional organization are conducted manually, they are too delicate and laborious, and even some categories did not appear in the corpus, only the former classification is quantitatively researched, i.e., paratactic and hypotactic elaboration are discussed in this section in detail.

5.2.1 General Distribution of Elaboration (Elaboration/Clause Nexuses)

In Section 5.1.4.1, the general information (descriptive statistics) of elaboration annotated in the corpus and the proportion of elaboration in expansion structures in two groups have been compared, but what is the condition of elaboration in clause nexuses, i.e., the proportion of elaboration in the whole clause complex system? From Table 5.2.1-1 (see Appendix 5.2), around 11.69% of clause complexes in group D and 12.43% in group M are elaborations that in some texts, this proportion ranges from 6.16% to 21.32% in group D and 4.23% to 24.71% in group M. These data are kinds of right-sided skew, a little bit platykurtic, and normally distributed in both groups as the skewness values in group D are 0.693 and in group M 0.393, the kurtosis values are -0.002 and -0.171 separately, the division of the skewness to the standard error of skewness in group D is 1.676 and in group M is 1.461 (within -2/+2), the division of the kurtosis to the standard error of kurtosis is -0.002 in group D and -0.321 in group M (also within -2/+2), the Shapiro-Wilk in both groups are close to 1 (0.955 of group D and 0.980 in group M), and the P-value of Shapiro-Wilk in both groups are larger than 0.05 (0.193 in group D and 0.245 in group M).

Independent Samples T-Test

	t	df	p	VS-MPR *	Mean Difference	SE Difference	Cohen's d
Elaboration-CC Perc.	-0.873	67.218	0.193	1.159	-0.742	0.850	-0.176

Note. For all tests, the alternative hypothesis specifies that group D is less than group M.

* Vovk-Sellke Maximum *p* -Ratio: Based on a two-sided *p*-value, the maximum possible odds in favor of *H*₁ over *H*₀ equals $1/(-e p \log(p))$ for $p \leq .37$ (Sellke, Bayarri, & Berger, 2001).

Note. Welch's t-test.

Table 49 The Welch's t-test for comparing the proportion of elaboration in clause nexuses in Master and PhD sub-corpora

For the test, the Welch's t-test (See Table 49) is hereby taken that the negative values of the statistic (-0.873), and the location parameter (-0.742) indicate that the proportion in group D is 0.8/0.7% smaller

than that in group M, and the SE difference value is 0.850, which also confirms this difference. As the p-value is 0.193, and the VS-MPR value is 1.159, possibilities still exist that the original hypothesis (group D > group M) can be accepted, and the probability of accepting the alternative hypothesis is only 1.159 times of rejecting it. This test has a small effect size as the Cohen's d value is -0.176. So, the proportion of group D can be less than that of group M, with more necessity to verify.

In the Bayesian Mann-Whitney U test (See Figure 78), the Bayesian factor (BF_{0+}) is calculated to be 7.257 with the alternative hypothesis specifying that the proportion of group D is larger than that of group M, which means that the probability of accepting the null hypothesis (group D < group M) is 7.257 times of rejecting it with a small effect size and the Rhat value 1.072. Thus, group D has a lower proportion of elaboration than group M in clause nexuses, which corresponds to the proportion of elaboration in expansions (Section 5.1.4.1).

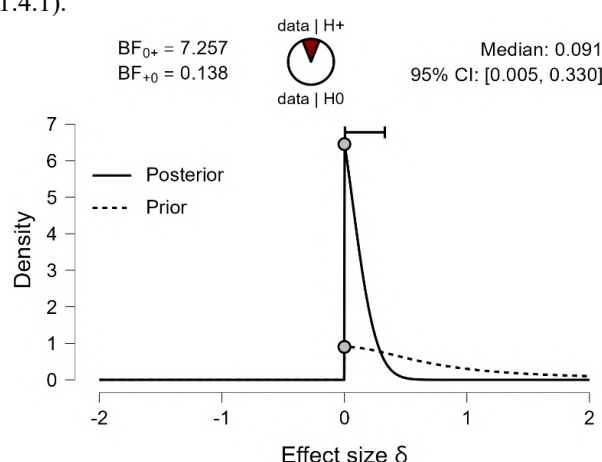


Figure 78 The inferential plots comparing the proportion of elaboration of clause nexuses in Master and PhD sub-corpora (with the original hypothesis group D < group M)

5.2.2 The Distribution of Paratactic Elaboration

In all clause nexuses and elaboration structures, 585 paratactic elaborations in group D and 661 in group M in total can be found, with a mean of 18 in the former and around 8 in the latter. (See Table 5.2.2-1 in Appendix 5.2) In some texts, a maximum of 52 are used in both groups, while in some texts, only two are applied in group D and even zero in group M. With the P-value of Shapiro-Wilk in both groups less than 0.001, there is no doubt that data are abnormally distributed. The boxplots (Figure 79) show how data are exactly distributed in two groups, in which group D has a large right-sided tail with two particular cases and group M has a narrow but long right-sided tail with three cases.

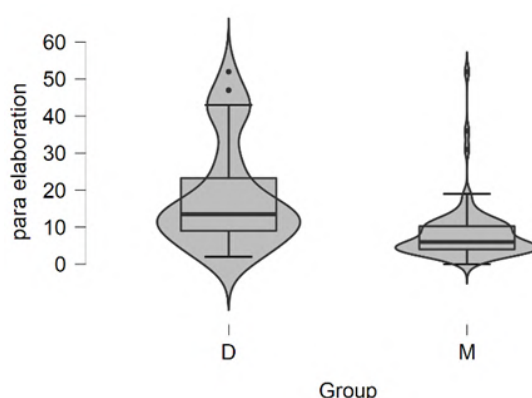


Figure 79 The boxplots of the number of paratactic elaborations in Master and PhD sub-corpora

5.2.2.1 Elaboration Parataxis/Clause Nexuses

The average proportion of paratactic elaboration in clause nexuses in each text is quite small, and the mean is only 2.3% (in group D) or 2.57% (in group M). (More general statistics can refer to Table 5.2.2.1-1 in Appendix 5.2) As the P-value of Shapiro-Wilk is smaller than 0.05 in both groups, data of this parameter are surely abnormally distributed, as shown in Figure 80. The shape of the violin plots in two groups is affected by the data of paratactic elaboration numbers, and the distribution in group D is even bimodal.

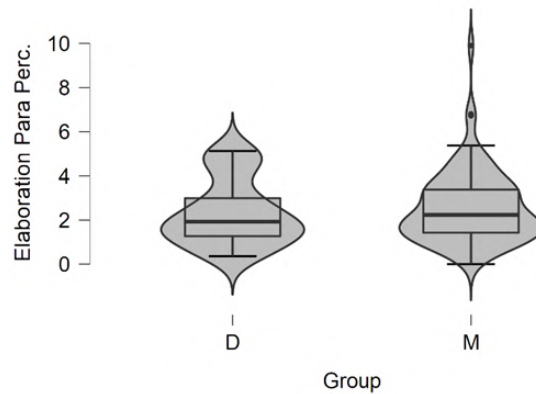


Figure 80 The boxplots comparing the proportion of paratactic elaboration in clause nexuses in Master and PhD sub-corpora

In the Mann-Whitney U test, the p -value is 0.23, (See Table 50) which shows the possibility of accepting the original hypothesis to some degree, and the VS-MPR value is 1.088, this indicates that the probability of accepting the alternative hypothesis is 1.088 times rejecting it, i.e., the alternative hypothesis is hard to be completely accepted. The Hodges-Lehmann value suggests that group D is only 0.235% smaller than group M on this parameter (can even be ignored). With almost no effect size (the Rank-biserial correlation value is only -0.09) of this test, the Bayesian Mann-Whitney U test is then conducted.

Independent Samples T-Test

	W	df	p	VS-MPR*	Hodges-Lehmann Estimate	Rank-Biserial Correlation
Elaboration Para Perc.	1165.000		0.230	1.088	-0.235	-0.090

Note. For the Mann-Whitney test, effect size is given by the rank biserial correlation.

Note. For all tests, the alternative hypothesis specifies that group D is less than group M.

Note. Mann-Whitney U test.

* Vovk-Sellke Maximum p -Ratio: Based on a two-sided p -value, the maximum possible odds in favor of H_1 over H_0 equals $1/(-e p \log(p))$ for $p \leq .37$ (Sellke, Bayarri, & Berger, 2001).

Table 50 The Mann-Whitney U test for comparing the proportion of paratactic elaboration in clause nexuses in Master and PhD sub-corpora

When the original hypothesis is set to be group D has a lower proportion of paratactic elaboration per clause nexuses than group M, the BF_{0+} value is 6.96 (See Figure 81) with a small to medium effect size and the R_{hat} value 1.130 (diversity does exist within two groups), which means that the probability of accepting the original hypothesis is almost 7 times of rejecting it. A conclusion can thus be reached that group M has a higher proportion than group D, though this difference can be small as the abnormally distributed data in the corpus and the small number of paratactic elaborations discovered.

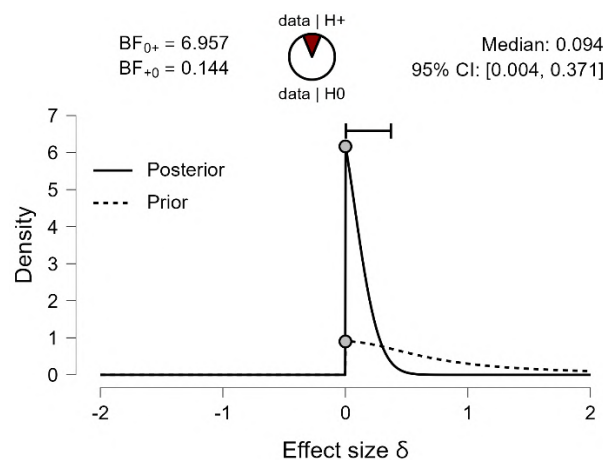


Figure 81 The inferential plots comparing the proportion of paratactic elaboration in clause nexuses in Master and PhD sub-corpora (with the original hypothesis group D<group M)

5.2.2.2 Elaboration Parataxis/Parataxis

Besides clause nexuses, the denominator can also be parataxis, which investigates the proportion of paratactic elaboration compared to paratactic extension, paratactic enhancement, paratactic locution, etc., in the corpus. In Table 5.2.2.2-1 (see Appendix 5.2), the proportion of paratactic elaboration in parataxis in two groups in each text is thus compared with the result that the mean of the two groups is almost the same (6.6%). Data in group D seems to be right-sided skew, leptokurtic, and normally distributed as the value of skewness is 0.395, of kurtosis is -0.946, the division of the skewness to the standard error of skewness is 0.954 (within -2/+2), the division of kurtosis to the standard error of kurtosis is -1.169 (also within -2/+2), the Shapiro-Wilk value is 0.947 (close to 1), and the P-value of Shapiro-Wilk is 0.12 (>0.05); and in group M is right-sided skew (the value of skewness is 1.244), leptokurtic (the kurtosis value is 1.945), the values of two divisions are 4.625, 3.656 separately, the Shapiro-Wilk value is 0.913 (slightly far away from 1), and the P-value of Shapiro-Wilk is smaller than 0.001. Figure 82 shows the violin plots of data distribution in Master and PhD sub-corpora.

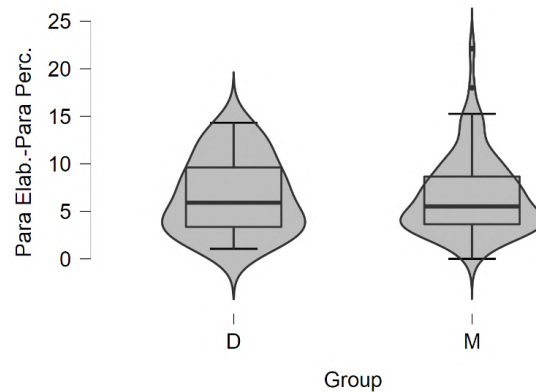


Figure 82 The boxplots comparing the paratactic elaboration in parataxis in Master and PhD sub-corpora. Again, the Mann-Whitney U-test (Table 51) is chosen. When the original hypothesis in group D does not differ from that in group M, the p-value is 0.964 with a VS-MPR value of 1, and the alternative hypothesis is totally rejected! The difference between the two groups is only 0.065%, as indicated by the Hodges-Lehmann estimate, and this test has a very small effect size, as suggested by the rank-biserial correlation value (0.006).

Independent Samples T-Test

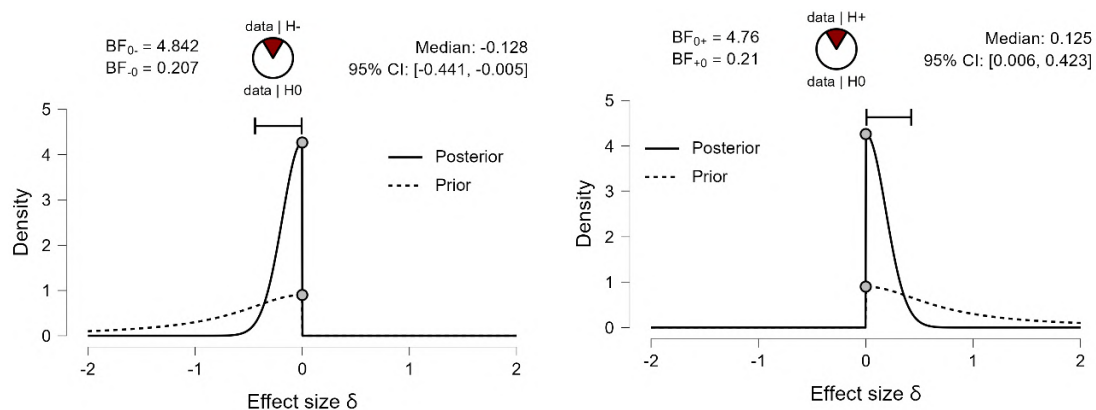
	W	df	p	VS-MPR*	Hodges-Lehmann Estimate	Rank-Biserial Correlation
Para Elab.-Para Perc.	1287.500		0.964	1.000	0.065	0.006

Note. For the Mann-Whitney test, effect size is given by the rank biserial correlation.

Note. Mann-Whitney U test.

* Vovk-Sellke Maximum p -Ratio: Based on a two-sided p -value, the maximum possible odds in favor of H_1 over H_0 equals $1/(-e p \log(p))$ for $p \leq .37$ (Sellke, Bayarri, & Berger, 2001).

Table 51 The Mann-Whitney U test for comparing the paratactic elaboration in parataxis in Master and PhD sub-corpora



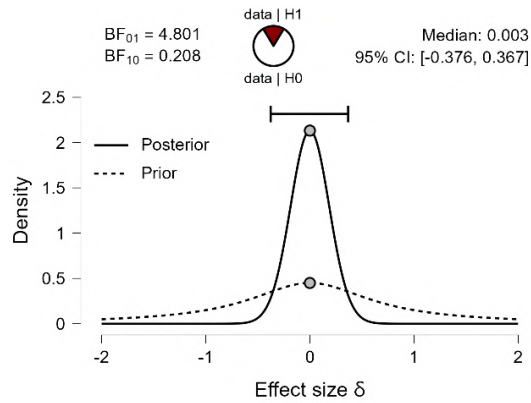


Figure 83 The inferential plots comparing the paratactic elaboration in parataxis in Master and PhD sub-corpora (with diverse original hypotheses)

For better presenting the differentiation between the two groups, the Bayesian Mann-Whitney U-test is then conducted. When the original hypothesis is that group D has a larger proportion than group M on the proportion of paratactic elaboration in parataxis in two groups, the probability of accepting this original hypothesis is 4.842 times (See Figure 83, left 1) of rejecting it with a medium effect size and the Rhat value 1.04; when the original hypothesis is that group D < group M, the probability of accepting this is 4.76 times (Figure 83, right 1) of rejecting it with a medium effect size; and when the original hypothesis is that two groups have no difference, then the possibility of accepting it is 4.801 times (Figure 83, left 2) of rejecting it with a large effect size. A conclusion can here be made that doctoral students apply almost the same proportion of paratactic elaboration in parataxis compared to Master students in thesis writing, though the former is slightly higher than the latter group.

5.2.2.3 Elaboration Parataxis/Expansion Parataxis

Being restricted to paratactic expansion (also includes paratactic extension and paratactic enhancement), the proportion of paratactic elaboration in the two groups does not increase too much as compared to the proportion in the former two sections. The mean is only around 6.8% (See Table 5.2.2.3-1 in Appendix 5.2), with the highest no more than 23% (in group M). Data are right-sided skew, leptokurtic, and normally distributed in group D as the value of the skewness is 0.419, the value of the kurtosis is -0.995, the division of the skewness to the standard error of skewness is 1.012 (within -2/+2), the division of the kurtosis to the standard error of kurtosis is -1.230 (also within -2/+2), the Shapiro-Wilk is 0.936, and the P-value of Shapiro-Wilk is 0.058; while are right-sided skew, leptokurtic, and abnormally distributed in group M as the value of the skewness is 1.242, the value of the kurtosis is 2.076, the division of the skewness to the standard error of skewness is 4.617 (larger than +2), the division of the kurtosis to the standard error of kurtosis is 3.902 (also larger than +2), the Shapiro-Wilk is 0.917, and the P-value of Shapiro-Wilk is less than 0.001. From the violin plots in Figure 84, it is apparent that group D has a short but broader tail to the right side while group M has a long and narrow tail because of two specific cases with very high percentages.

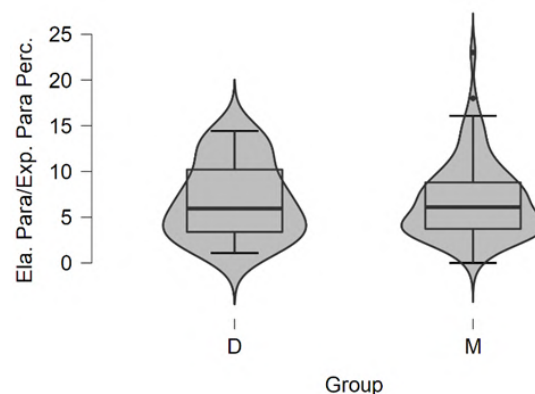


Figure 84 The boxplots comparing the proportion of paratactic elaboration in paratactic expansion in Master and PhD sub-corpora

In the Mann-Whitney U-test, the P-value is 0.913, (See Table 52) and the VS-MPR is 1 with the original hypothesis group D equals group M on the proportion of paratactic elaboration in paratactic expansion parameter. Thus, the alternative hypothesis is rejected with a large probability. The Hodges-Lehmann estimate indicates that group D is 0.079% larger than group M, with a small effect size for the test as the

rank-biserial correlation is only 0014. So, almost no difference can be found between these two groups.

Independent Samples T-Test

	W	df	p	VS-MPR*	Hodges-Lehmann Estimate	Rank-Biserial Correlation
Ela. Para/Exp. Para Perc.	1297.500	0.913	1.000	0.079	0.079	0.014

Note. For the Mann-Whitney test, effect size is given by the rank biserial correlation.

Note. Mann-Whitney U test.

* Vovk-Sellke Maximum p -Ratio: Based on a two-sided p -value, the maximum possible odds in favor of H_1 over H_0 equals $1/(-e p \log(p))$ for $p \leq .37$ (Sellke, Bayarri, & Berger, 2001).

Table 52 The Mann-Whitney U test of comparing the proportion of paratactic elaboration in paratactic expansion in Master and PhD sub-corpora

This doubtful result brings to the Bayesian Mann-Whitney U-test with diverse original hypotheses. The computation results indicate that the highest Bayesian factor supporting the original hypothesis is 5.492 (group D > group M), then 4.749 (group D < group M) and 5.093 (group D = group M) with medium effect size and the Rhat value 1.01. (See Figure 85) By comparing these probabilities, a statement can be affirmed that the proportion of paratactic elaboration in paratactic expansion in group D has almost no difference from group M, though the former is a little bit larger than the latter.

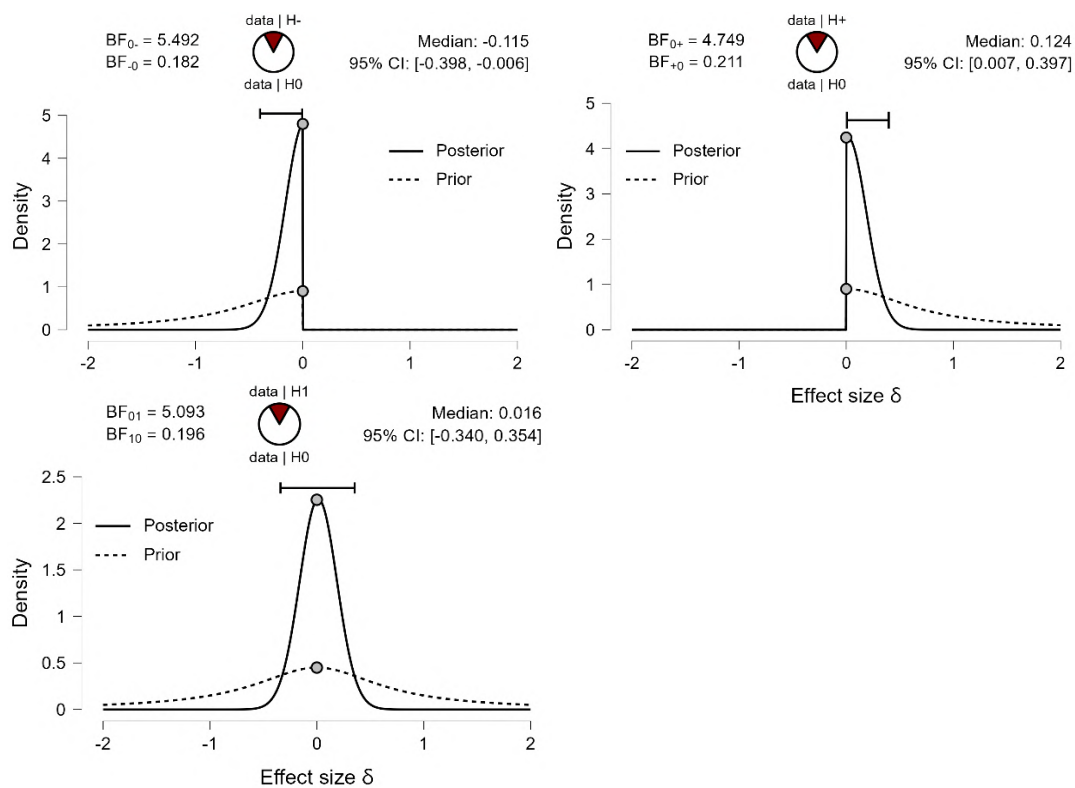


Figure 85 The inferential plots comparing the proportion of paratactic elaboration in paratactic expansion in Master and PhD sub-corpora (with diverse original hypotheses)

5.2.2.4 Elaboration Parataxis/Elaboration

Besides, the denominator of paratactic elaboration can also be the lowest superior in the whole system, i.e., elaboration, which still covers hypotactic elaboration. Hence, the proportion can be high or apparent that in some texts, it can even reach 75% in group M and 40% in group D. (See Table 5.2.2.4-1 in Appendix 5.2) The mean of paratactic elaboration in elaboration in group D is around 19.66% and 22.66% in group M. Data in group D is leptokurtic (the kurtosis value is -0.609), normally distributed (the division of skewness to the standard error of skewness is 0.229, the division of kurtosis to the standard error of kurtosis is -0.753, the Shapiro-Wilk value is 0.980, and the P-value of Shapiro-Wilk is 0.809) with almost no skewness (the skewness value is 0.095) while in group M is right-sided skew (the skewness value is 1.763), leptokurtic (the kurtosis value is 3.377), abnormally distributed as values of the two divisions are 6.554, 6.348 separately, and the P-value of Shapiro-Wilk is smaller than 0.001. The general layout can be seen from the boxplots in Figure 86 that four particular cases in group M make the data distribution abnormal.

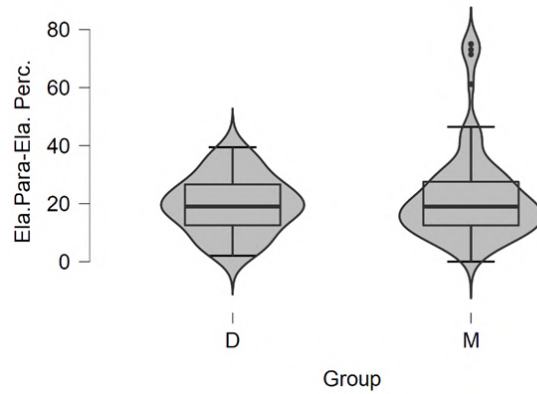


Figure 86 The boxplots comparing the proportion of paratactic elaboration in elaboration in Master and PhD sub-corpora

For the Mann-Whitney U-test (See Table 53), when the original hypothesis is that group D has the same proportion as group M on the proportion of paratactic elaboration in elaborations, the p-value is 0.837 with the VS-MPR equals to 1, this means that possibilities still exist in rejecting the original hypothesis. The difference indicated by the Hodges-Lehmann estimate shows that group D is 0.429% lower than group M on this parameter with a very small effect size (the rank-biserial correlation value is only -0.025).

Independent Samples T-Test

	W	df	p	VS-MPR*	Hodges-Lehmann Estimate	Rank-Biserial Correlation
Ela.Para-Ela. Perc.	1247.500		0.837	1.000	-0.429	-0.025

Note. For the Mann-Whitney test, effect size is given by the rank biserial correlation.

Note. Mann-Whitney U test.

* Vovk-Sellke Maximum p -Ratio: Based on a two-sided p -value, the maximum possible odds in favor of H_1 over H_0 equals $1/(-e p \log(p))$ for $p \leq .37$ (Sellke, Bayarri, & Berger, 2001).

Table 53 The Mann-Whitney U test for comparing the proportion of paratactic elaboration in elaborations in Master and PhD sub-corpora

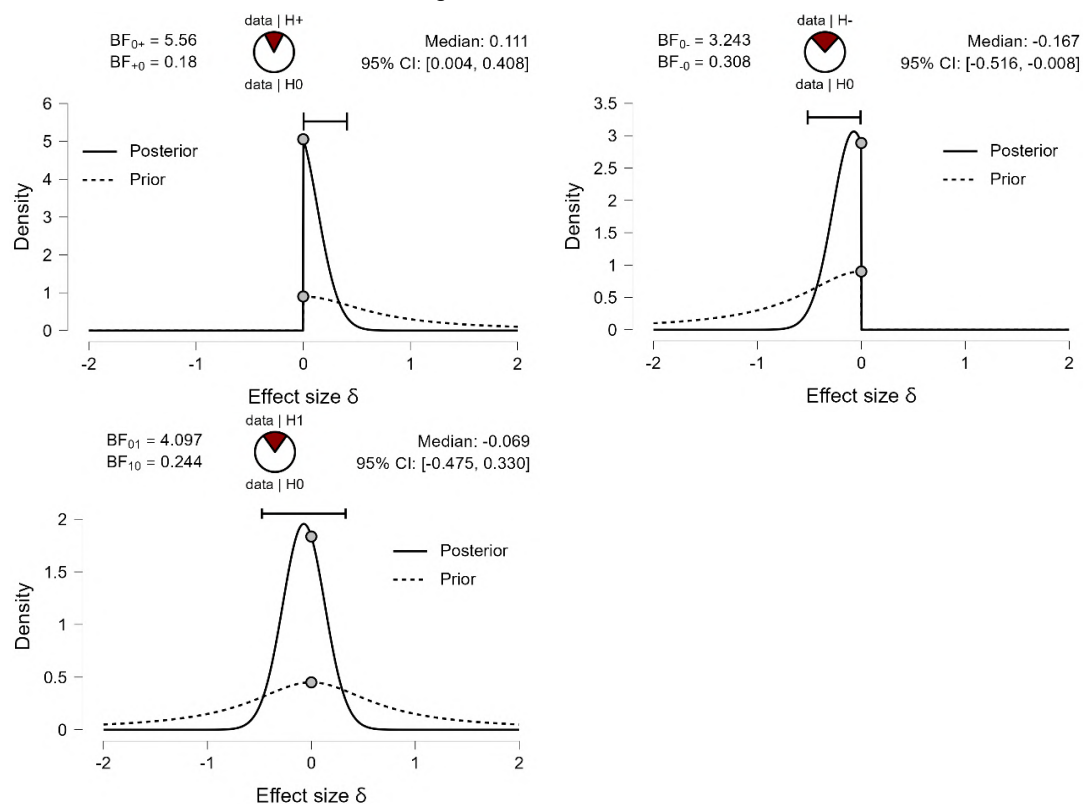


Figure 87 The inferential plots comparing the proportion of paratactic elaboration in elaborations in Master and PhD sub-corpora

To make diversities clearer, the Bayesian Mann-Whitney U-test is necessary. From Figure 87, it is interesting that the possibilities of supporting the original hypothesis are 5.56 times of rejecting it when the null is set as group D having a lower proportion than group M, with a small to medium effect size and the Rhat value 1.141; and are 3.243 times with a medium effect size when the null is group D being larger than group M; and 4.097 times with a large effect size when the original hypothesis is that two groups having an equal proportion. So, the comparison result tends to be group D's being lower than group M on the proportion of paratactic elaboration in elaborations in thesis writing, though this difference is not quite obvious because of some high percentages found in the Master's group.

5.2.3 The Distribution of Hypotactic Elaboration

In some long texts, more hypotactic elaborations would be chosen for conveying ideas compared to paratactic elaborations that the highest can be 130 in group D and 91 in group M; while in some texts, only 20 (in group D) or even 2 (group M) are chosen, the data are thus have particular cases and broad ranges. The mean number of group D is about 68, and of group M is 31. Other descriptive statistics are demonstrated in Table 5.2.3-1 (see Appendix 5.2) that data in both groups have a right-sided tail (the skewness values are 0.527 and 0.720 respective) and are leptokurtic distributed. Being different from the distribution of paratactic elaborations in the corpus, the data of hypotactic elaborations have two particular high cases in group D and one case in group M (See Figure 88) as more clause nexuses bring about a higher number of hypotactic elaborations.

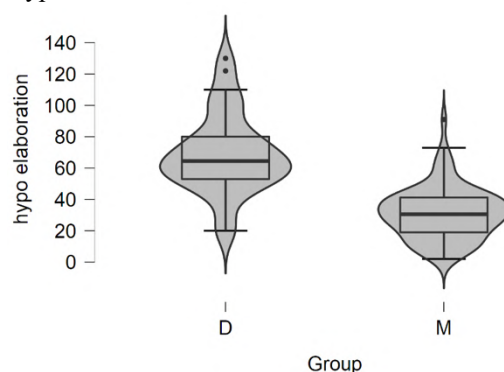


Figure 88 The boxplots comparing the number of hypotactic elaborations in Master and PhD sub-corpora

5.2.3.1 Elaboration Hypotaxis/Clause Nexuses

Likely, denominators of hypotactic elaboration can also be clause nexuses, hypotaxis, hypotactic expansion, and elaboration, in which the proportion of the first one is the smallest that the mean in both groups is no more than 10% (9.374% in group D and 9.858% in group M) and the proportion can range from 4% to 20.9% for doctoral texts and 1% to 21% for Master's texts. (See Table 5.2.3.1-1 in Appendix 5.2)

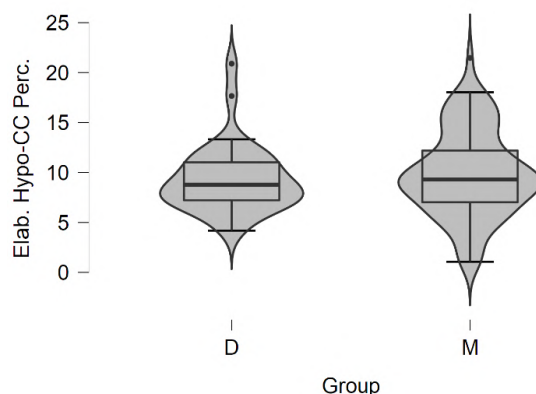


Figure 89 The boxplots comparing the proportion of hypotactic elaboration in clause nexuses in Master and PhD sub-corpora

For hypotactic elaboration in clause nexuses, data in group D is unexpectedly right-sided skew (the skewness value is 1.549), leptokurtic (the kurtosis value can be as high as 3.522!) and abnormally distributed (the value of skewness divided by the standard error of skewness is 3.742, the value of kurtosis divided by the standard error of kurtosis is 4.354, the Shapiro-Wilk is 0.884 and the P-value of

Shapiro-Wilk is only 0.002) while in group M is slightly right-sided skew, platykurtic and normally distributed seen from the statistics (the skewness value is 0.275, the kurtosis value is -0.162, the value of two divisions are 1.022, -0.305 for each parameter, the Shapiro-Wilk is 0.980 and the P-value of Shapiro-Wilk is 0.227). But boxplots in Figure 89 seem to conflict with the viewpoints above that except for two cases in group D, data are normally distributed; while are bimodal in group M.

In the Mann-Whitney U-test (Table 54), with the alternative hypothesis group D being smaller than group M, the p-value is 0.25, which suggests the possibility of accepting the original hypothesis. With the VS-MPR value being 1.061, the probability of accepting the alternative hypothesis is only 1.061 times of rejecting it. The specific difference is indicated by the Hodges-Lehmann estimate that group D is only 0.539% smaller than group M with an extremely small effect size (denoted by the rank-biserial correlation value -0.082).

Independent Samples T-Test

	W	df	p	VS-MPR*	Hodges-Lehmann Estimate	Rank-Biserial Correlation
Elab. Hypo-CC Perc.	1175.000		0.250	1.061	-0.539	-0.082

Note. For the Mann-Whitney test, effect size is given by the rank biserial correlation.

Note. For all tests, the alternative hypothesis specifies that group D is less than group M.

Note. Mann-Whitney U test.

* Vovk-Sellke Maximum p -Ratio: Based on a two-sided p -value, the maximum possible odds in favor of H_1 over H_0 equals $1/(-e p \log(p))$ for $p \leq .37$ (Sellke, Bayarri, & Berger, 2001).

Table 54 The Mann-Whitney U test for comparing the proportion of hypotactic elaboration in clause nexuses in Master and PhD sub-corpora

The Bayesian Mann-Whitney U test makes this statement clearer that the probability of accepting the hypothesis that group D is smaller than group M is 6.556 times (Figure 90, left 1) of rejecting it with a small to medium effect size and the Rhat value 1.069; the possibility of accepting the opposite hypothesis (group D > group M) is only 2.968 times (Figure 90, right 1) of rejecting it with a medium effect size; and the probability of accepting the original hypothesis (group D=group M) is 4.086 times (Figure 90, left 2) of accepting the alternative hypothesis with a large effect size. Thus, the first statement is preferred that the proportion of hypotactic elaboration in clause nexuses of group D is lower than that of group M, though this discrepancy is not quite evident.

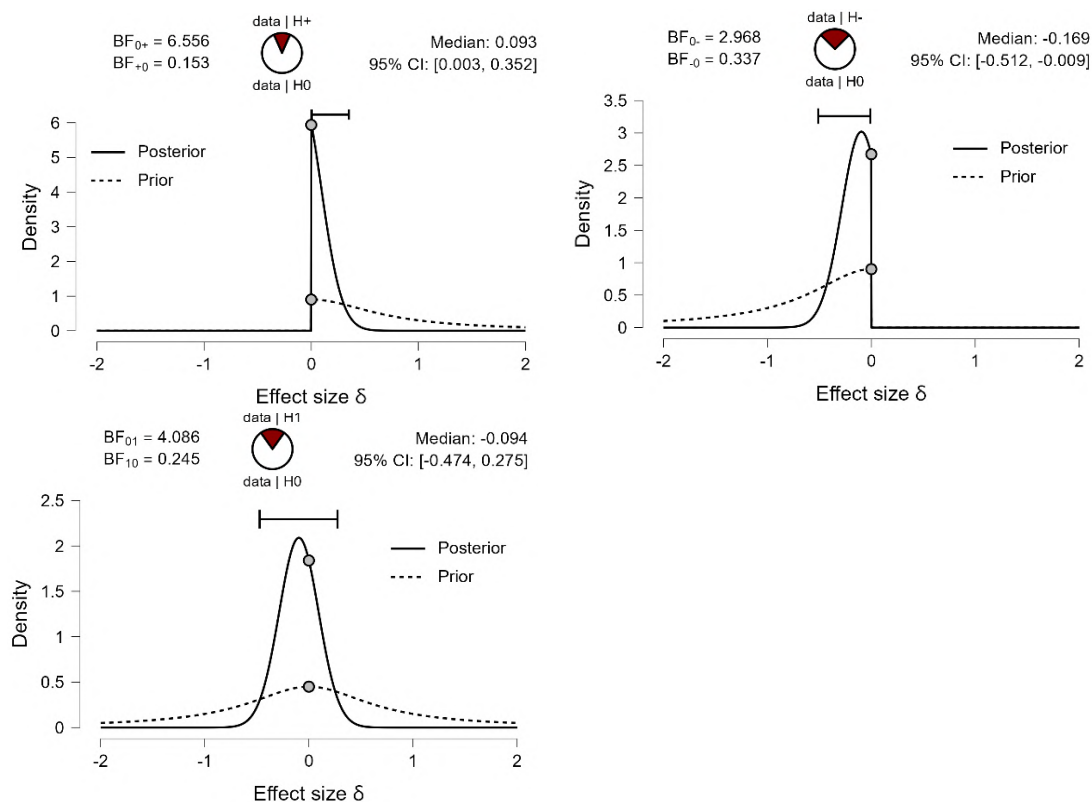


Figure 90 The inferential plots comparing the proportion of hypotactic elaboration in clause nexuses in Master and PhD sub-corpora (with diverse original hypotheses)

5.2.3.2 Elaboration Hypotaxis/Hypotaxis

For the proportion of hypotactic elaboration in hypotactic clause complexes, the doctor's corpus ranges from 6.67% to 25.52% with a mean of 14.34%, and the Master's corpus ranges from 2.12% to 35.10% with a mean of 15.91%. (See Table 5.2.3.2-1 in Appendix 5.2) Data distribution in the two groups is consistent, i.e., right-sided skew (the skewness values are 0.758 in group D and 0.168 in group M), a little bit leptokurtic (the kurtosis values are 0.673 in group D and 0.241 in group M), and normally distributed (values of skewness divided by the standard error of skewness are 1.831 in group D and 0.625 in group M, values of kurtosis divided by the standard error of kurtosis are 0.832 and 0.453 respectively, the Shapiro-Wilk are all close to 1, and the P-value of Shapiro-Wilk are 0.111 and 0.594, which are larger than 0.05). The boxplots (Figure 91) confirm the statement that except for one case in each group, other data are indeed normally distributed.

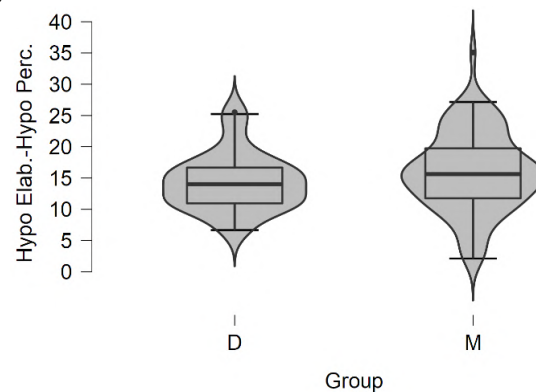


Figure 91 The boxplots comparing the proportion of hypotactic elaboration in hypotaxis in Master and PhD sub-corpora

So, Welch's t-test results can be a reference that negative values of t (-1.465) and mean difference (-1.572) show the degree of difference between the two groups, as well as the SE difference value (1.073). The p -value is 0.073, which means that the original hypothesis can be totally rejected, i.e., the alternative hypothesis (group $D <$ group M) is accepted. The probability of accepting the alternative hypothesis is 1.919 (see the VS-MPR parameter) times of accepting the alternative, with a small to medium effect size denoted by Cohen's d value -0.285. (See Table 55)

Independent Samples T-Test

	t	df	p	VS-MPR*	Mean Difference	SE Difference	Cohen's d
Hypo Elab.-Hypo Perc.	-1.465	79.329	0.073	1.919	-1.572	1.073	-0.285

Note. For all tests, the alternative hypothesis specifies that group D is less than group M .

* Vovk-Sellke Maximum p -Ratio: Based on a two-sided p -value, the maximum possible odds in favor of H_1 over H_0 equals $1/(-e p \log(p))$ for $p \leq .37$ (Sellke, Bayarri, & Berger, 2001).

Note. Welch's t-test.

Table 55 The Welch's t-test for comparing the proportion of hypotactic elaboration in hypotaxis in Master and PhD sub-corpora

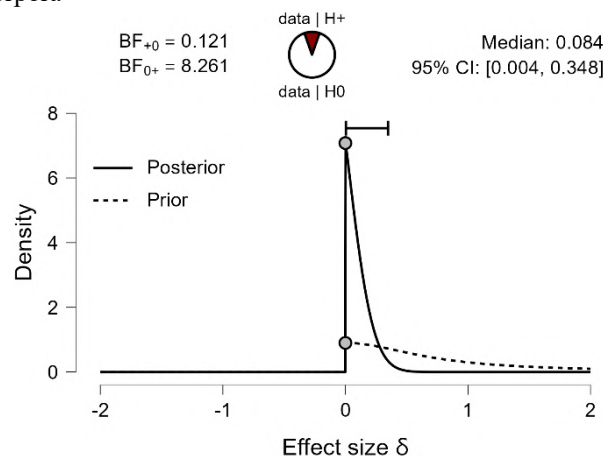


Figure 92 The inferential plots comparing the proportion of hypotactic elaboration in hypotaxis in Master and PhD sub-corpora

This conclusion can be affirmed by the Bayesian Mann-Whitney U test that the probability of accepting the original hypothesis (group D < group M) is 8.261 times of accepting the alternative with a small to medium effect size, as noted by Figure 92, and the Rhat value is 1.085, which suggests that this diversity indeed exists.

5.2.3.3 Elaboration Hypotaxis/Expansion Hypotaxis

In hypotactic expansion, hypotactic elaboration in some texts can account for 35% (in group D) and even 43.5% (in group M), while in some texts only takes up 8.9% (in group D) or even 2.6% (in group M), and the mean of two groups is 19.80% and 22.41% respectively. (See Table 5.2.3.3-1 in Appendix 5.2) Data in group D are mildly right-sided skew (the skewness value is 0.590), slightly leptokurtic (the kurtosis value is 0.587), and normally distributed (the value of skewness divided by the standard error of skewness is 1.425, the value of kurtosis divided by the standard error of kurtosis is, the Shapiro-Wilk is 0.965, and the P-value of Shapiro-Wilk is 0.38); and in group M are platykurtic (the kurtosis value is 0.174), normally distributed (values of two divisions are -0.037 and 0.327 for each, the Shapiro-Wilk is 0.99, and the P-value of Shapiro-Wilk is 0.775) with almost no skewness (the value is -0.01).

So, Welch's t-test (Table 56) can be applied in comparing the proportion of hypotactic elaboration in hypotactic expansion within two groups. The results indicated that group D has a smaller proportion than group M as the p-value is 0.034 (rejecting the original hypothesis and thus accepting the alternative one), and the probability of accepting this saying is 3.196 (the VS-MPR value) times rejecting it, the difference within two groups is around 1.849% (the t value) or 2.613% (noted by the mean difference) or 1.413% (from SE difference) under diverse calculating parameters. The test has a small to medium effect size, as suggested by Cohen's d value -0.357.

Independent Samples T-Test

	t	df	p	VS-MPR *	Mean Difference	SE Difference	Cohen's d
Ela. Hypo/ Exp. Hypo Perc.	-1.849	81.591	0.034	3.196	-2.613	1.413	-0.357

Note. For all tests, the alternative hypothesis specifies that group D is less than group M.

* Vovk-Sellke Maximum p -Ratio: Based on a two-sided p -value, the maximum possible odds in favor of H_1 over H_0 equals $1/(-e p \log(p))$ for $p \leq .37$ (Sellke, Bayarri, & Berger, 2001).

Note. Welch's t-test.

Table 56 The Welch's t-test for comparing the proportion of hypotactic elaboration in hypotactic expansion in Master and PhD sub-corpora

The conclusion that group D has a lower proportion of hypotactic elaboration in hypotactic expansion than group M can also be proved by the Bayesian Mann-Whitney U test result that the probability of accepting the original hypothesis (group D < group M) is 9.626 times (indicated by Figure 93) of rejecting it, with a small effect size and the Rhat value 1.06.

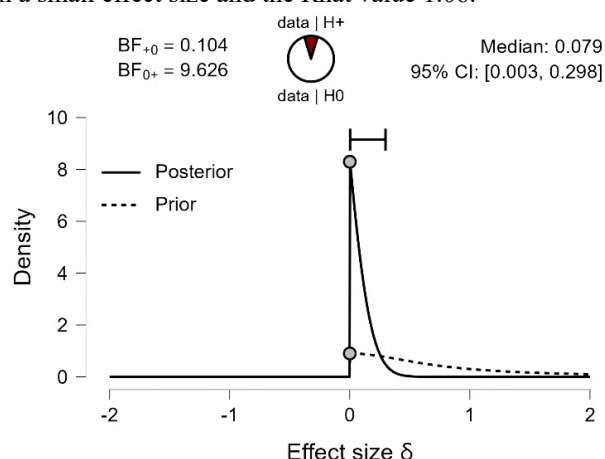


Figure 93 The inferential plots comparing the proportion of hypotactic elaboration in hypotactic expansion in Master and PhD sub-corpora (with the original hypothesis group D < group M)

This is an interesting phenomenon, as well as the proportion of hypotactic elaboration in hypotaxis that hypotactic elaboration may affected by a higher proportion of extension and enhancement, then the proportion of hypotactic elaboration in Doctors' must be lower than Masters', another effect can be the higher number of hypotaxis and hypotactic expansions in Doctors' theses compared to Masters' that less number of hypotactic elaboration divided by these structures indeed bring to lower proportion. To make

sure of this, further comparisons are still necessary.

5.2.3.4 Elaboration Hypotaxis/Elaboration

The last comparison for hypotactic elaboration is its proportion in elaboration as opposed to paratactic elaboration. So, the mean in the two groups is 80.35% (group D) and 77.34% (group M), and the distribution of data is normal in group D while abnormal in group M with a left-sided tail, which includes four particular cases with very small proportions. Being reversed from the comparison result of paratactic elaboration in elaborations, the proportion of hypotactic elaboration in elaborations in group D must be bigger than that in group M, with the same statistics in the Mann-Whitney U test. But the Bayesian factors are a little bit different: when the original hypothesis is that group D is larger than group M, the possibility of accepting this is 5.739 times (Figure 94, left 1) of rejecting it with a small effect size and the Rhat value is 1.044; when the original hypothesis is that group D is smaller than group M, the probability of accepting this is 3.824 times (Figure 94, right 1) of rejecting it with a medium effect size; and when the original hypothesis is that two groups have the equal proportion, the probability of accepting it is 4.59 times (Figure 94, left 2) of accepting the alternative one with a large effect size. Thus, how much the difference between the two groups is clearly described.

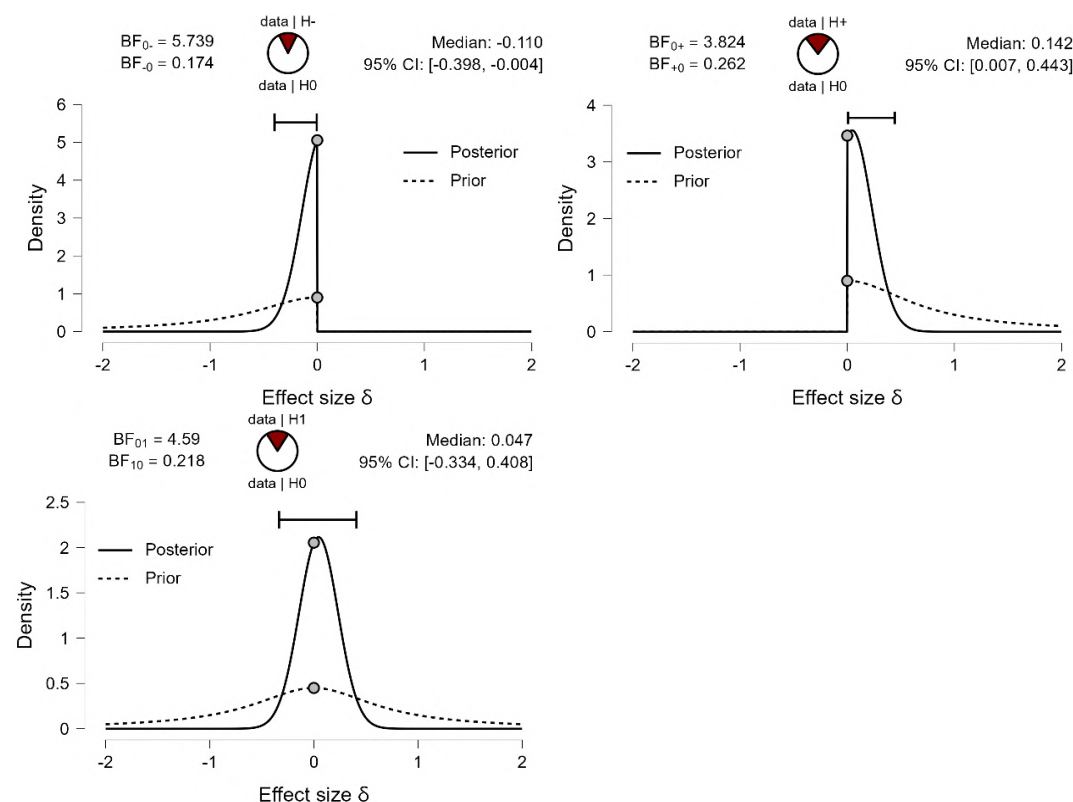


Figure 94 The inferential plots comparing the proportion of hypotactic elaboration in elaborations in Master and PhD sub-corpora (with diverse original hypotheses)

5.2.4 Summary of Elaboration Comparison

For comparing these parameters in elaboration (including paratactic and hypotactic elaboration), no matter whether the denominator is clause nexuses, parataxis, paratactic expansion, or elaboration itself, except for three parameters, doctoral students have a lower proportion than Master students in other parameters. One is the proportion of paratactic elaboration in parataxis that the difference between the two groups is too small, with group D having just a little higher proportion than group M; the second is the proportion of paratactic elaboration in paratactic expansion that group D is slightly higher than group M on this parameter; and the third is the proportion of hypotactic elaboration in elaborations that group D is higher than group M though this difference may be affected by some particular cases found in Masters' texts. One main reason may be that more extensions and enhancements used by doctoral students make the proportion of elaboration less; the other can be among the relatively few elaborations, doctoral candidates used hypotactic elaboration with a slightly higher relative frequency, while the frequency of paratactic elaboration was relatively lower, but the number of paratactic expansions was few, resulting in the situation of three almost equal parameters mentioned above.

5.3 Elaboration (Qualitatively Analysis)

Quantitative methods for elaboration discuss the relationship within clauses more from the grammatical level than the comparison between two groups only demonstrates how students have diversities in choosing grammatical structures in writing theses, leaving alone the lexical choice on the same cline of grammatical-lexical presentation that grammatical functions are realized by lexis and that is more suitable for qualitative analysis case by case.

5.3.1 General Information of Elaboration

As has been described in Section 3.1.3.1, elaboration includes three distinct kinds: exposition, exemplification, and clarification, in which exposition means that the secondary clause emphasizes the information of the first clause, restates the themes of it in another manner, or presents the first one from another aspect; exemplification indicates that the secondary clause exemplifies the information in the primary one; and clarification refers to the secondary clause clarifying or backing up the thesis in the primary clause by using an explanation or clarification comment. (Nagar & Fine, 2013) Thus, besides discussing elaboration clause complexes paratactically or hypotactically, they will also be classified as exposition, exemplification, or clarification in the following sections to probe into questions such as how many can be discovered in each corpus for each grammatical category, how many lexical categories are adopted in each corpus for realizing grammatical functions, and how do students actually choose and use in practical theses writing, etc.

5.3.2 Exposition (Paratactic Elaboration)

Most exposition relations are paratactic that altogether 339 are found in the Doctoral corpus and 306 in the Master's corpus, in which punctuation marks (colon, dash, semicolon) account for the most, phrases (that is, that is to say, in other words...), words and acronyms (i.e., namely, specifically...) take up the left. The detailed number of each lexis applied in thesis writing in two groups can be referred to Table 57:

Elaboration	M	D
Exposition (paratactic)	306	339
colon	99	144
dash	13	26
i.e.	26	33
in another saying	1	-
in other words	12	17
in particular	-	1
namely	12	15
or rather	1	-
put it another way	-	1
semicolon	5	7
simply put	1	-
that is to say	47	24
that is	88	71
to some extent	1	-

Table 57 The number of each lexis for paratactic exposition clause complexes applied in Master and PhD sub-corpora

Traditionally, punctuation marks are not subject to grammatical or lexical analysis in clause complex relations. However, they function in a manner similar to words and phrases that appear with considerable frequency in the corpus. Punctuation marks can be considered to fall between implicit and explicit clause complexes. They have clear indicators of clause relations, yet they are not straightforward words or phrases with simple lexical meanings. The relations indicated by punctuation marks require comprehension and construal through the contextual meanings of clauses.

E.g., 5.3.1 || Table 3.1 below shows the detailed information of the instructional arrangement: || study of the first two units can be seen as the Preparation Stage, || during which the Theory of Conversational Implicature is introduced || and examples are illustrated by the teacher. || ---- CMAC12LI_6_317

E.g., 5.3.2 || This example indicates one of the crucial facts ||[that syntactic approaches rarely attempt to

deal with: || that reporting is interpreted by means of structures other than quotes or reported clauses, || and that, as Volosinov (1929/1973/1986) argues, || any attempt to treat indirect speech solely in terms of mechanical transformations is bound to distort the picture]]. || ---- CPhD06_0004_359

The colon in E.g., 5.3.1 illustrates the contents of the table, while in E.g., 5.3.2, it points to important facts that are rarely dealt with in syntactic approaches. Both of these functions to explain the thesis of the first clause and to act as connecting links between two clauses. It is noteworthy that the initial colon can be replaced with *in which*, whereas the subsequent colon cannot, as the secondary clause's meaning is enclosed by the primary clause in the former instance. Conversely, in the latter instance, the secondary and third clauses possess an identical meaning and status to that of the primary clause, rendering the use of a colon obligatory. This is distinct from the function and meaning of a dash in a clause complex.

E.g., 5.3.3 || The Skopos theorie, <<which is one of the important Western translation theories,>> offers one possible solution to this problem—a solution [[that in practical terms may be seen to accommodate other theories and approaches ---- || to propose, rather than prescribe]], a mode of thought [[that will allow the translators to explore new avenues of approach to the task of producing a successful target text]]. || ---- CMAC07CU_35_559

E.g., 5.3.4 || This enables tree structures to be built ----|| it is a central claim of RST || that the structure of every coherent discourse can be described by a single rhetorical structure tree, || whose top schema application creates a span encompassing the whole discourse. || ---- CPhD05_0003_915

The dash in E.g., 5.3.3 denotes the solution that is to be proposed to accommodate other theories and approaches, by which the contents of the former clause are explained and even broadened, and this dash can be changed into *i.e.* or a colon, which would not affect the conveying and understanding of meanings. The dash in E.g., 5.3.4 fully demonstrates what enables tree structures to be built and can be replaced by a colon but not *i.e.*, as the dash as well as the colon can be followed by clauses with the scope of meanings unlimited; while the meaning of the clause after *i.e.* usually is almost the same as the former clause that it connects with.

E.g., 5.3.5 || For the words of type 1, both the Jiangjin early and the late groups were found to have the similar performance with the native speaker group ($p=.080>.05$; $p=.058>.05$), || *i.e.**, the difference between the native control group and the Jiangjin early group was not significant, || so was the native control group and the Jiangjin late group. || ---- CMAC09PH_21_402

E.g., 5.3.6 || One studies [[whether the candidate can convey language functions naturally, || *i.e.**, whether he/she can express ideas clearly.]] || ---- CPhD05_0008_626

The acronym, *i.e.*, in these two examples apparently restates the gist of the former clause paratactically with the meaning scope of the latter clause almost being the same as the former one, and the second example has the same clause structures before and after the *i.e.*, what the writer wants to convey is thus reinforced and interpreted. The same goes for the phrases *that is to say* and *that is*, which account for the highest percentage of notional words and phrases in transmitting paratactic exposition relations.

E.g., 5.3.7 || The main goal of English language teaching (ELT) is to help learners remove their language obstacles, || get to know another culture, || enhance their communicative competence, || improve their English knowledge || and draw conscious attention to the similarities with and differences from the learners' first language, || **that is***, to develop the ability to use English effectively for practical communication. || ---- CMAC08CU_33_139

E.g., 5.3.8 || Functional grammar might partly explain || why EFL texts are difficult to comprehend; || **that is***, the obligatory discourse elements in English texts are taken by the Chinese as optional, || and the implicitness [[Chinese people are inclined to use in information transmission]] leads to additional comprehension difficulties to native English speakers. || ---- CPhD05_0003_51

The phrase *that is* in E.g., 5.3.7 actually concludes the goals of English language teaching mentioned in all former clauses by concentrating on the main points of contents expressed, and in E.g., 5.3.8 explains the reasons for the question raised in the former clause: why functional grammar can make the difficulty of comprehending EFL texts clear. An alternative usage can be *i.e.* but not *that is to say* in the second example, as it is not equipped with the function of demonstrating the thesis of the former clause in detail.

E.g., 5.3.9 || The hearer cannot understand the speaker's intention from ostensive information, || **that is to say*** the implication of the speaker's utterance can only be sensed, || not explained by words. || ---- CMAC13CU_27_595

E.g., 5.3.10 || The values of affect are like other appraisal values such as the modals of probability or obligation in that || they are scalable, || **that is to say*** they operate on a sliding scale from low, through median to high intensity. || ---- CPhD07_0004_983

Typically, *that is to say* should be enclosed in parentheses with two clauses, followed by a comma. However, in these two instances, it is coincidentally followed by a clause directly, serving to emphasize

the meaning of the former clause. In addition to *that is*, the other option is *i.e.*, which can be used here to provide further elaboration on the content of the preceding clause, as in the former example, or to draw a conclusion, as in the latter example.

E.g., 5.3.11 ||| The referential meaning, linguistic meaning, and pragmatic meaning of language signs form a totality themselves, || however, in many cases of translation, it is impossible to achieve correspondence in the three aspects at the same time, || and the less important meanings in the text should give way to the most important meaning || to enable as much as possible, || **in other words***, the order of meanings to be conveyed should be in accordance with their respective importance. ||| ---- CMAC11PR_10_339

E.g., 5.3.12 ||| Textual patterns have been described with the purpose [[of accounting objectively for the way clauses or groups of clauses relate to each other in the discourse]], || **in other words***, the real nature of the pattern is the sense of order [[perceived by a reader]]. ||| ---- CPhD05_0003_1001

Another similar phrase is *in other words*, which is used to make a summary of what has been described in former clauses in E.g., 5.3.11 paratactically and to repeat the purpose of textual patterns in the latter example with the scope of meaning before the phrase being almost the same as that after the phrase.

E.g., 5.3.13 ||| Euphemism is a method to take care of people's feeling, || so usually it relates to the PP theory, || in this example, Sophie minimized the dispraise of herself, || **in another saying***, we can also consider || she maximized the praise, || which violated the modesty maxim in PP. ||| ---- CMAC14PR_4_621

Likely, the phrase *in another saying*, which is only used once in all annotated clause complexes in the corpus, just restates the meaning of the clause *Sophie minimized the dispraise of herself* before it instead of the contents of all clauses before to make sure that the writer her/himself has made euphemism clear and then convey its contents related to the PP theory to readers, and readers then indeed understand what is stated and conveyed by the writer.

E.g., 5.3.14 ||| Therefore, based on previous researches, || this thesis aims to discover a new pattern || to define the X-ABLE derivatives, || **namely*** category-based & WFR revealing approach of definition. ||| ---- CMAC14MO_11_54

E.g., 5.3.15 ||| The use of reporting verbs, <<combined with their tenses and voices in LRs,>> helps the writer enhance the probability of achieving his aims, || **namely*** the acceptance of the analysis presented in this dissertation recognizes not only the importance of the semantics of the main verb but also the properties of the situation the sentence encodes. ||| ---- CPhD06_0004_147

The word *namely* can be the same as *i.e.*, and phrases illustrated above to just give a full description of what has been mentioned in the former clause or former clauses with the scope of meaning unchanged, such as the usage in E.g., 5.3.15; but it can also refer to a more detailed subject that is contained in the things mentioned before as applied in E.g., 5.3.14.

E.g., 5.3.16 ||| They are non-English majors; || they major in different subjects: law, computer science, chemistry, biology, literature, mathematics, economics, history and so on. ||| ---- CMAC05PR_24_387

E.g., 5.3.17 ||| Taking functional grammar, contrastive rhetoric, and relation theories as a theoretical framework (which can be found in Chapter Two), || two types of analysis were mainly conducted; || one is, in Chapter Three, contrastive analysis of English, Chinese and EFL texts of the same genre (i.e., book prefaces), [[focusing on the divergent realizations of text relations at both macro-and micro-discourse levels]]. ||| ---- CPhD05_0003_17

The punctuation mark *semicolon* is seldom employed in the corpus and is challenging to identify, as it typically connects clauses for positive addition relations in extensions that provide more information and add greater substance and meanings to the preceding clause. In performing the elaboration: exposition function, such as in E.g., 5.3.16, the subsequent clause merely reiterates and elucidates the term *non-English majors* by enumerating numerous subjects. In 5.3.17, the semicolon is regarded as a colon or the acronym *i.e.*, which is used to explain the preceding clause. There are two types of this: one is used for research, and the other is more suitable here, as only one part is interpreted. This is because the meaning scope of the latter clauses is just half of the thesis indicated by the former clause.

E.g., 5.3.18 ||| As for the implementation process, much emphasis is laid upon the professional operation of its stakeholders, || **put it another way***, it is "everyone fulfils his/her duty". ||| ---- CPhD17_0004_2031

The phrase *to some extent* equals to *which means that* or, *i.e.*, in E.g., 5.3.21 connecting two clauses by presenting the first clause in another aspect in the exposition clause complex. The meaning or thesis of the secondary clause, which thinks that students are not familiar with web-based environments, is actually diverse from the first one, which emphasizes that students judge their learning effects by examinations, thus forming a kind of conclusion.

E.g., 5.3.19 ||| For instance, Xinhua native speakers will pronounce “农” (‘nong’ in Chinese, meaning

farming by English) the same as “龙” (‘lóng’ in Chinese, meaning dragon by English), || **simply put***, [n] and [l] are free allophones of the same phoneme /l/ in Xinhua dialect. || ---- CMAC15_04_379

Another phrase including the word *put* is *simply put*, which is more conclusive than *put it another way* in the exposition clause complex that the topic of former clauses is summarized by the latter with discussing subjects unchanged but information additional as demonstrated in the E.g., 5.3.19.

E.g., 5.3.20 || In addition, some students always view || that they can know about their learning effects just through the examinations, || **to some extent***, they are not adapting to the web-based environment. || ---- CMAC11ME_16_379

The phrase *to some extent* equals to *which means that* or *i.e.* in E.g., 5.3.21 connecting two clauses by presenting the first clause in another aspect in the exposition clause complex. The meaning or thesis of the secondary clause, which thinks that students are not familiar with web-based environments, is actually diverse from the first one which emphasizes that students judge their learning effects by examinations, thus forms a kind of conclusion.

E.g., 5.3.21 || According to Prince (ibid: 85), || Approximators refer to “the type of hedges that affect the Hedges propositional content but not the speakers’ commitment”, || **that is***, Approximators can change the truth condition of the proposition, || **or rather***, they may change or even cancel the original meaning of the proposition. || ---- CMAC12SP_005_92

In E.g., 5.3.21, *that is* as a phrase is used to demonstrate that the writer has already made clear what the author Prince discussed about approximators and then explained his understandings to readers, *or rather* shows that the writer actually adds more information to his/her own thinking with to some degree uncertainty.

5.3.3 Exemplification (Paratactic Elaboration)

Exemplification is easy to understand and find in the corpus as the secondary clause talks about more specific things that are described in the former clause by explicit conjunctions *for example*, *such as*, *in particular*, the acronym, *e.g.*, phrase *take ... for example*, words *like*, *particularly*, etc. In the corpus, altogether 81 annotated exemplifications can be discovered in Master’s theses and 56 in doctoral dissertations, in which phrases *for example* and *such as* take up the highest percentage. (See Table 58)

Elaboration	M	D
Exemplification	81	56
e.g.,	5	4
examples/instance	16	2
for example,	20	14
for instance,	5	6
in particular	-	2
like	4	2
particularly	-	3
such as	26	22
take ... as examples	5	1

Table 58 The number of each lexis for paratactic exemplification clause complexes applied in Master and PhD sub-corpora

E.g., 5.3.22 || For example, we can’t know || if some sub-meanings of the word “eat” includes the meaning of “drink”, || because in some ancient Chinese novels, we often use “吃” (eat) to refer to “喝” (drink), || **for example***, we say “吃酒” (drink wine) instead of “喝酒” (drink wine). || ---- CMAC11SE_19_173

By using the phrase *for example* for connection, how *eat* can be referred to as *drink* in some ancient Chinese novels is clearly stated in the latter clause for readers to comprehend, as shown in E.g., 5.3.22. The meaning scope and discussion subjects of the former clause are not narrowed down but concentrated and proved.

E.g., 5.3.23 || Attempting to explain the results she considers unexpected, || Wells makes the interesting point that ‘developing systems may have characteristics that are simply not evident in mature languages’ (p. 250); || **for instance***, child L1 learners demonstrate incipient knowledge of inflectional morphology || before reaching what researchers might consider to be a criterion of ‘acquisition’. || ---- CPhD15_0001_606

The same function can be performed by the phrase *for instance* that the meaning of the former clause is what the writer wants to reinforce and apply in the research and the example described in the

latter clause is usually what the writer is familiar with in his/her knowledge category cognition, as indicated by E.g., 5.3.23. The chosen example normally is also what readers are acquainted with or what the writer thinks that readers can accept and easily understand, though they may not know or have not heard of.

E.g., 5.3.24 ||| These are taken from different genres, || **e.g.***, the instructional text patterns are taken from instruction booklets and from recipes, the expository text patterns from textbooks and encyclopedias, and so on. ||| ---- CPhD09_0002_291

The acronym *e.g.* can be followed by many examples as it reduces the space of written materials and time and energy of the writer, such as that in E.g., 5.3.24.

E.g., 5.3.25 ||| Let's study some **examples***: || as we know, || there is one thing in the kitchen [[which is hard to define in western and Chinese families]]. ||| ---- CMAC11SE_19_653

The function of exemplification seems to be fulfilled by the word *examples*, but actually by the colon that the latter clause is apparently indicated by the punctuation mark in E.g., 5.3.24.

E.g., 5.3.26 ||| Platform was used in the experiment in the following way, << **take Sina Weibo as an example***: >> || the teacher created an account at this microblogging service. ||| ---- CMAC14RE_15_363

It seems that it is the phrase *take...as an example* in the clause *take Sina Weibo as an example* (it is a parenthesis and thus is enclosed in the double angle brackets) that leads to an exemplification in E.g., 5.3.26, however, it should be the punctuation mark that performs the signal within clauses.

E.g., 5.3.27 ||| Some explain from a cognitive perspective, || **such as*** Furukawa Yutaka (2001) who in one of his articles points out || that some verbs in presentational constructions must take a perfective marker *le* plus quantified nouns. ||| ---- CPhD07_0008_383

The conjunction *such as* seems to have no difference from *e.g.*, or *for example* in functioning as a connector of exemplification in a clause complex but outnumbers the latter two. It is strange that *such as* can only be followed by a noun or a noun phrase while *for example* does not have such limitations, the choice of the former is still more than the latter in the corpus. One explanation for this can be the high probability of nominalization of themes in a clause in thesis writing. As *such as* is followed by a noun or noun phrases, it emphasizes the noun or noun phrases that come up in the former clause and then gives examples; while the phrase *for example* or *for instance* can also refer to the whole matter described in the former clause as well as the subject only, by emphasizing from diverse angles.

E.g., 5.3.28 ||| The logical relations between words, sentences and paragraphs are usually indicated by the conjunctions, || **like*** “however” means the turning; || “all in all, in summary, in brief,” mean the summary; || “for, because, since” mean the reason, etc. ||| ---- CMAC13RE_10_495

The word *like* has a more general and broader scope of exemplification without emphasizing any noun (noun phrase) or any matter, as *etc.* can be used at the end of the related clause, which is diverse from the phrase *for example* or *for instance*. Thus, in E.g., 5.3.27, *like* is followed by many examples listed for only proving the thesis of the former clause: the logical relations between words, sentences, and paragraphs are indicated by conjunctions.

E.g., 5.3.29 ||| Here, although Susan is higher in the tree than herself, || and does not dominate herself, || not every category [[which dominates Susan]] dominates herself; || **in particular***, the DP containing friend dominates Susan, || but does not dominate herself. ||| ---- CPhD15_0001_705

The meaning scope of the latter clause connected by the phrase *in particular* in E.g., 5.3.28 is narrowed down as compared to that of the former with a specific example chosen, and the matter *not every category which dominates Susan dominates herself* is emphasized.

E.g., 5.3.30 ||| Chinese language learners desire for an affirmative evaluation when they are placed in a communication setting, **particularly*** when they get involved into conversation with the significant others using a new language(L2). ||| ---- CPhD13_0003_466

It is the same condition when the word *particularly* is chosen in E.g., 5.3.28 that the communication setting in which Chinese language learners need affirmative evaluation mentioned in the former clause is strengthened by the example situation described in the latter clause with the thesis unchanged but discussion subjects transformed into learners' desire for affirmative evaluations in communicating with important foreigners, which would bring to further investigations in the writer's research.

5.3.4 Clarification (Paratactic Elaboration)

The number of paratactic clarification clause complexes found in annotated corpus is not big enough and the distinction of paratactic clarification from paratactic exposition is not easy as both of them are used for explaining thesis of former clauses, though clarification can involve the shift of polarity of clauses marked by some words and phrases like *in fact*, *actually*, *indeed*, *at least*, *i.e.*, *viz.*, etc. (Halliday, 2014, p. 464)

E.g., 5.3.31 ||| As a result, conclusion can be drawn || that past tense marking in the 60 participants' IL has

stopped developing, || **at least*** during the period of five months between the two pretests. ||| ---- CPhD08_0012_1519

The clause *at least during the period of...* in E.g., 5.3.29 actually agnates with the clause *the past tense marking in ... has stopped developing at least during the period of...*, in which the Subject is left out to avoid redundancy of grammatical structures in a clause complex. The phrase *at least* provides more information for the condition under which the *past tense marking stops developing*, as mentioned in the second clause, and comments it with support instead of objection that the *past tense marking surely stops developing during this period of time*. Actually, the rare use of these phrases or words for marking clarification clause nexuses represents students' inability, to some degree, to criticize or comment on what they have experienced or known about the world or read from articles and books. Words like *actually*, *indeed*, etc. are often intersected in a clause instead of being treated as a conjunction which connects two clauses in fact.

5.3.5 Hypotactic Elaboration

For hypotactic elaboration, a large quantity of clause complexes is performed by finite non-defining relative clauses (in traditional grammar) marked by words such as *which*, *what*, *who*, etc., to introduce background information or characterization in discourse for the first clause or interpret some aspects of it and even make evaluations or comments on the thesis of former clauses, with the rest are unmarked (implicit) non-finite clauses, which should be judged from their agnates in annotation. (See Table 59)

	M	D
Hypotactic Elaboration	1641	1343
Non-finite (implicit)	276	328
Non-defining relative: WH-	1365	1015

Table 59 The number of hypotactic elaborations with markers annotated in Master and PhD sub-corpora E.g., 5.3.32 ||| This result was similar to the research of Maasum & Maarof (2012: 1255), || **who*** asserted || that EFL learners preferred to use "dictionary" strategies to enhance understanding and memory. ||| ---- CMAC14RE_15_591

E.g., 5.3.33 ||| A second possibility is offered by Jakubowicz (1984) and Solan (1987), || **who*** argue || that children initially treat pronouns like anaphors. ||| ---- CPhD15_0001_358

The domain of a non-defining relative clause can be the whole former clause or any of its constituents that the word *who* in these two examples both points to the name of the person in front of it (Maasum & Maarof, or Jakubowicz & Solan), i.e., just part of the former clause by emphasizing the importance of these scholars' viewpoints. This kind of clarification is normally found in the Literature review part of thesis writing, as students always refer to other researchers' academic points to support their own ideas or to provide a foundation for building up new theories by precisely listing researchers' names. Bigger names (scholars who have a great reputation in the academic circle with a long list of publications) appearing here usually indicate higher reliability of the thesis with students' diffidence of their own points of view, while non-significant names coming up sometimes show students' broad range of reading with losing the reliability of their theses.

E.g., 5.3.34 ||| To conclude, in spite of its short history of development, the pragmatic study of literature has made great progress abroad, || **which***, to a large degree, has contributed to the development in China. ||| ---- CMAC14LIT_25_127

E.g., 5.3.35 ||| In explaining these results || Vermetten et al. (2001) speculated || that Agreeableness involves compliance and cooperativeness, || **which*** makes the agreeable individuals more likely to consolidate their learning || and regulate their study habits in response to external demands. ||| ---- CPhD10_0001_456

In E.g., 5.3.32, the domain of the *which* clause is only part of the former clause, i.e., *the pragmatic study of literature*, while in E.g., 5.3.33, the domain is the whole clause in front of it, that is, *Agreeableness involves compliance and cooperativeness*. Both of the domains are reinforced and regarded as the Subjects of the latter clauses, their influences are then described, and the meaning scope of the former clauses is broadened.

E.g., 5.3.36 ||| Since in EEBO, the subcorpus size is quite unevenly distributed, << **ranging*** from 712,130 to 99,452,875 words,>> || the author used the normalized frequency per million words instead || to make the results comparable in each decade. ||| ---- CPhD19_0002_1509

The non-finite clause *ranging from ... to ... words* in E.g., 5.3.34, though embedded, agnates with *the subcorpus size is quite unevenly distributed that it ranges from ...to ... words* in which *that it* is left out with the subject still being *the subcorpus size*, the emphasize of *unevenly distributed* subcorpus is thus lightened while how large the size is of each text in the subcorpus is clarified, and then further

information or management of the corpus size provided by following clauses is enabled.

E.g., 5.3.37 ||| Two different evaluation methods are applied and conducted in the two classes during the whole process of the experiment, || teaching skills portfolio goes to EC and traditional Micro-teaching goes to CC. ||| ---- CMAC12ME_5_363

The secondary clause in E.g., 5.3.35 is actually connected with *that is, i.e.*, or a colon with the former one, by indicating a paratactic exposition of the domain of the whole former clause, *how two different evaluation methods are applied and conducted in the two classes* for the research is thus clearly described.

E.g., 5.3.38 ||| Reading for Understanding develops important reading comprehension skills, || **including*** critical-thinking skills, predicting outcomes, making inferences, drawing conclusions, comparing and contrasting, using cause and effect, and distinguishing specific from general. ---- CPhD14_0002_695

The second clause in E.g., 5.3.36 is an implicit exemplification that agnates with the clause *Reading for Understanding develops ... skills, which includes...* by omitting the word *which* and turns the finite non-defining relative clause into a non-finite clause initiating by the word *including*. By listing the kinds of skills in the secondary clause, the function of exemplification is fulfilled, and how important *Reading for Understanding can develop reading comprehension skills* is emphasized.

E.g., 5.3.39 ||| For example, everyone [[who knows || that Tom is an idiot]], says “Tom is a genius” to express the opposite meaning, || **implying*** || that “Tom is extremely silly.” ---- CMAC14LIT_25_182

In agnation, the last clause in E.g., 5.3.37 can be *which actually implies that ...*, in which the word *which* refers to the part *Tom is a genius*, the word *implying* thus performs the implicit hypotactic clarification function that the polarity of the former clause is shifted into a negative one with demonstrating *Tom is a genius* in fact means *Tom is extremely silly* to emphasize the thesis of former clauses, i.e., *to express the opposite meaning*.

E.g., 5.3.40 ||| On the other hand, the standard deviations of scores on traits 1C,2A,2B,3,4A, and 4B are quite large, || **showing*** heterogeneity of their ability on these traits. ---- CPhD05_0008_1128

Again, in E.g., 5.3.38, the secondary clause omits *which*, which originally can be *which shows ...* with *which* referring to *the standard deviations of scores on traits... are quite large*. What the writer wants to reinforce and explain is that *large standard deviations of scores* indicate *heterogeneity of students' ability*. The implicit hypotactic exposition function is thus realized.

E.g., 5.3.41 ||| Currently, it has been generally assumed || that context should be treated as a dynamic factor in construal (Cao Jing-yuan 2008, p. 33), || profoundly different from the static one in the previous studies. ---- CMAC10SE_18_443

In E.g., 5.3.39, the words *which is* in the secondary clause have not been used, but the meaning remains unchanged and has not been affected by this deletion. It is the dynamic factor in construal that the context should be regarded as instead of a static one, an implicit hypotactic clarification with a shift of polarity is clearly indicated.

5.3.6 Summary of Elaboration Marker Choices

In summary, the semantic domain of the first clause can be broadened or narrowed down in the secondary one with the discussion subjects unchanged for exposition and clarification clause complexes marked by some words or phrases (that is, i.e., punctuation marks), while remains almost the same for some markers such as *in other words*, *that is to say*, etc. Phrases in marking exemplification relations such as *for example*, *such as*, and *and like* seem to be the same but actually have a specific scope of application, which is not because of the academic writing genre but more attributed to grammatical logic. The division between clarification and exposition is easily mixed, though the former involves shifts of polarity within two clauses. For finite non-defining relative clauses beginning with “WH-” structures, the domain of the secondary clause can be the whole former clause or just part of it, more investigation can thus be conducted in digging out the reason for this differentiation. The hardest part can be the identification of implicit elaborations, especially denoted by non-finite clauses written by students, some of which can even be wrong. Though methods of clause agnation can be applied in clause complex relations' discovery, the differentiation of categories, whether a clause nexus is an exposition or a clarification, can be troublesome and, in most conditions, depends on the semantic and contextual environment in a text.

5.4 Extension (Quantitatively Comparison)

Extension, as described in Section 3.1.3.2, is the relationship of one clause that develops the former one by adding more information to it or extending its meaning in a clause complex. Usually, an extension can

be paratactic or hypotactic from the status of two clauses and can also be an addition, variation, or alternation when semantically or logically distinguished. Again, the number of the latter organization is not paratactically or hypotactically compared as they can be laborious and too delicate for research, only extension is thus quantitatively distinguished from Master students' texts and doctoral students' dissertations on parataxis and hypotaxis.

5.4.1 General Distribution of Extension (Extension/Clause Nexuses)

The comparison of general information (descriptive statistics) of extension annotated in the corpus and the proportion of extension in expansion structures in two groups can be found in Section 5.1.4.2. Here, the condition of extensions in clause nexuses, i.e., the proportion of extension in the whole clause complex system, will be discussed. As shown in Table 5.4.1-1 (see Appendix 5.4), the proportion of group D ranges from 15% to 45% with a mean of around 32%, and group M from 22% to 49% with a mean of 34%. An initiative conclusion is supposed that group D has a lower proportion than group M. Data of the former group has almost no skewness (the value is only 0.023) and is platykurtic (the value of the kurtosis is 0.172), normally distributed as judged from statistics (the division of the skewness by the standard error of skewness is 0.056, the value of the kurtosis divided by the standard error of kurtosis is 0.213, the Shapiro-Wilk value is 0.974 and the P-value of Shapiro-Wilk is 0.62) while of the latter is a little bit right-sided skew (the value of the skewness is 0.378), leptokurtic (the value of the kurtosis is -0.372), and normally distributed as values of two divisions are 1.405, and -0.699 respectively, the Shapiro-Wilk value is 0.973 and the P-value of Shapiro-Wilk is 0.093 (not far away from 0.05). Actually, data in group D has a short, narrow left-sided tail (See Figure 95), which indicates some small percentages of extension per clause nexuses.

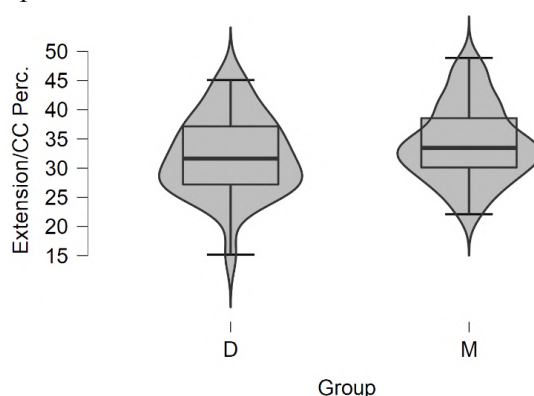


Figure 95 The boxplots comparing the proportion of extension in clause nexuses in Master and PhD sub-corpora

The normally distributed data with diverse variances of two groups can thus be compared by Welch's t-test results (See Table 60). Negative values of t (-1.609) and mean difference (-2.217) indicate that the proportion of extension in clause nexuses of group D is 1.6% or 2.2% smaller than that of group M in diverse calculating parameters, the same goes for SE difference that the value is 1.378. As the p -value is only 0.057 and the VS-MPR value is 2.261, the probability of accepting the alternative hypothesis (group D < group M) is 2.261 times rejecting it, and the original hypothesis can highly refuse possibility. The effect size of this test is small to medium, as noted by the value of Cohen's d (-0.339).

Independent Samples T-Test

	t	df	p	VS-MPR *	Mean Difference	SE Difference	Cohen's d
Extension/CC Perc.	-1.609	55.257	0.057	2.261	-2.217	1.378	-0.339

Note. For all tests, the alternative hypothesis specifies that group D is less than group M.

* Vovk-Sellke Maximum p -Ratio: Based on a two-sided p -value, the maximum possible odds in favor of H_1 over H_0 equals $1/(-e p \log(p))$ for $p \leq .37$ (Sellke, Bayarri, & Berger, 2001).

Note. Welch's t-test.

Table 60 The Welch's t-test for comparing the proportion of extension in clause nexuses in Master and PhD sub-corpora

The same conclusion can be reached from the Bayesian Mann-Whitney U test that when the alternative hypothesis specifies that the location of group D is larger than the location of group M, the possibility of accepting the original hypothesis is 8.89 times accepting the alternative one as indicated by the Bayesian factor value (See Figure 96) with the Rhat value 1.081 and the effect size small.

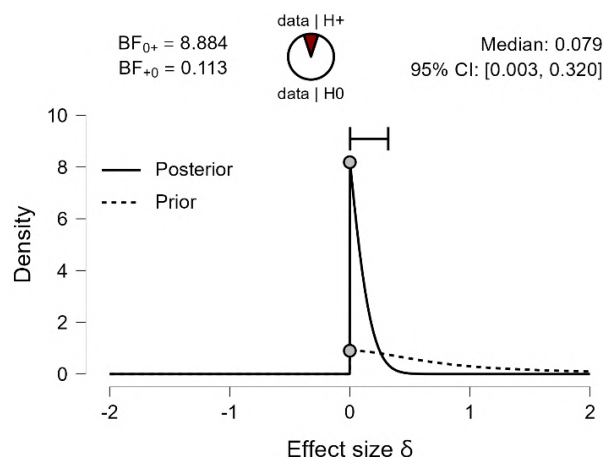


Figure 96 The inferential plots comparing the proportion of extension in clause nexuses in Master and PhD sub-corpora (with the original hypothesis group $D < \text{group M}$)

So, the proportion of extension in clause nexuses written by doctoral students is smaller than that by Master students. This makes the proportion of enhancement in clause nexuses in two groups deductive, that group D must have a higher one than group M as the proportion of elaboration of the former is also less than the latter group. It is the higher proportion of enhancement that results in a lower proportion of extension and elaboration in the doctoral group.

5.4.2 The Distribution of Paratactic Extension

In all extensions, in total, 6829 paratactic extensions are found in doctoral candidates' texts, with a range from 61 in some texts to 351, and 7865 in Master students', which ranges from 34 in some texts to 257. (See Table 5.4.2-1 in Appendix 5.4) The mean number of the former group is around 213, and of the latter is 98, and their magnification is smaller than 2.5. The data distribution in group D seems to be leptokurtic and normal with almost no skewness as the skewness value is 0.004, the kurtosis value is -1.195, the division of the skewness to the standard error of skewness is 0.01 (within -2/+2), the division of the kurtosis to the standard error of kurtosis is 1.477 (also within -2/+2), the Shapiro-Wilk value is 0.948 (closes to 1), and the P-value of Shapiro-Wilk is 0.126 (larger than 0.05); while in group M is right-sided skew, leptokurtic, abnormal as indicated by the skewness value 1.422, the kurtosis value 2.907, the divisions of two parameters 5.286 and 5.464 separately, the Shapiro-Wilk value 0.900 (not very close to 1), and the P-value of Shapiro-Wilk being smaller than 0.05 (<0.001 indeed). However, the data distribution of group D is bimodal, as indicated by Figure 97, and of group M is normal when two cases with a high number of paratactic extensions are excluded for calculation.

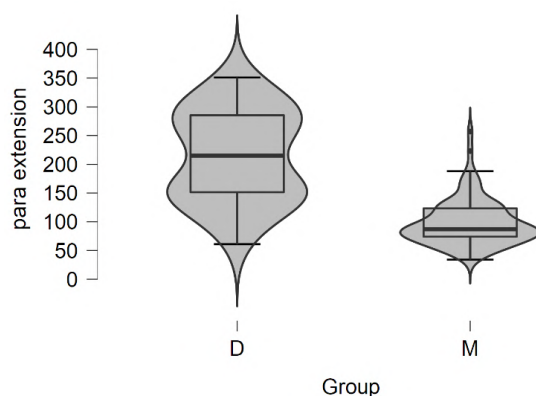


Figure 97 The boxplots of the number of paratactic extensions in Master and PhD sub-corpora

5.4.2.1 Extension Parataxis/Clause Nexuses

The proportion of paratactic extension in clause nexuses inspects the condition of paratactic extension in the whole clause complex system that it is lower in group D as compared to group M on only considering the range and the mean of two groups (13% to around 41% with a mean 28% in the former group and 16.6% to about 46% with a mean 31% in the latter). (See Table 5.4.2.1-1 in Appendix 5.4) Data in the two groups are concurrently a little bit right-sided skew, to some degree platykurtic, and normally distributed as values of skewness are 0.127 in group D and 0.368 in group M, values of kurtosis are

-0.260 and -0.379 respectively, values of the skewness divided by the standard error of skewness are 0.307 and 1.368 (within $-2/+2$), values of kurtosis divided by the standard error of kurtosis are -0.321 and -0.712 (also within $-2/+2$), the Shapiro-Wilk values are 0.975 and 0.976 (close to 1) for each group, the P-value of Shapiro-Wilk for two groups are 0.644 and 0.142 (larger than 0.05). The same conclusion can be obviously indicated by the boxplots in Figure 98 below:

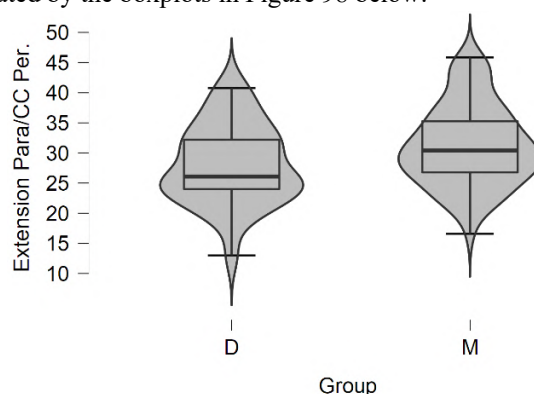


Figure 98 The boxplots comparing the proportion of paratactic extension in clause nexuses in Master and PhD sub-corpora

Thus, Welch's t-test is suitable for comparing this proportion in two groups as data are normally distributed with different variance values. Results in Table 61 show that group D is 2.42% or 3.1% lower than group M, as indicated by the t-value or the mean difference value. The SE difference value (1.381) also proves a kind of diversity. The p-value is 0.014, demonstrating high disapproval of the original hypothesis (group D is larger than group M). With the VS-MPR value of 6.026, the possibility of accepting the alternative hypothesis is 6.026 times of rejecting it, i.e., accepting the original one. This test has a medium effect size as Cohen's d value is -0.467, in which the negative sign only means group D is less than group M on the proportion of paratactic extension in clause nexuses.

Independent Samples T-Test

		t	df	p	VS-MPR*	Mean Difference	SE Difference	Cohen's d
Extension Per.	Para/CC	-2.242	58.143	0.014	6.026	-3.097	1.381	-0.467

Note. For all tests, the alternative hypothesis specifies that group D is less than group M.

* Vovk-Sellke Maximum p -Ratio: Based on a two-sided p -value, the maximum possible odds in favor of H_1 over H_0 equals $1/(-e p \log(p))$ for $p \leq .37$ (Sellke, Bayarri, & Berger, 2001).

Note. Welch's t-test.

Table 61 The Welch's t-test for comparing the proportion of paratactic extension in clause nexuses in Master and PhD sub-corpora

The same comparison result can be proved by the Bayesian Mann-Whitney U test that the probability of accepting the original hypothesis (group D < group M) is 14.531 times accepting the alternative one with a Rhat value of 1.034 and small effect size as indicated by Figure 99.

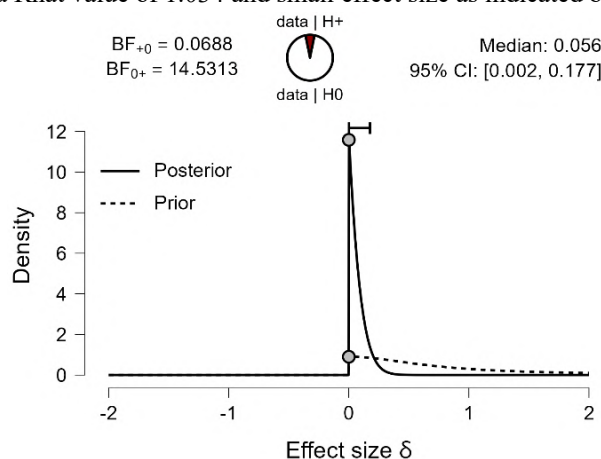


Figure 99 The inferential plots comparing the proportion of paratactic extension in clause nexuses in Master and PhD sub-corpora (with the original hypothesis group D < group M)

If taking paratactic extension as an easier structure than the hypotactic extension in academic writing for acquiring and applying, then doctoral students' preference for choosing other more complex structures instead of the least paratactic extension in conveying ideas needs further exploration.

5.4.2.2 Extension Parataxis/Parataxis

In parataxis, besides paratactic elaboration (in Section 5.2.2.2), paratactic extension, and paratactic enhancement, still paratactic locution and idea (in projection part) clause complexes can be found that the comparison of the proportion of paratactic extension in parataxis investigates it with other paratactic sub-categories. From Table 5.4.2.2-1 (see Appendix 5.4), it is interesting that the mean of paratactic extension in parataxis in two groups is almost the same with the value around 80%, the range of which is 57% to 91% in group D and 57% to more than 95% in group M. Data in the former group is left-sided skew (the skewness value is -0.614), slightly leptokurtic (the kurtosis value is 0.917), and seems to be normally distributed as the value of skewness divided by the standard error of skewness is -1.483 (within -2/+2), the value of kurtosis divided by the standard error of kurtosis is 1.133 (also within -2/+2), the Shapiro-Wilk value is 0.945 (closes to 1), and the P-value of Shapiro-Wilk is 0.106 (just a little bit bigger than 0.05); and in group M is also left-sided skew, slightly leptokurtic, and abnormally distributed as seen from the statistics (the skewness value is -0.613, the kurtosis value is 0.525, values of two divisions are -2.279 and 0.987 respectively, the P-value of Shapiro-Wilk is 0.974 and the P-value of Shapiro-Wilk is 0.107, which is not very large than 0.05). But boxplots in Figure 100 indicate that except for one case in each group, data can be normally distributed.

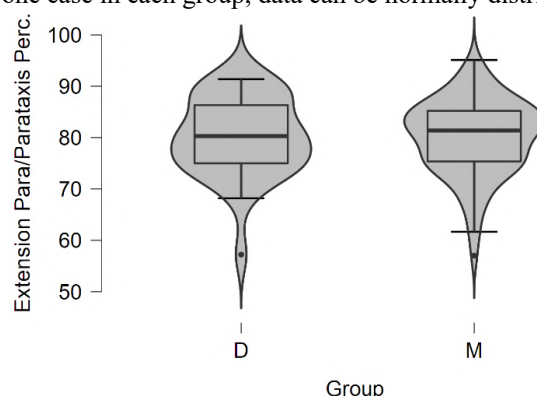


Figure 100 The boxplots comparing the paratactic extension in parataxis in Master and PhD sub-corpora

For the Mann-Whitney U test (Table 62), when the original hypothesis is that group D has the same proportion as group M, the P-value is 0.920, with the VS-MPR value equals 1, which means that the alternative hypothesis should be totally rejected and the original one is accepted. As the Hodges-Lehmann Estimate is -0.145, group D is just 0.145% bigger than group M on the proportion of paratactic extension in parataxis; the difference can actually be quite small. The effect size of this test, indicated by the rank-biserial correlation value -0.012, is very small.

Independent Samples T-Test

	W	df	p	VS-MPR*	Hodges-Lehmann Estimate	Rank-Biserial Correlation
Extension Para/Parataxis Perc.	1264.000		0.920	1.000	-0.145	-0.012

Note. For the Mann-Whitney test, effect size is given by the rank biserial correlation.

Note. Mann-Whitney U test.

* Vovk-Sellke Maximum p -Ratio: Based on a two-sided p -value, the maximum possible odds in favor of H_1 over H_0 equals $1/(-e p \log(p))$ for $p \leq .37$ (Sellke, Bayarri, & Berger, 2001).

Table 62 The Mann-Whitney U test for comparing the paratactic extension in parataxis in Master and PhD sub-corpora

The Bayesian Mann-Whitney U test is conducted then as the independent samples t-test above cannot exactly reveal how big the difference is between the two groups. Results in Figure 101 show that the possibility of accepting the original hypothesis is 4.595 times (left 1) of rejecting it when the alternative hypothesis is that group D has a smaller proportion than group M with the Rhat value 1.1 and a small to medium effect size; the probability of accepting the original one is 4.43 times (right 1) of accepting the alternative hypothesis when the latter is group D being bigger than group M, with a small to medium effect size; and the Bayesian factor value is 4.511 (left 2) with a large effect size when the original hypothesis means that no diversity exists in two groups. By comparing the Bayesian values,

the conclusion tends to be that group D has a slightly larger proportion of paratactic extension in parataxis than group M, though this difference is quite small.

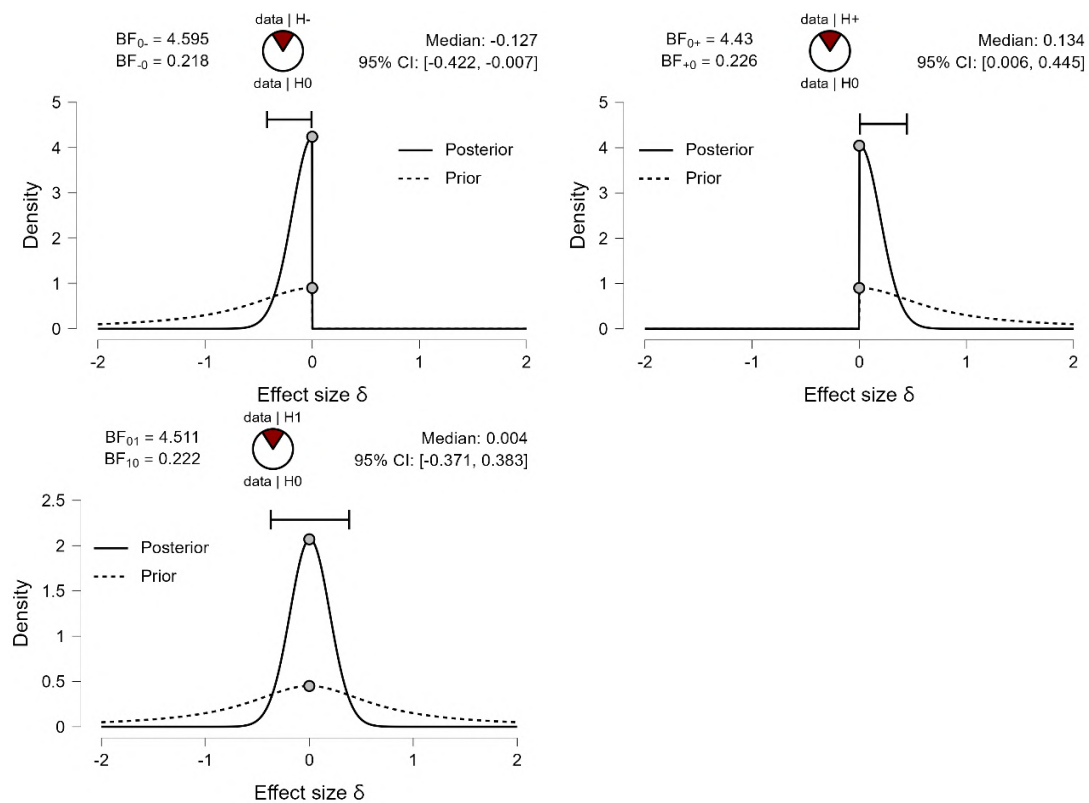


Figure 101 The inferential plots comparing the paratactic extension in parataxis in Master and PhD sub-corpora (with diverse original hypotheses)

5.4.2.3 Extension Parataxis/Expansion Parataxis

The parameter paratactic extension in paratactic expansion, in fact, contrasts paratactic extension with paratactic elaboration and paratactic enhancement in two groups that the second has been proved to be almost the same, as indicated by Section 5.2.2.3. The first of which has a large proportion in two groups ranging from 71.77% to as high as 93.5% with a mean of 82.25% in group D and from 59.29% to 95.12% with a mean of 82.09% in group M. (See Table 5.4.2.3-1 in Appendix 5.4) Data in the former group seems to be right-sided skew (the skewness value is 0.126), leptokurtic (the value of kurtosis is -1.331), normally distributed as the division of the skewness by the standard error of skewness is 0.304 (within -2/+2), the value of kurtosis divided by the standard error of kurtosis is 1.645(also within -2/+2), the Shapiro-Wilk value is 0.934, and the P-value of Shapiro-Wilk is 0.052 (nearly to 0.05); and in group M is left-sided skew (the skewness value is -0.845), leptokurtic (the kurtosis value is 0.917) and abnormally distributed as values of two divisions are -3.141 and 1.724 for each, the Shapiro-Wilk value is 0.957 and the P-value of Shapiro-Wilk is only 0.009 (smaller than 0.05).

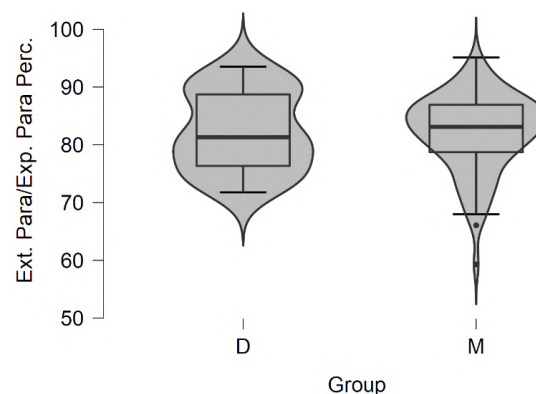


Figure 102 The boxplots comparing the proportion of paratactic extension in paratactic expansions in Master and PhD sub-corpora

A more detailed distribution of data in the two groups can be read from boxplots in Figure 102 that it is bimodal in group D and has a long, narrow tail to the left side in group M with two particular cases.

With data being abnormally distributed in both groups, the Mann-Whitney U test (Table 63) is firstly conducted with the result that the P-value (0.926) and the VS-MPR value (1) indicate total abjection for alternative hypothesis and the original hypothesis that group D has the same proportion as group M is accepted with high probability. The Hodges-Lehmann Estimate value (-0.192) shows that the difference between the two groups can only be 0.192%, with a very small effect size, as demonstrated by the rank-biserial correlation value (-0.012). But which group has a related higher proportion remains unanswered.

Independent Samples T-Test

	W	df	p	VS-MPR*	Hodges-Lehmann Estimate	Rank-Biserial Correlation
Ext. Para/Exp. Para Perc.	1265.000	0.926	1.000	-0.192	-0.192	-0.012

Note. For the Mann-Whitney test, effect size is given by the rank biserial correlation.

Note. Mann-Whitney U test.

* Vovk-Sellke Maximum p -Ratio: Based on a two-sided p -value, the maximum possible odds in favor of H_1 over H_0 equals $1/(-e p \log(p))$ for $p \leq .37$ (Sellke, Bayarri, & Berger, 2001).

Table 63 The Mann-Whitney U test for comparing the proportion of paratactic extension in paratactic expansions in Master and PhD sub-corpora

Again, the Bayesian Mann-Whitney U test is necessary for comparison. When the alternative hypothesis is that the location of group D is smaller than that of group M, the probability of accepting the original hypothesis (group D > group M) is 5.014 times (See Figure 103, left 1) of rejecting it with the Rhat value 1.064 and a small to medium effect size; when the alternative hypothesis is that the group D has a higher proportion than group M, the possibility of accepting the original as compared to the alternative is 4.35 times (Figure 103, right 1) with a small to medium effect size; and when the original is that group D has the same proportion as group M, the possibility of accepting the original is 4.66 times (Figure 1, left 2) of refusing it with a large effect size. Thus, a conclusion can be reached that doctoral students have a higher proportion of paratactic extension in paratactic expansions than Master students, though this difference can also be small.

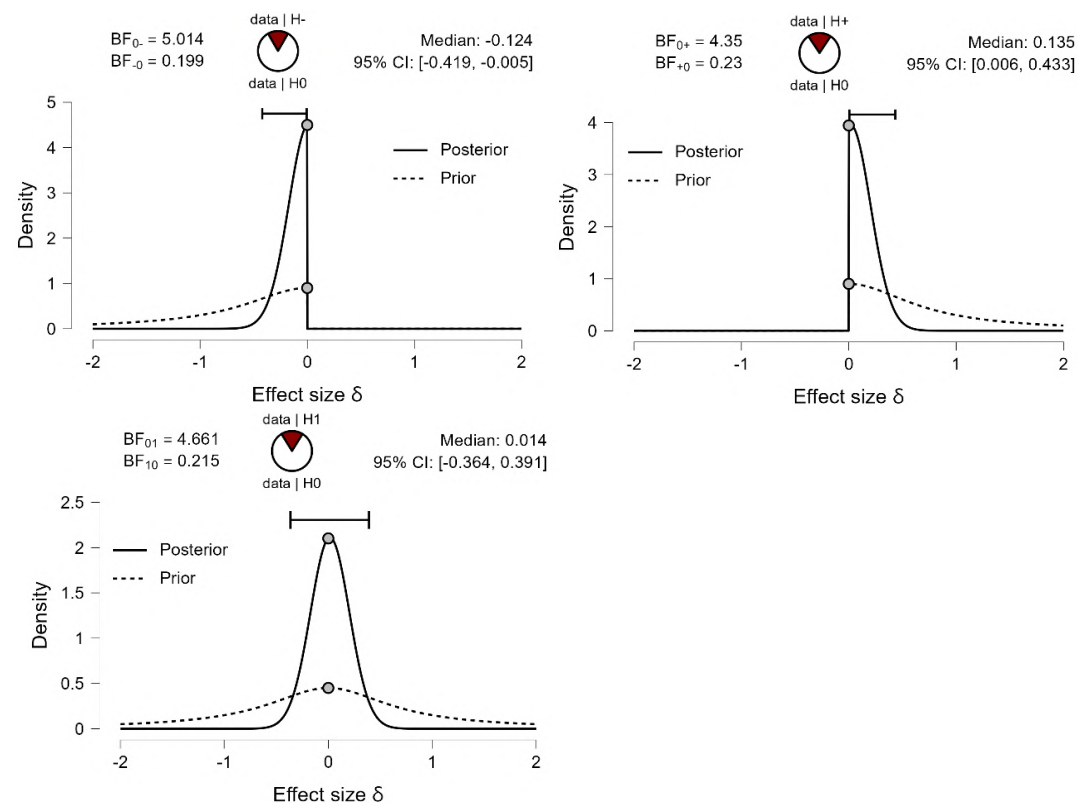


Figure 103 The inferential plots comparing the proportion of paratactic extension in paratactic expansions in Master and PhD sub-corpora

5.4.2.4 Extension Parataxis/Extension

To compare paratactic extension with hypotactic extension used by students, the proportion of paratactic extension in extensions is calculated. In some doctoral texts, this parameter can account for 80% to 96% with a mean of 87.55%; and in some Masters' texts, it can take up to 65.39% to even 100% with a mean of 90.98%. (See Table 5.4.2.4-1 in Appendix 5.4) It is no wonder that data in group D is leptokurtic (the kurtosis value is -0.97) and normally distributed (the value of skewness divided by the standard error of skewness is -0.188 (within -2/+2), the value of kurtosis divided by the standard error of kurtosis is 1.199 (also within -2/+2), the Shapiro-Wilk value is 0.962, and the P-value of Shapiro-Wilk is 0.311) with almost no skewness (the skewness value is -0.078); while in group M is left-sided skew (the skewness value is -1.306), leptokurtic (the kurtosis value is 3.248), and non-normally distributed (values of two divisions are 4.855, and 6.105 respectively, the Shapiro-Wilk value is 0.921, and the P-value of Shapiro-Wilk is smaller than 0.001). But except in two cases with small percentages of paratactic extension in all extensions, as noted by boxplots in

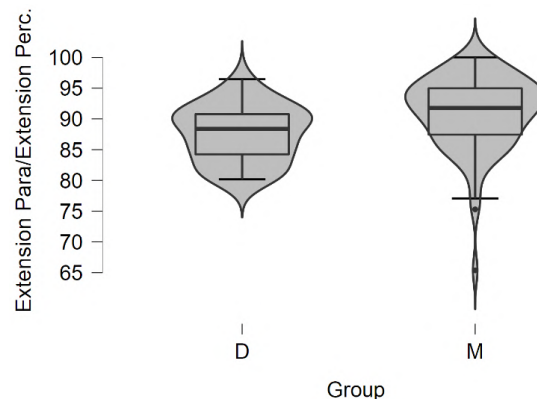


Figure 104, group M can be normally distributed.

Figure 104 The boxplots comparing the proportion of paratactic extension in extensions in Master and PhD sub-corpora

Independent Samples T-Test

	W	df p	VS-MPR*	Hodges-Lehmann Estimate	Rank-Biserial Correlation
Extension Para/Extension Perc.	754.500	< .001	128.698	-3.905	-0.411

Note. For the Mann-Whitney test, effect size is given by the rank biserial correlation.

Note. For all tests, the alternative hypothesis specifies that group D is less than group M.

Note. Mann-Whitney U test.

* Vovk-Sellke Maximum p -Ratio: Based on a two-sided p -value, the maximum possible odds in favor of H_1 over H_0 equals $1/(-e p \log(p))$ for $p \leq .37$ (Sellke, Bayarri, & Berger, 2001).

Table 64 The Mann-Whitney U test for comparing the proportion of paratactic extension in extensions in Master and PhD sub-corpora

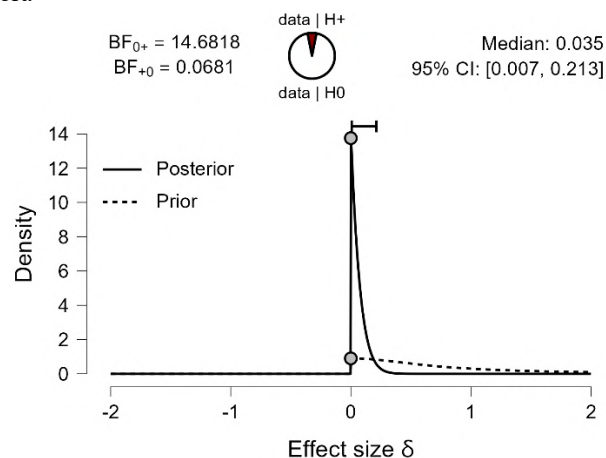


Figure 105 The inferential plots comparing the proportion of paratactic extension in extensions in Master and PhD sub-corpora (with the original hypothesis group D < group M)

Thus, only the Mann-Whitney U test result can be referred to. From Table 64, it is quite obvious that the

proportion in group D is smaller than group M as the p-value is smaller than 0.001, which means that the original hypothesis (group D > group M) is totally refused, and the alternative one is fully accepted. The VS-MPR value (128.698) indicates that the probability of accepting the alternative hypothesis is 128 times of rejecting it. And the proportion of group D can be 3.9%; smaller than that of group M as suggested by the Hodges-Lehmann Estimate value. This test has a medium effect size as read from the rank-biserial correlation value -0.411.

The same conclusion that doctoral students use a lower proportion of paratactic extension per extension than Master students can also be confirmed by the Bayesian Mann-Whitney U test, as can be seen from Figure 105: the probability of accepting the original hypothesis (group D < group M) is 14.68 times of rejecting it as indicated by the Bayesian factor value with a small effect size. The Rhat value is 1.025, which proves the existence of differences between the two groups.

5.4.3 The Distribution of Hypotactic Extension

The number of hypotactic extensions is not too much that it ranges from 10 to 63 in some doctoral texts and from even 0 to 31 in some Masters'. The mean of group D is around 29 while of group M is only 9. (See Table 5.4.3-1 in Appendix 5.4) Data are both right-sided skew (the skewness values are 0.919 of group D and 1.096 of group M) and abnormally distributed (values of the skewness divided by the standard error of skewness are 2.220 in group D and 4.074 in group M, which are both larger than +2; values of the kurtosis divided by the standard error of kurtosis are 0.307 in group D and 3.406 in group M, the former of which is within -2/+2 while the latter is bigger than +2; the Shapiro-Wilk values are 0.911 and 0.932 respectively, which are not far away from 1; and the P-value of the Shapiro-Wilk in group D is 0.012 while in group M is less than 0.001) in two groups, and group D is platykurtic as the kurtosis value is 0.248 and group M is leptokurtic as its value is 1.812. From boxplots in Figure 106, except in two cases in group M, data can be normally distributed; and except in one case in group D, some texts still have high numbers of hypotactic extensions.

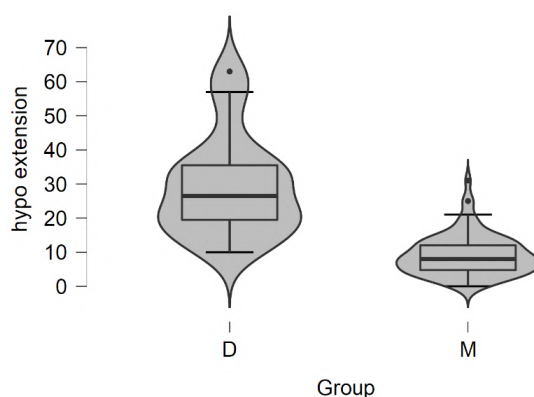


Figure 106 The boxplots comparing the number of hypotactic extensions in Master and PhD sub-corpora

5.4.3.1 Extension Hypotaxis/Clause Nexuses

Though the proportion of paratactic extension in all clause nexuses has been compared, the condition of hypotactic extensions in clause nexuses in two groups cannot be predicted.

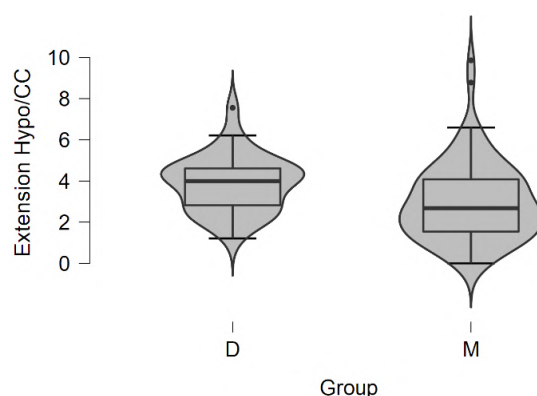


Figure 107 The boxplots comparing the proportion of hypotactic extension in clause nexuses in Master and PhD sub-corpora

From Table 5.4.3.1-1 (see Appendix 5.4), the distributions are then clear as they range from 1.2% to 7.6% with a mean value of 3.86% in the doctoral group and from 0 to 9.86% with a mean value of 2.98% in the Master's group. It seems that data of group D are right-sided skew (the skewness value is 0.356), platykurtic (the kurtosis value is 0.401) and normally distributed (the division of skewness to the standard error of skewness is 0.860, the value of the kurtosis divided by the standard error of kurtosis is 0.496, the Shapiro-Wilk value is 0.971 and the P-value of the Shapiro-Wilk is 0.54) while of group M are right-sided skew (the skewness value is 1.011), leptokurtic (the kurtosis value is 1.733) and abnormally distributed (values of two divisions are 3.758 and 3.258 separately, the Shapiro-Wilk value is 0.943 while the P-value of Shapiro-Wilk is 0.001). But when boxplots (Figure 107) are referred to, group D has one particularly high case, and group M has two, making data in both groups abnormally distributed.

For comparing data in the two groups, only the Mann-Whitney U test is suitable. Results in Table 65 show that the p-value is only 0.002, with the original hypothesis of group D having a lower proportion than group M. Thus, the original one is totally rejected, and the alternative hypothesis that group D is greater than group M is accepted with the probability of 28.184 times (indicated by the VS-MPR value). Their difference can be 1.016%, as seen from the Hodges-Lehmann estimate value. This test has a small to medium effect size, as indicated by the rank-biserial correlation value of 0.347.

Independent Samples T-Test

	W	df	p	VS-MPR*	Hodges-Lehmann Estimate	Rank-Biserial Correlation
Extension Hypo/CC	1724.500		0.002	28.184	1.016	0.347

Note. For the Mann-Whitney test, effect size is given by the rank biserial correlation.

Note. For all tests, the alternative hypothesis specifies that group D is greater than group M.

Note. Mann-Whitney U test.

* Vovk-Sellke Maximum p -Ratio: Based on a two-sided p -value, the maximum possible odds in favor of H_1 over H_0 equals $1/(-e p \log(p))$ for $p \leq .37$ (Sellke, Bayarri, & Berger, 2001).

Table 65 The Mann-Whitney U test for comparing the proportion of hypotactic extension in clause nexuses in Master and PhD sub-corpora

The testing result can be confirmed by the Bayesian Mann-Whitney U test as the Bayesian factor value is 13.18 (See Figure 108) when the original hypothesis is group D being greater than group M, with the probability of accepting the original one being 13.18 times of accepting the alternative hypothesis, a small effect size and the Rhat value being 1.021. Thus, doctoral students prefer hypotactic extensions (in clause nexuses) to Master students. But which exact category of hypotactic extensions is the most favorable to both groups can be a good research question.

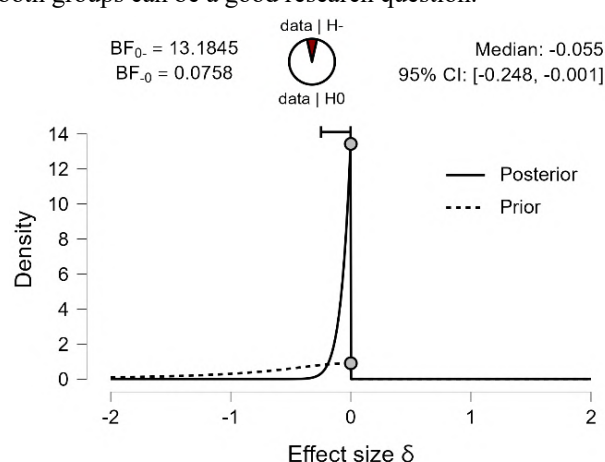


Figure 108 The inferential plots comparing the proportion of hypotactic extension in clause nexuses in Master and PhD sub-corpora (with the original hypothesis group D > group M)

5.4.3.2 Extension Hypotaxis/Hypotaxis

Again, though the proportion of hypotactic extension per clause nexuses is already known (see Section 5.4.3.1), it gives no hints or signs at all of the proportion tendency of hypotactic extension in hypotaxis. Table 5.4.3.2-1 (see Appendix 5.4) suggests that the range of hypotactic proportion in hypotaxis ranges from only 2% to 13% in group D with a mean of around 6%, and ranges from 0 to 14.79% with a mean value of 4.85% in group M. Data of the former group is right-sided skew (the skewness value is 0.658), leptokurtic (the kurtosis value is 1.433), and normally distributed (the value of skewness divided by the

standard error of skewness is 1.589, which is within $-2/+2$; the value of kurtosis divided by the standard error of kurtosis is 1.771, which is also within $-2/+2$; the Shapiro-Wilk value is 0.954; and the P-value of Shapiro-Wilk is 0.189) when statistics are investigated; and of the latter is also right-sided skew (the skewness value is 0.777), slightly leptokurtic (the kurtosis value is 0.752) but abnormally distributed (values of two division are 2.888 and 1.413 respectively, the Shapiro-Wilk value is 0.96 while the P-value of Shapiro-Wilk is only 0.013, which is smaller than 0.05). But boxplots in Figure 109 demonstrate that except for one case with a high percentage in each group, others are seemingly normally distributed.

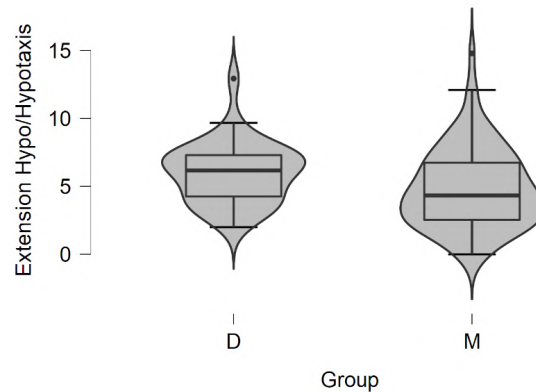


Figure 109 The boxplots comparing the proportion of hypotactic extension in hypotaxis in Master and PhD sub-corpora

The Mann-Whitney U-test is thus adopted to compare two groups of abnormally distributed data with diverse variance values. When the original hypothesis is that group D has a lower proportion than group M, the p-value is 0.009 (See Table 66), which means that the alternative hypothesis is totally accepted; and the VS-MPR value is 8.411, the possibility of favoring the alternative hypothesis over the original one is 8.411 times. The Hodges-Lehmann estimate value indicates that the difference in proportion is 1.316%. This test has a small to medium effect size, as the rank-biserial correlation value is 0.286.

Independent Samples T-Test

	W	df	p	VS-MPR*	Hodges-Lehmann Estimate	Rank-Biserial Correlation
Extension Hypo/Hypotaxis	1645.500	0.009	8.411	1.316	0.286	

Note. For the Mann-Whitney test, effect size is given by the rank biserial correlation.

Note. For all tests, the alternative hypothesis specifies that group D is greater than group M.

Note. Mann-Whitney U test.

* Vovk-Sellke Maximum p -Ratio: Based on a two-sided p -value, the maximum possible odds in favor of H_1 over H_0 equals $1/(-e p \log(p))$ for $p \leq .37$ (Sellke, Bayarri, & Berger, 2001).

Table 66 The Mann-Whitney U test for comparing the proportion of hypotactic extension in hypotaxis in Master and PhD sub-corpora

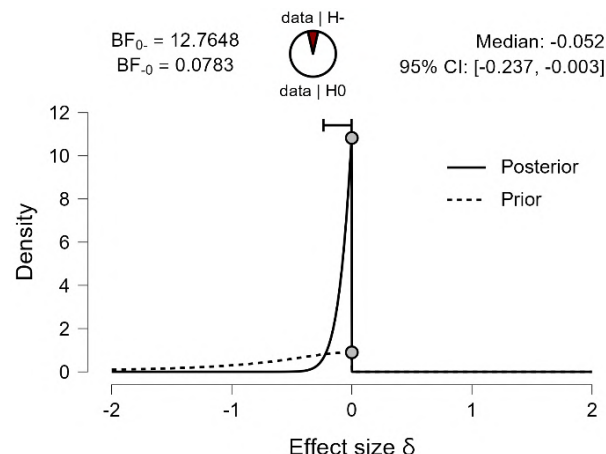


Figure 110 The inferential plots comparing the proportion of hypotactic extension in hypotaxis in Master

and PhD sub-corpora (with the original hypothesis group D > group M)

This result is correspondence to the Bayesian Mann-Whitney U test result that the possibility of accepting the original hypothesis (group D > group M) is 12.76 times (

Figure 110) of accepting the alternative one with a small effect size and the Rhat value is 1.007.

So far, the proportion of hypotactic elaboration in the hypotaxis of group D is already known to be less than that of group M, while the proportion of hypotactic extension in the hypotaxis in the former is larger than that in the latter, how is the condition of hypotactic enhancement in hypotaxis in two groups then?

5.4.3.3 Extension Hypotaxis/Expansion Hypotaxis

Before answering the question above, the denominator can be narrowed down to hypotactic expansion, and the proportion of hypotactic extension in hypotactic expansion is more useful to predict the result of the question. For the doctors' group, the parameter ranges from 3% to 16.49% with a mean value of 8.26%; and for the Masters' group, it ranges from 0 to 24.71% with a mean value of 6.88%. (See Table 5.4.3.3-1 in Appendix 5.4) Data in the former group seems to be slightly right-sided skew (the skewness value is 0.385), platykurtic (the kurtosis value is 0.462), and normally distributed (the value of the skewness divided by the standard error of skewness is 0.93, the value of the kurtosis divided by the standard error of kurtosis is 0.571, the Shapiro-Wilk value is 0.967, and the P-value of Shapiro-Wilk is 0.42) while in the latter group is right-sided skew (the skewness value is 1.1), leptokurtic (the kurtosis value is 2.673), and abnormally distributed as values of two division are 4.089 and 5.024 for each, the Shapiro-Wilk value is 0.94, while the P-value of Shapiro-Wilk is less than 0.001. Actually, group D has a short, narrow tail of a high percentage, and group M has a particular case also with a high proportion (as shown in Figure 111) that data in both groups are abnormally distributed.

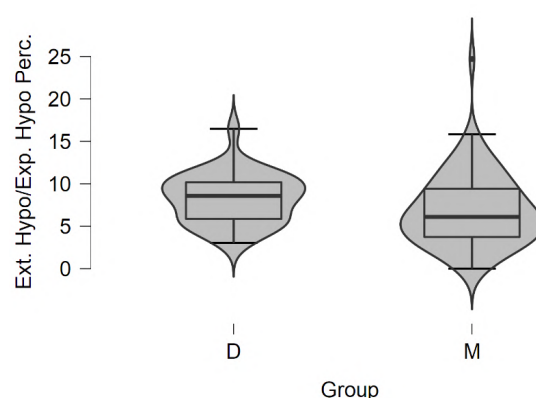


Figure 111 The boxplots comparing the proportion of hypotactic extension in hypotactic expansion in Master and PhD sub-corpora

The Mann-Whitney U test is thus referred to for abnormally distributed data with diverse variance values. It indicates that the original hypothesis should be rejected, as the P-value is 0.017, (See Table 67) i.e., the alternative hypothesis that group D has a larger proportion than group M is accepted, and this probability of acceptance is 5.43 times of rejection as suggested by the VS-MPR value. The difference between the two groups is 1.653% under the Hodges-Lehmann estimate, and this test has a small to medium effect size as the rank-biserial correlation value is 0.259.

Independent Samples T-Test

	W	df	p	VS-MPR*	Hodges-Lehmann Estimate	Rank-Biserial Correlation
Ext. Hypo/Exp. Hypo Perc.	1611.500	0.017	5.430	1.653	0.259	

Note. For the Mann-Whitney test, effect size is given by the rank biserial correlation.

Note. For all tests, the alternative hypothesis specifies that group D is greater than group M.

Note. Mann-Whitney U test.

* Vovk-Sellke Maximum p -Ratio: Based on a two-sided p -value, the maximum possible odds in favor of H_1 over H_0 equals $1/(-e p \log(p))$ for $p \leq .37$ (Sellke, Bayarri, & Berger, 2001).

Table 67 The Mann-Whitney U test for comparing the proportion of hypotactic extension in hypotactic expansion in Master and PhD sub-corpora

When the original hypothesis specifies that the value location of group D is larger than group M, the Bayes factor value is 10.606 (see Figure 112) with a Rhat value of 1.081 and a small effect size. This means that the probability of accepting the original hypothesis is 10.606 times of rejecting it.

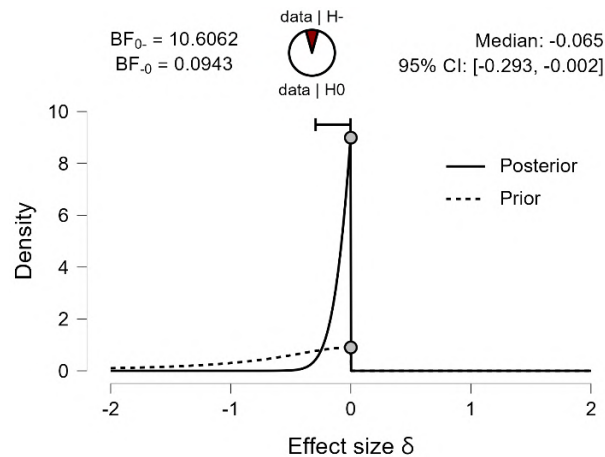


Figure 112 The inferential plots comparing the proportion of hypotactic extension in hypotactic expansion in Master and PhD sub-corpora (with the original hypothesis group D> group M)

5.4.3.4 Extension Hypotaxis/Extension

From Section 5.4.2.4, i.e., the comparison of the proportion of paratactic extension in all extensions, doctoral students have a totally lower value than Master students, which means that the result is contrary for the comparison of the proportion of hypotactic extension, i.e., the doctoral group has a higher value than the Master group. However, detailed descriptive statistics for both groups are still needed to demonstrate the basic conditions of data distribution in each text. In Table 5.4.3.4-1 (see Appendix 5.4), the proportion of hypotactic extension in extension in group D ranges from 3.5% to 19.8% with a mean value of 12.46%; and in group M from 0 to 34.62% with a mean value 9.02%. Data in the former group is leptokurtic (the kurtosis value is -0.97), normally distributed (the value of skewness divided by the standard error of skewness is 0.188, the value of kurtosis divided by the standard error of kurtosis is -1.199, the Shapiro-Wilk value is 0.962, and the P-value of Shapiro-Wilk is 0.311) with almost no skewness (the skewness value is 0.078) while in group M is right-sided skew (the skewness value is 1.306), leptokurtic (the kurtosis value is 3.248), and abnormally distributed as values of two divisions are 4.855, and 6.105 separately, the Shapiro-Wilk value is 0.921, and the P-value of Shapiro-Wilk is less than 0.001). More information can be received from the boxplots in Figure 113 that except for two particular cases with high percentages, data in group M can be normally distributed.

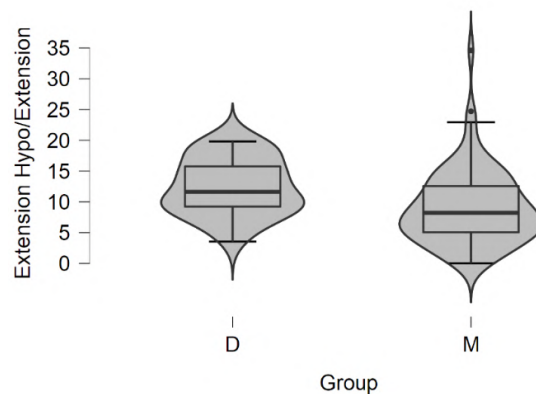


Figure 113 The boxplots comparing the proportion of hypotactic extension in extensions in Master and PhD sub-corpora

The statistics of the Mann-Whitney U test can be the same as that from the independent samples t-test of the proportion of paratactic extension in extensions (see Section 5.4.2.4) by only changing the original hypothesis into group D being less than group M. But the Bayesian Mann-Whitney U test result is a little bit different as the Bayes factor value is 32.284, indicated by Figure 114, which means that the probability of accepting the original hypothesis (group D>group M) is 32.28 times of rejecting it with the Rhat value 1.014 and a large effect size. So, doctoral students have a higher proportion of hypotactic extension in extensions than Master students.

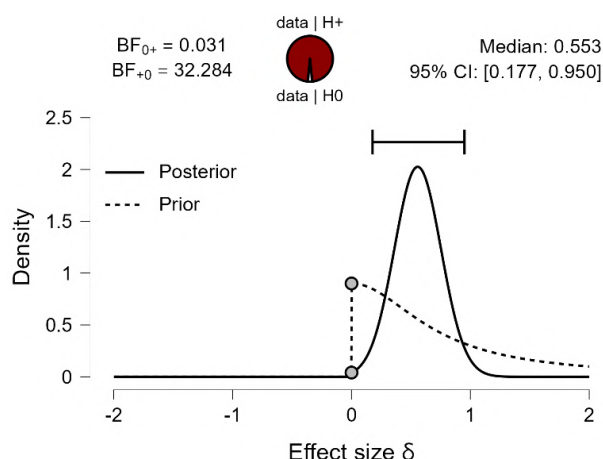


Figure 114 The inferential plots comparing the proportion of hypotactic extension in extensions in Master and PhD sub-corpora

5.4.4 Summary of Extension Comparison

To sum up, Master students prefer extensions as well as paratactic extensions in thesis writing compared to doctoral students as the proportions of extension in clause nexuses, paratactic extension in clause nexuses, and paratactic extension in extensions are all higher; while hypotactic extensions are more favored by doctoral students as other parameters such as hypotactic extension in clause nexuses, hypotactic extension in extensions, etc. account for higher percentages. Though doctoral dissertations have larger probabilities in the proportions of parameters relating to hypotactic extensions, this has no effect on the proportion of extension in clause nexuses of the doctoral group as hypotactic extensions only take up no more than 20% of all extensions.

5.5 Extension (Qualitatively Analysis)

Quantitative methods for extension explored the paratactic and hypotactic extensions from within- and around- grammatical structures in the clause complex system comparatively in doctoral and masters' groups. Not only paratactic vs. hypotactic extensions have been demonstrated, but also the diversities in each text within paratactic or hypotactic extensions have been presented. However, how students differ in their choice of lexical structures for extensions in academic writing remains unknown. This is an area where qualitative rather than quantitative methods can provide insight.

5.5.1 General Information of Extension

From Section 3.1.3.2, extension has been divided into three exact kinds: addition, variation, and alternation, in which addition means that two clauses are positively (A and B), negatively (not A and not B), or adversatively (A and conversely B) added together to form a clause complex with meaning of the former clause expanded by the secondary one; variation indicates that the subject in the former clause is replaced (not A but B) or subtracted (A but not all A) by the subject mentioned in the latter clause; and alternation is selection (A or B) within subjects that are described in two clauses. By intersecting with taxis, more choices for writing an extension clause complex are brought to students that how many would they use for each grammatical category, what kind of lexical categories would be applied in for each grammatical category in each group (i.e., Master and doctoral), would more similarities or diversities be discovered for their practical theses writing, and whether hypotactic or paratactic addition/variation/alternation clause complexes signed by some lexical markers are more favored by doctoral students, etc. are all choices related questions that to be investigated in this part.

5.5.2 Addition

For additional extensions in clause complexes chosen by students, most of them are positive ones (account for at least 75% for both groups) marked by *and* and co-ordination phrases like *not only... but also...*, *as well as*, *apart from*, etc.; around ¼ are adversative ones marked by words *but*, *while*, *whereas*, etc.; and left are negative additions signed by *nor* or the phrase *neither... nor...*

5.5.2.1 Positive

Unsurprisingly, the majority of clauses in positive additional extensions are connected by the word *and* (see Table 68). This is because it is the most straightforward option, and it is often used without conscious consideration of the consistency of subjects, polarity, modality, tense, and other aspects of the subsequent clause when it is linked to the preceding one. Then is the phrase *not only... but also...* and *as well as*, though the number is not as high as predicted. Around 10% to 20% of positive additions in each group are implicit so that no obvious markers can be found for annotating the relationship within clauses but to semantically judge, such as a list of additional clauses with the last one starting with the word *and*.

Extension	M	D
Addition: Additive	4140	3223
additive	575	533
and	3368	2503
apart from	7	8
as well as	25	32
besides	7	14
not only... but also...	137	127
both... and...	1	4
on the one hand... on the other (hand)	20	2

Table 68 The number of each lexis for positive addition clause complexes applied in Master and PhD sub-corpora

E.g., 5.5.1 ||| One of the limitations of the study is [[that the number of the proverbs chosen as data in this thesis is not very large]], || **and*** the other limitation is [[that all the data are taken from one book]]. ||| ---- CMAC15_02_780

E.g., 5.5.2 ||| Semiotics acknowledges || that semiotic systems have a conventional character, || **and*** that they can be used to make objective classifications of the elements populating our reality. ||| ---- CPhD10_0015_1251

The marker for paratactic addition relationship in E.g., 5.5.1 seems to be the phrase *one... and the other...*, but actually is the word *and* which just connects two clauses to demonstrate the second limitation as an additional one for that mentioned in the first clause. Subject (referred to as *limitation*) and tense remain unchanged, the meaning, however, is expanded with more information. It is the same condition with clauses in E.g., 5.5.2 that others remain the same in the third clause as compared to the secondary one, except that the voice has been changed from active to passive, and additional characteristics of semiotic systems have been described.

E.g., 5.5.3 ||| Learners should be asked **both** to identify the speech act within the discourse in terms of its linguistic features || **and** to comment on the sociolinguistic environment [[that causes the speech act to occur]]. ||| ---- CMAC05PR_24_711

E.g., 5.5.4 ||| As a result of these two factors, then, this kind of it-clauses **both** enables the writer to make explicit statements of perspectives || **and*** allows him to describe his perspectives in as much detail || as he wishes and in ways [[that cannot be represented by single lexical items.]] ||| ---- CPhD06_0004_1915

Another marker for paratactic addition: additive: positive is the phrase *both... and...* As in the two examples above, the secondary clause leading by the word *and* is connected with the former one, which includes the word *both* with the same subject *learners are asked to perform something* in the first example and what the *it-clause* can enable the write do because of *these two factors* in the second example, and with the same finiteness (non-finite clause *to do something* in E.g., 5.5.3) and tense (third-person singular present tense in E.g., 5.5.4).

E.g., 5.5.5 ||| To do this will get only a little amount of information of the score, || it usually needs a series of score conversion process || and make the final report **not only** have a large amount of information || **but also*** can be explained reasonably. ||| ---- CMAC13ME_11_260

E.g., 5.5.6 ||| This clarification will **not only** supplement research on personality, language learning strategies and language performance, || **but also*** help second language acquisition researchers develop a unified theory on second language acquisition. ||| ---- CPhD10_0001_72

Being the same as the phrase *both... and...*, the phrase *not only... but also...* adds two clauses together with the same subject, which can also be stated as that the secondary clause provides more information about the subject of the first one, as can be demonstrated in E.g., 5.5.5 (the subject is the *final report*) and E.g. 5.5.6. (the subject is *this clarification*). The phrase *not only in not only... but also...* and the word *both* in *both... and..* would not be in the leading status in the first clause (at the beginning of it) with high probabilities in forming up a clause complex.

E.g., 5.5.7 || Such *langue items*, on the one hand, help the Trump administration establish a positive image || and win the hearts of the people, || **on the other hand***, give the Trump administration sufficient reasons and rationality to provoke trade fictions, || because it is for the people to do so. || ---- CMAC19_05_462

E.g., 5.5.8 || Cooreman and Kilborn conclude || that L1 processing strategies are more or less transferable: || when there is a high degree of overlap in form-function mappings across the learner's L1 and L2, || direct transfer occurs; || **on the other hand***, when there is incongruity between L1 and L2 form-function mappings, || indirect transfer (transfer at the functional level alone) occurs. || ---- CPhD08_0005_627

Normally, for positively extending a clause by adding information to it, the phrase *on the other hand* would follow after *on the one hand*, such as suggested by E.g., 5.5.7 with the subject of the secondary clause deleted, as that of both clauses is *langue items*, and the latter clause is providing more reasons of the advantage of using these *langue items* in reaching some purposes. But the phrase *on the one hand* may also disappear in demonstrating the additive: positive relation in two clauses by the application of only the phrase *on the other hand* as described in E.g., 5.5.8. The structure can still signify another reason why *L1 processing strategies are more or less transferable*, as concluded by Cooreman and Kilborn.

E.g., 5.5.9 || At this time, the body will immediately relax with a sincerely smile, adrenaline and Cortisol reduce the secretion of heart, liver, gastrointestinal activities tend to benign, gas exchange, digestion and absorption, balance blood pressure and so on, || thereby strengthening physiological functions || **as well as*** solving the harm of the pressure with ease. || ---- CMAC13CU_27_353

E.g., 5.5.10 || Therefore, raters should pay special attention to the candidate's effectiveness [[in getting across the intended message || **as well as*** giving appropriate responses to the message passed on to him/her, i.e., communicativeness]]. || ---- CPhD05_0008_1256

In the two examples above, the relationship between two clauses is paratactic addition: additive connected by the phrase *as well as*, with the finiteness (non-finite clauses start with V.+ing.) and tense of the secondary clause remaining the same as the former, and the clause order can be turned without changing the meaning of the whole clause complex or what the writer wants to emphasize. The phrase *as well as* can actually be replaced by the word *and*, which does not have such a kind of restriction for consistency.

E.g., 5.5.11 || Langacker (2004) considers || **apart from*** sense image, <<including the visual image and audio image, etc.,>> || there lies an important image—motor image. || ---- CMAC06MO_18_231

E.g., 5.5.12 || **Apart from*** being the brand name, || “more” suggests || that the quality of More cigarette is more terrific than any other kind ---- || being longer, better and milder probably. || ---- CPhD07_0011_708

The same phrase for indicating positive additive relation between clauses as former ones is *apart from*, which however, can be put at the beginning of the whole clause complex or within two clauses, i.e., the location of this phrase is flexible. Another difference comes from the emphasis that by using *apart from*, the clause with this phrase has known information while without this indicates the stress of new information such as *motor image* instead of *sense image* in E.g., 5.5.11 and *being longer, better and milder* rather than *being a brand name* of the word *more* in *More cigarette*.

E.g., 5.5.13 || **Besides*** rejecting traditional views of class, || she also denigrates society's attempts to restrict women's activities. || ---- CMAC06LIT_31_443

E.g., 5.5.14 || That is to say, **besides*** expressing comments on || and showing attitudes towards propositions, || this type of perspectival metadiscourse can perform the function of evaluation. || ---- CPhD06_0004_2347

A relative word of the phrase *apart from* is *besides*, which has the same meaning, function (emphasizing the information provided in the clause without *besides*), location in clauses (flexible), and following structures (non-finite verb) as the former one. In addition, these two markers do not have any limited condition on the structure of two clauses that consistency is not obligatory. The only diversity is that *besides* can be a conjunctive marker for adding two clause complexes together by being the first word with a comma in the second one, while *apart from* cannot. So, in E.g., 5.5.13, the people's action of *denigrating society's attempts to restrict women's activities* is what the writer wants to reinforce, and in E.g., 5.5.14, the real *function of evaluation* that performed by *the specific type of perspectival metadiscourse* is what really surprises the writer, by choosing the word *besides* for marking hypotactic extension: addition relations.

E.g., 5.5.15 || They are frank enough to tell their personal stories, || disclose their secrets, || talk about scandals about them, || share their experience and thoughts with all the audience (both those on the spot and those in front of the television). || ---- CMAC11CU_15_62

E.g., 5.5.16 || In collaborative learning, group leaders are also encouraged to take responsibilities for their groups, || organize and direct the learning process || and give his own comment and assessment together with his peers'. || ---- CPhD11_0007_341

In E.g., 5.5.15, no marker is present within the clauses, yet the relation remains paratactic addition: positive additive. This is evidenced by the continuous provision of explanations for why *people are frank enough*, as previously described in the initial clause. However, the writer has erroneously deleted the word *and* between the third and fourth clauses. In E.g., 5.5.16, it is the same story with the first and second clauses that *and* is omitted in case of repeating (as it has already appeared within the second and the third clauses). This kind, i.e., the relation of clauses without any obvious marker (word or phrase), is an implicit clause complex vs. an explicit one (See Section 7.3).

5.5.2.2 Negative

The number of negative additional relations is small, with only 11 being dug out in Doctoral dissertations, and six are discovered in master's theses. Mainly, two markers are used to indicate this kind: one is the phrase *neither... nor...* (4 cases in Doctor's and 3 in Master's), and another is the word *nor* (7 in Doctor's and 3 in Master's).

E.g., 5.5.17 ||| In spite of the fact [[that he **neither** receives any proper education from his birth || **nor*** gets the right moral guidance from anyone]], || Oliver isn't revealed as a complicated figure torn between good and evil. ||| ---- CMAC10LIT_7_95

E.g., 5.5.18 ||| Without such links academics could **neither** justify their arguments || by connecting their research activities to significant work in the field, || **nor*** use this disciplinary knowledge to establish the novelty of their position (Gilbert, 1976; Berkenkotter & Huckin, 1995). ||| ---- CPhD06_0004_1137

Normally, the phrase *neither... nor...* is regarded as the negative form of the phrase *either... or...*, but in SFG, it has a closer relation to the phrase *both... and...* that themes in both clauses are emphasized and denied with the second clause adding more information to the former one. The Subject, tense, voice, aspect, mood, etc., of the second clause, usually remain the same as the former one, just as what has been demonstrated in the two examples above that traits or characteristics being not possessed by Oliver in the former example or by academics in the latter are clearly described. The phrase *either... or...*, however, involves selecting one within two, thus performs an alternation function (see Section 5.5.4), and is not the agnation item of *neither... nor...*

E.g., 5.5.19 ||| Dora can't live up to David's image of her, || **nor*** can she offer [[what he wants || and needs]]. ||| ---- CMAC10LIT_7_298

E.g., 5.5.20 ||| However, no things are exactly the same with each other, || **nor*** are English and Chinese E/P constructions. ||| ---- CPhD07_0008_439

Compared to the phrase *neither...nor...*, the word *nor* puts more stress on the theme of the clause that it leads to a clause complex by mentioning a new trait, as shown in E.g., 5.5.19. Though it can be written as *Dora can neither live up to ..., nor can offer...*, the meaning has to some degree changed as the Subject (she or Dora) has been deleted in the second clause, and the semantic core has been shifted to the former one. The second example cannot use *neither... nor...* as the Subject of the former clause (*things*) has been changed to *Chinese and English E/P constructions* in the latter, which more likely is clarifying by giving a specific example instead of merely providing more information for the theme of the former clause.

5.5.2.3 Adversative

The majority of adversative relations are connected by markers *but* in paratactic clause complexes and *while* in hypotactic ones (see Table 69). The remaining relation is led by *without* and *whereas* in hypotaxis and (*and*) *yet* in parataxis. The total number of adversative additions found in Doctoral dissertations and Master theses are almost the same, with that of hypotactic ones in the former group outnumbering the latter (679 vs. 521).

Extension	M	D
Addition: Adversative	1235	1214
but	655	475
while	430	503
without	45	24
whereas	46	152
(and) yet	16	19
however	37	36
on the contrary	6	5

Table 69 The number of each lexis for adversative addition clause complexes applied in Master and PhD sub-corpora

E.g., 5.5.21 ||| The subjects have learned English for many years || and possess amount of prior

knowledge, || **but*** most of them have problems in imaging prior knowledge || when meeting a new knowledge point. || ---- CMAC13RE_10_409

E.g., 5.5.22 || What Hoey mentions here might be understood in another way [[that language resources, either clues or inter-sentential relations at paragraph levels, are finite, || **but*** the possible patterns built out of these resources are infinite]]. || ---- CPhD05_0003_993

When the converse of things is discussed, either part of one theme or the whole theme in the first clause can be reversed, and the third way is to hold a negative attitude by mentioning a new one. In the examples above, the first strategy is adopted by choosing the word *but* that *part of the subjects has problems in imaging prior knowledge* in E.g., 5.5.21 and *part of patterns built out of language resources are infinite* in E.g., 5.5.22. This does not mean that *but* can only be used in the first strategy.

E.g., 5.5.23 || In the mid-term examination, the passing rate of the experimental class is only 67.86%, || which is lower than that of the controlled class (71.43%), || **and yet***, in the final examination, the passing rate reaches 85.71% || which is much higher than that of the controlled class (78.57%). || ---- CMAC07LI_20_487

E.g., 5.5.24 || On the other hand, rating according to OARS is time-consuming, || **yet*** it is worthwhile using considering the detailed information it can offer. || ---- CPhD05_0008_882

Another marker for paratactic adversative clause complex is the word (*and*) *yet*, which seems to be the same as the word *but* by both emphasizing the theme of the following clause. In E.g., 5.5.23, the content of the whole former clause is denied though the theme remains the same (the passing rate of the examination) in the first and the third clauses, and the words *and yet* can be substituted by *however* or *whereas*, which can actually lead a separate adversative clause complex; or by words like *but* and *while* without the comma. In E.g., 5.5.24, the theme of the former clause is similarly negated, and a new topic is mentioned, i.e., the theme has been transformed from *rating* to *the worth of using rating*.

E.g., 5.5.25 || Therefore, the prosodification way of adding regular past inflection—ed to the stem in English as an adjunction is difficult for Chinese L2es to acquire; || **however***, for irregular past morphology, inflection is added internal to the stem, similar to uninflected and monomorphic forms. || ---- CMAC14PH_17_416

E.g., 5.5.26 || Usually, English E/P constructions are used singly in discourse; || **however***, they can be used in a succession. || ---- CPhD07_0008_277

When the word *however* is used to present the addition: adversative relation in two clauses, the reverse condition described in the secondary clause can be quite different from those in the initial one, as shown in E.g., 5.5.25, when the subject of the former clause is *the prosodification way of adding regular past inflection—ed to the stem in English as an adjunction* while that of the latter clause is *irregular past morphology with the inflection being added internal to the stem*, the meaning that the writer intends to convey seems to be the easier acquirement of the latter structure as compared to the former, though it has not been clearly demonstrated in words. But the adversative information of the secondary clause can also be part of the characteristics of the subject of the initial one, just as what E.g., 5.5.26 writes: the latter clause connected by the word *however* is providing another condition of how *English E/P constructions* can be used.

E.g., 5.5.27 || It means || that we are not able to control the nature; || **on the contrary*** we are overcome by the variety and beauty of the nature. || ---- CMAC10LIT_15_601

E.g., 5.5.28 || Obviously, both Langacker and Lakoff think || that there carries with it some abstract locative meaning, || yet it is thought in this dissertation || that there does not designate an abstract location, || **on the contrary***, it merely serves as an introductory word [[whose function is to introduce the following nominal phrase]]. || ---- CPhD07_0008_681

The phrase *on the contrary* in linking two clauses with an addition: adversative relation usually describes another condition of the subject mentioned in the primary clause, as indicated by E.g., 5.5.27 and 5.5.28: the opposite condition of *we control the nature* is *we are overcome by the variety and beauty of the nature*, the subject of both clauses is *we*, and the thing discussed by the writer is the relation of *we* and *the nature*; the reverse condition of the content *there carries with it some abstract locative meaning* is *it merely serves as an introductory word*, in which the latter is accepted by the writer, and the subject of both clauses is *there*.

E.g., 5.5.29 || In the example above, we would see || that the first speaker talks about a lot of advantages of Japan, || **while*** the second speaker seems to disagree with the first one more or less. || ---- CMAC12SP_005_456

E.g., 5.5.30 || The difference between Preliminary Level and Advanced Level and that between Preliminary Level and Intermediate Level are the most significant with their value $p=0.000$ and $p=0.003$, || **while*** the difference between Intermediate Level and Advanced level is $p=0.118(p>0.05)$, || indicating statistical insignificance. || ---- CPhD08_0005_1065

The word *while* in marking hypotactic addition: adversative relation can only be found within two

clauses, wherein the theme usually remains unchanged. Rather than stressing more about the content of the clause it leads to, the comparison between two groups described under the theme is what the word *while* actually indicates. In E.g., 5.5.29, whether *the second speaker agrees or disagrees with what the first speaker talks about* is compared by the writer, and in E.g., 5.5.30, *the difference between Preliminary and Advanced Levels and that between Preliminary and Intermediate Levels* is contrasted with *the difference between Intermediate Level and Advanced Level*, from which groups or things that can be compared must be relative.

E.g., 5.5.31 ||| The more specific options for disclaim are deny and counter, || **whereas*** those for proclaim are concur, pronounce and endorse. ||| ---- CMAC17_03_141

E.g., 5.5.32 ||| **Whereas*** path analyses have not shown significant influence for the emotionality component, || the path from worry to academic achievement has been found to be significant for adolescents (Williams 1991) and postsecondary students (Bandalos et.al.1995). ||| ---- CPhD08_0001_842

Also, for hypotactic adversative clause complexes in making comparison, the word *whereas* can be put at the beginning or the end of the whole clause complex or within two clauses, i.e., the location of it is very flexible. But no matter where it is located, the content of the second clause is always what the writer wants to emphasize. In E.g., 5.5.31, it is the *specific options for proclaim* instead of *for disclaim* that really matters, and in E.g., 5.5.32, *the path from worry to academic achievement* being significant is a more important point by choosing *whereas*.

E.g., 5.5.33 ||| And the classification of hedges on the semantic level seems to only involve the vagueness of words or phrases || **without*** thinking about the vagueness of the truth-value of propositions and the relationship between the speakers and the propositional contents, || and thus to some degree can't cover all of the hedges. ||| ---- CMAC12SP_005_82

E.g., 5.5.34 ||| The teacher focuses on the ideas in the teaching instructions || **without*** paying much attention to the needs of the individuals in the class. ||| ---- CPhD08_0001_1338

For leading a non-finite hypotactic adversative clause, the word *without* can be located at the beginning of the clause complex or between two clauses as the content or the theme of the clause without this marker is usually more important. The theme of the clause beginning with the word *without* can be more detailed (some characteristics of the vagueness), as compared to that of the former clause being connected to, as shown in E.g., 5.5.33 or can be a new one such as in E.g., 5.5.34 (from *teaching instruction ideas* to *individual needs*).

No matter the second clause positively adds some contents to the primary clause by choosing words *and*, *while*, *whereas*, phrases like *not only... but also...*, *apart from*, *as well as* or double negated, i.e., deny things described in both clauses by using the phrase *neither... nor...*, or deny only part of the themes conveyed in the main clause or even the whole clause followed by mentioning new themes in selecting words *but*, *yet*, *while*, *whereas*, and *without*, more information is provided for the first clause, the relation between clauses are thus extension: addition.

5.5.3 Variation

More than 90% of variation relations (210 and 185 respectively) chosen by both Master and Doctoral students found in the corpus are replacive and marked by words like *instead*, *but*, *not*, and *rather* and phrases such as *instead of*, *not... but...*, and *rather than*; left is subtractive led by *except*, *except for*, *except that*, etc.

5.5.3.1 Replacive

Altogether, 191 extension: variation: replacive relations are discovered in the master's group and 172 in the Doctoral group from the corpus that students in the former group prefer to use *instead (of)* (72 of group M and 39 of group D), while in the latter group more likely to choose *rather (than)* (47 in group M and 71 in group D). (See Table 70) The number of the structure *not... but...* (*not* included) (57 in both groups) and of the phrase *but not* (4 in two groups) is the same, leaving alone some Master students apply *but* in expressing replacive clause functions.

E.g., 5.5.35 ||| Teachers should let students know || that they learn a foreign language and foreign culture || **not** because they have to take in all of foreign cultures or follow their patterns, || **but*** because they need to communicate better with foreign friends, || take in the advantages of other cultures, || and most important, to understand their culture better and more deeply. ||| ---- CMAC08CU_33_235

E.g., 5.5.36 ||| Studies on MA should **not** be totally confined to linguistic realm, || **but*** be extended to embodied experience. ||| ---- CPhD17_0001_33

The phrase *not... but...* indicates choice that what is described (usually common sense or general opinions) in the clause led by *not* is disapproved while the content in clauses beginning with *but* is

emphasized and suggested by the writer for readers to reconsider about sayings and even accept them. In E.g., 5.5.35, students normally think that *learning a foreign language is for taking in foreign cultures or following patterns*. But the writer's opinion is obvious by using *not... but...* to advocate the *need for better communication, advantages of other cultures and a better understanding of the culture*. In E.g., 5.5.36, as studies on MA used to be *confined to linguistic realm*, the writer calls on extended research of *embodied experience*, which may broaden our horizons.

Extension	M	D
Variation: Replacive	191	172
but	11	-
but not	4	4
instead of	59	27
instead	13	12
not	6	4
not... but...	51	53
rather than	37	51
rather	10	20

Table 70 The number of each lexis for variation: replacive clause complexes applied in Master and PhD sub-corpora

E.g., 5.5.37 ||| But the sea is **not*** merely a purified world; || it is the place for ceaseless fighting || since Santiago wants to defeat nature with his courage and determination. ---- CMAC10LIT_15_391

E.g., 5.5.38 ||| Some items were **not*** answered; || they were treated as missing values [[which were replaced with “999”s]]. ---- CPhD10_0018_752

When the word *not* appears alone for variation: replacive relation within clauses, the clause usually ends with a semicolon that the word *but* is unnecessary in the following clause as indicated by examples above. In E.g., 5.5.37, what the writer wants to convey is that *the sea is not a purified world but a place for ceaseless fighting*, i.e., the first two clauses can be united into only one as a replacive phrase marked by *not... but...* But the writer purposely takes them apart into two clauses connected by a semicolon; the theme of the secondary clause is thus deeply stressed and described in detail. The same goes for E.g., 5.5.38, where two conditions of *some items* are independently discussed.

E.g., 5.5.39 ||| Prefix just changes the meaning of the base || **but not*** alter the word class of the base. ||| ---- CMAC08SE_25_281

The phrase *but not* can only be located between two clauses with or without a comma in front of it for highlighting the theme of the clause without it in an extension: variation: replacive clause complex, such as in E.g., 5.5.33. The whole clause is agnate with that uses the phrase *not... but...*, which can be written as *Prefix does not alter the word class of the base, but just changes the meaning of the base*. The degree of emphasis on the theme of the agnation, however, is not so deep as the original one by using *but not* and putting the thematic clause in front of it.

E.g., 5.5.40 ||| The second important contribution is the insistence that this variation is neither trivial nor a pale reflection of “real” language, || **but*** that it is systematic || and that systematicity of linguistic variation is an imperative object of study in itself. ||| ---- CMAC19_03_135

The word *but* alone for an extension: variation: replacive clause complex relation usually follows up semantically negative clauses without the word *not*, thus it can be seen as a derivative marker of the phrase *not... but...*, as suggested by E.g., 5.5.40. This kind has yet not been discovered in the Doctoral corpus. But the number 11 in the Master's group can be explained by a larger sample size on the one hand, and peculiar writing habits in that some implicit structures would be repeatedly used by some Master's students on the other. In the example, no matter whether *that it* is deleted or not because of redundancy, *but* indeed leads to another clause for a replacive relation that *this variation is systematic instead of trivial or a pale reflection of “real” language*, which is the place that the writer may require readers to pay attention to.

E.g., 5.5.41 ||| In the above situation, Manny is going to play golf tomorrow, || but he prepares for it with his carefully selected garments, || **instead of*** practicing in advance or some other things related to playing golf. ||| ---- CMAC15_01_521

E.g., 5.5.42 ||| So, **instead of*** allowing them to assume || that academic discourse conventions are self-evidently impersonal, || we need to lead students towards an understanding [[that there is no single set of rules or practices we can apply to writing in all fields, || and that the explicit writer presence and stancetaking is often an effective rhetorical option]]. ||| ---- CPhD13_0010_1743

For hypotactic extension: variation: replacive relations of clause complexes, the phrase *instead of* can be put at the beginning of the dependent clause or between two clauses to lead to a dependent one. The theme of the primary clause, i.e., the clause without this phrase, is usually more important than, it is what the writer intends to transmit to readers after comparison and selection within themes described in both primary and dependent clauses. Hence, *Manny's preparing for carefully selected garments instead of practicing in advance before going to play golf* in E.g., 5.5.41 is reinforced, and the contrast is mixed with sneer because of choosing *instead of* by the writer, and *help students understand some characteristics of writing instead of allowing them to assume something* is considered to be right and necessary in writing teaching by the writer.

E.g., 5.5.43 ||| However, one should note || that the suggestions for teaching are not intended to be exhaustive, exemplary, or very specific; || **instead***, they are intended to provide teachers with guidelines and cursory examples of ways [[in which they might proceed to develop their own teaching procedures]]. ||| ---- CMAC07LI_20_493

E.g., 5.5.44 ||| Language characterizes the collective semiotic register of our human society, and without causing confusion; || **instead***, it grants us a special way of understanding the world surrounding us. ||| ---- CPhD10_0015_795

The word *instead* is located within two clauses by following a semicolon or a comma and is being followed by a comma or at the end of the latter clause in a clause complex for indicating hypotactic variation: replacive relations that the clause without *instead* normally has markers for negative meanings and then would be disagreed by the writer of its theme, while with it is what the writer prefers to highlight semantically as demonstrated in two examples above. Using *instead* rather than *instead of* provides wider choices of tense, modality, voice, Subject, mood, etc., for a clause.

E.g., 5.5.45 ||| Participants preferred sending private messages to the teacher || **rather than*** submitting their comments in the regular way in a SNS website. ||| ---- CMAC14RE_15_511

E.g., 5.5.46 ||| In Flege, Yeni-Komshian and Liu (1999), musical ability was touched upon, || but it was self-rated in conjunction with mimicry ability || **rather than*** independently and rigorously measured through a standardized test. ||| ---- CPhD10_0009_308

Another phrase that is similar to *instead of* is *rather than*, though the latter has a broader range of verb forms (not only gerundial structure) following it. Usually, this phrase has no comma in front of it when it is located within two clauses and used to connect them, though it can also be located at the beginning of the whole clause complex. The theme of the clause without *rather than*, such as *participants prefer to send private messages to the teacher* in E.g., 5.5.45 and *musical ability was self-rated in conjunction with mimicry ability* in E.g., 5.5.46 is what the writer emphasizes that only some facts are stated rather than selections are made.

E.g., 5.5.47 ||| From this point, it is clear || that discourse is not negative and passive, || and also power is not a kind of controlling and repressing force, || **rather*** they are the necessary and positive factors in the construction of society. ||| ---- CMAC16_02_277

E.g., 5.5.48 ||| In companies with healthy cultures, employees aren't kept in the dark; || **rather**, they are supported in the belief [[that they are part of an exciting future]]. ||| ---- CPhD07_0004_829

Not quite the same as the word *instead*, *rather* can only be put between two clauses with or without a comma after it in denoting an extension: variation: replacive relation in a clause complex that the theme of the clause with it is always being emphasized after comparing it with the theme of the clause without it and then is advised for readers to accept. Thus, *power is necessary and a positive factor in constructing the society* in E.g., 5.5.47 and *employees are supported by a belief* in E.g., 5.5.48 are key points that should be paid attention to by readers.

5.5.3.2 Subtractive

The number of extension: variation: subtractive relations found in the corpus is limited in that most of them are *except* related, in addition to *but* in the Master's group, *only* in the doctoral group, and *rather than* for both groups. (See Table 71)

E.g., 5.5.49 ||| Although the names reflect colorful cultural flavours, || it is usually translated literally to their pinyin forms, || **except*** when the authors mean to add something to it to reveal some characteristics of the person. ||| ---- CMAC07CU_35_263

E.g., 5.5.50 ||| Schifffrin (Ibid.) argued || that oh occurs || as speakers shift their orientation to information || when used in conversation || **except*** being used as an emotional marker. ||| ---- CPhD17_0002_1037

In most cases, the word *except* would be located within two clauses to connect them for marking a paratactic variation: subtraction clause complex by mentioning the specific condition described in the clause with it. It can also be the head of the whole clause complex, i.e., the sequence of two clauses (the initial and the second clause) is reversible. The exception is important, but the theme of the clause without *except* seems to be more stressed that its existence won't be affected by any exception. Hence,

names are usually translated to pinyin forms except for one condition in E.g., 5.5.49 and *oh* occurs when speakers shift their orientation to information except that it has another use in E.g., 5.5.50.

Extension	M	D
Variation: Subtractive	19	13
but	5	-
except	8	6
except for	5	-
except that	-	5
only	-	1
other than	1	1

Table 71 The number of each lexis for variation: subtractive clause complexes applied in Master and PhD sub-corpora

E.g., 5.5.51 ||| The phenomena are widely attested in Wu, || are all **but*** unknown among Mandarin and Min dialects. ||| ---- CMAC12PH_1_17

The agnate clause complex of E.g., 5.5.45 above can be: *The phenomena are all widely attested in Wu, except being unknown among Mandarin and Min dialects*, in which the word *but* performs the same function as *except* for denoting a paratactic variation: subtractive relation within two clauses that the characteristics accord to *Wu but not Mandarin and Min dialects*, which then calls for attention from readers. Exception thus normally does not require choice but only notice.

E.g., 5.5.52 ||| **Only*** when writing is connected with known non-verbal iconic and indexical signs || and thus given an intersemiotic dimension, || will we begin to grasp the language written to us. ||| ---- CPhD10_0015_1377

The word *only* is extraordinarily hard to identify for paratactic extension: variation: subtractive relation that only one has been found in the corpus. In most cases, *only* would not lead to a clause or even a clause complex and is usually used to refer to a condition. In E.g., 5.5.52, to be specific, *when* is an obvious marker that the clause complex can easily be regarded as a temporal or conditional relation. But semantically, the clause complex agnates with the sentence: *Except when writing is connected with ..., will we not begin to...* that *only* and *except* both indicate subtractive function and the former is positive. It can also be written as: *But for the condition when writing is..., will we not begin to...* that *but*, and *except* all show negative subtraction with *not* being necessary in the latter clause.

E.g., 5.5.53 ||| It is hardly to see the studies of CET on social impacts || **except for*** How to View CET Correctly (Xiao Xiao, 2010) discussing the feimess and social effect of CET. ||| ---- CMAC13ME_11_104

E.g., 5.5.54 ||| **Except for*** giving some special guidance, || the teacher may collect some common problems in reading || and give guidance publicly, || and some special parts of reading materials could be picked out to share. ||| ---- CMAC13RE_10_679

Master students have a preference for the phrase *except for* than *except that* in signifying hypotactic extension: variation: subtractive relation in clause complexes in academic writing that only 5 cases have been discovered from the corpus. Though non-finite verbs have to be added after the preposition *for*, the location for *except for* in a clause complex (it only leads to a dependent clause) is selectable, i.e., it can appear at the beginning of the whole clause nexus or between two clauses as a connection. The theme of primary clauses that don not lead by *except for* is more important than that in dependent clauses. Thus, *studies of CET on social impacts that are rarely seen* in E.g., 5.5.53 make the research conducted by the writer necessary, and *the teacher's duty to give guidance publicly more than some special guidance* in E.g., 5.5.54 is what the writer more focus on.

E.g., 5.5.55 ||| **Except that*** the difference between group 1 and the error correction group for articles is significant at the 0.051 level on the immediate posttest, || the difference between group 1 and the error correction group on both posttests for the three features is significant at 0.01 or 0.05 level, || which provides robust evidence [[that the treatment of error correction gives rise to the development of the three target features]]. ||| ---- CPhD08_0012_1423

E.g., 5.5.56 ||| At the dictogloss stage, the task demands for the participants in the oral production group were the same as those of the other two experimental groups || **except that*** the participants were expected to make an oral production || after the original passage were read twice. ||| ---- CPhD08_0012_803

The phrase *except that* is more favored by doctoral students in that it has no requirement for non-finite verbs as followers as the phrase *except for* does, it can be followed by any noun (works as Subjects in a clause) instead. The theme of the clause beginning with *except that* does matter to some extent, that of the primary clause is of more importance that the writer catches readers' attention by

choosing *except that* in a clause complex. Semantically, the dependent clause is not merely a subtraction of the primary one (See E.g., 5.5.55), it can also give more information to the theme of the latter in mentioning particular conditions that E.g., 5.5.56 belongs to this kind as *participants in the oral production group were expected to make an oral production under certain circumstances* in addition to other operations as the other two experimental groups.

E.g., 5.5.57 ||| The Cooperative Principle and its attendant maxims by Grice help exchange information maximally effectively and efficiently, || better convey and understand || what the speaker says, || **other than*** giving rise to misunderstanding. ||| ---- CMAC14LIT_25_213

E.g., 5.5.58 ||| **Other than*** segmenting experiences of the world into sequences and entities, || Halliday and Matthiessen (2004:29) point out || that language is always enacting personal and social relationships with other participants. ||| ---- CPhD13_0010_419

The phrase *other than* in manifesting a hypotactic extension: variation: subtractive relation in a clause complex should be able to be replaced by *except for* in most cases, but in the two examples above, students seem to mix *other than* with the phrase *rather than* which can be replaced by *instead of*, i.e., the clause relationship in these two examples are more replacive than subtractive that writers convey selective points of view to support one aspect but not the other. Given the current corpus data, which reveals only one instance of *other than* in each group, the assertion that students encounter challenges in differentiating this construction from *rather than* necessitates further evidence and an investigation into the underlying reasons, including linguistic and cultural influences.

5.5.4 Alternation

The third category of extension is alternation whether the writer is stating the selective characteristics of objects being described or matters being discussed, is not sure of what is being described/discussed, or is transferring the right of choosing to readers in deciding which point or matter should be a better choice is alternative that diverse words or phrases such as *or*, *either... or...*, *if... not (...then)*, *or else* can be used in theses writing, in which *or* takes the highest proportion in our corpus (204 in group M and 272 in group D), only 25 (5 in group M and 20 in group D) clause complexes altogether are marked by *either... or...*, and left two phrases though have listed as principal markers for alternation, it is more potential that they are applied in indicating enhancement: condition: negative relations (See Section 5.7.6.2) rather than alternations in the corpus.

E.g., 5.5.59 ||| During the post production of sitcoms, the producers and technicians may add to recorded laughter || **or*** manipulate the sound level || to get the most desirable humorous effects. ||| ---- CMAC14PR_4_277

E.g., 5.5.60 ||| As long as the teacher ignored development of students' autonomous learning abilities, || **or*** deprived them of their interaction and opportunities of cooperative learning, the important components of language development, || the LSSs would surely stay low not only in language competence but also in language performance. ||| ---- CPhD11_0006_1216

When the word *or* is chosen, the Subject of two clauses connected by it normally remains the same, with similar grammatical structures following it. In a clause complex, the primary clause can be extended with numerous clauses beginning with *or*, but it is still regarded as an alternation between the primary one and the second clause and within the secondary clause and other clauses in this condition. In the two examples above, both writers are stating the facts that as long as one condition (the theme of the clauses connected by *or*) is established, some aims can be reached, i.e., *the most desirable humorous effects can be got* in E.g., 5.5.59 and *the LSSs would surely stay low* in E.g., 5.5.60.

E.g., 5.5.61 ||| In the equipollent patterns, the properties are symmetrical, || **either** arranged end-to-end and completely disjunct, || **or*** completely overlapping. ||| ---- CMAC10SE_18_125

E.g., 5.5.62 ||| Although some scholars (e.g., Bunton 1999, 2002; Paltridge 2002; Kwan 2006) have made contributions to the study of dissertations, || they **either** focus on the generic structure, || **or*** concentrate on the linguistic features of the discourse. ||| ---- CPhD06_0004_815

As the phrase *either... or...* confines the number and the sphere of choices that only two are inclusive and provided for consideration, and the themes of these two clauses are about two attributes of one thing or one matter, it is not quite commonly used in thesis writing. In the two examples above, the paratactic clauses exhibit a symmetrical structure with identical subjects, namely *symmetrical properties* in the first instance and *scholars* in the second. Additionally, they adhere to the same verb form, encompassing both passive and active voice with the present tense. By choosing *either... or...* in E.g., 5.5.62, scholars who *made contributions to the study of dissertations* in the writer's understanding are dichotomous that his/her novel research may be a third kind, which is of great necessity and values.

5.5.5 Summary of Extension Marker Choices

In extension clauses complexes, the use of the conjunction *and* gives rise to a notable prevalence of the additive relation, which accounts for over 70% of the total relations. It should be noted, however, that there are also numerous clause complexes of the adversative relation. These findings indicate that when students extend their clauses, they either add positive information to the preceding clause or provide opposing to them. These two principles represent the foundation of clause expansion. While various types of conjunctions and phrases can indicate variation and alternation, students often avoid expressing the meaning of these changes in academic writing, frequently replacing them with simpler connector like *and*, or the connector *but*. This section examines the limited lexical and grammatical choices that students make in English academic writing with regard to extending clause relationships and related lexical realization.

5.6 Enhancement (Quantitatively Comparison)

Likewise, in quantitatively comparing the enhancement clause complex relations in Master theses and doctoral dissertations, two directions are feasible: one is intersected with taxis that paratactic or hypotactic enhancements can take diverse proportions in all clause nexuses calculated in the corpus, parataxis or hypotaxis respectively, paratactic or hypotactic expansions, and enhancements; the other is divided into categories that the dependent clause functions as an environment of the primary one from either temporal, spatial, manner, causal or conditional perspective. Though the latter can be more delicately discussed from taxis, the data of some kinds are not representative enough because of their small scales, and the comparative results may not be as convincing as predicted, research are thus conducted qualitatively in the next section without separating them apart into parataxis or hypotaxis.

5.6.1 General Distribution of Enhancement (Enhancement/Clause Nexuses)

The general information (descriptive statistics) of annotated enhancement clause complexes found in the corpus and the proportion of enhancement in expansion structures in two groups have been compared in Section 5.1.4.3 that doctoral students prefer enhancement in expansions than Master students, then how is the condition of the proportion of enhancement in the whole clause complex system? It seems that the same conclusion can be received from Table 5.6.1-1 (see Appendix 5.6) as the mean of group D is around 37% and of group M is 34.77%. The proportion ranges from 26.3% to 49% in the former group and from 18.62% to 57.26% in the latter. Data of group D are platykurtic, normally distributed with almost no skewness as the value of kurtosis is 0.376, the skewness value is 0.167, the division value of skewness to the standard error of skewness is 0.403 (within -2 to +2), the division value of kurtosis to the standard error of kurtosis is 0.465 (also within -2 to +2), the Shapiro-Wilk value is 0.991 (near to 1), and the P-value of Shapiro-Wilk value is 0.993 (also near to 1); while of group M are leptokurtic, abnormally distributed with a slightly right-tail skew as the value of the kurtosis is 1.583 (larger than 1), the skewness value is 0.457, the division value of skewness to the standard error of skewness is 1.70, the division value of kurtosis to the standard error of kurtosis is 2.98 (larger than +2), the Shapiro-Wilk value is 0.968 while the P-value of Shapiro-Wilk is only 0.04 (smaller than 0.05). Figure 115 shows that except for one particular case with high proportion and one with low enhancement in clause nexuses, the distribution curve of data in group M is bimodal, that most numbers concentrate on 30% to 40% and then 25% to 30%.

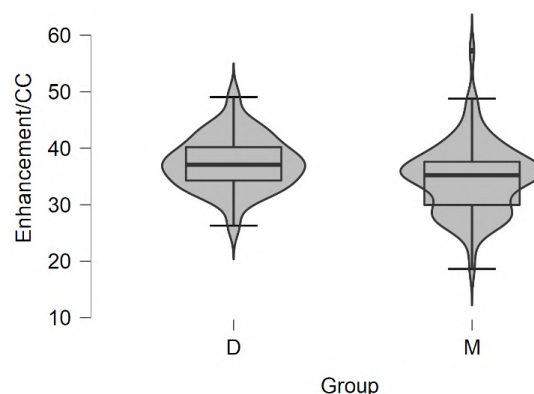


Figure 115 The boxplots comparing the proportion of enhancement in clause nexuses in Master and PhD sub-corpora

The test for comparing abnormally distributed data is then conducted by the Mann-Whitney U test when the alternative hypothesis is set as group D having a larger proportion than group M. The Hodges-Lehmann Estimate value is 2.549 (See Table 72), indicating that the difference between the two groups indeed exists. As the P-value is only 0.013 (smaller than 0.05), the original hypothesis is thus totally rejected. This means that the alternative one is accepted with a probability of 6.673 times as indicated by the VS-MPR value, i.e., group D is tested to have a larger proportion than group M, though this statement has a small effect size as noted by the Rank-Biserial Correlation value 0.272.

Independent Samples T-Test

	W	df	p	VS-MPR*	Hodges-Lehmann Estimate	Rank-Biserial Correlation
Enhancement/CC	1628.000	0.013	6.673	2.549	0.272	

Note. For the Mann-Whitney test, effect size is given by the rank biserial correlation.

Note. For all tests, the alternative hypothesis specifies that group D is greater than group M.

Note. Mann-Whitney U test.

* Vovk-Sellke Maximum p -Ratio: Based on a two-sided p -value, the maximum possible odds in favor of H_1 over H_0 equals $1/(-e p \log(p))$ for $p \leq .37$ (Sellke, Bayarri, & Berger, 2001).

Table 72 The Welch's t-test for comparing the proportion of enhancement in clause nexuses in Master and PhD sub-corpora

The same conclusion is suggested by the Bayesian Mann-Whitney U test result when the alternative hypothesis specifies that the location of group D is smaller than that of group M, the probability of accepting the original one (group D > group M) is 11.59 times rejecting it as seen from the Bayesian factor value (See Figure 116) with the Rhat value 1.122 and a small effect size.

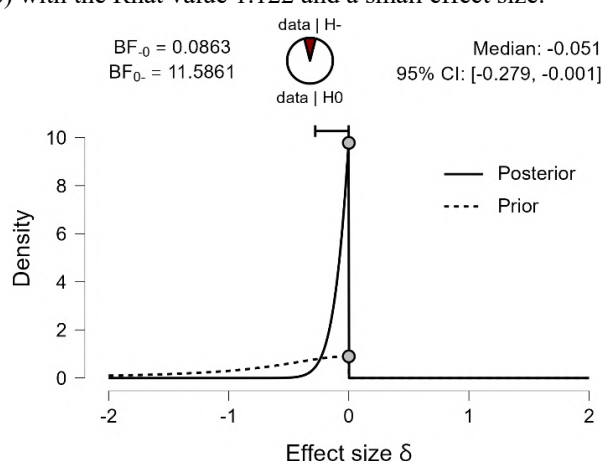


Figure 116 The inferential plots comparing the proportion of extension in clause nexuses in Master and PhD sub-corpora (with the original hypothesis group D > group M)

5.6.2 The Distribution of Paratactic Enhancement

For the number of paratactic enhancement clause complexes, 816 are discovered altogether, ranging from 6 to 50 in the doctoral candidates' dissertations and 1107 with a scope from 1 to 60 in Master students' texts. (See Table 5.6.2-1 in Appendix 5.6) The mean number of the former group is 25.5, and of the latter is 13.84, the division of these two values is smaller than 2.5, which denotes that the average proportion of paratactic enhancements in group D might not be larger than that in group M. The data seems to be platykurtic, normally distributed with a slightly right-tail skew of group D judged by the value of the kurtosis 0.816, that of the skewness 0.515, the value of the skewness divided by the standard error of skewness 1.24 (within $-2/+2$), that of the kurtosis divided by the standard error of kurtosis 1.01 (also within $-2/+2$), the Shapiro-Wilk value 0.969 and the P-value of Shapiro-Wilk 0.485; while is leptokurtic, abnormally distributed with a long right-sided tail of group M as shown by the kurtosis value 5.221, the skewness value 1.916, the division value of the skewness by the standard error of skewness 7.12 (larger than $+2$), the division value of the kurtosis by the standard error of kurtosis 9.81 (also larger than $+2$), the Shapiro-Wilk value 0.83 (not very close to 1), and the P-value of Shapiro-Wilk smaller than 0.001. Figure 117 shows the distribution more clearly that, except for two particular cases, most data rest within 5 to 25 in group D, and group M has three peculiar cases with high numerical values that its data distribution curve is even bimodal.

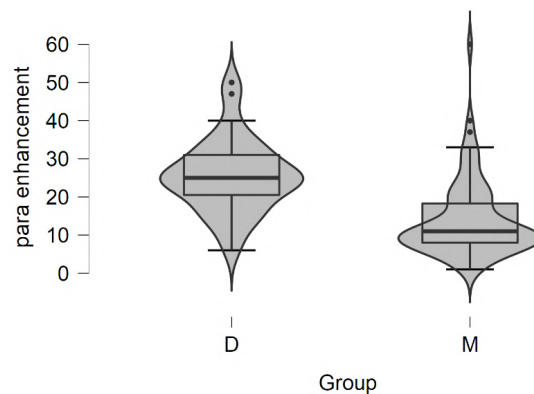


Figure 117 The boxplots of the number of paratactic enhancements in Master and PhD sub-corpora

5.6.2.1 Enhancement Parataxis/Clause Nexuses

The proportion of paratactic enhancement in the whole clause complex is quite small as the average in the doctoral group is only 3.54% and ranges from 1.39% to 6.75%, and in the Master's group is 4.23% with a range from 0.37% to 12.44%. (See Table 5.6.2.1-1 in Appendix 5.6) Data in group D are a little bit right-sided skew, platykurtic, and normally distributed when the skewness value is 0.146, the kurtosis value is -0.131, the value of skewness divided by the standard error of skewness is 0.35, the value of kurtosis divided by the standard error of kurtosis is -0.16, the Shapiro-Wilk value is 0.958, and the P-value of Shapiro-Wilk is 0.244; in group M are leptokurtic and abnormally distributed with a long right-sided tail as the skewness value is 1.067, the kurtosis value is 1.898, the division value of skewness by the standard error of skewness is 3.97 (larger than +2), the value of the kurtosis divided by the standard error of kurtosis is 3.57 (also larger than +2), the Shapiro-Wilk value is 0.941 and the P-value of Shapiro-Wilk is however 0.001 (smaller than 0.05). Actually, data of group D are bimodally distributed, with most of which located within 3% - 6%, and of group M are normally distributed except for three exceptions with high proportions as indicated by Figure 118 below.

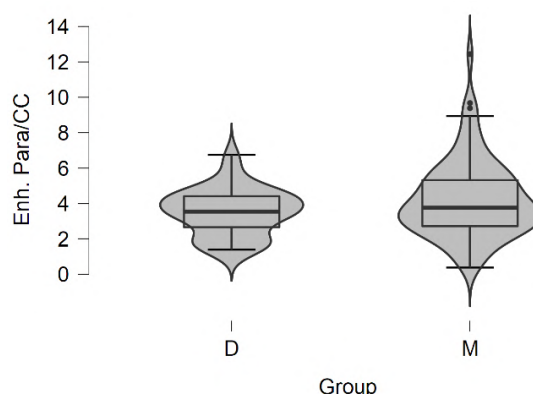


Figure 118 The boxplots comparing the proportion of paratactic enhancement in clause nexuses in Master and PhD sub-corpora

Independent Samples T-Test

Independent Samples T-Test						
	W	df	p	VS-MPR*	Hodges-Lehmann Estimate	Rank-Biserial Correlation
Enh. Para/CC	1079.000	0.098	1.613	-0.467		-0.157

Note. For the Mann-Whitney test, effect size is given by the rank biserial correlation.

Note. For all tests, the alternative hypothesis specifies that group D is less than group M.

Note. Mann-Whitney U test.

* Vovk-Sellke Maximum p -Ratio: Based on a two-sided p -value, the maximum possible odds in favor of H_1 over H_0 equals $1/(-e p \log(p))$ for $p \leq .37$ (Sellke, Bayarri, & Berger, 2001).

Table 73 The Mann-Whitney test for comparing the proportion of paratactic enhancement in clause nexuses in Master and PhD sub-corpora

With abnormally distributed data and diverse variances between the two groups, the Mann-Whitney test is thus conducted for comparison. When the original hypothesis is set as that group D is larger than group

M, the p-value is 0.098, and the VS-MPR value is 1.613 (see Table 73), the original hypothesis is therefore rejected. The probability of accepting the alternative hypothesis is 1.613 times that of accepting the original one, i.e., group D has a smaller average proportion compared with group M, and this kind of difference is 0.467, as suggested by the Hodges-Lehmann Estimate value. The test, however, has a small effect size, i.e., 0.157, as shown by the Rank-biserial correlation value.

This result is confirmed by the Bayesian Mann-Whitney U test when the alternative hypothesis specifies that the location of group D is greater than that of group M, the possibility of accepting the original hypothesis is 9.049 times rejecting it with a small to medium effect size, and the Rhat value is 1.046 as indicated by Figure 119.

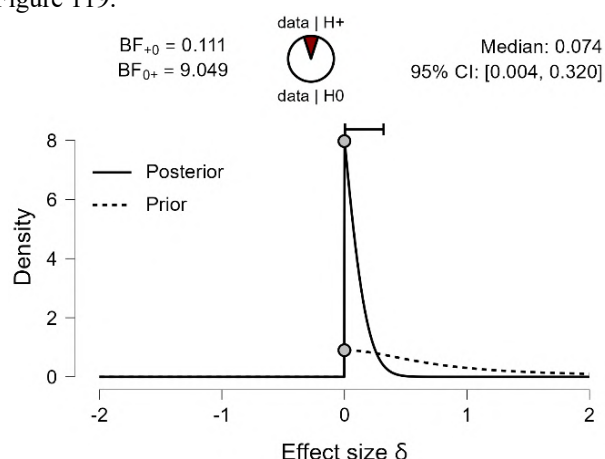


Figure 119 The inferential plots comparing the proportion of paratactic enhancement in clause nexuses in Master and PhD sub-corpora (with the original hypothesis group $D < \text{group M}$)

So, the proportion of paratactic enhancements in the whole clause nexuses in the doctoral group is smaller than that in the Master's group, though this difference is not obvious as not so many structures of this kind come up in the annotated corpus.

5.6.2.2 Enhancement Parataxis/Parataxis

Except for investigating the proportion of paratactic enhancements in all clause nexuses, how much they account for in all paratactic clause complexes also represents the weight and is of great importance. In Table 5.6.2.2-1 (see Appendix 5.6), this parameter in group D takes up around 10.65% and in group M 10.82%, which does not have big diversity. The range of the former group is from 4.13% to 25.88%, and of the latter group is from 1.22% to 24.76%. Data in group D are leptokurtic, abnormally distributed with a long right-sided tail as the skewness value is 1.132, the kurtosis value is 1.661, the division of skewness to the standard error of skewness is 2.72 (larger than +2), the division of kurtosis to the standard error of kurtosis is 2.05 (also larger than +2), the Shapiro-Wilk value is 0.897 and the P-value of Shapiro-Wilk is only 0.005; in group M are platykurtic, also abnormally distributed with a short right-sided tail as its skewness value is 0.565, the kurtosis value is 0.458, the two divisions are 2.10 (larger than +2) and 0.86 respectively, the Shapiro-Wilk value is 0.975 and the p-value of Shapiro-Wilk is 0.111. Actually, the distribution of data in the doctoral group is in line with that of the proportion of paratactic enhancement in clause nexuses that they are bimodal with two peculiarities of high proportions, and data distribution in the Master's group can be normal except for two cases with high proportions. (See Figure 120)

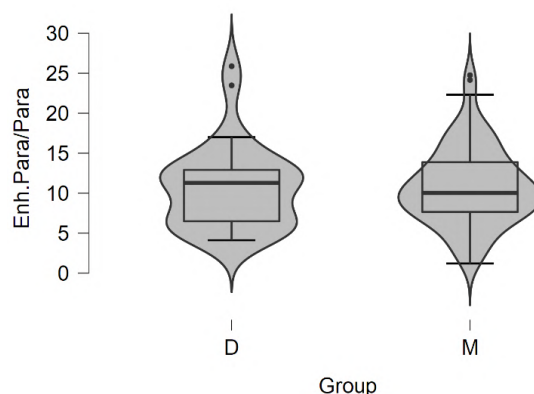


Figure 120 The boxplots comparing the paratactic enhancement in parataxis in Master and PhD sub-corpora

Again, the Mann-Whitney U-test is conducted for abnormally distributed data with diverse variances. Results in Table 74 show that the alternative hypothesis, i.e., group D has a smaller proportion than group M on this parameter, seems to be accepted as the P-value is 0.359 and the VS-MPR is exactly 1. The Hodges-Lehmann Estimate is smaller than 0 with the value -0.356, indicating that group D indeed has a smaller average proportion than group M. But this test has a small effect size as the Rank-biserial correlation value is only -0.044.

Independent Samples T-Test

	W	df	p	VS-MPR*	Hodges-Lehmann Estimate	Rank-Biserial Correlation
Enh.Para/Para	1223.500	0.359	1.000	-0.356	-0.356	-0.044

Note. For the Mann-Whitney test, effect size is given by the rank biserial correlation.

Note. For all tests, the alternative hypothesis specifies that group D is less than group M.

Note. Mann-Whitney U test.

* Vovk-Sellke Maximum p -Ratio: Based on a two-sided p -value, the maximum possible odds in favor of H_1 over H_0 equals $1/(-e p \log(p))$ for $p \leq .37$ (Sellke, Bayarri, & Berger, 2001).

Table 74 The Mann-Whitney U test for comparing the paratactic enhancement in parataxis in Master and PhD sub-corpora

To make the comparison clearer, the Bayesian Mann-Whitney U-test is then performed that when the alternative hypothesis specifies that the location of group D is greater than that of group M, the possibility of accepting the original hypothesis is 5.11 times (See Figure 121, left 1) of accepting the alternative one; when the alternative hypothesis is that group D has a smaller proportion than group M, the possibility of accepting the original hypothesis is 3.33 times (ibid, right 1) of rejecting it; and when the original hypothesis is that group D has the same proportion compared with group M, the probability of accepting the original hypothesis hereby is 4.03 times (ibid, left 2) of accepting the alternative one. The test has a small to medium effect size with the Rhat value of 1.154. Hence, the conclusion here tends to be that the doctoral group has a slightly smaller proportion of paratactic enhancement in parataxis than the Master's group from this corpus, though this diversity is not as obvious as other parameters.

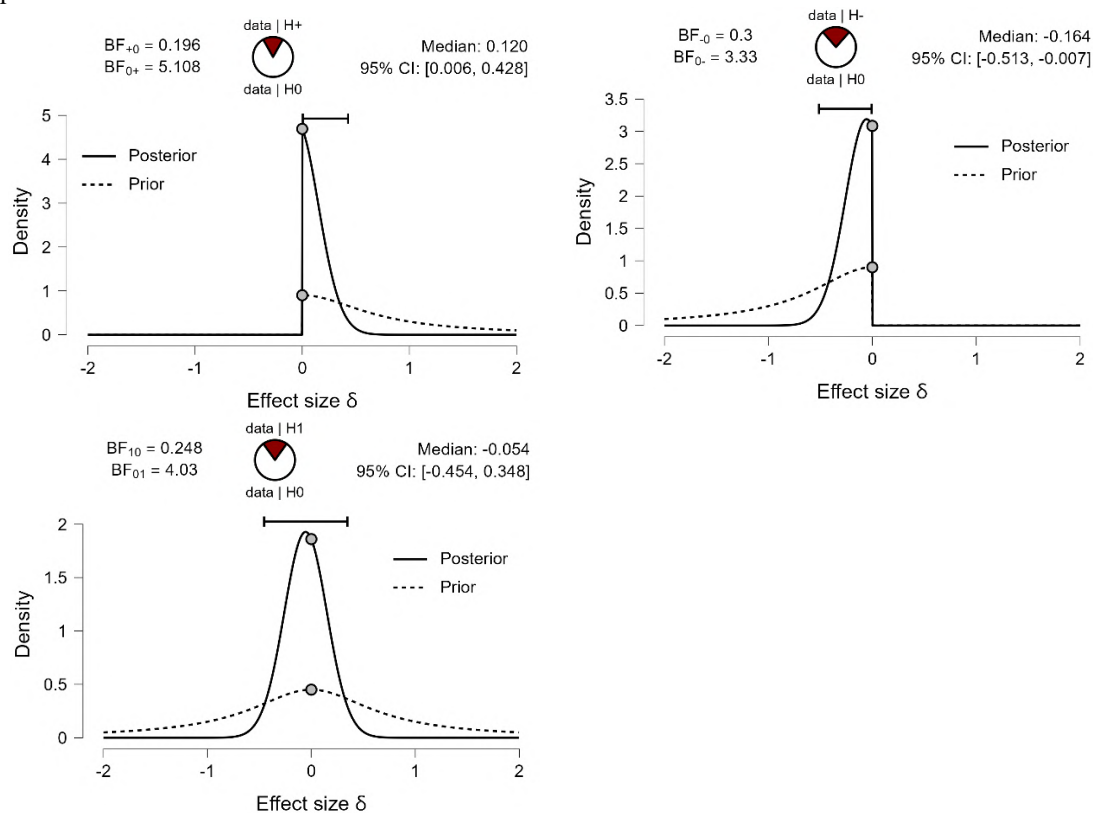


Figure 121 The inferential plots comparing the paratactic enhancement in parataxis in Master and PhD sub-corpora (with diverse original hypotheses)

5.6.2.3 Enhancement Parataxis/Expansion Parataxis

One point can be deductive here that in all paratactic expansions, the proportion of paratactic enhancement of the doctoral group must be smaller than that of the Master group as two other parameters, i.e., paratactic elaboration and paratactic extension of group D, have been proved to account for slightly higher proportions than that of group M. (See Sections 5.2.2.3 & 5.4.2.3) But how exactly different they are? The average proportion of paratactic enhancement in paratactic expansion in group D is around 10.90%, with a range from 4.28% to 25.88%, and in group M is 11.10%, ranging from 1.22% to 27.10%, as indicated by Table 5.6.2.3-1 (see Appendix 5.6). From the skewness value 1.047, the kurtosis value 1.504, the value of skewness divided by the standard error of skewness 2.53 (larger than +2), the value of kurtosis divided by the standard error of kurtosis 1.86, the Shapiro-Wilk value 0.894, and the P-value of Shapiro-Wilk 0.004, it can be stated that data in group D are leptokurtic, abnormally distributed with a right-sided skew; and from the skewness value 0.699, the kurtosis value 0.936, two division values 2.60 (also larger than +2) and 1.76 separately, the Shapiro-Wilk value 0.967 and the P-value of Shapiro-Wilk 0.038, data in group M are also considered to be leptokurtic, abnormally distributed with a short right-sided tail. In fact, data in group D have two exceptions with high proportions and are bimodally distributed, most of which concentrate in 5% to 15%, and in group M have three peculiarities with high proportions. (See Figure 122)

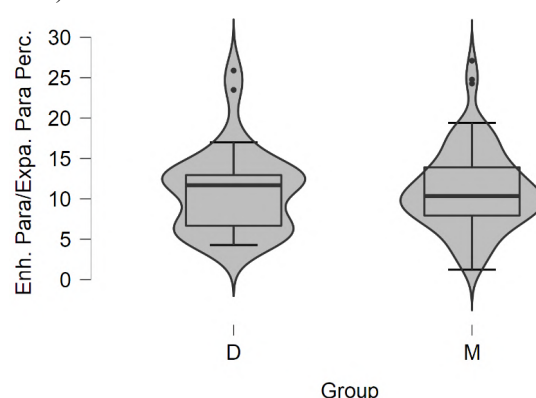


Figure 122 The boxplots comparing the proportion of paratactic enhancement in paratactic expansions in Master and PhD sub-corpora

For abnormally distributed data with diverse variances, the Mann-Whitney U test (See Table 75) is first conducted. When the alternative hypothesis is that group D has a lower proportion than group M, the P-value is 0.398 (not very far away from 0.37) with the VS-MPR value being 1, the alternative hypothesis is reluctantly accepted, and the diversity within two groups is only 0.35% as noted by the Hodges-Lehmann Estimate value. As this test has an extremely small effect size, i.e., 0.032, as indicated by the Rank-biserial correlation, further tests are therefore necessary.

Independent Samples T-Test

	W	df	p	VS-MPR*	Hodges-Lehmann Estimate	Rank-Biserial Correlation
Enh. Para/Expa. Para Perc.	1239.500	0.398	1.000	-0.345	-0.032	

Note. For the Mann-Whitney test, effect size is given by the rank biserial correlation.

Note. For all tests, the alternative hypothesis specifies that group D is less than group M.

Note. Mann-Whitney U test.

* Vovk-Sellke Maximum p -Ratio: Based on a two-sided p -value, the maximum possible odds in favor of H_1 over H_0 equals $1/(-e p \log(p))$ for $p \leq .37$ (Sellke, Bayarri, & Berger, 2001).

Table 75 The Mann-Whitney U test for comparing the proportion of paratactic enhancement in paratactic expansions in Master and PhD sub-corpora

In Bayesian Mann-Whitney U test, when the alternative hypothesis specifies that the location of group D is greater than the location of group M, the probability of accepting the original hypothesis is 5.59 times (Figure 123, left 1) of accepting the alternative one with small to medium effect size and the Rhat value is 1.069; when the alternative hypothesis is that group D has smaller proportion than group M, the possibility of accepting the original one is 3.86 times (ibid, right 1) of rejecting it; and when the alternative hypothesis is set as that two groups have the same proportion, the probability of accepting the original one is 4.57 times (ibid, left 2) of accepting the alternative one. By comparing the numerical values of these possibilities, the first statement is more likely to be accepted, i.e., the doctoral group has

a lower proportion than the Master group on the parameter of paratactic enhancement in paratactic expansions, though diversities in the two groups are not obvious.

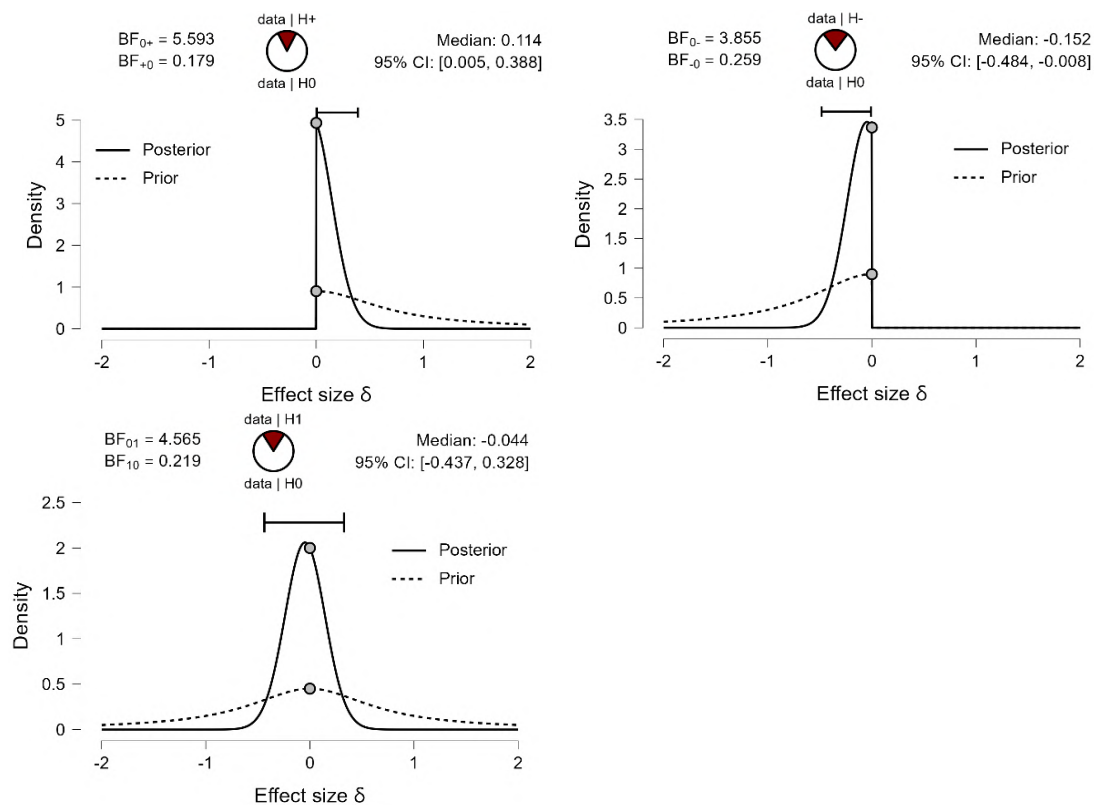


Figure 123 The inferential plots comparing the proportion of paratactic enhancement in paratactic expansions in Master and PhD sub-corpora

5.6.2.4 Enhancement Parataxis/Enhancement

The fourth dimension for paratactic enhancement comparison is its proportion in enhancement clause relations, and the result reflects the condition of paratactic versus hypotactic enhancements. In general, the average proportion of this parameter in the doctoral group is 9.53%, with a range from 3.26% to 17.30%, and in the Master group, 12.37%, with a range from 1.12% to 33.33%. (See Table 5.6.2.4-1 in Appendix 5.6)

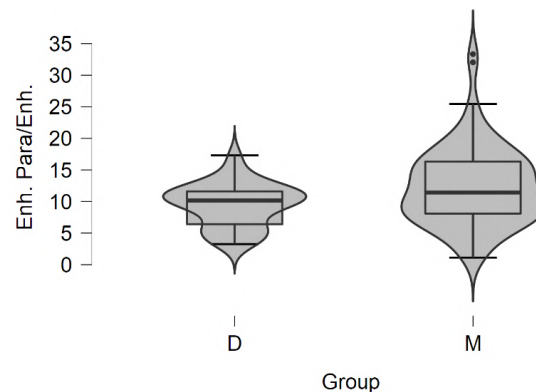


Figure 124 The boxplots comparing the proportion of paratactic enhancement in enhancements in Master and PhD sub-corpora

Data in the doctoral group seem to be platykurtic, normally distributed with slightly left-sided skew as indicated by its skewness value of -0.137, its kurtosis value of -0.311, the value of skewness divided by the standard error of skewness -0.33 (within -2/+2), the value of kurtosis divided by the standard error of kurtosis -0.38 (also within -2/+2), the Shapiro-Wilk value 0.958, and the P-value of Shapiro-Wilk 0.24; and in Master group can be leptokurtic, abnormally distributed with a right-sided skew because its skewness value is 0.875, its kurtosis value is 1.515, the value of two divisions are 3.25 (larger than +2) and 2.85 (also larger than +2) separately, the Shapiro-Wilk value is 0.954, and the P-value of Shapiro-Wilk is only 0.006 (smaller than 0.05). It is more obvious that data in group D are bimodally

distributed, with most of which located within 5% to 15%, and group M has two exceptions with high proportions (larger than 30%) when the boxplots of two groups are referred to. (See Figure 124)

Again, the Mann-Whitney U-test is performed to compare abnormally distributed data in two groups with different variances. When the alternative hypothesis is that group D has an average lower proportion than group M, the p-value after the test is only 0.016, and the VS-MPR value is 5.53 (See Table 76), which fully confirms the statement of the alternative hypothesis that the probability of accepting it is 5.53 times of accepting the original one. The Hodges-Lehmann Estimate value of -2.457 indicates that the diversity in the two groups is 2.457%. This test has a small to medium effect size, as the rank-biserial correlation value is 0.26.

Independent Samples T-Test

	W	df	p	VS-MPR*	Hodges-Lehmann Estimate	Rank-Biserial Correlation
Enh. Para/Enh.	947.000		0.016	5.530	-2.457	-0.260

Note. For the Mann-Whitney test, effect size is given by the rank biserial correlation.

Note. For all tests, the alternative hypothesis specifies that group D is less than group M.

Note. Mann-Whitney U test.

* Vovk-Sellke Maximum p -Ratio: Based on a two-sided p -value, the maximum possible odds in favor of H_1 over H_0 equals $1/(-e p \log(p))$ for $p \leq .37$ (Sellke, Bayarri, & Berger, 2001).

Table 76 The Mann-Whitney U test for comparing the proportion of paratactic enhancement in enhancements in Master and PhD sub-corpora

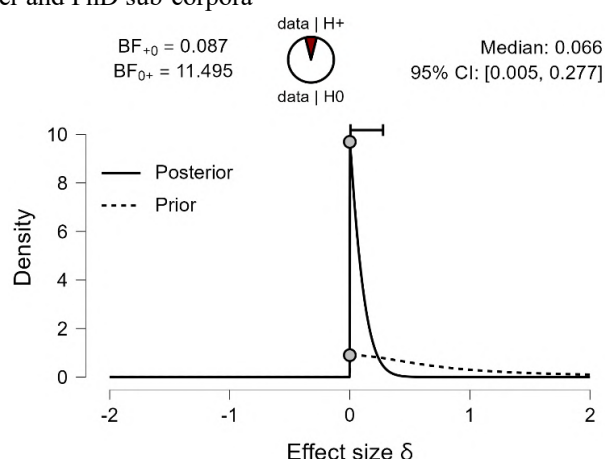


Figure 125 The inferential plots comparing the proportion of paratactic enhancement in enhancements in Master and PhD sub-corpora (with the original hypothesis group D < group M)

To verify the testing results above, the Bayesian Mann-Whitney U test is also referred to that the possibility of accepting the original hypothesis is 11.50 times (Figure 125) of accepting the alternative one when the latter specifies that the location of group D is greater than the location of group M, with the test effect size being small and the Rhat value being 1.063. Hence, the doctoral group has a lower average proportion of paratactic enhancement enhancements than the Master group in the corpus.

5.6.3 The Distribution of Hypotactic Enhancement

The number of hypotactic enhancements can be as high as 505 in some doctoral texts and can be as low as 50 in some Master texts in the corpus, with the mean number in group D being around 256 and in group M being 97 (the division of two mean values are 2.64, which is larger than 2.5, i.e., the multiple of the number of clause nexuses in two groups, the average proportion of doctoral group can hereby be assumed to be higher than that of Master group.) (See Table 5.6.3-1 in Appendix 5.6) Data in group D seem to be platykurtic, normally distributed with a short right-sided tail seeing from its skewness value of 0.64, its kurtosis value of 0.554, the value of skewness divided by the standard error of skewness of 1.55 (within $-2/+2$), the value of kurtosis divided by the standard error of kurtosis 0.68 (also within $-2/+2$), the Shapiro-Wilk value 0.965, and the P-value of Shapiro Wilk 0.373; while in group M are leptokurtic, abnormally distributed with a long right-sided skew when its skewness value is 1.553, its kurtosis value is 3.34, values of two divisions are 5.77 and 6.28 respectively, the Shapiro-Wilk value is 0.882, and the P-value of Shapiro-Wilk is smaller than 0.001. From Figure 126, it is interesting that the data distribution of the two groups is analogous to those of the hypotactic extension numbers that,

except for one case in group D and four cases in group M with high proportions, they can be normally distributed.

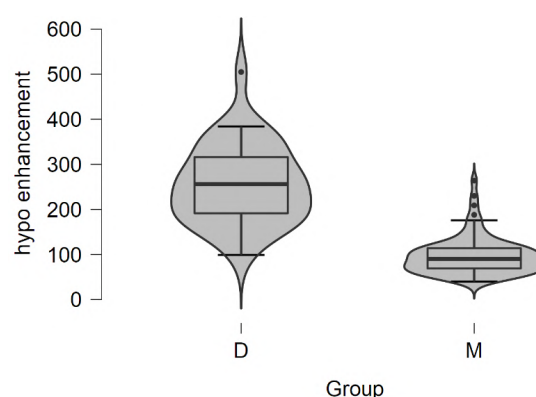


Figure 126 The boxplots comparing the number of hypotactic enhancements in Master and PhD sub-corpora

5.6.3.1 Enhancement Hypotaxis/Clause Nexuses

The comparison results of paratactic enhancement in clause nexuses are not predictable for the condition of hypotactic enhancement in clause nexuses, as enhancement is not the only category of clause nexus. In Table 5.6.3.1-1 (see Appendix 5.6), around 23.24% to 45.87%, with an average of 33.83% of clauses nexuses are hypotactic enhancements scattered in doctoral texts; and 16.21% to 53.99% with an average 30.53% are in Master's texts. Data in group D are platykurtic, normally distributed with slightly right-sided skew as its skewness value is only 0.202, its kurtosis value is 0.535, the value of skewness divided by the standard error of skewness is 0.49, the value of kurtosis divided by the standard error of kurtosis is 0.66, the Shapiro-Wilk value is 0.986, and the P-value of Shapiro-Wilk is 0.938; but in group M are leptokurtic, abnormally distributed with a short right-sided tail when its skewness value is 0.565, its kurtosis value is 1.746, values of two divisions are 2.1 (larger than +2) and 3.28 (also larger than +2) separately, the Shapiro-Wilk value is 0.971, and the P-value of Shapiro-Wilk is only 0.069. More detailed information can be received from the boxplots in Figure 127 that data in the former group are bimodal with one case with high proportion and in the latter concentrate in 26% to 34% with one peculiarity which exceeds 50%.

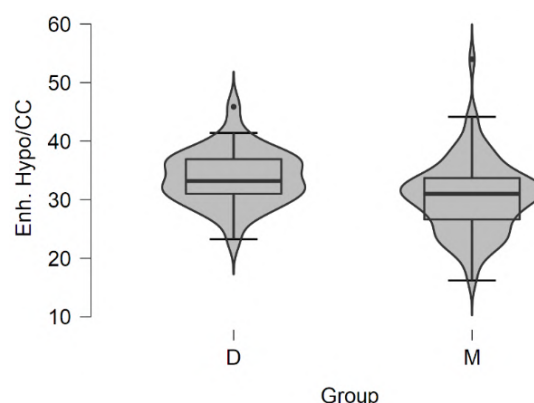


Figure 127 The boxplots comparing the proportion of hypotactic enhancement in clause nexuses in Master and PhD sub-corpora

Independent Samples T-Test

	W	df	p	VS-MPR*	Hodges-Lehmann Estimate	Rank-Biserial Correlation
Enh. Hypo/CC	1731.000		0.002	31.505	3.432	0.352

Note. For the Mann-Whitney test, effect size is given by the rank biserial correlation.

Note. For all tests, the alternative hypothesis specifies that group D is greater than group M.

Note. Mann-Whitney U test.

* Vovk-Sellke Maximum p -Ratio: Based on a two-sided p -value, the maximum possible odds in favor of H_1 over H_0 equals $1/(-e p \log(p))$ for $p \leq .37$ (Sellke, Bayarri, & Berger, 2001).

Table 77 The Mann-Whitney U test for comparing the proportion of hypotactic enhancement in clause nexuses in Master and PhD sub-corpora

For comparing abnormally distributed data with different variances, the Mann-Whitney U test is carried out. When the alternative hypothesis is that group D has a lower average proportion than group M, the p-value is only 0.002 after the test with the VS-MPR value 31.505 (See Table 77), which indicates that the odds in favor of the alternative hypothesis over the original one is 31 times. The Hodges-Lehmann Estimate value of 3.432 suggests that the differentiation between the two groups is just 3.432%. This test has a small to medium effect size, as noted by the Rank-biserial correlation value of 0.352.

This result can be approved by the Bayesian Mann-Whitney U test that when the alternative hypothesis specifies that the location of group D is smaller than the location of group M, the Bayes factor is tested to be 12.37 (See Figure 128), which denotes that the possibility of accepting the original hypothesis is 12.37 times of accepting the alternative one. This test has a small effect size, though its Rhat value is 1.037. Diversity, hence, indeed exists, and the doctoral group is confirmed to choose more hypotactic enhancements in clause nexuses than the Master's group in this corpus.

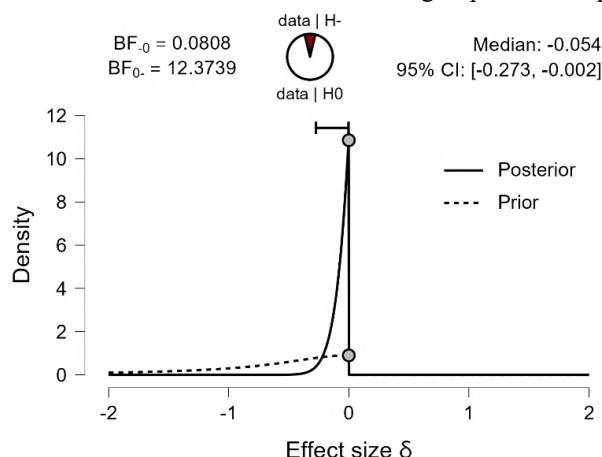


Figure 128 The inferential plots comparing the proportion of hypotactic enhancement in clause nexuses in Master and PhD sub-corpora (with the original hypothesis group D > group M)

5.6.3.2 Enhancement Hypotaxis/Hypotaxis

The proportion of hypotactic enhancement in clause nexuses above cannot provide any indication for the proportion of hypotactic enhancement in hypotaxis. In some texts written by Master students, this parameter can reach up to 78.31% or be only 29.20%, with an average proportion of 50.40%; and in some doctoral texts, it can be from 42.58% to 63.67% with a mean of 51.97%. (See Table 5.6.3.2-1 in Appendix 5.6) It is apparent that data in group D are platykurtic, normally distributed with slightly right-sided skew when the skewness value is 0.298, the kurtosis value is -0.176, the value of skewness divided by the standard error of skewness is 0.72 (within -2/+2), the kurtosis divided by the standard error of kurtosis is -0.22 (also within -2/+2), the Shapiro-Wilk value is 0.976, and the P-value of Shapiro-Wilk is 0.678; in group M are platykurtic (the kurtosis value is 0.407), right-sided skew (the skewness value is 0.516), and abnormally distributed (values of two divisions are 1.92 and 0.77 separately, the Shapiro-Wilk value is 0.977 and the P-value of Shapiro-Wilk is only 0.166). From the boxplots in Figure 129, the data distribution of this parameter in group M has two exceptions and, in both groups, tends towards being bimodal, with most of the data concentrated in 40% to 55%.

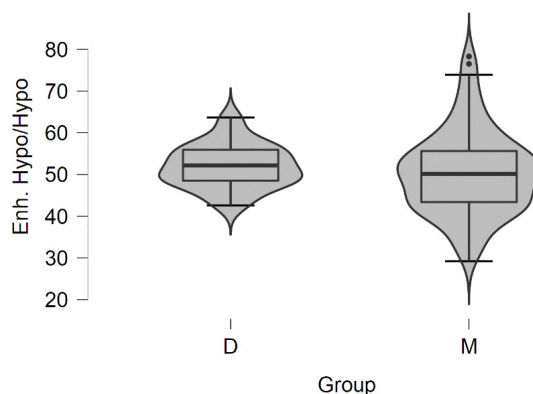


Figure 129 The boxplots comparing the proportion of hypotactic enhancement in hypotaxis in Master and PhD sub-corpora

After performing the Mann-Whitney U test, the p -value from Table 78 is just 0.102 when the alternative hypothesis is that group D has an average bigger proportion than group M. The alternative hypothesis, therefore, is accepted, and the possible odds in favor of it over the original one is 1.578 times as indicated by the VS-MPR value. The Hodges-Lehmann estimate means that the diversity within the two groups is 2.06%. Though this test has a small effect size, as noted by the rank-biserial correlation value 0.154, it provides a reference for the alternative hypothesis of the Bayesian Mann-Whitney U test below.

Independent Samples T-Test

	W	df	p	VS-MPR*	Hodges-Lehmann Estimate	Rank-Biserial Correlation
Enh. Hypo/Hypo	1477.500	0.102	1.578	2.060	0.154	

Note. For the Mann-Whitney test, effect size is given by the rank biserial correlation.

Note. For all tests, the alternative hypothesis specifies that group *D* is greater than group *M*.

Note. Mann-Whitney U test.

* Vovk-Sellke Maximum p -Ratio: Based on a two-sided p -value, the maximum possible odds in favor of H_1 over H_0 equals $1/(-e p \log(p))$ for $p \leq .37$ (Sellke, Bayarri, & Berger, 2001).

Table 78 The Mann-Whitney U test for comparing the proportion of hypotactic enhancement in hypotaxis in Master and PhD sub-corpora

From Figure 130, the Bayes factor value is 8.85 when the alternative hypothesis specifies that the location of group D is smaller than that of group M hereby, the possibility of accepting the original hypothesis is 8.85 times of accepting the alternative one, which indicates that doctoral students prefer to choose hypotactic enhancements in hypotactic clause nexuses than Master students from this corpus. The test effect size is small, but the Rhat value is 1.003, which means that the between- and within-chain variance of the MCMC estimates for the delta parameter is convergent, i.e., the difference between the two groups indeed exists.

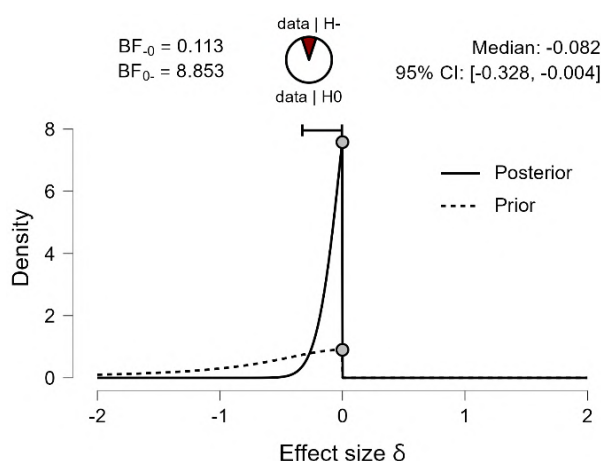


Figure 130 The inferential plots comparing the proportion of hypotactic enhancement in hypotaxis in Master and PhD sub-corpora (with the original hypothesis group $D >$ group M)

5.6.3.3 Enhancement Hypotaxis/Expansion Hypotaxis

Though it has been proved that in all hypotactic expansion clause relations, doctoral students prefer to choose less hypotactic elaborations (on average) while more hypotactic extensions (also on average) than Master students from this corpus in former sections (Sections 5.2.3.3 & 5.4.3.3), the condition of hypotactic enhancements still remains unclear. In doctoral texts, the proportion of this parameter can be 61.43% to 84.45% with an average of 71.95%; and in Master texts, it can be as low as 47.06% or as high as 96.73% with a mean of 70.71%. (See Table 5.6.3.3-1 in Appendix 5.6) Proportions in doctoral group are platykurtic, normally distributed with slightly right-sided skew as its skewness value is 0.276, its kurtosis value is -0.466, the value of skewness divided by the standard error of skewness is 0.67 (within $-2/+2$), the value of kurtosis divided by the standard error of kurtosis is -0.58 (also within $-2/+2$), the Shapiro-Wilk value is 0.969, and the P-value of Shapiro-Wilk is 0.471; and in Master group are also platykurtic, normally distributed with almost no skew as its skewness value is 0.204, its kurtosis value is only 0.033, values of two divisions are 0.76 and 0.06 respectively, the Shapiro-Wilk value is 0.991 (nearly to 1), and the P-value of Shapiro-Wilk is 0.832. Boxplots in Figure 131 proved this statement that except for one case with a high proportion in group M, data are normally distributed in both groups.

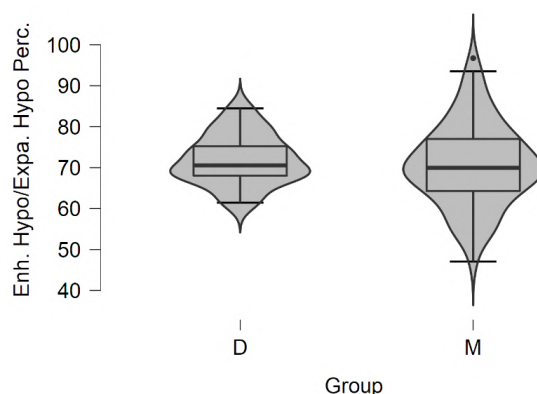


Figure 131 The boxplots comparing the proportion of hypotactic enhancement in hypotactic expansion in Master and PhD sub-corpora

For normally distributed data with diverse variances within two groups, the Welch's t-test is more suitable. When the alternative hypothesis is that group D has a larger average proportion than group M, the P-value is 0.215, and the VS-MPR value is only 1.114 (See Table 79), which means that the p-value is at most 1.114 times more likely to occur under the alternative hypothesis. The mean difference value of 1.236 and standard error difference value of 1.557 all suggest that the two groups have some diversity in their average proportions, i.e., 1.24% or 1.56%. The test, though, has a small effect size, as noted by Cohen's d value of 0.148, which at least gives hints to the original hypothesis of the Bayesian Independent Samples T-test.

Independent Samples T-Test

	t	df	p	VS-MPR*	Mean Difference	SE Difference	Cohen's d
Enh. Hypo/Expa. Hypo Perc.	0.794	92.846	0.215	1.114	1.236	1.557	0.148

Note. For all tests, the alternative hypothesis specifies that group D is greater than group M.

* Vovk-Sellke Maximum p -Ratio: Based on a two-sided p -value, the maximum possible odds in favor of H_1 over H_0 equals $1/(-e p \log(p))$ for $p \leq .37$ (Sellke, Bayarri, & Berger, 2001).

Note. Welch's t-test.

Table 79 The Welch's t-test for comparing the proportion of hypotactic enhancement in hypotactic expansion in Master and PhD sub-corpora

From the results of the Bayesian Mann-Whitney U test, the Bayes factor is 7.09 (see Figure 132) when the alternative hypothesis specifies that the location of group D is smaller than the location of group M, suggesting that the probability of accepting the original hypothesis is 7.09 times of accepting the alternative one, with a small to medium effect size and the Rhat value being 1.047. Thus, doctoral students are more likely to use hypotactic enhancements in hypotactic expansions than Master students.

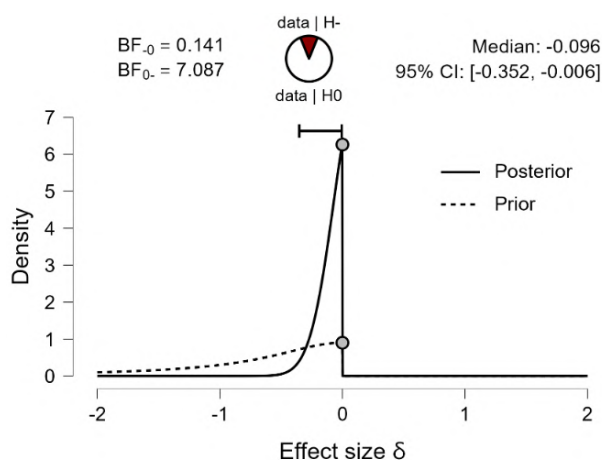


Figure 132 The inferential plots comparing the proportion of hypotactic enhancement in hypotactic expansion in Master and PhD sub-corpora (with the original hypothesis group D > group M)

5.6.3.4 Enhancement Hypotaxis/Enhancement

It can be inferred from Section 5.6.2.4 that the average proportion of hypotactic enhancement in enhancements of the doctoral group must be larger than that of the Master group as this is contrary to the paratactic ones. But how about the detailed descriptive statistics of both groups, and how the comparison results of the Bayesian test can still be interesting? From Table 5.6.3.4-1 (see Appendix 5.6), it is known that the average proportion of this parameter in the doctoral group is about 90.47% with a range from 82.70% to 96.74%; and in the Master group, is 87.63% with a range from 66.67% to as high as 98.88%. The data distribution conditions are of course the same as those in paratactic enhancement per enhancements only with adverse diagrams that it seems to be platykurtic, normally distributed with a slightly right-sided skew in group D when its skewness value is 0.137, its kurtosis value is -0.311, the division value of skewness by the standard error of skewness is 0.33, the value of kurtosis divided by the standard error of kurtosis is -0.38, the Shapiro-Wilk value is 0.958, and the P-value of Shapiro-Wilk is 0.24; and is leptokurtic, abnormally distributed with a left-sided skew in group M as its skewness value is -0.875, its kurtosis value is 1.515, values of two divisions are -3.25 (larger than +2) and 2.85 (also larger than +2) respectively, the Shapiro-Wilk value is 0.954, and the P-value of Shapiro-Wilk is only 0.006 (smaller than 0.05). The boxplots in Figure 133 show that data in the former group are bimodal and, in the latter, have two peculiarities with low proportions.

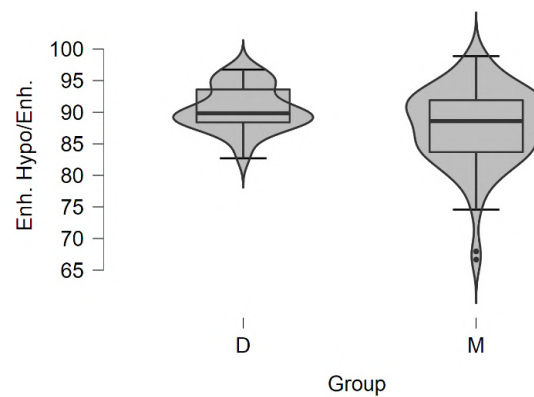


Figure 133 The boxplots comparing the proportion of hypotactic enhancement in enhancements in Master and PhD sub-corpora

The Mann-Whitney U test results of this parameter are the same as those of paratactic enhancement in enhancements, as Table 76 has shown, though the alternative hypothesis has changed into “group D has a smaller average proportion than group M.” But the Bayesian Mann-Whitney U test result is not quite the same as the Bayes factor value is 12.37, as shown by Figure 134, which indicates that the possibility of accepting the original hypothesis is 12.37 times accepting the alternative one when the latter is set as that the location of group D is smaller than that of group M, with a small effect size and the Rhat value being 1.032. Hence, doctoral students prefer hypotactic enhancement in enhancements to Master students in this corpus.

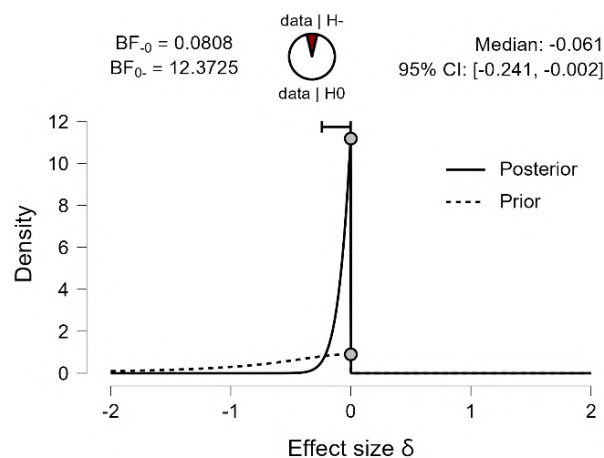


Figure 134 The inferential plots comparing the proportion of hypotactic enhancement in enhancements in Master and PhD sub-corpora

5.6.4 Summary of Enhancement Comparison

For enhancement structures in this corpus, it is unified that no matter how the denominator is for calculating and comparing changes, doctoral students have a higher tendency to choose hypotactic enhancements, while Master students prefer paratactic enhancements, and the preference for hypotactic ones in the former group leads to the higher average proportion of enhancements in clause nexuses, as indicated by the highest Bayes factor value of the proportion of hypotactic enhancement dividing clause nexuses. The proportions of paratactic enhancement in parataxis or paratactic expansions, though they do have obvious diversities within two groups from this corpus after tests, the abundance of data in more texts written by more students and the higher number of paratactic enhancements in each text may strengthen these testing differences.

5.7 Enhancement (Qualitatively Analysis)

The quantitative method effectively illustrates the distinctions in the paratactic and hypotactic combination of enhancement clause complexes in master's and doctoral dissertations in the corpus, the differences of various enhancement categories within the system, and the diversities in taxical and functional intersection linguistic features. However, it is currently unclear which lexis specifically reflects these differences and how this realization is carried out in the clause complexes of master's and doctoral dissertations. Therefore, this section employs a qualitative method to analyze the choice of relevant marking lexical items and the differences in their realization in specific clause complexes to better explain the lexicogrammatical strata of language.

5.7.1 General Information of Enhancement

Section 3.1.3.3 identifies five distinct strategies through which one clause may be enhanced by another: temporal, spatial, manner, causal, and conditional. Among the aforementioned categories, those about temporal enhancement encompass simultaneous (i.e., occurring at the same time), earlier, and later; manner enhancement is divided into manner or comparison; causal enhancement involves cause/result/purpose, that is, the content described by the dependent clause is the cause or result or purpose of the independent clause; conditional enhancement can be positive/negative and concessionary, that is, the condition described by the dependent clause has a certain polarity.

The intersection of paratactic/hypotactic structure and enhancement categories provides students with a more expansive range of options when composing enhancement clause complexes. In light of these considerations, it is pertinent to inquire as to the number of enhancement clause complexes that will be observed in each grammatical category, the specific lexical categories that will be employed in each grammatical category within two groups (i.e., master's and doctoral), the extent to which similarities or differences will be discerned in the context of practical thesis writing, the marking lexical features that doctoral students will utilize in realizing parataxis or hypotaxis of temporal/manner/causal/conditional clause complexes, and so forth in this section.

5.7.2 Temporal

Of the 1,841 temporal enhancing clause complexes selected by master's students, approximately two-thirds are *same time*, 26% are *later*, and only 6.25% are *earlier*. Among the 1,559 temporal enhancing clause complexes chosen by doctoral candidates, around three-quarters are *same time*, 17% are *later*, and the proportion of *earlier* clause complexes is marginally higher than that of master's students, at 7.57%. In general, master's students tend to utilize temporal enhancement clause complexes more frequently than doctoral candidates when describing phenomena. This preference may be attributed to a number of literary theses in the corpus available for analysis. Additionally, the inclination of master's students toward *later* CCs further substantiates their emphasis on the sequence of events.

5.7.2.1 Same Time

The conjunction *when*, *while*, *whenever*, (*and*) *meanwhile*, the preposition *in* with its related *in the process of*, *as* with relevant *as soon as* and *once* are among the most common words or phrases used to express the idea of *same time* in enhancing relationships between clauses. (see Table 80) The conjunction *when* is employed in the overwhelming majority of instances. However, doctoral students demonstrate a proclivity for utilizing *in* and *while* to a greater extent than their master's-level counterparts, although these two words also account for a certain proportion of their usage.

Enhancement	M	D
Temporal: Same time	1244	1175
as	18	30
as soon as	3	4
in	86	168
in the process of	18	9
(and) meanwhile	19	11
once	20	19
when	1004	802
whenever	4	10
while	72	122

Table 80 The number of each lexis for temporal: same time clause complexes applied in Master and PhD sub-corpora

E.g., 5.7.1||| **When*** dealing with affective side of language learners, || attention needs to be given both to how we can overcome problems created by negative emotions || and to how we can create and use more positive, facilitative emotions. || ---- CMAC07RE_19_353

E.g., 5.7.2||| Austin argued || that **when*** a person speaks, || he or she is performing three acts simultaneously. || ---- CPhD17_0002_1088

When the clause leading by the word *when* is put at the beginning of a clause complex, i.e., when the dependent clause is located before the independent one, the subjects of the latter clause are changeable in a confined and strengthened environment built in the former clause; and when the dependent clause is located after, i.e., at the end of the whole clause complex, the subjects in the independent clause now are unchangeable and thus reinforced by mentioning the abstract or specific time chosen in the dependent one. In the two examples above, the time is definitive, and actions or matters in the time are chosen by writers for statement, explanation, description, representation, etc.

E.g., 5.7.3||| **Whenever*** she doesn't comply with Murdstone's demands, || she will end up || in yielding to him || as long as he threatens to stop loving her. || ---- CMAC10LIT_7_280

E.g., 5.7.4||| For example, learners try to relax || **whenever*** they feel anxious about of using the target language, || and they will often talk to someone about [[how they feel || when they are learning the target language]]. || ---- CPhD10_0001_420

Though the word *whenever* functions almost the same as the word *when* in structures, it has more restrictions on the time things or matters happen semantically that *this time and the subjects* in the independent clauses do not like the dependent clause beginning with *when* that the time can be flexible, are closely binding, as shown in two examples above. *She will yield to Murdstone* not at this or that time, but at the exact time that *she doesn't comply with his demands* in E.g., 5.7.3; and the time that learners want to relax is certainly when *they feel anxious about using the target language* in E.g., 5.7.4. The word *when* in the embedded clause nexus cannot be changed into *whenever* as learners sometimes do not have any feel in learning the target language.

E.g., 5.7.5||| Some relative authorities do not use effective means to improve and promote the authority of other foreign language to social candidates || **while*** reducing the social impact of CET-4. || ---- CMAC13ME_11_374

E.g., 5.7.6||| **While*** designing the LPBL module, || teachers have to consider carefully students' learning characters and interests, with the purpose of promoting students' knowledge, skills, and dispositions. || ---- CPhD11_0007_899

The word *while* though is almost the same as *when* semantically in denoting temporal relationship and functionally in leading the dependent clause with either non-finite verbs or Subjects, it is not the first choice compared to *when* because of its connotative function for connecting an extension: addition: adversative relation that it would be chosen when the writer wants to convey both meanings. In the two examples above, *while* is adversative to some degree in expressing somewhat opposite ideas but is more temporal in a dependent clause either at the beginning or at the end of a clause complex followed by non-finite verbs *reducing* or *designing*.

E.g., 5.7.7||| Giles' translation realizes referential meaning transference with partial correspondence, || there is no denial of the fact [[that his application of generalization makes the rendition expressive and intelligible]], || **meanwhile*** the readers are informed of the original message appropriately. || ---- CMAC11PR_10_399

E.g., 5.7.8 ||| Fruitful findings indicate || that learners at the similar initial language proficiency level and the related skill level have, in general, improved all the skills, || **meanwhile*** their critical thinking disposition has been fostered, too. || --- CPhD11_0007_231

Meanwhile looks similar to the word *while* but connects two clauses with equal status, i.e., it is a maker for paratactic enhancement: temporal relation in a clause complex. Usually, it follows a comma and should also have a comma after it; sometimes, it follows a semi-colon with the mark *and* in front of it. The time, abstract or concrete, indicates the same point by this word that things or matters described in two clauses happen simultaneously. E.g., 5.7.7 describes Gile's techniques for translation and how quickly readers understand and absorb the message that Gile transfers; and in E.g., 5.7.8, the writer obviously depicts advantages of some methods that can improve learners' skills and at the same time *foster their critical thinking disposition*. Thus, *meanwhile* is not merely an indication of time but also a function for addition or supplementary.

E.g., 5.7.9 ||| **In*** addressing each one of these questions, || the purpose of this thesis is to yield insight into the growing body of discovering characteristics of evidentiality employed in the academic writing of scholars from the journal and of master graduates in the two domestic normal universities. || --- CMAC17_05_238

E.g., 5.7.10 ||| The theoretical frameworks of systemic-functional grammar (henceforth SFG) and cognitive grammar (henceforth CG) will be used here || **in*** analysing reporting and various perspectives in reporting. || --- CPhD06_0004_80

For signifying temporal relation in a clause complex, the preposition *in* gets close to the *on*, which both lead to non-finite dependent clauses, though the latter is rare in this corpus. *In* usually refers to a period, which extends from its spatial connotation with a scope. So, in E.g., 5.7.9, the time refers to the whole period of *addressing each question*, and in E.g., 5.7.10, the time points to the process of reporting analysis.

E.g., 5.7.11 ||| **In the process** [[of* doing literature survey]], the author of this thesis finds || that although critical essays on Wilde's works cover many respects, || the studies on death theme in Wilde's works are either too general, || lacking in convincing details and proof, or too narrow in discussion. || --- CMAC10LIT_21_99

E.g., 5.7.12 ||| The fact [[that students were constrained from asking the teachers [[who were unfamiliar to them]]]] highlights the role of the classroom teacher **in the process** [[of* fostering students' autonomy]]. || --- CPhD12_0009_1326

The phrase *in the process of* though derives from the word *in* in indicating the temporal relationship: the following non-finite clause, however, is embedded in the word *process* with a preposition *of* in between. Here, the structure *in the process of doing literature survey* in E.g., 5.7.11 and *in the process of fostering students' autonomy* in E.g., 5.7.12 are all circumstantial in one clause that they are not regarded as separated clauses.

E.g., 5.7.13 ||| Thus, **as*** the developments go further, || few people can even detect the relationships between the meanings now and its prototypical meaning. || --- CMAC11SE_19_729

E.g., 5.7.14 ||| **As*** we come to more fully understand the relationship between personality and the learning process in real world settings, <as regards practical problem solving,> || we may find ways to bring the impact of our basic tendencies and dispositions into a level of awareness [[that could help to facilitate a change in an individual's repertoire of characteristic adaptations]]. || --- CPhD10_0001_968

The word *as* for a temporal relationship can easily be mixed with a cause: reason relation in a clause complex that the boundary between them is sometimes indistinct and hence cannot be judged except for comparing clause complexes with their agnation to estimate whether they are closer to a temporal indication *when* or to reasonable markers such as *because*, *due to*, *considering (that)*, etc. In E.g., 5.7.13, *as* might be substituted by *when*, and in E.g., 5.7.14, it is similar to the word *while* for both emphasizing the importance of the point either overt or metaphorical that things described by the Subject of the main clause(s) occur. But still, *as* instead of *when* or *while* is used, which largely relates to its connotation of indicating reasons.

E.g., 5.7.15 ||| Then CF would be given in response to learners' failure of correcting their own errors || and retreated || **as soon as*** the learners can independently solve the problem. || --- CMAC17_02_292

E.g., 5.7.16 ||| The tonal and rhythm patterns on the original test were trialled || **as soon as*** they came into existence, || and then a practical issue arose. || --- CPhD10_0009_752

When *as soon as* is applied in marking time, it usually refers to a point (exact time), like the word *as*, instead of a sphere or a scope of time (marked by *in*, *when*, *while*, etc.). The phrase has no limitation on its location in a clause complex, i.e., the dependent clause can be put before or after the independent one. Interestingly, the matters described in the independent clause always happen after those depicted in the

dependent one led by *as soon as*, meaning that *response would be given by CF after the point that learners solve the problem* in E.g., 5.7.15; and *tonal and rhythm patterns would be trialed after they came into being* in E.g., 5.7.16. But the relation between two clauses is annotated by temporal: same time rather than temporal: after/before as the time difference is quite short and things that happen in two clauses are in quick succession.

E.g., 5.7.17 ||| **Once*** they find themselves distracted or confused by certain unknown content, || they should adjust their concentration at once || so as to follow the thinking of the speakers. ||| ---- CMAC13LI_12_444

E.g., 5.7.18 ||| In other words, **once*** the EDV is settled, || all the slots of an event frame will be filled with such a default value || and then affect the evaluative meaning of ideation element in this slot. ||| ---- CPhD17_0001_897

A resemblance of the phrase *as soon as* is the word *once* in both, suggesting the point of time that things come up as depicted in two clauses, having changeable locations in leading to a dependent clause in the whole clause complex, being followed by finite clauses with Subjects and Objects, and connecting independent clauses discussing on subsequent affairs. Matters depicted in the clause beginning with *once* must have been realized or fulfilled that perfect tense would usually be chosen for the clause; while it is not the same case when *as soon as* leads to a clause that can also talk about things to be completed or to come. So, in E.g., 5.7.17, *they* must have been *distracted or confused by certain unknown content*, and in E.g., 5.7.18, *the EDV must have been settled*, then strategies should be adopted, or the next steps can be taken for further and wider influences.

Halliday categorizes conjunctions that indicate “same time” enhancement between clauses into three categories: “extent,” “point,” and “spread.” Indeed, they represent a synthesis of the “point,” “linear,” and “covering” range of time. However, I would rather prefer a categorization of these words into three groups: “point,” “sphere,” and “scope,” for indicating the temporal range covered by each. To illustrate, the temporal range of certain words does not align with Halliday’s proposed divisions in specific contexts, for example, the word *when* can indicate a brief period in addition to “point.” In certain instances, the ranges indicated by specific words overlap or merge, as is the case with *while*, which can signify either a specific time point or a broader range of time. Additionally, some words do not frequently express ranges in written texts, particularly in the corpus selected for this thesis. For instance, the use of *whenever* to denote “spread” or “scope” is rarely observed in academic papers. However, the usage of *whenever* to signify “point” or “sphere” is more prevalent. It would, therefore, be beneficial to investigate whether there are any notable differences in the lexis used by master’s and doctoral students to express temporal ranges.

5.7.2.2 Earlier

Words for expressing an enhancement: temporal: earlier relation in two clauses are usually *before* and *until*, in which the former is the most welcomed, though the number of these two words is not so frequent as that of other temporal markers in both masters and doctoral theses. (see Table 81)

Enhancement	M	D
Temporal: Earlier	115	118
before	99	89
until	16	29

Table 81 The number of each lexis for temporal: earlier clause complexes applied in Master and PhD sub-corpora

E.g., 5.7.19 ||| Examples of such contextualization include, for example, showing pictures of a city || **before*** asking the students to discuss some topics about that city, || or playing a video clip from a film adaptation of the topics and culture information the class is about to study. ||| ---- CMAC08SP_18_393

E.g., 5.7.20 ||| **Before*** handing out the materials to participants, || the notion of computer adaptive test and short context reading were firstly explained. ||| ---- CPhD14_0002_927

The choice of *before* in connecting clauses with temporal: earlier relationships would not be confined by finiteness or the location of the dependent clause. Hence, students are more likely to apply it in writing hypotactic temporal clause complexes. Though *before* can be replaced by *after* by changing clause sequence, it is still highly used as the information of the independent clauses is what students attempt to emphasize. This can be supported by E.g., 5.7.19 that *showing pictures of a city* is better deemed as *an example of contextualization* and E.g., 5.7.20 that the explanations of some notions are important as they may affect understandings of materials that are handed to participants.

E.g., 5.7.21 ||| The statement of meanings is therefore the weak point in language-study, || and will remain so || **until*** human knowledge advances very far beyond its present state. ||| ---- CMAC11SE_19_307

E.g., 5.7.22 ||| The whole process works in this way || **until*** the topic of the text is completely explored. ||| ---- CPhD09_0002_1409

The dependent clause led by *until* can also be located before the independent clause, but this is normally not the case, especially when the subjects of the dependent clause are to be reinforced. And *until* denotes a point of time that matters described in the independent clause end, before which the status stands still. It is obvious from E.g., 5.7.21 that *the statement of meanings* continuously be *the weak point in language studies* before *human knowledge advances*; and from E.g., 5.7.22 that *the whole process* remains unchanged before *the text topic is fully explored*, objects in the independent clause in both examples have stable status before an exact point call them to end.

5.7.2.3 Later

Among the conjunctions that convey temporal relations in clauses, *and then* is the most frequently used, with nearly 200 out of 482 choices by master's students and nearly half of the 266 choices by doctoral students using this conjunction. (see Table 82) *After* is the second most frequently used conjunction, with nearly 186 master's student choices and 83 doctoral student choices. The former hypotactically connects clauses, while the latter paratactically connects clauses. Other conjunctions with similar functions are *then*, *since*, and *finally*. While *afterwards*, *ultimately*, and *whereafter* are also present in the theses, they are used in only a few instances.

Enhancement	M	D
Temporal: Later	482	266
after	186	83
afterwards	2	1
and then	198	125
(and) finally	22	18
since	31	17
then	41	21
(and) ultimately	1	1
whereafter	1	-

Table 82 The number of each lexis for temporal: later clause complexes applied in Master and PhD sub-corpora

E.g., 5.7.23 ||| Individual writing is advised to supervise every student's strive for accuracy, || **and then*** the students can exchange their works with group members. ||| ---- CMAC09ME_24_487

E.g., 5.7.24 ||| It makes frequency counts first || **and then*** carefully analyses the discourse features and perspectives of the writers in LR's in the 100 dissertations. ||| ---- CPhD06_0004_201

For paratactically connecting clauses in temporal: later relation, the Subjects of the clause led by *and then* can be consistent with that of the other as in E.g., 5.7.24 that the center of two clauses remains as *it*, or be totally different as suggested by E.g., 5.7.23 in which *individual writing* has changed into *the students*. Sometimes, the affair described in the secondary clause (with *and then*) should be based on that in the primary one that it would not happen until the object in the former clause (without *and then*) is realized.

E.g., 5.7.25 ||| During the cognitive process, "sign" creates "concept" in one's brain, || **then*** "concept" creates "image" || after mixing with people's subjective feeling. ||| ---- CMAC08SE_25_459

E.g., 5.7.26 ||| In the longer lyric poem, the speaker initially visits the real world [[where the Beadsman suffers bitter cold and absolute solitude]]; || **then*** he turns to the past world of fairy tales, [[in which Porphyro ventured into the icy castle || and saved the beautiful Madeline]]; || in the last stanza, the speak travels back to the present: ... ||| ---- CPhD17_0003_1981

By using *then* in paratactic enhancement: temporal: later relationship, the subject of the secondary clause usually changes into a diverse but related one like in E.g., 5.7.25 from *sign* to *concept*, and the time interval is not so transient as expressed by *and then* that one thing happens after another, not in a short period. Here, *then* more denotes the event sequence described in clauses and can thus be followed by clauses beginning with conjunctive *at last*, *in the last*, and *finally*, etc., just as in E.g., 5.7.26.

E.g., 5.7.27 ||| It is hoped || that, this analysis, though limited and imperfect in many aspects, can help readers to get a better understanding of Wilde's death theme, || **and finally*** transcend death. ||| ---- CMAC10LIT_21_739

E.g., 5.7.28 ||| To put it another way, || representation is a cover term [[which has to do with Thibault's (2004) semiotic hierarchy for meaning representation]]; || viz. embodiment is to be

represented as iconic semiosis in the first place, || then as indexical semiosis || **and finally*** as symbolic semiosis. || ---- CPhD17_0001_773

Now comes to the conjunctive *and finally* that in E.g., 5.7.27, the conjunction (*and*) *then* is not a must for choosing *and finally* in conveying the ultimate goal of the analysis conducted by the writer, and in E.g., 5.7.28, though *be presented* has been deleted in last two clauses, how *embodiment is represented to be* in sequence or importance is clarified by using *then* and the marker *and finally*.

E.g., 5.7.29 || Any tourism text provides its readers with necessary traveling information about the tour destination, || and also it can widen the readers' knowledge, || and then arouse their interests, || **and ultimately*** persuade them to travel to the very place. || ---- CMAC09PR_28_247

E.g., 5.7.30 || According to Wang Zhenhua, the sentence patterns a writer employs in literature, << be it a statement, an interrogation or an imperative; || be it a simple and short sentence or a long and complicated sentence; || be it subordination, coordination, or ellipsis,>> have their respective purposes || and consequently reflect the writer's attitude, ways and extent of engagement, || **and ultimately*** are worthy of attention in appraisal. || ---- CPhD07_0004_525

The conjunctive *and ultimately* is the same as *and finally* in function and location but has only been chosen once in each group from the corpus, probably due to Zipf's *the principle of least effort* that students prefer to employ easier or shorter words and those that are more familiar to them in writing when alternatives are longer and more complex to them. More than underlining temporal arrangement or the sequence of matters' development, *and ultimately* contains the highlight on the essential meaning or uttermost importance of the affair in the clause headed by it, just as shown by examples above.

E.g., 5.7.31 || Its real lyricist was the American singer-songwriter, Curly Putman, and **afterwards***, it was recorded by several singers, such as Porter Wagoner and Bobby Bare. || ---- CMAC20_01_1415

E.g., 5.7.32 || Plato and Aristotle treated sign as the object of philosophical investigation, and **afterwards***, more and more thinkers had various elaborations of sign. || ---- CPhD10_0015_296

By choosing the conjunctive *and afterwards*, both the origin of affairs discussed in the primary clause and the follow-up development of these affairs under the influence of the origin depicted in the secondary clause is reinforced by the writer that *the real lyricist* was mentioned and other names of cover versions were also focused on in using *and afterwards* in E.g., 5.7.31 and it was *Plato and Aristotle's regarding sign as an object of philosophical investigation* that gives supports for studying on the sign by following thinkers as suggested by the writer in E.g., 5.7.32.

E.g., 5.7.33 || **After*** studying the two English examples, || let's check another language for further study. || ---- CMAC11SE_19_231

E.g., 5.7.34 || The LSSs are weak in application of learning strategies, || but **after*** receiving strategy training, || they perform better than before || in applying strategies in learning || and become more sensitive to the English language. || ---- CPhD11_0006_56

In E.g., 5.7.33, *after* is used for retrospect of past studies by the writer and then introduction of further studies to the reader: this is normal in academic writing, especially when the dependent clause headed by *after* is located before the independent clause. But sometimes, it is selected for explaining the effects of some methods or strategies on participants in experimental research in this form as indicated by E.g., 5.7.34 that the comparison of students' behaviors before and after the training is what the writer intends to focus on.

E.g., 5.7.35 || In the second phase, Jakobson brought forward distinctive markedness of phonemes, **whereafter*** he extended it to the morphological domain. || ---- CMAC10SY_9_110

The word *whereafter* is rare, and only one has been discovered in the annotated corpus, as shown in E.g., 5.7.35. The whole clause complex can be agnate with *and then he extended it ...* or *after that, he extended it ...*, all of which are paratactic enhancements: temporal: later relations. But compared to *after* or *then*, the choice of *whereafter* represents a closer connection between affairs depicted in two clauses, i.e., without *the propose of distinctive markedness of phonemes*, Jakobson could not able to *extend it to the morphological domain*.

E.g., 5.7.36 || **Since*** lexicology returned to be the focus of western linguistics in the midst of 1980s, || Lexical acquisition has been considered as the central task in the study of second language acquisition. || ---- CMAC12MO_3_103

E.g., 5.7.37 || Discourse analysis at "character-to-character" level draws most attention from stylisticians || **since*** modern drama stylistics came into being. || ---- CPhD17_0002_548

Being different from other markers of enhancement: temporal: earlier relation, the word *since* connects two clauses within which the dependent one is on the time nodes an affair began or ended (usually the tense of the clause is past or perfect tense) and the independent one is about the effects on following affairs or the development of the original matter or the consequence. In E.g., 5.7.36, the writer highlights both the time that *lexicology returned to be the focus of western linguistics* and this fact as well by using *since*; and in E.g., 5.7.37, without the existence of *modern drama stylistics*, *discourse analysis*

will not be able to continuously absorb knowledge or methods from stylisticians, the significance of the dependent clause would not thus be embodied.

5.7.3 Spatial

5.7.3.1 Same Place

In contrast with temporal relations, doctoral candidates prefer spatial enhancement in clause complexes when their choice is 129 while that of master's is 54, as shown in Table 83, in which most of the choice in the former group is *as far as* to indicate the extent of the space and then is *where* for denoting a point, while that in the latter group is *where* and then *as far as*. Although the phrase *in so far as* resembles *as far as* is just chosen by doctoral candidates for a few cases, and the word *wherever* for suggesting the "spread" of space only appears once in the master group.

Enhancement	M	D
Spatial: Same place	54	129
as far as	21	64
in so far as	-	6
where	32	59
wherever	1	-

Table 83 The number of each lexis for spatial: same place clause complexes applied in Master and PhD sub-corpora

E.g., 5.7.38 ||| More interestingly, a fake reflexive pronoun of *ziji* (self) is optional in (i), || which is contrast with English as in (iia) and (iib), || **where*** the fake reflexive pronoun with coreferential to the subject is obligatory. ||| ---- CMAC13SY_7_505

E.g., 5.7.39 ||| In nonmilitary settings, the strategy concept has been applied to clearly non-adversarial situations, || **where*** it has come to mean a plan, step, or conscious action toward achievement of an objective. ||| ---- CPhD10_0001_316

When *where* in a clause refers to a word rightly before it, no matter whether a comma rests within them or not, the clause headed by *where* is downranked, i.e., embedded into the word (often a noun) instead of having the same taxis as the former clause. In E.g., 5.7.38, *where* substitutes the situation of fake reflexive pronoun in English as indicated by examples (iia) and (iib) rather than only examples; and in E.g., 5.7.39, *where* indicates *non-adversarial situations* with the *strategy concept been applied to* more than simply situations, the status of two clauses in each example, therefore, is hypotactic that the clause with *where* is dependent.

E.g., 5.7.40 ||| **Wherever*** Oliver goes, || evil thrives around him. ||| ---- CMAC10LIT_7_93

The word *wherever* means *everywhere* and the clause complex in E.g., 5.7.40 is agnate with the saying *Evil thrives around Oliver no matter where he goes*. The focus is always on the theme of the independent clause be it in the secondary or the first position in a clause complex.

E.g., 5.7.41 ||| **As far as*** the Chinese culture is concerned, || the Chinese people usually have a tendency for literature, || although there existed a military examination system, || which the readers can find in Chapter 26 of the novel. ||| ---- CMAC07CU_35_517

E.g., 5.7.42 ||| As indicated in Table 6-19 and Table 6-20, || the experimental group mean is much lower than the comparison group mean || **as far as*** old items are concerned || and the group difference is significant at the 0.01 level. ||| ---- CPhD08_0012_1533

Strangely is that when the phrase *as far as* is employed in clause complex for an enhancement: spatial: same place relation, it always appears in the saying *as far as ... is concerned*, which does not reveal its literal meaning in indicating the actual distance of two affairs but the connotative one for suggesting the rarely mentioned matters or turning the topic for discussion into a new one. So, *the Chinese culture* in E.g., 5.7.41 is a new topic that has not been referred to by the writer before; and *old items* in E.g., 5.7.42 are seldom paid attention to and hereby raised for consciousness.

E.g., 5.7.43 ||| A discourse mapping might occur || **in so far as*** the slot values are specific enough to instantiate the emotive values of Vehicle. ||| ---- CPhD17_0001_1659

E.g., 5.7.44 ||| **In so far as*** connotation is concerned, || there always has to get contextual knowledge involved in || for interpreting the connotative meaning, || which in turn makes attitude invoked by lexis || rather than inscribed by lexis. ||| ---- CPhD17_0001_471

The phrase *in so far as* is not "that far" in the distance (either literally or metaphorically) as the former phrase *as far as* specifically when an affair is put forward, though these two are changeable on functions and structural locations (See E.g., 5.7.44). It is not discovered in the annotated corpus of the master's group and is not so common compared to *as far as* in the Doctoral group. By employing *in so*

far as in spatial relation, the distance of the writer's recognition or cognition on the clause theme is reinforced, like in E.g., 5.7.43, *as the slot values are ... Vehicle* within a certain scope or to some extent, *a discourse mapping might occur*.

5.7.4 Manner

Although the choice of enhancement: manner relation in clauses by doctoral students (2,105 altogether) outnumbers that by master's students (1,368 in total), the proportion of manner: means in two groups is almost the same (68.22% in doctor's and 66.74% in master's), so does the proportion of manner: comparison. This means that in each manner enhancement chosen by students in thesis writing, almost one is manner: comparison, and two are manner: means.

5.7.4.1 Means

Together, 1,436 manner: means enhancement clause complexes chosen by doctoral candidates, around 60% are realized by the word *by* for a hypotactic relation, followed by 15.64% by the conjunction (*and*) *thus* for a paratactic one; and in total 913 choices by master's students, 63.2% are realized by the preposition *by* and then 15.77% by (*and*) *thus*. (see Table 84) Amazingly, the proportion of these two markers is almost the same in two groups! Other markers for the manner: means relation can be the preposition *through*, *via*, *whereby*, and the connection *in that way*.

Enhancement	M	D
Manner: Means	913	1436
by	577	859
in that way	-	1
(and) thus	144	226
via	5	3
whereby	3	3
through	14	9

Table 84 The number of each lexis for manner: means clause complexes applied in Master and PhD sub-corpora

E.g., 5.7.45 ||| Topic change employed by Dede is involved in this dialogue, || **thus*** the speaker Dede violates the maxim of relation accordingly. ||| ---- CMAC15_01_787

E.g., 5.7.46 ||| However, in small group or dyadic interaction with familiar partner(s), there may be some security [[in knowing || that other students are being faced with similar communicative tasks, || **and thus***, their face-protected concern would diminish || and they won't be so focused on the evaluation of their partner(s) || and so concerned about losing face because of improper communicating behaviors]]. ||| ---- CPhD13_0003_1053

In paratactic enhancement: manner: means relation, and the conjunction *thus* is similar to *in this way* that both lead to the secondary clause in a clause complex with the primary one depicting a kind of method or behavior. So, in E.g., 5.7.45, it is clear that *in the way of involving topic change in the dialogue, the speaker Dede ...*; and in E.g., 5.7.46, students' face-protected concern would diminish by way of knowing that *other students are facing similar communicative tasks*, suggesting that clauses are all embedded in the noun *security* rather than having the same taxis as the first clause.

E.g., 5.7.47 ||| The author places the two contradictory situations in the same paragraph || so that readers have a clearer idea of the consequences leaders' manoeuvres might lead to, || **and in this way*** she succeeds in convincing them of the necessity of being cautious about their deeds on the leaders' part. ||| ---- CPhD07_0004_653

In fact, *and in this way* is useful in clearly demonstrating the relation between two clauses but is often ignored by students in academic writing, and only one case has been found from the annotated corpus. The clause complex above can also be written as: *In the way of placing..., she succeeds in ...*, which not only emphasizes the strategy that the author used for reaching her aim but also attracts readers' attention to the effects of this strategy.

E.g., 5.7.48 ||| In addition, **by*** conducting the questionnaire to the students in experimental group || after the experiment of the image schema theory being applied to vocabulary teaching, || it can be figured || that the participants taught by image schema theory show a positive attitude towards the new vocabulary teaching strategy from cognitive perspective. ||| ---- CMAC16_05_770

E.g., 5.7.49 ||| The team had some agreed measures, || such as monitoring students' learning process || **by*** reading their learning logs, || providing them feedbacks in the form of comments in the logs || and

having ten minutes' discussion of the solutions to problems in the self-access learning in the classroom instruction each week. ||| ---- CPhD12_0009_1522

The preposition *by* is flexible in its location in a clause complex and hence is quite common for marking the hypotactic manner: means relation though a finite verb is required after it in the dependent clause such as *conducting* and *reading* above. When the clause beginning with *by* is put in the front, ways or strategies adopted by study objects are stressed by the writer, as demonstrated in E.g., 5.7.48 that *conducting questionnaire on students in the experimental group after ...* is a useful method and receives positive evaluation on presenting the experimental results from the writer. The desired effect could not be highly reached when the dependent clause (headed by the marker *by*) is put afterward, as shown in E.g., 5.7.49.

E.g., 5.7.50 ||| For them, participants had the highest frequency of using the "prior knowledge" strategy || **whereby*** its mean score was above 3.6. ||| ---- CMAC14RE_15_577

E.g., 5.7.51 ||| Kellerman (1983) makes psychotypology, the learner's perception of language typology, central to his perspective on transfer, || **whereby*** the learner's recognition of congruent forms between the native and target languages either facilitates or interferes with L2 acquisition. ||| ---- CPhD08_0005_481

From the annotated corpus, the number of enhancement: manner: means clause complexes that use *whereby* is limited. Actually, more are employed in clauses cited by the writer, i.e., not written by students but by scholars or advanced researchers. Though it seems to mark spatial: same place relation as *where* is included in the word, the meaning tends to imply how things are realized through a manner or a method indicated by the marker *by* however. So, the clause complex in E.g., 5.7.50 agnates with *By using the ...strategy, participants had the highest frequency, the mean score of which was above 3.6*, including both *by* and *which* to connect clauses; and in E.g., 5.7.51, *whereby* can absolutely turn into *by* which for suggesting manner relation.

E.g., 5.7.52 ||| The theory of typological universals, <<first brought forward by Greenberg (1966),>> attempts to confirm the markedness || **via*** investigating different language families, || and further to bring their common features to light. ||| ---- CMAC10SY_9_136

E.g., 5.7.53 ||| Debates continue in this regard || as researchers find no fixed pattern later on || to sustain this over generalized summary || **via*** examining vast language data [[including mostly expositive and argumentative essays]]. ||| ---- CPhD20_0002_78

The clause complex in E.g., 5.7.53 can be understood as: *Via examining vast language data, researchers find no fixed pattern...* that these two clauses are on the same clause level, i.e., have the same taxis though the former is dependent. From the original clause order, what is more focused on by the writer is the result that *researchers find no fixed pattern* no matter what strategy is adopted. In E.g., 5.7.52, *via* can totally be changed into *by* without any influence.

E.g., 5.7.54 ||| **Through*** understanding examples the teacher explains, || practicing the inferential model by themselves, and assistance from the teacher, || the students in the experimental class gradually master the inferential skills based on the Theory of Conversational Implicature || and form the habit and subconscious mind in thinking and working out the implied meaning in the conversation. ||| ---- CMAC12LI_6_385

E.g., 5.7.55 ||| When there is no such strong personal element, || the passives of reporting verbs are used to enable general presentations || **through*** avoiding identification with the individual researchers. ||| ---- CPhD06_0004_1819

E.g., 5.7.56 ||| **Through*** analyzing the orders of the items listed by students from both classes in Table 15, || we found the differences between the two classes in this section. ||| ---- CPhD11_0006_726

The word *through* functions the same as *via* and *by* in that it can lead to the dependent clause being in either the first place or the secondary place in denoting the manner: means relation within clauses. In E.g., 5.7.54, students are able to improve themselves with the means of a list of activities described after the marker *through*; and in E.g., 5.7.56, differences are able to be discovered by analyzing students' diversified activities, both examples present the writer's intention to make ways or activities as foregrounds and then stress more on the following affairs or effects brought about by these steps by turning the depended clause (led by *through*) into the first place. However, when the depended clause beginning with *through* is placed afterward, measures taken are more important than the effects they cause in the writer's point of view, as suggested by E.g., 5.7.55. Compared with *by*, the choice of *through* attracts readers' attention to the process of how activities are conducted instead of only mentioning these activities.

5.7.4.2 Comparison

The most frequent word for manner: comparison relation in enhancement clause relation is *as*, which takes up 85.93% in the master's group (391/455) and even as high as 92.23% in the doctor's group

(617/669). (see Table 85) Other markers for this relation can be *like*, *(and) so*, *as if*, *as though*, *as much as*, *likely*, and *similarly*, though most of these words or phrases are seldom chosen by students in two groups.

Enhancement	M	D
Manner: Comparison	455	669
as	391	617
as if	8	4
as much as	-	3
as though	1	4
like	23	12
(and) so	9	2
likely	1	-
similarly	-	1

Table 85 The number of each lexis for manner: comparison clause complexes applied in Master and PhD sub-corpora

E.g., 5.7.57 ||| As can be seen from all the above, || the understanding of listening comprehension develops in a scientific and systematic way, || **so*** does the listening teaching, || as the actual practice always improves based on the research development. || ---- CMAC12LI_6_105

E.g., 5.7.58 ||| Among the thousands of personal narratives [[that have been recorded and studied]], variability in the degree of subjectivity has been revealed, || **so*** do their study results: in social status, narrativity, or literary aesthetics. || ---- CPhD20_0002_1562

For paratactic enhancement: manner: comparison relation, the conjunction *so* has to lead the secondary clause and be followed by a tense (usually presented by auxiliary verbs *do*, *does*, *did*, etc., and copula *is*, *are*, *was*, *were*, etc.) that is similar to the primary one, as demonstrated in examples above ..., *so does the listening teaching*, and ..., *so do their study results*. By choosing this kind of comparison, the writer, in fact, is seeking out similarities of diverse behaviors or affairs, such as *listening comprehension* versus *listening teaching* or the *variability of subjectivity degree* to *study results*.

E.g., 5.7.59 ||| This comparison of Levinson's is very vivid, || imagine || that when you push a gun's trigger, || the gun fires, || **similarly*** when you push presupposition-triggers, || presupposition will be generated or introduced by these triggers from constructions, just like in a gun fire. || ---- CPhD07_0008_157

Usually, the word *similarly* would be followed by a comma in a paratactic enhancement: manner: comparison relation, though it was forgotten in the example above. And in choosing this marker, the secondary clause can have the same structure as the primary one because of its connotation as indicated by *when you push a gun's trigger* and *when you push presupposition-triggers* above that these two conditions of the "push"-behavior are compared.

E.g., 5.7.60 ||| While for this contrast in word-final position, XH informants can better produce this contrast than SH informants, || very **likely***, it has much to do with lack of a similar word-final sound in XH dialect. || ---- CMAC15_04_481

A parallel word to *similarly* is *likely*, which can connect clauses with diverse structures as shown by E.g., 5.7.60, even though the writer intends to explain the meaning of the former clause rather than compare the subjects within two clauses.

E.g., 5.7.61 ||| After the first time listening of the material, || the teacher asks students to talk about [[what they have been heard in pairs]], || just **like*** what the teacher and all the students do in the first stage. || ---- CMAC12LI_6_361

E.g., 5.7.62 ||| When reading the ad, || people easily tend to think || that Martell Spirit, paralleled with the national spirit of France, must be really wonderful, || just **like*** what we Chinese people might think of Maotai, the spirit of China. || ---- CPhD07_0011_697

Though *like* can be followed by non-finite clauses of hypotactic relation, no clause complex of this kind has ever been found from the corpus, and almost all are finite hypotaxis led by the conjunction *like* as described in *just like what ...* in two examples above. By comparing *what the teacher asks students to talk about* with *what the teacher and all the students do in the first stage*, the writer tells readers that these two behaviors are the same in essence as suggested by E.g., 5.7.61; and through mentioning the Chinese people's attitude towards the spirit Maotai with the marker *like* in E.g., 5.7.62, the attitude of French people to Martell spirit is clearly demonstrated. The comparative seemingly intention results actually in categorization.

E.g., 5.7.63 ||| **As*** discussed in the last section, || memes are just like genes, || so the evolution of memes is also similar to the evolution of genes definitely. || ---- CMAC15_02_204

E.g., 5.7.64 ||| Iron and steel manufacture, the production of steam engines, and textiles were all powerful influences, || **as*** was the rise of a machine-building sector able to spread mechanization to other parts of the economy. ||| ---- CPhD07_0011_547

Normally, the conjunction *like* can be substituted by *as* in finite clauses in hypotactic enhancement: manner: comparison relation that the Subject of the clause is deleted, making the theme of the clause a highlight that the writer tends to attract readers' attention, as demonstrated in E.g., 5.7.63, where the first clause can agnate with *Just like what we have discussed in the last section*, ..., the writer chooses *as* instead to keep a distance from readers and seek for an objective attitude toward the clause theme. When the dependent clause with the head *as* is located afterward, the theme of the independent one is stressed. That is why *as* but not *like* is chosen because the writer intends to focus more on *powerful influences such as iron and steel manufacture, steam engines, etc.* rather than *the rise of a machine-building sector* when the clause is written as *just like the rise of ... was able to ...*

E.g., 5.7.65 ||| And **as if*** to insist upon this Cinderella equation, || the cruel stepsisters turn up wealthy and favoured by nature, with a cruel mama in tow, evil female mother/ daughter triads of Reeds, Brocklehursts and Ingrams, fortune's darlings, all of them marriageable and promising in the accepted terms of the day and of the novel. ||| ---- CMAC06LIT_31_711

E.g., 5.7.66 ||| The foreman in the joke behaves || **as if*** he were asked to evaluate the state of the guilt || and not merely to indicate a decision. ||| ---- CPhD07_0011_1067

Themes of the independent and dependent clauses in a hypotactic enhancement: manner: comparison clause complex connected by the conjunction *as if* are not compared with each other, but are either concluded or explained by another one as demonstrated in E.g., 5.7.65 that the theme of the independent clause is summarized by the dependent one by using *as if* that all staffs happened to Cinderella's cruel sisters seem to follow up an equation; and in E.g., 5.7.66 that the theme of the independent clause, i.e., *the behavior of the foreman in the joke* is exemplified by the dependent one, to be specific, *be asked to evaluate... and to indicate...* The conjunction *just like*, however, does not possess this sort of contextual differentiation.

E.g., 5.7.67 ||| When Santiago was steering on his watch, || —he could not see the green of the shore now but only the tops of the blue hills [[that showed white || **as though*** they were snow-capped]] and the clouds [[that looked like high snow mountains above them]]. ||| ---- CMAC10LIT_15_393

E.g., 5.7.68 ||| In linguistics texts have often been discussed || **as though*** their meanings were constant for all users. ||| ---- CPhD05_0003_329

The conjunction *as though* seems to be the same as the phrase *as if*, as the clause in E.g., 5.7.67 can be written into *the blue hills showed white as if they were snow-capped*, and in E.g., 5.7.68 as the clause *as if meanings of texts were constant...*, but it usually locates afterward in leading a dependent clause in a manner: comparison clause complex. When *as if* or *as though* is chosen, the theme is assumed to possess some characteristics, i.e., *the blue hills seem to be snow-capped* or *the meanings of texts seem to be constant*, but, in fact, this kind of imagination is inexistent. Interestingly, *like* can be used in the former example because hills can be snow-capped, but it cannot be applied in the latter one as meanings were originally not *constant for all users*.

E.g., 5.7.69 ||| They are sometimes unsure of what it means to critique, || thinking || that critiquing means finding fault || **as much as** they can || so as to appear knowledgeable || and relates to knowledge exhibitionism. ||| ---- CPhD13_0010_1562

Though the phrase *as much as* is rare in the corpus, it is obvious that it functions as a connection for hypotactic enhancement: manner: comparison relation, as suggested by E.g., 5.7.69 that more faults that people can find are compared with fewer faults to indicate that *finding too much faults* does not mean criticize and is not advocated by the writer.

5.7.5 Causal

The causal enhancement chosen by students in the two groups has almost the same number (3,370 in the master's group and 3,315 in the doctoral group), in which the cause: purpose relation takes the highest proportion in the former group (53.65%), and then is the cause: reason relation (43.47%), leaving the cause: result relation the lowest proportion (only 2.88%); while the cause: reason relation (48.99%) and the cause: purpose relation (48.11%) have almost the same proportion in the doctoral group, leaving the cause: result relation has a proportion of 2.90%. Hence, master's students tend to clarify the purpose of other scholars' intentions or his/her aims, doctoral candidates tend to explain stuff or phenomena in thesis writing. Though the cause: result relation seems to be an *argumentum e contrario* of the cause: reason relation, its low proportion is highly decided by the limited choice as only the phrases such as *so that*, *(and) thus* and *as a result* are available for meaning conveying.

5.7.5.1 Reason

In presenting the cause: reason relation of a clause on another one in a clause complex, the most common word chosen for students in both the master's and doctoral groups is *because* for hypotactically connecting finite clauses (401 in group M and 476 in group D). (see Table 86)

Enhancement	M	D
Cause: Reason	1465	1624
and so	3	6
and therefore	29	49
as	245	288
because	401	476
because of	7	8
considering	22	7
for	103	130
in that	28	112
seeing that	2	2
since	219	315
so	322	126
therefore	71	73
thereby	8	16
(and) thus	4	15
hence	-	1
for this reason	1	-

Table 86 The number of each lexis for cause: reason clause complexes applied in Master and PhD sub-corpora

Then is the word *so* for paratactically combining two clauses by master's students (322) and *since* for hypotactically connecting finite clauses by doctoral candidates (315). This is followed by the word *as* for hypotactic finite clauses, though its number in the doctoral group (288) is a little higher than that in the master's (245). The choice of *since* (219) by master's students also has a large proportion in all markers. Then is the word *for* as a preposition for paratactically presenting the effects before causes in clause complexes in two groups (103 in group M and 130 in group D), and doctoral candidates also have a preference for the word *so* (126 in total) in connecting cause-effect clauses paratactically. The phrase *in that* is more favored by doctoral candidates compared to master's students (112 in the former and only 28 in the latter). Leaving the word *therefore* (71 in group M and 73 in group D) and its relevant structure *and therefore* (29 in group M and 49 in group D) an important status. Then are some uncommon markers such as *considering*, *thereby*, and *seeing that* for finite clauses, *(and) thus*, *and so*, *hence* and *for this reason* for paratactic clauses, and *because of* for non-finite hypotactic clauses. All in all, doctoral candidates prefer to provide reasons for phenomena or stuff compared to explaining the purposes of researchers when their choice of the cause: reason enhancement relation is compared to that by master's students (1,624 in the former and 1,465 in the latter).

For paratactic enhancement: causal: reason relation, conjunctions can be *so*, *therefore*, *thereby*, *thus*, *hence*, etc., with *and* in front of it in connecting two clauses. Usually, these conjunctions have to be in the secondary clause in a clause complex to denote a "cause-effect" sequence and reinforce the effect of an affair because of the reasons stated before, and the reasons are generally concerned with facts, aims, consequences, status, viewpoints, etc.

E.g., 5.7.70 ||| The final goal of any language learning is to communicate, || **so*** the students should learn to cooperate || in order to improve their competence of English reading. ||| ---- CMAC13RE_10_153

E.g., 5.7.71 ||| First, the low proficiency level students' target language was more limited than the high proficiency students' || and they did not have so many interlanguage resources || as the high proficiency students did to use, || **so*** they tended to employ avoidance or L1-based strategies. ||| ---- CPhD10_0018_1082

The connector *so* is very common in indicating a paratactic cause-effect relationship in clause complexes. In thesis writing, in addition to listing the causes, *so* is also used after clauses to state the purpose or make comparisons. For example, in E.g., 5.7.70, a goal of language learning is described, while in E.g., 5.7.71, a state is shown, or a fact is stated before presenting the results. Therefore, although *so* is chosen to

connect clauses with the results, the essence of causing these results is different.

E.g., 5.7.72 ||| Much research has been done in this domain on the items [[which constitute little or no difficulty for translation and language learning]] || **and so*** actually do not have much value for comparison. ||| ---- CMAC06SY_17_641

E.g., 5.7.73 ||| Learners with high motivation and self-confidence but low anxiety have low filters || **and so*** obtain || and let in plenty of input. ||| ---- CPhD08_0001_450

It is strange to use *and so* to connect two clauses for conveying a cause-effect relationship, as shown in E.g., 5.7.72 & 5.7.73. If the word *and* is removed, the second clause in both cases should then be preceded by a comma, and the subject of the first clause be repeated; while if the word *and* is kept, the relation within clauses may seem fine, is indeed mixed up. The causal clause in the former example is a given fact or a result, while that in the latter example is a state. The use of *and so* makes it unclear to the reader whether the writer is expressing an additional relation or a causal relation between clauses.

E.g., 5.7.74 ||| Obviously, Amir considers his father as his possession, || **therefore*** he thinks || his father can only love him || and no one has the right to share it with him. ||| ---- CMAC15_05_318

E.g., 5.7.75 ||| Chinese speakers, on the contrary, pay much attention to the pursuing of becoming of things in the world, || and through their consistent pursuing of it, Chinese speakers want to seek for the doings [[that should be followed by themselves]], || **therefore*** making a moral judgment of the natural instincts of themselves and the world. ||| ---- CPhD07_0008_1383

Compared to *so*, the word *therefore* is more formal when expressing the paratactic enhancement: cause: reason relationship between clauses. Unlike *so*, which focuses on the causal clause, *therefore* emphasizes the result or the consequences/effects caused by the former clause as demonstrated in E.g., 5.7.74 and 5.7.75, in which that in the former example is similarly a fact, while that in the latter is a descriptive state.

E.g., 5.7.76 ||| Compared with explicit metaphor, || implicit metaphor conveys a strong sense of judgement, || **for this reason***, it is mainly used to express strong emotions (ibid: 207). ||| ---- CMAC20_01_171

The conjunctive *for this reason* resembles the phrase *as a result* in leading to a cause: reason relationship in clauses, which can also be substituted by the word *so* or *therefore*, though *for this reason* emphasizes more on the effects in a sense. The content of the causal clause in E.g., 5.7.76 is a fact for stating the difference between *implicit metaphor* and *explicit metaphor*.

E.g., 5.7.77 ||| The current relevance of the present perfect construction refers to the relevance [[established between the event profiled by the past participle in the construction and the reference point (i.e., the position of the speaker and the hearer)], || **and therefore***, this relevance is subjective. ||| ---- CMAC10SY_9_310

E.g., 5.7.78 ||| It's understandable why the titles of business texts seldom employ attitudinal items denoting affect, || which involves a conscious participant || and how he feels || since feelings and emotions are highly personal || **and therefore*** lack objectivity || **and so*** are less likely to be used to persuade a serious and sensible business reader. ||| ---- CPhD07_0004_481

Though the connector *and therefore* is regular and common for indicating a cause: reason relationship within clauses, it seems to be misused in thesis writing by both master's students and doctoral candidates, as shown in E.g., 5.7.77 and 5.7.78. The word *and* is redundant in the former example, as only *therefore* is enough to convey the causal relation. And the choice of *and therefore* in the latter example is repetitive as both the word *since* in its former clause and the mark *and so* in its latter clause are applied in signifying the cause: reason relationship. It's hard to tell that the whole clause complex is not translated by computational translating software from Mandarin to English. Though the writer in E.g., 5.7.78 is stating a viewpoint, it is the *attitudinal items* rather than *feelings and emotions* that *lack objectivity* according to the writer's meanings. This *lacking objectivity* is what the writer intends to reinforce as it would have impacts on those being described in the following clause, i.e., *attitudinal items denoting affect are less likely to be used to persuade....*

E.g., 5.7.79 ||| **Seeing that*** EAA is a complex phenomenon, || it has several reflections concerning its meaning, || and **thereby*** we just postulate a broken line [[to indicate the ME axis covering varied subtypes]]. ||| ---- CMAC10SE_18_381

E.g., 5.7.80 ||| The development of awareness of language, culture, and language learning strategies broadens students' horizons || and helps them learn language more effectively, || **thereby*** playing an important role in students' overall development, especially in their language learning process. ||| ---- CPhD11_0006_431

The same redundancy can be found in E.g., 5.7.79 when both the phrase *seeing that* and the preposition *thereby* are chosen for signifying the cause: reason relationship between two clauses, but *seeing that* is a hypotactic connector whereas *and thereby* seems to be a paratactic one when the word *and* is added, though not necessary. The reason in this example more resembles a fact, but E.g., 5.7.80 turns out to be a viewpoint of the *development of awareness of language, culture, and language learning strategies*. The

word *thereby* functions the same as *therefore* or *hence* and *thus* in connecting clauses for causal relations, though it may be considered as a spatial one due to the word *by* within.

E.g., 5.7.81 ||| Jay points out this wrong explanation with his awareness || that it is not true, || **thus*** what he says does not comply with the sub-maxim of quality. || ---- CMAC15_01_567

E.g., 5.7.82 ||| It presupposes the result of the narrating act, || **thus*** abiding by the linguistic economy principle, || the narrative finale resorts to an implicit representation. || ---- CPhD20_0002_1302

Though the word *thus* is usually chosen to denote a manner: means or manner: comparison relationship in two clauses, sometimes would be applied in connecting two clauses with a cause: reason relationship as suggested by E.g., 5.7.81 & 5.7.82. The agnation can be the words *so* and *therefore* for a paratactic status. The reason in the former example is a description, while in the latter, it is a kind of state, and the choice of *thus* makes the explanation more formal.

E.g., 5.7.83 ||| As is noted, || text is a written record of language production, || **and thus*** is generally considered as static (rather than dynamic) in nature, || though the interest in dynamic features of text can be sporadically located in some literature of text generation and perception (e.g., S. Stoddard, 1991). || ---- CPhD05_0003_63

The linkage *and thus* is similar to the word *thus* as it would also be seen as a connector for a manner: means relationship in clauses, with the word *and* can be deleted. Normally, the agnation can be *as a result* or *therefore*, though in E.g., 5.7.83 *therefore* is more suitable as the subject of the secondary clause remains to be the *text*. Compared to the word *thereby*, the content linked by *and thus* seems to be more serious, as shown in the example. The choice of *and*, though it makes two clauses closer, does not influence the relationship decided by the word *thus*.

E.g., 5.7.84 ||| It is through communication || that individuals learn to relate effectively and appropriately to the social environment || and, **hence***, are able to fulfill various needs and desire (Kim, 1988, 2001, cited in Yashima, 2004). || ---- CPhD13_0003_1569

The word *hence* in denoting a cause: reason relationship is rarely seen as it may be regarded as a sign for temporal: later time when the reason described in the former clause must happen before the result described in the latter clause as found in E.g., 5.7.84. To present the reason as a viewpoint, *hence* can be substituted by *thus* or *therefore* for a paratactic relation.

E.g., 5.7.85 ||| In his words, the Chinese authors writing about demonization of China in the West were not only nationalistic but opportunistic || **because*** they “attempted to cash in on the popular sentiments that the U.S., Britain and the West are seeking to ‘contain’ or restrain China’s emergence on the world stage”. || ---- CMAC18_01_181

E.g., 5.7.86 ||| **Because*** the students have the knowledge of English vocabulary and grammar, || the teacher could relegate error correction to the students. || ---- CPhD16_0001_1943

The most common binder for a hypotactic enhancement: cause: reason relationship within clauses is *because*, the emphasis of the content of which (the dependent or independent clause) is highly influenced by the location, i.e., when the independent clause is in the initial status, it is the result that is emphasized (see E.g., 5.7.85); while when the dependent clause with the word *because* is in the initial status, it is the reason that is focused on by the writer (see E.g., 5.7.86). This interchangeable character of clause location makes the writer’s intention, as well as the subject of the clauses, outstanding.

E.g., 5.7.87 ||| Just **because of*** their poor ability to express Chinese culture in English, || their intercultural communicative competence is poor as well. || ---- CMAC14CU_8_379

E.g., 5.7.88 ||| They are less likely to get involved in classroom communication for fear of being ridiculed || **because of*** making mistakes || or not communicating properly. || ---- CPhD13_0003_1105

The prepositional phrase *because of* in connecting clauses or elements of a cause: reason relation would often be followed by non-finite clauses or phrases or just nouns as indicated by E.g., 5.7.88 and 5.7.87. The reason being emphasized can be the description of subjects’ or stuff’s characters, properties, peculiarities or status, etc. Again, the location of the dependent clause being led by *because of* would be affected by the writer’s tendency to reinforce.

E.g., 5.7.89 ||| **As*** different grades meet different standard, || corrective feedback should be given in individual’s zone of proximal development. || ---- CMAC17_02_50

E.g., 5.7.90 ||| Globalization forces are becoming increasingly evident in medical education || **as*** medicine is a global profession || and medical knowledge and research have traditionally crossed national boundaries, || which makes physicians have studied medicine and provided services in various countries of the world. || ---- CPhD11_0007_67

No matter where the conjunction *as* is in connecting two clauses with a cause: reason relationship, it is the effect that is emphasized by the writer as suggested by E.g., 5.7.89 (a cause-effect clause complex with the “effect” being focused on) and 5.7.90 (an effect-cause clause complex with the “effect” being emphasized) as the word *as* is not so strong in tone compared to other words such as *because* and *since*.

The content follows *as* is usually facts or objective reasons, the binder *as* can thus be grammatically replaced by *because* but not semantically or pragmatically in practice.

E.g., 5.7.91 ||| As far as the law of causality is concerned, || it is indeed consistent with the law of temporal sequence, || **for*** things always develop from cause to effect, from condition to result and from hypothesis to possibility. || ---- CMAC06SY_17_397

E.g., 5.7.92 ||| Eight interviewees felt lucky to be the subjects of LPBL model, || and all declared || that this sort of model went over quite well with young people, || **for*** it provided them with more opportunities to demonstrate their abilities. || ---- CPhD11_0007_1457

E.g., 5.7.93 ||| **For*** Item 8, 9, 10, and 11 are mainly about the instructors in the training program, || the author labeled the first factor as instructor effectiveness. || ---- CPhD17_0004_773

The word *for* in paratactically connecting two clauses with a cause: reason relationship together is common, and the clause led by *for* is usually located in the secondary place (E.g., 5.7.91 & 5.7.92) rather than the initial one (E.g., 5.7.93) though some writers would do it on purpose. No matter where the “cause” clause is in a clause complex, it is always the clause that is reinforced by writers as the reason would otherwise be omitted if it is not more important than the “effect” clause. The choice of *for* instead of *as* thus lies in the importance of “cause” instead of “effect,” and another reason for the choice of *for* rather than *because* comes from either the informality or weak tone or the writer’s preference.

E.g., 5.7.94 ||| English writing is the most difficult part for learners, even for English majors || **since*** other skills (listening, speaking and reading) level cannot result in the mistake in the comprehension while communication with the help of other auxiliary factors such as tone, facial expression etc. || ---- CMAC14WR_04_67

E.g., 5.7.95 ||| **Since*** not all the metaphors are qualified to contain attitudinal meanings, || the identification for MA in this research is carried out by a MIP+AT operation. || ---- CPhD17_0001_113

When the writer intends to emphasize the effects of facts on persons or stuff, the choice of the connector for combining two clauses with a cause: reason relationship would usually be *since*, which is followed by negative meanings as demonstrated in E.g., 5.7.94 and 5.7.95, to exclude part of the reasons in explanation. The word *as* is not so strong as *since* in focusing on fixed reasons, not even to mention the word *for* is not the most welcomed choice by doctoral candidates in thesis writing. The agnation of *since* resembles *considering that* in most cases, such as *considering that not all the metaphors are ...* though the latter is not so familiar to students as judged by the number of choices.

E.g., 5.7.96 ||| Although the sub-meanings are extended from the same core, || they share different dimensions to the prototype || **in that*** their relations with prototype are different. || ---- CMAC11SE_19_739

E.g., 5.7.97 ||| However, transfer is inadequate || **in that*** the effect of the consequent misclassification of English XP reflexives as X0 anaphors is also shown in the high group’s acceptance of long-distance antecedents for reflexives in sentences of the PRNP condition. || ---- CPhD15_0001_1414

The phrase *in that* is unique in connecting cause: reason clauses hypotactically as it always leads to the dependent clause and is located in the secondary status to emphasize the reasons being described. Though the phrase agnates with the word *because* most of the time, the location of the dependent clause is unchangeable, and the tone of *in that* is stronger than *because* in signifying the particular reason that the writer intends to inform readers of. Hence, it is the only reason for *sub-meanings having different relations with a prototype* that makes them *share different dimensions to the prototype* in E.g., 5.7.96 and *the effect being shown in ...* makes the *transfer inadequate* in E.g., 5.7.97.

E.g., 5.7.98 ||| **Seeing that*** EAA is a complex phenomenon, || it has several reflections concerning its meaning, || and thereby we just postulate a broken line || to indicate the ME axis covering varied subtypes. || ---- CMAC10SE_18_381

E.g., 5.7.99 ||| There is no limit to the number of metaphors that can be embedded in this way (Charteris-Black, 2005, p. 54), **seeing that*** in some cases, one single frame is inadequate to invoke comprehensive attitudes which the Topic is intended to express. || ---- CPhD17_0001_1818

The conjunctive phrase *seeing that*, though is similar to *as* or *because* in denoting a cause: reason relationship between clauses, emphasizes more the factuality or objectivity of reasons. And what is located at the initial status in a clause complex is what the writer intends to convey. That is to say, as there is a sequence of events temporarily, when the dependent clause led by *seeing that* is at the initial place, it is the reason that is more important; but when the independent clause is at the initial place, it is the effect caused by the described reason that is more outstanding. Hence, the reason in E.g., 5.7.98 cannot be ignored, with its impact on subsequent events being mentioned; while the explanation in E.g., 5.7.99 seems to be what the writer wants to convey.

E.g., 5.7.100 ||| All in all, **considering that*** Chang Hen Ge falls into the sentimental poem, || it is of great significance for the readers [[to catch the implied pragmatic meaning in the original work]]. ||| ---- CMAC11PR_10_567

E.g., 5.7.101 ||| Yet, when it comes to practical lexicography, || this automatic feature can bring excessive workload on the part of lexicographers, || **considering that*** a great majority of lexical bundles have high frequency but limited lexicographic value || and thus have to be removed (Granger & Paquot 2015; Verdaguer et al. 2013; see also Section 2.1.1). ||| ---- CPhD19_0001_809

The phrase *considering that* is normally followed by a point of view or a perspective in signifying the cause: reason relationship in clauses as decided by its connotation semantically. Such as in E.g., 5.7.100, it is the point that *Chang Hen Ge falls into the sentimental poem* that makes it *of great significance*...; and in E.g., 5.7.101, it is the fact that *a great majority of lexical bundles*.... makes the *automatic feature bring excessive workload on*.... Hence, the reason is more important than the effects of events, the phrase *considering that* is agnate with *in case that* or *in that* in this sense, though the former relates to embedded clauses.

5.7.5.2 Result

It is amazing that the number of choosing cause: result relation in two clauses is almost the same in two groups, in which most of the choices are *so that* for connecting a finite clause hypotactically, leaving some cases connected by (*and*) *thus*. (see Table 87)

Enhancement	M	D
Cause: Result	97	96
so that	85	87
such that	-	1
(and) thus	10	5
as a result	2	3

Table 87 The number of each lexis for cause: result clause complexes applied in Master and PhD sub-corpora

The connection *as a result* for paratactically combining two clauses only appears twice or third, and the phrase *such that*, though resembles *so that* structurally, is seldom found in theses.

E.g., 5.7.102 ||| This may be justifiable || as far as English is concerned, || for the structure of an English complex sentence is in essence the same as that of an English clause || and, **as a result***, the basic framework of an English complex sentence will show itself || as long as the SV structure of its main clause is identified. ||| ---- CMAC06SY_17_21

E.g., 5.7.103 ||| Theoretically, starting from the Systemic Functional Linguistics (Henceforth SFL), it investigates the semantic relationship between reporting and the various perspectives realized by the linguistic structures of reporting in discourse, || and **as a result***, establishes the hypothesis [[that perspective exists at different levels of reporting language, from reporting verbs to reporting clauses and discourse]]. ||| ---- CPhD06_0004_145

The paratactic enhancement: cause: result relation marked by the conjunctive phrase *as a result* is often connected by *and*, which seems to be the additional relation within two clauses; otherwise, clauses are separated and should be intersected with punctuation marks like a comma or a full stop in between. By choosing paratactic rather than hypotactic relation in the two examples above, both reasons and results are weighted by the writer in meaning-making.

E.g., 5.7.104 ||| Some students are “imaginative readers”: || especially if they know something about the subject, || or have strong opinions about it, || they may interpret the text in the light of their own experience and viewpoints, || **so that*** they find it difficult to separate what the writer says from what they feel themselves. ||| ---- CMAC13RE_10_231

E.g., 5.7.105 ||| Hewings (2004) has found || that novice writers conflate subject and theme at the beginning of a clause, || **so that*** their texts take on features of narratives and descriptions (for instance, there are; this is). ||| ---- CPhD13_0010_785

In denoting a hypotactic enhancement: cause: result relation, dependent clauses beginning with the conjunctive phrase *so that* have to be placed in the secondary place (see examples above) as reasons always happen before results in the sequential order, i.e., results cannot exist without any reason that could cause them logically. Interestingly is that *so that* can also indicate the hypotactic: enhancement: cause: purpose relation when it is followed by modal verbs (see Section 5.7.5.3). Thus, it is hard to distinguish these two relations. As not all dependent clauses with modal verbs suggest the purpose relation, the structural difference is not determinacy for relations division or judgment, which means that some clauses taking modal verbs as Rheme indicate the result relation semantically. Hereby, clauses describing affairs that have already happened and that can agnate with *for the result that*... in meaning

potential is usually connected by enhancement: result relation just like in E.g., 5.7.104 that *students indeed have difficulty in separating writer's sayings from their own opinions* as they may understand or have understood *texts with their own experience* and in E.g., 5.7.105 that *texts have features of narratives and descriptions* because of *green writers' mixing subject with theme in a clause*.

E.g., 5.7.106 ||| If, for example, a Claim-Argument relation is said to hold between two segments A and B, || **such that*** A expresses a claim and B an argument favoring that claim, || then prior to this analysis and according to some criterion, A and B must have been labeled as "discourse segments." ||| ---- CPhD05_0003_791

The phrase *such that* for marking the hypotactic enhancement: cause: result relation is analogous to *so that* in a clause complex, but unlike the latter one, it can also attract the reader's attention to the reasons before it because of the denotation of *such* semantically. So, besides describing the results of what A and B stand for in E.g., 5.7.106, the writer emphasizes the reason that it is the Claim-Argument relation signed by some words that makes these two segments have identified expressions.

E.g., 5.7.107 ||| In-depth investigations and studies have yet to be conducted on the details and particularity of L2 Lexical acquisition, || and the methods of empirical research have yet to be riched, || **thus*** there is great potentiality for its perfection and continual improvement. ||| ---- CMAC12MO_3_126

E.g., 5.7.108 ||| About ten native speakers of Chinese (from Taiwan and mainland China) participated in the studies of Knickerbocker (2007) and Tahta et al. (1981), || but they occupied a fraction of the sample, || **thus*** unable to be representative of the whole group of Chinese EFL learners. ||| ---- CPhD10_0009_451

Though *thus* is normally used to connect clauses as a hypotactic enhancement: manner: means or comparison relations as have been explained in Section 5.7.4, it would more likely be used by students in indicating the hypotactic enhancement: cause: result relation that the former clause plays the reason part of the latter one, which is obviously demonstrated from examples above that further studies and empirical research methods on L2 Lexical acquisition make the *perfection and continual improvement* essential in E.g., 5.7.107; and the sample fraction occupied by ten native speakers of Chinese mentioned by the writer cause these speakers' non-representativity of the whole group of Chinese EFL learners in the writer's point of view in E.g., 5.7.108. These independent clauses before *thus* are actually more cause-result inclined instead of being manner-means or comparison tended, which originates from the meaning of *thus* and can then agnate with those followed by *hence*.

E.g., 5.7.109 ||| The lack of an atmosphere of English language is one of the obstacles in learning English for students || so that it is hard for students to grab the background knowledge of target language || and to think in a western way, || **and thus*** the speed and accuracy of reading are affected. ||| ---- CMAC13RE_10_681

E.g., 5.7.110 ||| Accordingly, a confident individual may be inclined to overestimate his or her proficiency || **and thus*** has less anxiety in interaction, || whereas an unconfident individual tends to underestimate his or her competence || **and thus*** is more anxious about communication. ||| ---- CPhD13_0003_805

E.g., 5.7.111 ||| Yet, the speaker recognizes the absence of growth, change, and maturing in the latter || **and thus*** he returns to the reality || and perceives it with an insight. ||| ---- CPhD17_0003_1999

The usage of *and thus* is a little bit strange in that students prefer to apply it in suggesting cause: result relation rather than manner: means within clauses, which resembles *thus*, but the additional meaning is endowed with clause relationship because of *and* in between. One explanation comes from the language transfer of 也因此(ye yin ci) in Mandarin Chinese that either official documents or colloquial materials would choose it as a linkage; the other originates from the variety of *and so* that the misuse of these markers is quite common for non-native students of English. If *and thus* tends to indicate by means of in E.g., 5.7.109 that *By means of grabbing the background knowledge of English and thinking in their way, the speed and accuracy of reading* abilities of students are able to be improved, these two words in E.g., 5.7.110-111 demonstrate more reason-result inclination that *a confident individual may have less anxiety in interaction because of his/her overestimate of proficiency while an unconfident one feels more anxious due to his/her underestimate* in E.g., 5.7.110 and *the speaker returns to reality for the reason that he recognizes the absence of ...* in the following example.

5.7.5.3 Purpose

In choosing the cause: purpose relation in clauses, the number of master's students (1,808) outdistances that of doctoral candidates (1,595), as suggested by Table 88, in which the choice of the preposition *to* has no big difference (1,097 in master's and 1,004 in the doctor's) and takes the highest proportion (60.67% in the former and 62.95% in the latter). Followed by the phrase derived from *to*, i.e., *in order to* (396 in group M and 298 in group D). Then, the phrases *so that* for finite clauses and *so as to* for non-finite

clauses are also welcomed by students in two groups. Therefore, it is the preference of choosing *to*, *in order to*, *so as to*, and *so that* that accounts for a higher proportion of cause: purpose relation in causal enhancement clause complexes by master's students. As to the preposition *for*, doctoral candidates seem to tend to choose it in connecting non-finite clauses. Other phrases like *in order that* and *such that* for finite clauses and two prepositional phrases for non-finite clauses, i.e., *for the sake of* and *with the aim of* can also be found in thesis writing. Generally, using cause: purpose markers for connecting finite clauses compared to non-finite ones takes up no more than 10%.

Enhancement	M	D
Cause: Purpose	1808	1595
for	84	144
for the sake of	1	-
in order that	7	2
in order to	396	298
so as to	106	60
so that	114	84
such that	1	2
to	1097	1004
with the aim of	2	1

Table 88 The number of each lexis for cause: purpose clause complexes applied in Master and PhD sub-corpora

E.g., 5.7.112 ||| Moreover, **in order to*** explain exactly the definition, || there are some examples given by Aristotle. ||| ---- CMAC12SE_05_69

E.g., 5.7.113 ||| The former refers to a desire to learn the language || **in order to*** relate to and even become part of the target language culture; || and the latter has to do with practical reasons for language learning. ||| ---- CPhD08_0012_995

For marking the cause: purpose relation, the prepositional phrase *in order to* is common and would not limit the position of the dependent clause that it leads to. When the dependent clause is in the initial position, such as in E.g., 5.7.112, the event described in the independent clause usually has been fulfilled or is going to be finished; while when the dependent clause is in the secondary position, as in E.g., 5.7.113, the tense of the independent clause can be simple present, which would not be limited by the content described in the dependent clause.

E.g., 5.7.114 ||| Students can make full use of various resources and reference materials || to complete high-quality works in collaboration with partners, || **so as to*** demonstrate students various skills, such as cooperation abilities, social skills and culture consciousness, and so on. ||| ---- CMAC12ME_5_169

E.g., 5.7.115 ||| Students may also be encouraged to watch English films or videos in the original often || **so as to*** imitate the performance of native speakers || and therefore improve their proficiency in these aspects. ||| ---- CPhD05_0008_1268

In conveying the cause: purpose logical meaning within two clauses, the prepositional phrase *so as to* has to be in the secondary position when it leads to a dependent clause. In this circumstance, the tense of the dependent clause (at the initial position) can be past tense or future following modal verbs in signifying events that have not been fulfilled, as demonstrated in E.g., 5.7.114 & 5.7.115, which is quite common in most cases. Compared to the word *to*, the goal or the aim after *so as to* seems to be a final or ultimate one.

E.g., 5.7.116 ||| **To*** acquire the attitude of the students to feedback of development zone, || the researcher made interviews with the participants. ||| ---- CMAC17_02_590

E.g., 5.7.117 ||| On the other hand, ZPD-based corrective feedback draws attention to the analysis of internalized linguistic information instead of the automatic access to relevant L2 information || **to*** communicate meanings in real time. ||| ---- CMAC17_02_526

E.g., 5.7.118 ||| Though interlanguage studies have been characterized by a concern || **to*** formulate and test hypothesis of a universal order of SLA, || typology (and psychotypology) appears to be the most important variable in determining the likelihood of language transfer. ||| ---- CPhD08_0005_499

The most common word for connecting clauses with the cause: purpose relationship is *to*, which can either be located at the initial place (as in E.g., 5.7.116) or be located at the secondary place (in E.g., 5.7.117 & 5.7.118) when it leads to a dependent clause, with the purpose being described an either big or a small one. Though the tense of the independent clause would not be limited, the secondary position of the dependent clause can cause a mixture of the component of elements after *to* be either a dependent one or an embedded one. This can be judged by the changeability of the dependent clause, i.e., if the element

after the word *to* can be transferred from the secondary place to the initial one, then the element with *to* is a dependent clause; otherwise, it is an embedded one. As demonstrated in E.g., 5.7.117, as the structure *to communicate meanings in real time* can be moved in the front of the clause complex, it is a dependent clause in describing a purpose.

E.g., 5.7.119 ||| **In order that*** the English-speaking readers can be attracted by the view “Evening Bell at Nanping Hill”, || this part should not be omitted || when translated into English. ||| ---- CMAC09PR_28_381

E.g., 5.7.120 ||| Thissen and Mislevy (2000) suggested || a mixture of “target precision” and “maximum number of items” should be used in practice || **in order that*** certain measurement accuracy could generally be obtained without the item pool [[running out of appropriate items to administer]]. ||| ---- CPhD14_0002_785

The difference between the phrase *in order that* from the proposition *in order to* in connecting cause: purpose clauses is the introduction to new subjects in the former, with modal verbs usually in the dependent clause to signify events that have not happened yet. So, the purpose in E.g., 5.7.119 is *the English-speaking readers can be attracted by the view...*, and in E.g., 5.7.120 is *certain measurement accuracy could generally be obtained....* When the dependent clause is in the initial position (as in the former example), the purpose is reinforced, and when the independent clause is in the initial position (in the latter example), the thing to be fulfilled is emphasized.

E.g., 5.7.121 ||| For any pair of conflicting constraints, the higher-ranked constraint takes absolute precedence, || **such that*** violations of a lower constraint are minimally permitted to the extent [[that they minimize violations of the higher constraint]]. ||| ---- CMAC12PH_1_130

E.g., 5.7.122 ||| Mauranen (1996) argues || that relationships between adjacent sentences are therefore the most important, || **such that*** an unsatisfactory of appearance a text passage can often be traced back to its thematic lack of fit within the paragraph. ||| ---- CPhD09_0002_1189

In signifying the cause: purpose relationship in clauses, the phrase *such that* in leading to a dependent clause can only be located at the secondary position, being followed by either a perfect tense (as in E.g., 5.7.121) or a future tense plus the modal verbs (in E.g., 5.7.122). No matter what the tense of the dependent clause is, it is only the purpose described that the writer intends to emphasize. And this kind of purpose seems to be narrower than that led by the phrase *so that* when the scope is considered.

E.g., 5.7.123 ||| The so-called Nominalization refers to the process [[in which a verb or adjective in a sentence converts to a noun or noun phrases]], || **so that*** the noun or noun phrase gains the meaning of the verb or adjective || and possess the grammatical function of noun. ||| ---- CMAC12SY_13_377

E.g., 5.7.124 ||| Cooperative learning tasks both inside and outside the classroom are frequently assigned || **so that*** students have access to various learning environments [[where they can participate and find strategies to use English || to complete the tasks]]. ||| ---- CPhD11_0006_1616

To judge cause: result relation within clauses from cause: purpose relation of *so that*, one condition is that modal verbs take the finite place of the clause led by *so that*, and the writer intends to convey a purpose by presenting the events that have not yet happen in the dependent clause. Hence, *so that* is usually followed by modal verbs for a purpose relation and is agnate with the phrase *for the purpose that*, though not all clauses with *so that* at the beginning and then have modal verbs signify a purpose relationship. This can be demonstrated by E.g., 5.7.124, when the modal verb *can* is deleted and the dependent clause is located at the secondary position. In fact, the verb in E.g., 5.7.123 after the phrase *so that* should be (can/may) gain instead of *gains* for a cause: purpose relation, which is parallel with the verb *possess* hereafter.

E.g., 5.7.125 ||| Moreover, **for*** improving the appreciation of verbal humor in different countries [[which reflects distinguishable culture in cross-cultural communication]], || the thesis also puts forward some advice to help people appreciate exotic humors in a better way. ||| ---- CMAC13CU_27_727

E.g., 5.7.126 ||| In the course of cross-domain mapping, attitudes towards entities and structure of source domain can be substantially taken along || **for*** inventing a similar attitude towards entities and structure of target domain. ||| ---- CPhD17_0001_477

The word *for* in signifying a cause: purpose relationship in two clauses has no big difference as compared to the word *to*, though the former must be followed by a non-finite verb with *-ing* form. The location of the dependent clause is unlimited, though the initial position indicates a reinforcement of reason as indicated by E.g., 5.7.125. But *for* is not so welcomed by students as *to* for leading to a purpose, which is a strange phenomenon to be investigated.

E.g., 5.7.127 ||| **For the sake of*** avoiding taking those undesired words accounted in, || all of those undesired words are marked with the token < > manually || and set as Hidden in Tag Settings of Global Settings; ||| ---- CMAC12SP_005_262

The phrase *for the sake of* performs the same function as the word *for*, though the following clause has to

begin with a non-finite verb because of the preposition *of*. Though the dependent clause can also be located in the secondary place in a clause complex, it is usually in the initial status, for especially emphasizing the purpose by the writer, such as in E.g., 5.7. 127. But most cases would be chosen *for* rather than *for the sake of* for the flexibility and effort-saving of the former in a clause complex.

E.g., 5.7.128 ||| In addition to the above specific fields in which studies on hedges are involved, another corpus-based research effort is found to be made on the comparison of hedges between native speakers and non-native speakers, || **with the aim of*** revealing the features of hedges used by EFL learners and native speakers || and finding out the differences between the learners and native speakers.

||| ---- CMAC12SP_005_173

E.g., 5.7.129 ||| Inspired by this, more and more educators in China began to pay attention to students' individual variables || **with the aim of*** making English teaching more effective. ||| ---- CPhD08_0001_707

Although the prepositional phrase *with the aim of* leading to a cause: purpose dependent clause can also be located at the initial position in a clause complex, it would usually be located at the secondary place to emphasize the events that have been done before introducing to the purpose. Hence, the writer intends to inform readers of the action that *another corpus-based research effort is found to be ...* in E.g., 5.7.128, and *more and more educators in China began to pay attention to...* in E.g., 5.7.129. Usually, the choice of this phrase would keep the subject of the independent clause in the dependent ones, and the phrase, being agnate with *for* or *for the sake of*, is seldom chosen in thesis writing.

5.7.6 Conditional

The number of conditional enhancement clause complexes chosen by students in two groups has no big difference (1,023 in group M and 1,174 in group D), in which more than half of the choices are concessive ones in the doctoral group (takes up 51.36%) and nearly half of them are also concessive ones in the master's group (48%). The proportion of positive conditional clause complexes in all conditional ones selected by masters' students (46.04%) is a little higher than that by doctoral candidates (43.44%), leaving the proportion of negative conditional relation in two groups not larger than 6% (5.96% exactly in group M and 5.2% in group D). Although the markers for negative conditions are numerous, the small number in two groups may indicate a specific language characteristic. The higher proportion of concessive conditions in all clause complexes derives from more markers being available for meaning conveying.

5.7.6.1 Positive

The number of condition: positive clause complexes chosen by masters' students and doctoral candidates is almost the same (471 in the former group and 510 in the latter group), though more than 80% of them are marked by the word *if* for a finite or a non-finite hypotactic relation in clauses (411 in the former group and 424 in the latter group). (see Table 89)

Enhancement	M	D
Condition: Positive	471	510
as long as	26	22
assuming that	1	1
if	411	424
if...then...	25	51
in case	1	2
(and) in that case	3	2
provided that	1	5
supposing that	1	-
Observing that	2	2
then...	?	?
In case of	-	1

Table 89 The number of each lexis for condition: positive clause complexes applied in Master and PhD sub-corpora

The relevant structure *if...then...* is more favored by doctoral candidates (51) compared to master's students (25), and the phrase *as long as* for connecting finite clauses also has a certain preference by students in two groups (26 in group M and 22 in group D). Without considering the word *then* as it would

usually be mixed with that indicating a temporal: later relation in clauses, other structures for connecting condition: positive related clauses can be *assuming that*, *provided that*, *supposing that*, *observing that*, and *in case* for finite clauses, *(and) in that case* for paratactic clauses, and *in case of* for non-finite hypotactic clauses.

E.g., 5.7.130 ||| **If*** students can do reflective learning, || they will do autonomous learning || and know || how to learn English writing. ||| ---- CMAC14WR_04_229

E.g., 5.7.131 ||| What we can do is to describe some typical or commonly used patterns as well as their variations || **if*** we want to do anything with discourse patterns. ||| ---- CPhD05_0003_995

The word *if* in marking a positive conditional relation in clauses is frequently seen in thesis writing as it is flexible for choice that the dependent clause with *if* can either be in the initial position as E.g., 5.7.130 does or in the secondary position as E.g., 5.7.131 does. When it is the former choice, the condition that is mentioned in the dependent clause is reinforced; and when it is the latter choice, the effect of the condition on the event then is what the writer intends to emphasize. As to the word *then*, when the dependent clause is in the primary position in a clause complex, it can appear or not. When *then* is not chosen by the writer, this may be because of his/her writing behavior. But when the word *if* is semantically excluding some conditions, the word *then* would have to be chosen in the clause complex. Another situation for the deletion of *then* comes from the modal verbs after *if* for possibilities and the maintaining of the subject in two clauses, just as that shown in E.g., 5.7.130.

E.g., 5.7.132 ||| **If** this is so, || **then*** it was Canale and Swain (1980) [[who first suggested || that there was more to “competence” than just an ability to create, or even appropriately use, grammatically correct sentences]]. ||| ---- CMAC05PR_24_143

E.g., 5.7.133 ||| Furthermore, **if** the relationship between a signifier and its signified is relatively arbitrary, || **then*** it is clear that interpreting the conventional meaning of signs requires familiarity with appropriate sets of conventions. ||| ---- CPhD10_0015_989

Hence, for the structure *if... then...* for a positive condition relationship in clauses, the description that is presented by the dependent clause, i.e., after the word *if* is then excluded by the choice of *then*. And because of the existence of *then*, the dependent clause led by the preposition *if* has to be in the initial position in a clause complex to emphasize the impact of conditions on the event afterward. Being affected by the meaning of the word *then*, which usually appears in the temporal relation for describing a different time of events, the event mentioned in the independent clause usually happens after that in the dependent one. Therefore, *if the relationship between a signifier and its signified is relatively not arbitrary*, then whether *interpreting the conventional ...requires familiarity with... or not* would become unclear when E.g. 5.7.133 is interpreted.

E.g., 5.7.134 ||| Third, the reflective learning had positive effects on learners' vocabulary mostly, || **then*** the structure had ranked second place of the most positive effects of reflective learning. ||| ---- CMAC14WR_04_15

Sometimes, the word *if* would be missing for a condition: positive relation, leaving only a *then* in the clause complex. In such a condition, the relation would easily be mixed up with temporal: later relation as it is not clear whether the former clause is stating a condition for the latter one or happens before the latter one temporarily. In E.g., 5.7.134, the logical meaning is not *the reflective learning had positive effects on learners' vocabulary mostly* earlier, and then *the structure had ranked the second place*, but is that the condition of *the reflective learning having positive effects on learners' vocabulary mostly* decides the secondary place of another kind of structure. Therefore, though this kind of structure is not easy to find in doctoral dissertations, it is worthy of further studies for distinction with conditions, principles, or norms because of its importance.

E.g., 5.7.135 ||| **As long as*** they hold correct attitude toward Chinese culture, || they can express it properly sooner or later. ||| ---- CMAC14CU_8_565

E.g., 5.7.136 ||| As a “not-there” phenomenon or “unfalsifiable” phenomenon, absolute fossilization is better understood as a hypothesis || and it is possible to guard against fossilization || **as long as*** the fossilized features are noticed and taken into account of rational contingency learning || and there is enough frequency tallying after noticing. ||| ---- CPhD08_0012_487

The adverbial conjunction *as long as* in connecting two clauses with a condition: positive relation indicates the uniqueness of the condition because the satisfaction of the only condition being mentioned by *as long as* can be enough in leading to subsequent events. Therefore, when the dependent clause is located in the initial position, the condition is then emphasized (as shown in E.g., 5.7.135); and when it is located in the secondary or third position in a clause complex, the effects being affected by the condition are thus focused on by the writer (see E.g., 5.7.136).

Besides the “if” related structure for positive conditional relation in clauses, there are a group of finite verbs for the same function, which are considered verbal conjunctions in connecting clauses such as *assuming that*, *provided that*, *supposing that*, *observing that*, etc., though the number of these

conjunctions is units digit in thesis writing by students. In Halliday's (2014: 484) words, these structures make clauses semantically overlap between expansion and projection and are therefore special for discussion.

E.g., 5.7.137 ||| **Assuming that*** language learners can consult a dictionary about [[how to spell the wanted word correctly, || and how to use it in the proper way in certain context]], || whether the sentence they construct is appropriate or not seems much more troublesome, || because most of the college students may not possess the judgment as to a self-generated sentence, especially some complex one, is target-like or not (Zhao, 2000: 81). ||| ---- CMAC05WR_20_340

E.g., 5.7.138 ||| In Research 3, we provided students with a self-access learning environment || **assuming that*** students would be able to make use of the affordance [[that the system offered]] || and take charge of their own learning process. ||| ---- CPhD12_0009_1692

The first verbal conjunction is *assuming that* in presenting the writer's prediction as a condition. No matter whether it leads to a dependent clause in the initial or the secondary position in a clause complex, the finite verb usually is a modal verb such as *can* in E.g., 5.7.137 and *would* in E.g., 5.7.138. However, slight differences can still be discovered, as the event described in the independent clause in the former example is not so expected by the writer, while that in the latter is highly expected. Whether the expectation exists or not depends probably on the choice of modal verbs after *assuming that*.

E.g., 5.7.139 ||| He aims to mediate the contradiction of science and art, || and puts forward a semiotic translation criterion [[which echoes his argument made in 1988]], || that is, **provided that*** the most important meaning of the text enjoys the priority to be conveyed, || other related meanings should be transferred at the same time as much as possible, || and it is necessary to make adaptation and compensation for meaning loss. ||| ---- CMAC11PR_10_221

E.g., 5.7.140 ||| **Provided that*** the evaluative knowledge in conceptual frames is more inclined to be acknowledged and accessible by a group of people, || emotive values in them are always interpersonally defaulted and familiarized within a specific community, || viz. there has a degree of familiarity for EDV. ||| ---- CPhD17_0001_909

The conjunction *provided that*, though, can lead to a dependent clause in the initial or the secondary position in a clause complex for a condition: positive relation, usually is located in the initial status as shown in E.g., 5.7.139 and E.g. 5.7.140. Being different from the word *if*, which is mainly realized in the diverse choice of modal verbs afterward, the event that is described in the independent clause based on the condition led by *provided that* is of a high probability of happening. This means that the event in the dependent clause conditioned by *if* may happen or not, but that conditioned by *provided that* must happen, which is decided by the meaning of the conjunction *provided that*. Only based on this possibility can the event being described in the independent clause be negotiable. Hence, only on the condition that *the evaluative knowledge in conceptual frames can the emotive values in them be interpersonally defaulted...* as suggested in E.g., 5.7.140. In addition, the event mentioned in the independent clause might have no relation with that in the dependent one when the word *if* is chosen; but the situation is quite converse as to the conjunction *provided that*, i.e., the event described in the independent clause is highly relevant to that in the dependent one when *provided that* is chosen for a connection in two clauses. That is why the subject of the dependent clause in E.g., 5.7.139 is *the most important meaning of the text*, and that of the independent clause is *other related meanings*.

E.g., 5.7.141 ||| **Supposing that*** the meta-linguistic process can be described as "L2 input to L2 output" as mentioned previously, Zhao found that there was the interruption of L1 between the L2 input and the L2 output, which caused the meta-linguistic breakdown in Chinese students' college English writing. ||| ---- CMAC05WR_20_349

The conjunction *supposing that* for a condition: positive relation is interchangeable with *assuming that* though the former is usually located in the initial position leading to a dependent clause as shown in E.g., 5.7.141. This structure only provides a possible condition for the event afterward, which means that *the meta-linguistic process be described as ...* is a possibility for Zhao to find that *there was the interruption...*

E.g., 5.7.142 ||| **Observing that*** the retained object is related to the DP in initial sentence, || some researchers (Xu 1999; Xu 2001; Han 2000, 2001; Chen & Wen 2001, etc.) share the same syntactic properties of unaccusativity from the aspect of possessor movement, || though their explanations differ a lot in syntactic operations such as case assignment. ||| ---- CMAC13SY_7_309

E.g., 5.7.143 ||| In the first two months, the two technicians were in charge of the technical operation of the self-access center, || but in the middle of the program, **observing that*** students might need help in the Center, || it was arranged that seven teachers work as a counselor on a two-hour shift every working day || to give help || when needed. ||| --- CPhD12_0009_1189

The conjunction *observing that* well explains its function in connecting clauses as it resembles *on observing that* or *on this condition* in leading to a dependent clause in an initial position in a clause

complex. What is mentioned in the dependent clause is the basis for the happening of the event in the independent clause, that is why the subject in two clauses remains the same, with that of the independent clause usually being deleted. As suggested by E.g., 5.7.143, the subject of the dependent clause is *students*, and that of the independent clause is also *students*, who are *arranged seven teachers*... In E.g., 5.7.142, it is on the condition *the retained object is related to*... that *some researchers share the same*... Without the condition, the description in the independent clause would be impractical.

E.g., 5.7.144 ||| Moreover, we know || that in English reading process, students may encounter many difficulties in comprehension || so the mistakes may occur frequently, || **and in this case*** it is advisable for teachers tolerate the students' classroom errors. || --- CMAC13RE_2_439

E.g., 5.7.145 ||| These ink spots form three almost closed circles **and in that case*** we see no longer merely several ink spots on a homogeneous background, but three surface circles bounded by these ink spots. || --- CPhD18_0002_567

The nominal conjunction *and in this/that case* in paratactically joining two clauses together for a positive conditional relation has no limitation on the choice of subjects, objects, finite verbs, etc., in the secondary clause after it. What is depicted in the clause before it is just a case or a kind of condition for the event after it. However, this freedom on other linguistic features does not increase students' preference for it in thesis writing.

E.g., 5.7.146 ||| They think they would feel more comfortable, more secure and have more freedom in this way; when they are invited to answer or ask questions in front of the class, they will lose face **in case*** they cannot have correct or effective responses. || --- CMAC05ME_21_366

E.g., 5.7.147 ||| Then the whole process of each interview was recorded with the informants using two digital voice recorders, || one Panasonic RR-US300 recorder || and the other is Shinco RV-10 recorder, || **in case*** one is out of work. || --- CPhD17_0004_831

Similar nominal conjunction *in case* in connecting condition: positive clauses are flexible in a clause complex as its dependent clause can either be after or before the independent one. When *in case* is in a postposition, the modality of the clause is normally a negative one as indicated by the modal verb *cannot* in E.g., 5.7.146, and the event being described would not happen with a high probability, though it may still happen sometimes. The post-posed clause with *in case* indicates that the writer has considered the importance of this kind of case, while the pre-posed clause suggests the emphasis on some measures being taken or behaviors being conducted by the subject. This may be the reason for being directly connected to the subject of a clause by the conjunction *in case*.

E.g., 5.7.148 ||| **In case of*** lacking of positive emotion, the learner's cognitive ability may be discounted. || --- CPhD08_0001_628

The structure *in case of* though is derived from the conjunction *in case* and can only be followed by finite verbs because of the preposition *of*. Usually, the subject of the dependent clause and that of the independent one is the same, with the former being deleted. As shown in E.g., 5.7.148, the condition is the learner's *lacking of positive emotion*, on which his/her *cognitive ability may be discounted*. The clause complex can also be interpreted as *in case that the learner is lack of positive emotion, ... or in case the learner is lack of positive emotion*... By emphasizing this condition, the dependent clause is thus pre-posed.

Though the conjunction *in the event of* and *to the extent that* also performs the same function in clause combination has never been found in thesis writing in the corpus by students. Hence, they are out of discussions here.

5.7.6.2 Negative

In limited choice of condition: positive enhancement clause complexes, the most frequent word can be the preposition *without* for connecting non-finite hypotactic clauses (20 in group M and 13 in group D), then is the phrase *no matter* (19 in group M and 15 in group D, though this is controversial) and the conjunction *unless* for non-finite or finite hypotactic clauses (most favored by group D with the number 22 and only 6 in group M). (see Table 90) The word *otherwise* for paratactic clauses also has a certain weight, and it is more favored by master's students (13 while 6 in group D). Other words or phrases in connecting condition: negative relation appear in the corpus are *if not...then...*, *if not*, *or else*, and *or* for paratactic clauses.

E.g., 5.7.149 ||| His free translation with annotation provide target readers further illustration on the cultural implication, || **or else*** the analogically lavished royal love and care of the monarch centred on Lady Yang cannot be perceived by those being devoid of relevant background knowledge. || --- CMAC11PR_10_405

E.g., 5.7.150 ||| Both writers have strong conviction of their ideas, || **or else*** they wouldn't try to spread them to others. || --- CPhD07_0004_995

Though the phrase *or else* has no great difference from *otherwise* in combining clauses with a negative

condition relationship, and is fixed in its location, i.e., in the secondary clause in a clause complex, it is usually followed by negative modal verbs such as *cannot* in E.g., 5.7.149 and *wouldn't* in E.g., 5.7.150. In fact, this relates to double negation as the clause complex in the latter example means the condition of *both writers having strong conviction of their ideas* leads to the result *they try to spread them to others*. The rare choice of *or else* thus comes from the double negation of meanings though the subject of the secondary clause may be changed into another one as demonstrated in E.g., 5.7.149, i.e., from *his free translation with annotation to the analogically lavished royal love and care of the monarch centered on Lady Yang*.

Enhancement	M	D
Condition: Negative	61	61
if not...then...	1	2
if not	-	2
no matter	19	15
or else	1	1
otherwise	13	6
unless	6	22
without	20	13
or	1	-

Table 90 The number of each lexis for condition: negative clause complexes applied in Master and PhD sub-corpora

E.g., 5.7.151 ||| Secondly, water must be limited in a certain boundary, || **or*** it may cause a disaster that its container cannot take into control. ||| ---- CMAC18_05_463

The choice of *or* in denoting a condition: negative condition is quite peculiar as clauses connected by *or* would usually be annotated as an extension: alternation relationship. In E.g., 5.7.151, *or* can totally be replaced by the word *otherwise* for a negative condition relation, and the clause being led by *or* has to be located at the secondary position rather than the initial one. The whole clause complex means *if water is not limited in a certain boundary, then it may cause a disaster that...* with *if not...then...* in two clauses for connection. This kind of case may not be alone in thesis writing and is thus worthy of more discussion.

E.g., 5.7.152 ||| In this way, the first language could be used to access prior knowledge || and at the same time teachers must introduce the relevant vocabulary during the discussion, || **otherwise*** a schema has been activated || but learning the second language has not been facilitated. ||| ---- CMAC07LI_20_543

E.g., 5.7.153 ||| Students must observe the rules strictly; || **otherwise*** he or she may ultimately be expelled from the group. ||| ---- CPhD11_0007_963

The word *otherwise* for paratactically linking two clauses with a negative condition relationship together has to lead to a clause in the secondary position in a clause complex because of the mention of the condition before. In this way, the intention of asking for the condition is stronger, as what is described after *otherwise* would not exist without the condition of the content being presented in the former clause. So, in E.g., 5.7.152, *the schema would not have been activated...if the first language is not used to access prior knowledge and at the same time teachers do not introduce the relevant ...*; and in E.g., 5.7.153, *he or she would not be expelled from the group if observing the rules strictly*. This means that the condition is a must for people or events with an intensification by choosing *otherwise*.

E.g., 5.7.154 ||| These verbs construe the impact as imminent, || **unless*** there is an immediate pre-emptive action from the US IDC group. ||| ---- CMAC19_05_468

E.g., 5.7.155 ||| **Unless*** one or more of these conditions pertain, || there is no good reason for leaders to apologize. ||| ---- CPhD07_0004_721

The conjunction *unless* in signifying a condition: negative relation in clauses can make the dependent clause be located at the initial (in E.g., 5.7.155) or the secondary position (in E.g., 5.7.154), in order to confirm the uniqueness of the condition. But when it is the former condition, what is being described in the clause is excluded by the choice of *unless*, i.e., when *one or more of these conditions* do not pertain, then *there is no good reason for...* When the dependent clause is at the secondary position, the condition described after *unless* is more likely to be overlooked by the writer as normally *these verbs construe the impact as imminent*, except for some specific condition, i.e., *there is an immediate preemptive ...*

E.g., 5.7.156 ||| If it does, || speaker's intention can be reflected by it, || **if not***, || the recipients will figure out some possible implicated meanings according to the speaker's utterances. ||| ---- CMAC15_01_123

E.g., 5.7.157 ||| As to verbs used in these two constructions, though they differ in their syntactic usage in these two constructions, || they have similar semantic components, || that is, verbs used in both English TtFG constructions and Chinese GtF constructions must be those [[that designate existence or presentation of some entity]], || **if not***, they will not be allowed to appear in such constructions in either language. ||| ---- CPhD07_0008_1285

Though the structure *if not* is rare in thesis writing, it is controversial in E.g., 5.7.156 as it is hard to judge whether the relation within two clauses is an addition: adversative one by *if... but if not...* or a condition: negative one which is agnate with the word *otherwise*. As the former clause is not a condition for the negation of the content being described in the latter clause, the whole clause complex is rather considered as being connected by an addition: adversative relation. But in E.g., 5.7.157, *if not* can be replaced by *otherwise* to indicate the former clause as a condition. Hence, the two clauses are paratactically connected, with the secondary clause being unable to be moved into the front.

E.g., 5.7.158 ||| **If** there are **not**, || **then*** they show more individualistic orientation. ||| ---- CMAC11CU_15_694

E.g., 5.7.159 ||| **If** the sender is **not** familiar with the current language, its codes and its culture || **then*** he or she will not be able to express anything at all. ||| ---- CPhD10_0015_1369

Derived from the word *if* for conditional relation, the phrase *if...not...then...* has a fixed sequence in paratactically connecting clauses as shown in E.g., 5.7.158 & 5.7.159. Though it resembles the word *otherwise* semantically, it cannot be replaced by it unless the initial clause, such as in E.g., 5.7.159 is changed into a positive one as *the sender must be familiar with ..., otherwise he or she will ...* Hence, this kind of structure is highly restricted by the word *if* and the negative word *not* before. Sometimes, the polarity of the condition would be mixed up, especially when the word *not* is a modifier to finite words instead of a structural indicator of the clause, like what E.g., 5.7.158 demonstrated. Then, the whole clause complex is just a condition: positive one.

E.g., 5.7.160 ||| But **no matter*** what kind of property does the test has, || score interpretation must be compared with a reference system || so as to express the meaning of the score. ||| ---- CMAC13ME_11_288

E.g., 5.7.161 ||| Owing to this kind of emotion and feeling, || so long as his father is concerned with him, || he will be cheerful || **no matter*** what happens around. ||| ---- CMAC15_05_342

The structure *no matter* in signifying the condition: negative relationship in clauses followed usually by *WH-* words have the meaning of *without any condition*. Such as in E.g., 5.7.160, the clause complex means *without any condition of what kind of property the test has, ...*; and in E.g., 5.7.161, it means *without any condition of what happens around*. Whether the dependent clause is in the initial position (see E.g., 5.7.160) or in the secondary position (E.g., 5.7.161), some facts that are described in the independent clauses would not be affected.

5.7.6.3 Concessive

Doctoral candidates prefer the choice of condition: concessive relation in clause complexes as indicated by Table 91 with its number 603, while that by master's students is 491, the difference lies in the choice of the preposition of *though* in conveying realizing this function (217 in group D while 128 in group M).

Enhancement	M	D
Condition: Concessive	491	603
although	244	267
but	?	?
despite	12	27
even if	20	24
even though	56	46
in spite of	5	4
nevertheless	3	-
though	128	217
while	2	-
without	20	13
yet	1	5

Table 91 The number of each lexis for condition: concessive clause complexes applied in Master and PhD sub-corpora

Though the preposition *although* is the most frequent structure chosen by students in both groups (267 in group D and 244 in group M) and functions almost the same as the word *though*, the difference in

choosing *though* may be explained by writing behavior from doctoral candidates instead of other reasons. Then the conjunction *even though* (56 in group M and 46 in group D) and *even if* (20 in group M and 24 in group D) are also available for connecting non-finite clauses like the word *although* does, are not so frequently discovered in the corpus as imagined. In choosing a preposition for connecting non-finite clauses for concessive relation, *despite* is more welcomed by doctoral candidates (27 by D vs. 12 by M), while *without* is more favored by master's students (20 by M vs. 13 by D). Still, the prepositional phrase *in spite of* for connecting non-finite hypotactic clauses, the word *while* for connecting finite or non-finite hypotactic clauses, and words such as *yet* and *nevertheless* for connecting paratactic clauses are useful in conveying a condition: concessive relation within clauses.

It is strange that the word *still* in paratactically combining two clauses with a condition: concessive relation has not been found in the corpus. As to the word *but*, the extension: addition: adversative *but* is hard to distinguish with the enhancement: condition: concessive *but* when it paratactically links two clauses together. The word *while* in hypotactically binding two clauses with a condition: concessive relation is unique as it can also present an extension: addition: adversative relation as well as an enhancement: temporal: same time relation, which makes the identification of the first relation difficult in practice.

E.g., 5.7.162 ||| Knowing this, Nick gets angry, || **yet*** was still half in love with her. ||| ---- CMAC14LIT_25_427

E.g., 5.7.163 ||| The semantic scope of image-G is not generally demonstrably bounded || or assuredly indicated, || **yet*** human being's construal possesses appreciable constructional magnitude. ||| ---- CPhD18_0002_642

As a marker for paratactic conditional concessive relation within two clauses, the clause beginning with *yet* would usually be located at the secondary place in a clause complex even though it can also be at the first place, which is rarely seen. The latter clause, either depicts activities or processes, represents a kind of result that is contrary to the status being described in the previous clause or the former clause, as obviously shown in E.g., 5.7.162: the former clause is a status of Nick's being angry and the secondary one is a result of his still loving to her. The word *yet* is similar to *but* at the same location or conjunctions like *even though*, *though*, *although*, etc., being placed at the beginning of the first clause. No matter what happened before, *yet* can be a prediction to the reader that the result of the mentioned event remains unchanged. Thus, *human being's construal still possesses constructional magnitude* no matter what kind of property the semantic scope of image-G has, as demonstrated in E.g., 5.7.163.

E.g., 5.7.164 ||| She initially sets up a sociosemiotic translation framework || but fails to extend it to a much wider scope, || **nevertheless*** her ideas are quite enlightening. ||| ---- CMAC11PR_10_249

When *nevertheless* is chosen to denote a paratactic enhancement: conditional: concessive relation within two clauses, it can only lead to the secondary clause rather than the initial one, with the semantic domain of the latter being broadened and the content being emphasized. Activities or status described in the former clause are not totally negated or rejected, which can be judged from the example here that the *sociosemiotic translation framework*, though restricted in scope from the author's perspective, is still preponderant when considering its brightening ideas. The author recognized the disadvantage of the mentioned framework, in fact, which would otherwise be neglected if *nevertheless* is substituted by *even though* and is put at the beginning of the whole clause complex.

E.g., 5.7.165 ||| I know it now || that you could still understand the main idea of the listening passage || **even if*** there are some new words or phrases you couldn't understand. ||| ---- CMAC13LI_12_548

E.g., 5.7.166 ||| **Even if*** "strong-willed" or "pig-headed" has the same literal meaning "being stubborn", || the former connotes "admiration for intensity of someone's will", viz. [judgement: +propriety], || while the latter connotes "frustration in dealing with someone's will", viz. [judgement: -propriety], || so negative evaluation towards the "country" is actually mapped by the connotation in the source. ||| ---- CPhD17_0001_1031

No matter where the conjunction *even if* is located in connecting two clauses and functioning as the marker for a hypotactic enhancement: conditional: concessive relation, the results of the whole stuff depicted in clauses would always be stressed that almost nothing can change or affect these kinds of results though factors described in dependent clauses led by *even if* are excluded not considered. Thus, the fact of *still understand the main idea of the listening passage* remains unchanged except for *some new words or phrases* in E.g., 5.7.165, and the connotation of *strong-willed* is still what the author explains in E.g., 5.7.166 though it also contains the same meaning *being stubborn* as another word *pig-headed*.

E.g., 5.7.167 ||| **Even though*** these variables are controlled in a limited range, || it has to be acknowledged || that these factors are important in a detailed and integrated research. ||| ---- CMAC14RE_15_667

E.g., 5.7.168 ||| Unless it is preceded by a circumstantial element, || the Hallidayan framework analyses there as the topical Theme, || **even though*** it has no representative function in transitivity. ||| ---- CPhD09_0002_465

When authors choose the phrase *even though* in either the first or the secondary clause (both are dependent ones) in denoting the concessive relation in a clause complex, normally facts would follow up the phrase, and these facts sometimes have to be accepted by people with some results discussed in the independent clauses being reached to basing on these facts after consideration. This can be told from the examples above that after accepting the fact of *controlling variables within a limited range*, a result is obvious that *some factors* are indeed important in the author's opinion (see E.g., 5.7.167) and the fact of *having no representative function in transitivity* has no effects on the result of *analyzing there as the topical Theme* after exploration by the author (see E.g., 5.7.168).

E.g., 5.7.169 ||| So the author concludes || that **though*** there is obvious field cognitive style differences among students with high English level, || the "field" difference doesn't have much influence on their choose or application of reading strategies. ||| ---- CMAC17_04_467

E.g., 5.7.170 ||| Therefore, in his study, the subjects who consistently answer all the sentences correctly (i.e., three out of three) are regarded to be consistent || **though*** he admits || that some misinterpretations may be caused by performance factors. ||| ---- CPhD15_0001_944

The mood of the word *though* in marking a concessive relation in two clauses is not so strong as *although*, even though both of them have no limit of location in leading to a dependent clause in a clause complex. Specifically, the content of the independent clause would usually not be influenced by affairs described in the dependent one with the beginning of the marker *though*, which, interestingly, is acceptable to or has already been accepted by the author. Hence, the author accepted the fact that *different students in high English level have diversities in filed cognitive styles*, but *their choice or application of reading strategies* turns out to be not affected (see E.g., 5.7.169); and the possibilities of *misinterpretations caused by performance factors*, though is acceptable by the author, have no effects on the criteria of regarding subjects as consistent or not (see E.g., 5.7.170).

E.g., 5.7.171 ||| **Although*** metonymy is considered an independent theory in cognitive linguistics, || the conceptual nature of metonymy can also be reflected by the structure of categories. ||| ---- CMAC11SE_19_1093

E.g., 5.7.172 ||| In this study, the researcher has used L2 and foreign language interchangeably to refer to English, || **although*** for Chinese students, English language is their foreign language. ||| ---- CPhD10_0001_90

Being different from conditions of the word *though*, contents mentioned in the dependent clause led by *although* most of the time would be ignored or dis-considered by people, but are in reality significant in the author's cognition, thus *although* instead of other words or conjunctions are chosen to demonstrate the concessive relation between clauses. In E.g., 5.7.171, without referring to the independence of metonymy as a *theory in cognitive linguistics*, at least the word *also* in the independent clause is redundant, and no comparison or supplement can be made to the property described in the dependent one. In E.g., 5.7.172, details in the depend clause are more explanatory for the author's using the name *L2* or *foreign language* to refer to English. So, *although* in revealing concessive relation is more about the author's concern of importance.

E.g., 5.7.173 ||| **While*** online teaching model has been recognized by most students, || it still needs improvements. ||| ---- CMAC11LI_9_88

A very unusual and even strange usage is the word *while* in denoting to a conditional: concessive relation, which would lead to a whole clause complex and can be seen from E.g., 5.7.173 that *while* is not conditional: temporal or addition: adversative here. Suppose it is substituted by *though* or *although*, this is just what the author intends to convey semantically. One explanation can be the meaning transfer from temporal relation that the improvement of the online teaching model is necessary based on the fact of students' recognition by the time when the author mentioned it; another is the misuse of *while* by the Master student, which is reluctantly to acknowledge here.

E.g., 5.7.174 ||| Besides, the genres of continuation tasks are relatively inadequate, || since researchers mainly focus on the alignment effect on narration writing, || paying less attention to argumentative writing, || **despite*** argumentative writing is of great importance to English learners and teachers in China. ||| ---- CMAC19_02_262

E.g., 5.7.175 ||| **Despite*** many efforts (extensive review of the English literature and proofreading by native writers of English, for instance) had been taken to bridge the gap between real-life communications in Chinese and the English counterparts in this dissertation, || the author would like to admit || that the gap is very difficult to clear out. ||| ---- CPhD17_0004_2231

Normally, the word *despite* would appear in *despite that* in a concessive relation within clauses with nouns being the Subject, though sometimes only *despite* is applied. When the dependent clause

beginning with *despite* is after the independent one, the contents in the former are what the author considers to be more significant. So, the emphasis in E.g., 5.7.174 is the appeal by the author or the necessity of researchers focusing more on *argumentative writing* instead of the *alignment effect on narration writing*. When the sequence of two clauses is switched just like where *despite* is located in E.g., 5.7.175, the discussions in the independent clause can be a supplement to those in the dependent ones. Hence, the author chooses *despite* to clarify his insufficient efforts in *bridging the gap between real-life communication within Chinese and English languages*, which further provides space for the limitation of his dissertation in later parts. Besides connecting with finite verbs as the preposition *in spite of* does, subsequent nouns of *despite* enable a wider scope of Subject choice and then more frequent usage in theses of dissertations.

E.g., 5.7.176 ||| **In spite of*** obtaining an IELTS score of 6.5 || when entering university in Australia, || the five participants did have different language proficiencies. || ---- CMAC20_04_539

E.g., 5.7.177 ||| **In spite of*** the fact [[that knowledge of the patterns can be considered culturally bound to the Western World and, to that extent, familiar enough for the learner]], || Hoey insists || that their teaching is not superfluous || even if their sole utility is to help students organize their thoughts. || ---- CPhD05_0003_1923

Unlike *despite*, the phrase *in spite of* usually appears at the beginning of a clause complex with finite verbs or the noun *fact* following it. When finite verbs are chosen, the Subject of the dependent clause is omitted in case of repetition and will not come up until the independent clause is connected. One character is that the contents of the dependent clause would be an addition or even negation to those of the independent one, the latter of which is what the author intends to emphasize. In E.g., 5.7.176, the same IELTS score for entering the University of the five mentioned participants has no effects on their diverse language proficiencies in the author's perspective by choosing *in spite of*; and in E.g., 5.7.177, the researcher Hoey's standpoint is stressed by using *in spite of* no matter how the fact is.

E.g., 5.7.178 ||| The common sense of one culture even directly translated to common sense for the world || **without*** concerning other thinking ways. || ---- CMAC14CU_8_215

E.g., 5.7.179 ||| To solve this problem, || Praat 4.4.20 was used to limit the duration of each pattern to 5 seconds || **without*** distorting its tonal and rhythm features. || ---- CPhD10_0009_756

The preposition *without* for a condition: concessive relation in clauses is easily mixed up with the addition: adversative relation as the dependent clause is always located at the secondary position, with a non-finite verb being after the word *without* in both relations. Being different from the addition: adversative relation, the conditional *without* excludes some conditions though it adds information to those being described in the independent clauses. Semantically, the phrase *in spite of* would include conditions being mentioned in its dependent clauses, the word *without* directly excludes all those conditions presented afterwards. The clause complex in E.g., 5.7.178 can be interpreted as *if concerning other thinking ways, the common sense of one culture would not be directly translated to common sense for the world*. And that in E.g., 5.7.179 can be *if distorting its tonal and rhythm features, Praat ... would not be used to limit the duration of each pattern to 5 seconds*. No matter how the clause complex can be interpreted, the relation between two clauses connected by the word *without* can be subjective in corpus annotation.

5.7.7 Summary of Enhancement Marker Choices

This section presents a detailed qualitative analysis of the five strategies for enhancement clause complexes, namely temporal/spatial/manner/causal/concessive, in the master's and doctoral theses and their realization in specific annotated lexis from the perspectives of the position of independent and dependent or the initial and secondary clauses, the writer's emphasis on the content of clauses, the difference in the functional realization of each category and its lexicogrammatical choices within and between systems, the writer's writing inclination or preference, the modality of verbs or whether the theme can be followed by modal verbs, the tense of verbs, the voice of clauses, and the mood of clauses, etc. Generally speaking, the causal relationship is the most frequently used strategy for enhancing clauses to express the writer's interpretation of things or opinions; followed by the temporal and the manner relationship, in which the former is widely applied in special genres such as literature, while the latter is mainly adopted in data comparison or method description in clause complexes; although there are numerous choices of annotating words for the concessive relationship, the use of these words to indicate condition: concession clause complexes is not a popular enhancement strategy among thesis written by students, this can be attributed to the directness and assertiveness of student opinions expressed in academic writing genres; the rarity of the spatial strategy is not difficult to explain, given that the academic writing register does not have high requirements for spatial description. In general, the enhancement is the most important relation of clause connection in clause complexes.

5.8 Conclusion for Expansion Conditions

This chapter examined the discrepancies in the utilization of expansion clause complexes in master's and doctoral theses by employing a mixed analytical approach. Through quantitatively comparing and analyzing the intra-system paratactic and hypotactic expansion choices, the three major categories of the inter-system, i.e., elaboration, extension, and enhancement, and the intersection of parataxis and hypotaxis with the aforementioned three categories, this chapter confirms that some categories are under-selected or wrongly selected in master's texts, while some are over-selected. The qualitative analysis of the grammatical selection and corresponding specific lexical realization in clause complexes of the three parts of the elaboration system (exposition, exemplification, and clarification), the three parts of the extension system (addition, variation, and alternation), and the five sub-categories of the enhancement system (temporal, spatial, manner, causal, and concessive) indicate that grammar and vocabulary are not independent systems, but rather are situated at opposite ends of a continuum. Given the intricate and multifaceted nature of the cross-sectional selection of categories within the three major systems and taxis, this particular data set is not included in the present analysis. But it can be examined as a discrete subject. A detailed comparison and summary of the specific selection of expansion clause complexes will be presented in Section 7.1. As the presence of multiple functions for a single marker presents a challenge in the process of data annotation, it will be elaborated upon in greater detail in Section 7.3. The complex recursion phenomenon and patterns of clause complexes will be described specifically in Section 7.4. In general, this chapter is comprehensive and intricate. The findings of the research demonstrate the distinctions in the grammatical features associated with the expansion of master's and doctoral dissertations. Moreover, the research methods employed in this chapter serve as a valuable reference and guiding framework for the investigation of other grammatical categories in SFG.

Chapter 6 Projection

6.0 Introduction

To answer questions raised in Section 1.3: 1. how are logico-semantic relations realized by the lexicogrammatical continuum in detailed texts; 2. do Master theses and PhD dissertations have variances in each grammatical category, and where do the diversities reflect in; can comparison results of these categories prove that PhD dissertations are more complex than MA theses in choosing logico-semantic relations; and 3. do these theses have a preference in choosing words/phrases that marking logico-semantic relations in clause complexes; what meanings do these markers convey and what functions do they perform in academic discourse, this chapter will be centered around the quantitative and qualitative analysis of projection choices in thesis writing from master's and doctoral students.

In detail, Section 6.1 will focus on the comparison of some basic indexes around projection in two groups, such as the general information of projection, the proportion of paratactic and hypotactic projections in clause nexuses, in corresponding superordinate structures and in projections, and the composition of projection, i.e., the proportion of locutions and ideas, to draw forth following sections.

Sections 6.2 & 6.3 are on locution, the former of which will compare the general condition of locution in all clause nexuses in two groups, followed by paratactic or hypotactic locution selections in categories such as clause nexuses, paratactic/hypotactic structures, paratactic/hypotactic projections, and locutions; and the latter concerns about lexical choices of projecting verbs with diverse kinds and relative frequencies, the classification of different types on the subjectivity-objectivity cline that these verbal verbs perform in projecting proposals or propositions by writers to interact with authors or readers in academic writing.

Sections 6.4 & 6.5 will discuss the analogous condition in the idea that the former contrasts the general adoption of ideas in all clause nexuses in the Master group with that in the doctoral group, the proportion of paratactic or hypotactic ideas in denominators like clause nexuses, paratactic/hypotactic structures, paratactic/hypotactic projections, and all ideas; and the latter emphasizes on lexical choices of projecting verbs with their kinds and corresponding frequencies, as well as the classification of various types on the subjectivity-objectivity cline that mental verbs function in projecting proposals or propositions by students to communicate with authors in spiritual or to demonstrate viewpoints to readers in thesis writing.

This chapter aims to reveal grammatical-lexical choices of projections in meaning conveyed by Master and doctoral students in their theses.

6.1 Projection Comparison

In this section, the comparison between texts written by Master and doctoral students will concentrate on projection the general conditions of projection, paratactic projection proportion, hypotactic projection proportion, and the composition of projection, i.e., the proportion of locution and idea will all be quantitatively compared to investigate diversities of students' choosing these grammatical structures in theses writing, thoughts conveying, knowledge transmitting and academic functions performing.

6.1.1 General Distribution of Projection

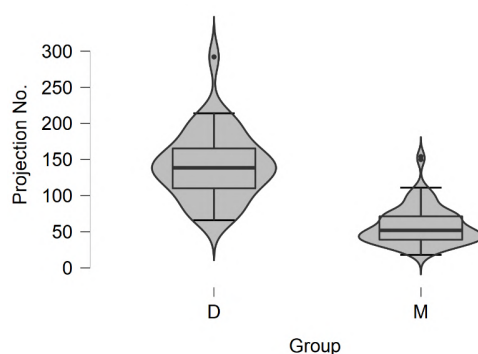


Figure 135 The boxplots of projection distribution in Master and PhD sub-corpora
The average number of group D (142), however, is also around 2.4 times that of group M (58.6), which is

the same condition as that of expansion in the two groups. From Table 6.1.1-1 (see Appendix 6.1) and Figure 135, data in group D are normally, leptokurtic (the kurtosis value is 1.642) distributed with a skew to the right tail (the skewness value is 0.828) because of one exception while in group M are abnormal, leptokurtic (the kurtosis value is 2.342) distributed with a short skew to the right tail (the skewness value is 1.298) and even trimodal except for two cases with high numbers.

6.1.1.1 Projection/Clause Nexuses

The percentage of clause nexuses per chosen sentences in each text in two groups (see Section 4.3.3.3) is almost the same, and the number of clause nexuses in the two groups also has no big difference, the proportion of projection per clause nexuses in each text in two groups are hence of necessity in calculating and comparing.

Seeing from the mean of the two groups (Table 6.1.1.1-1 in Appendix 6.1), only 19% of clause nexuses are projection structures, and the difference is quite small. Being opposite from the condition of expansion, the highest proportion of projection is only 35.63%, and the lowest can be 7.64% (both in group M). The data distribution is the inverse condition of that described in the expansion part that data in group D is platykurtic (the kurtosis value is -0.869) and left-sided skew (the skewness value is -0.176) while in group M is a little bit leptokurtic (the kurtosis value is 0.269) and right-sided skew (the skewness value is 0.571). When the division of skewness to the standard error of skewness is performed, the value in group D is -0.425 (within $-2/+2$) while in group M is 2.122 (larger than $+2$), the data in group D is thus normally distributed while in group M is abnormal; when the division of kurtosis to the standard error of kurtosis is considered, the value in group D is again -0.174 (within $-2/+2$) and in group M is 0.515 (also within $-2/+2$), and the Shapiro-Wilk value in two groups are all near to 1, and the P-value of Shapiro-Wilk is larger than 0.05, it is not easy to judge whether the data in two groups are normally distributed or not. Thus, the boxplots (Figure 136) are referred to except in one case in group M, other data are normally distributed in two groups, which is the same with the condition of expansion.

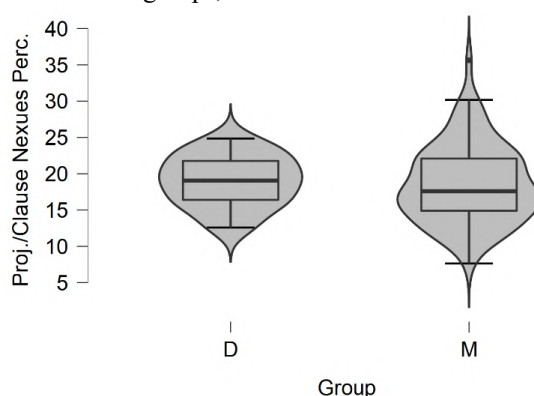


Figure 136 The boxplots of expansion per clause nexuses proportion in Master and PhD sub-corpora For the independent samples t-test, the results can be the opposite of that of expansion (see Section 5.1.1.1) though test results change slightly (Table 92): the p-value is 0.222 (the alternative hypothesis that group D is greater than group M is certified), the VS-MPR value is 1.102 (the possibility of accepting the alternative hypotheses is 1.102 times of neglecting it), and the Hodges-Lehmann Estimate value is 0.769 (the variance between two groups is only 0.769%, which can be quite small). Again, the Rank-Biserial Correlation is 0.093, which means that the effect size of this test is almost zero.

Independent Samples T-Test

	W	df	p	VS-MPR*	Hodges-Lehmann Estimate	Rank-Biserial Correlation
Proj./Clause Nexuses Perc.	1399.500		0.222	1.102	0.769	0.093

Note. For the Mann-Whitney test, effect size is given by the rank biserial correlation.

Note. For all tests, the alternative hypothesis specifies that group D is greater than group M.

Note. Mann-Whitney U test.

* Vovk-Sellke Maximum p -Ratio: Based on a two-sided p -value, the maximum possible odds in favor of H_1 over H_0 equals $1/(-e p \log(p))$ for $p \leq .37$ (Sellke, Bayarri, & Berger, 2001).

Table 92 The Mann-Whitney U test for comparing the projection per clause nexuses proportion in Master and PhD sub-corpora

In the Bayesian Mann-Whitney U test, when the alternative hypothesis is that the location of group

D is smaller than that of group M, the probability of accepting the alternative one is only 0.149 times of neglecting it (See Table 93) with a small effect size (See Figure 137), and the Rhat value is 1.082, which means that the between- and within- chain variance of the MCMC estimates for the delta parameter is not convergent, i.e., a difference exists between two groups.

Bayesian Mann-Whitney U Test

	BF ₀	W	Rhat
Proj./Clause Nexuses Perc.	0.149	1399.500	1.082

Note. For all tests, the alternative hypothesis specifies that the location of group D is smaller than the location of group M.

Note. Result based on data augmentation algorithm with 5 chains of 1000 iterations.

Table 93 The Bayesian Mann-Whitney U test for comparing the projection per clause nexuses proportion in Master and PhD sub-corpora

To demonstrate the detailed difference between the two groups, inferential plots with different alternative hypotheses are also referred to: when the alternative hypothesis is group D being unequal to group M, the possibility of accepting the original hypothesis (group D = group M) is 3.725 times (See Figure 137, left 1) of rejecting it; when the alternative hypothesis is that group D > group M, the possibility of accepting the original hypothesis is only 2.58 times (See Figure 137, right 1) of rejecting it; and when the alternative hypothesis is set as group D < group M, the probability of accepting the original hypothesis (group D > group M) can be as high as 6.698 times (See Figure 137, left 2) compared to rejecting it. So, the highest value is accepted, i.e., the hypothesis that group D has a larger projection per clause nexuses proportion than group M is accepted.

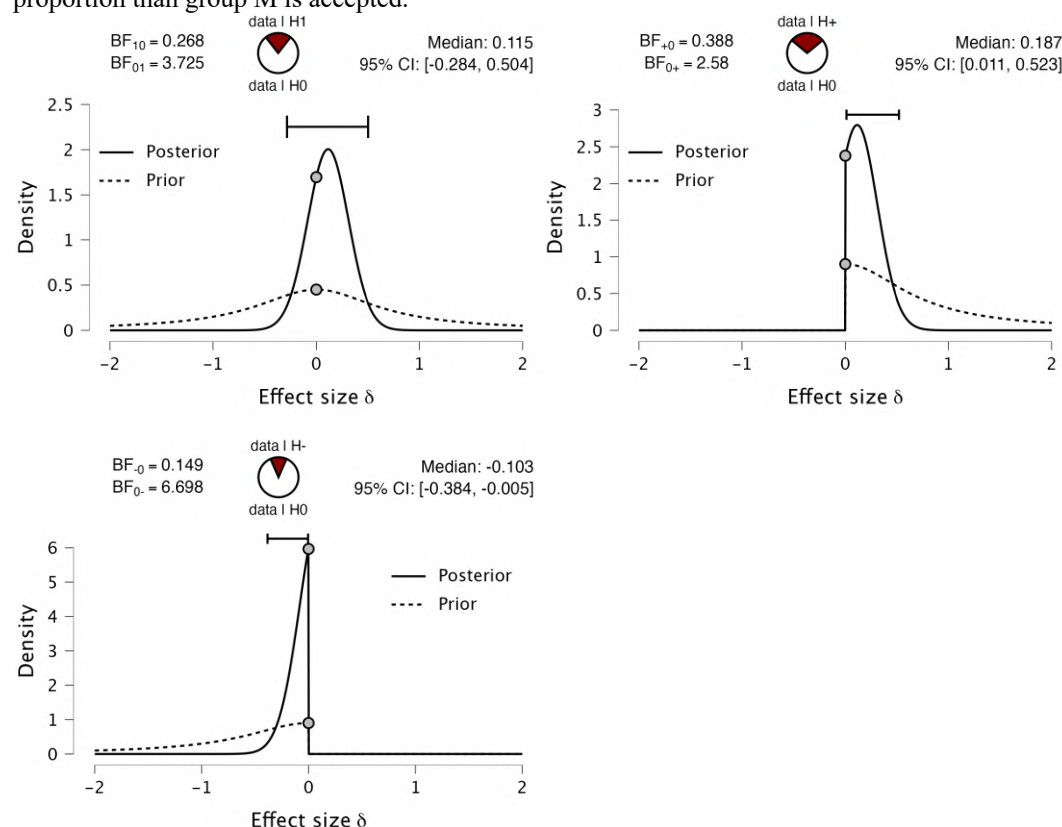


Figure 137 The inferential plots comparing the projection per clause nexuses proportion in Master and PhD sub-corpora with different alternative hypotheses

6.1.2 The Distribution of Paratactic Projection

Though the choice of projection in all clause nexuses by doctoral students outnumbers that of Master students, the condition of paratactic projection is still unknown. Thus, in this section, the number of paratactic projection choices and the proportion of it in all clause nexuses, in parataxis, and in projection structures chosen by two groups will be demonstrated clearly and then compared one by one.

The total number of paratactic projection structures found in the two groups is almost the same and

small (the sum in group D is 269 and in group M is 242) when the highest number in the Master group is only 19 and in the doctoral group can reach up to 83 according to personal choice, the lowest number, of course, is zero (Table 6.1.2-1 in Appendix 6.1) in two groups. The mean number of paratactic projections in chosen texts in two groups gets close to each other (8.4 in group D and 3 in group M), but data are right-sided skewed distributed as judged from the P-value of Shapiro-Wilk ($<.001$) above and Figure 138 that except one extremely high number (larger than 80) and one case that outnumbers 40 in group D, and two cases being larger than 20 in group M, most are normally and leptokurtic distributed.

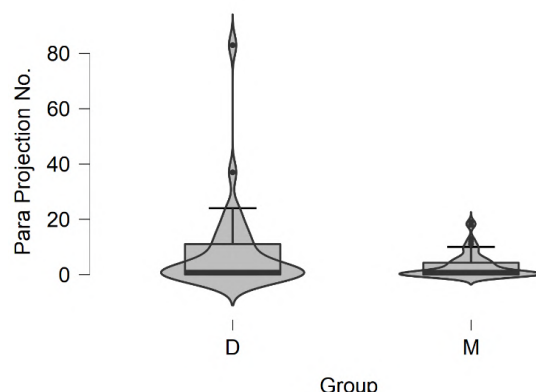


Figure 138 The boxplots of the number of paratactic projections in Master and PhD sub-corpora

6.1.2.1 Projection Parataxis/Clause Nexuses

As the number of paratactic projections is quite small, the proportion of it in all clause nexuses in two groups would be hard to compare. From Table 6.1.2.1-1 (see Appendix 6.1), the proportion can only be 8.27% in group D and 6.55 % in group M, and the mean proportion in the two groups is no larger than 1%. Data in the two groups are both right-sided skewed (the skewness value is 3.222 in group D and 1.846 in group M) and leptokurtic (the kurtosis value of group D is 12.097 and of group M is 3.123) distributed when judging from the P-value of Shapiro-Wilk that they are smaller than 0.001 in both groups, and the Shapiro-Wilk values are both far away from 1 (0.607 in group D and 0.722 in group M). Another method is to see the division of skewness to the standard error of skewness, and kurtosis to the standard error of kurtosis that values in group D are 7.78 and 16.19 (larger than +2), respectively, and in group M are 6.86 and 5.87 (also larger than +2) correspondingly. A clearer description can be seen in Figure 139 below that data are bimodal distributed in two groups except for one peculiar case in group D and 5 cases in group M.

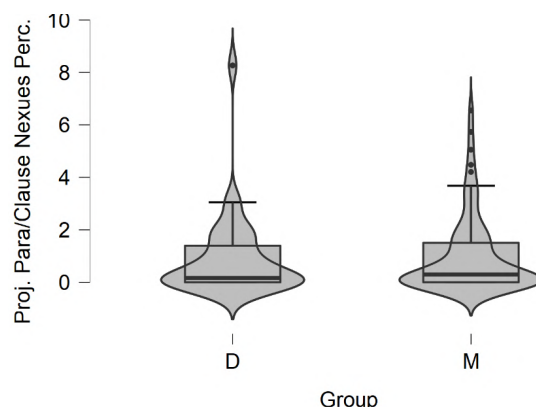


Figure 139 The boxplots comparing the proportion of paratactic projection per clause nexuses in Master and PhD sub-corpora

For abnormally distributed data, the suitable test for comparison can be the Mann-Whitney U test that when the alternative hypothesis is set as group D being unequal to group M, the p-value turns out to be 0.916 (Table 94), meaning that the original hypothesis (group D = group M) would rather be accepted. The Hodges-Lehmann Estimate value (-1.936×10^{-5}) indicates that the difference between the two groups can be ignored, though group D has a slightly smaller value compared to group M. The Rank-Biserial Correlation is only -0.012 , the effect size of the test is thus almost zero.

Independent Samples T-Test

	W	df	p	Hodges-L ehmann Estimate	Rank-Biserial Correlation	SE Rank-Biserial Correlation
Proj. Para/Clause Nexuses Perc.	1264.000		0.916	-1.936×10^{-5}	-0.012	0.121

Note. For the Mann-Whitney test, effect size is given by the rank biserial correlation.

Note. Mann-Whitney U test.

Table 94 The Mann-Whitney U test for comparing the paratactic projection per clause nexuses proportion in Master and PhD sub-corpora

When the Bayesian Mann-Whitney U test is conducted, the Rhat value is 1.027, suggesting that the between- and within-chain variance of the MCMC estimates for the delta parameter is not convergent, i.e., a difference between two groups deed exists. So, the test with diverse hypotheses is referred to: when the alternative hypothesis is set as group D being unequal to group M, the possibility of rejecting it is 4.395 times (See Figure 140, left 1) accepting it; when the alternative hypothesis is that group D having larger average proportion than group M, the probability of rejecting it is 5.047 times (Figure 140, right 1) accepting it; and the chance of rejecting the alternative hypothesis (group D < group M) is 3.893 times (Figure 140, left 2) accepting it. So, doctoral students have a smaller proportion of paratactic projection in all clause nexuses than Master students, though the difference is slight and comes from the preference for paratactic projections by some Master students.

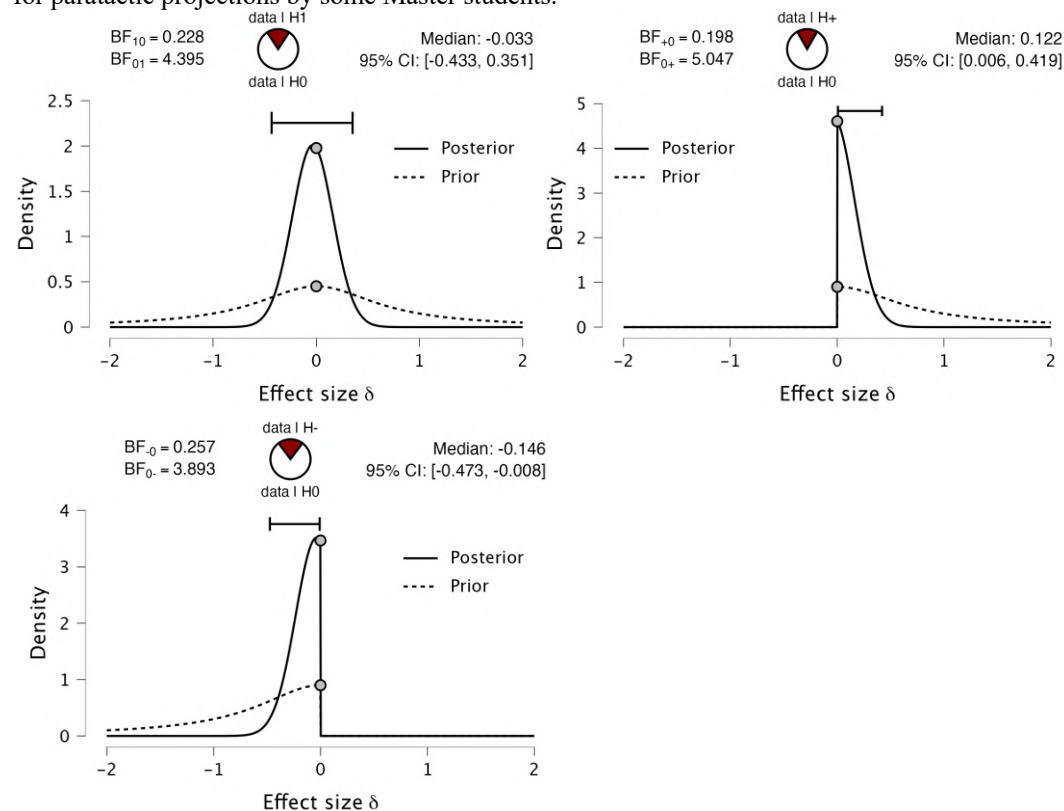


Figure 140 The inferential plots comparing the paratactic projection per clause nexuses proportion in Master and PhD sub-corpora

6.1.2.2 Projection Parataxis/Parataxis

The index paratactic projection per parataxis investigates the proportion of paratactic projection versus paratactic expansion that from Table 6.1.2.2-1 (see Appendix 6.1), only around 2.5% (2.64% in group D and 2.52% in group M) structures in all parataxis are projections. But in some cases, the proportion can reach up to 22.19% (group D) or 15.94% (group M). Data are both right-sided skew (the skewness value in group D is 2.924 and in group M is 1.855) and leptokurtic (the kurtosis value in group D is 10.388 and in group M is 2.928) distributed, which can be clearer viewed from Figure 141 that except three peculiar cases with high proportions, data in group D is bimodal; and except 5 cases with high possibilities, data in group M tends to be normal. When values of skewness are divided by the standard error of skewness, and of kurtosis are divided by the standard error of kurtosis, the former are both larger than +2 (7.06 in

group D and 6.90 in group M), and the latter are also both larger than +2 (12.84 in group D and 5.50 in group M), in addition, the Shapiro-Wilk value in two groups are far away from 1 (0.627 in group D and 0.708 in group M), with the P-value of Shapiro-Wilk being both smaller than 0.001, data are hence abnormally distributed in two groups.

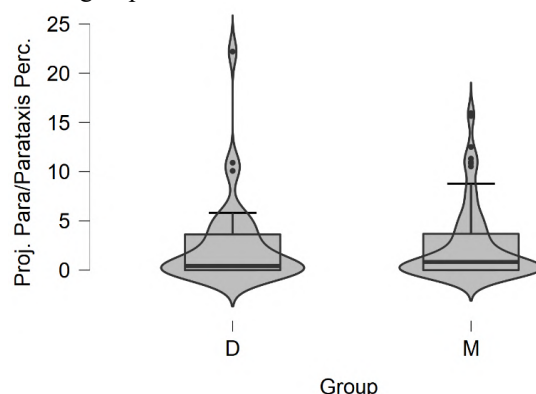


Figure 141 The boxplots comparing the paratactic projection in parataxis in Master and PhD sub-corpora

For the independent samples T-test, when the alternative hypothesis specifies that group D is larger than group M, the p-value is 0.481 (see Table 95) with VS-MPR being equal to 1, the hypothesis cannot be rejected but can also not be accepted as the effect size of this test is only 0.006 (indicated by the Rank-Biserial Correlation) and the Hodges-Lehmann Estimate value (3.511×10^{-6}) is extremely small, the difference within two groups are thus hard to tell.

Independent Samples T-Test

	W	df	p	VS-MP R*	Hodges-L ehmann Estimate	Rank-Biseri al Correlation	SE Rank-Biseri al Correlation
Proj. Para/Parataxis Perc.	1287.500		0.481	1.000	3.511×10^{-6}	0.006	0.121

Note. For the Mann-Whitney test, effect size is given by the rank biserial correlation.

Note. For all tests, the alternative hypothesis specifies that group D is greater than group M.

Note. Mann-Whitney U test.

* Vovk-Sellke Maximum p -Ratio: Based on a two-sided p -value, the maximum possible odds in favor of H_1 over H_0 equals $1/(-e p \log(p))$ for $p \leq .37$ (Sellke, Bayarri, & Berger, 2001).

Table 95 The Mann-Whitney U test for comparing the paratactic projection in parataxis in Master and PhD sub-corpora (with the null hypothesis group D < group M)

Again, the Bayesian Mann-Whitney U test is conducted for detailed comparison within two groups. The testing result, i.e., the Rhat value is 1.085 (larger than 1), indicating that a difference exists between the two groups. When the alternative hypothesis is chosen as group D being unequal to group M, the possibility of rejecting it is 4.637 times (See Figure 142, left 1) than accepting it; the possibility of accepting the hypothesis that group D is smaller than group M is 4.492 times (Figure 142, right 1) of rejecting it; and the chance of accepting the hypothesis that group D being greater than group M is 4.793 times (Figure 142, left 2) of rejecting it, with a medium effect size. Thus, the proportion of paratactic projection in all parataxis written by doctoral students can be slightly larger than that of group M, which corresponds to the testing result of paratactic expansion per parataxis proportion described in Section 5.1.2.2 that the latter has a reverse outcome though differences within two groups are minimal. If peculiarities in two groups are not taken into consideration, the result might maintain that two groups are the same in choosing paratactic projection (or paratactic expansion) in parataxis as projection has a limited ratio in parataxis no matter its grammatical structures, logico-semantic markers choices, by personal writing preference or because of writing requirement for text chunks.

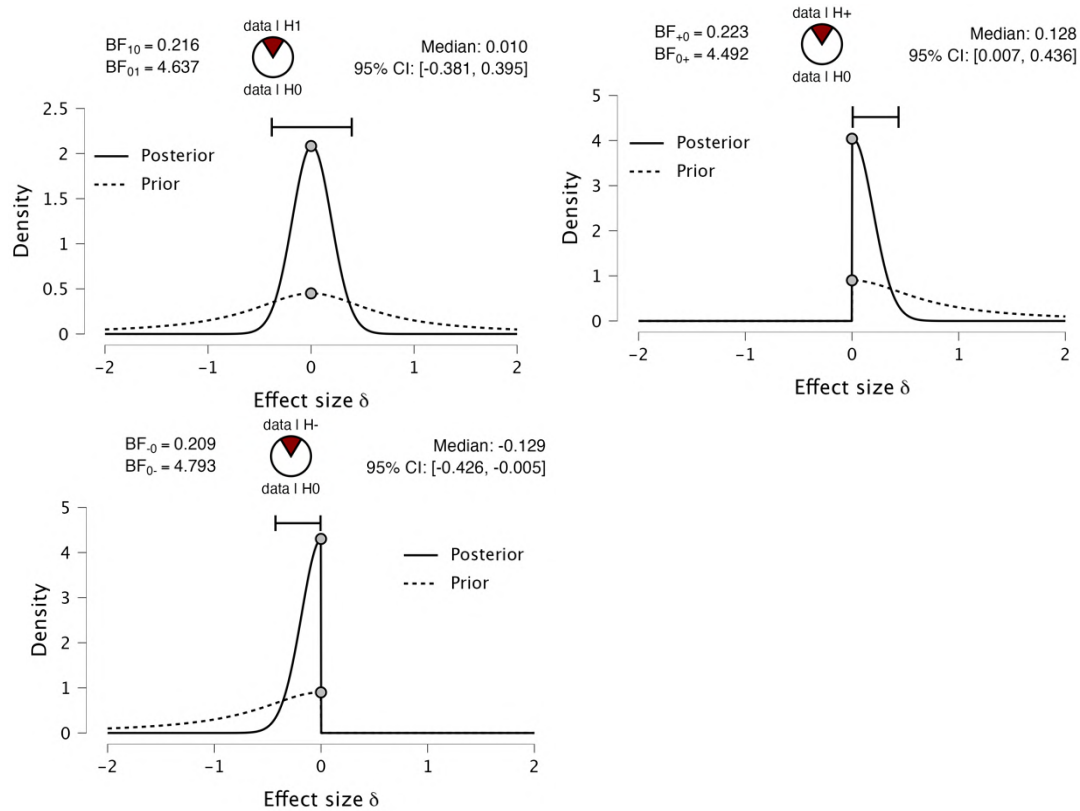


Figure 142 The inferential plots comparing the paratactic projection per parataxis in Master and PhD sub-corpora (with different alternative hypotheses)

6.1.2.3 Projection Parataxis/Projection

The index paratactic projection per projection indicates the proportion of paratactic projection versus hypotactic projection structures in all projection choices because of the small number of paratactic projections; the average proportion is only 4.88% in group D and 5.75% in group M, though the proportion can reach up to 44.62% in some doctoral dissertations or 40.91% in some Master theses (See Table 6.1.2.3-1 in Appendix 6.1). Data in both groups are right-sided (the skewness value is 3.387 in group D and 2.127 in group M), leptokurtic (the kurtosis value in group D is 14.261 and in group M is 5.001), and abnormally distributed as the division of the skewness value by the standard error of skewness (8.18 in group D and 7.91 in group M) and the kurtosis value by the standard error of kurtosis (17.63 in group D and 9.4 in group M) in two groups are both larger than +2, the Shapiro-Wilk value in both groups is far away from 1 (0.595 in group D and 0.703 in group M), and the P-value of Shapiro-Wilk are both smaller than 0.001. Figure 143 confirms the description that except for one case in group D (larger than 40%) and three cases in group M (larger than 20%), most of the proportions are smaller than 20% and normally distributed.

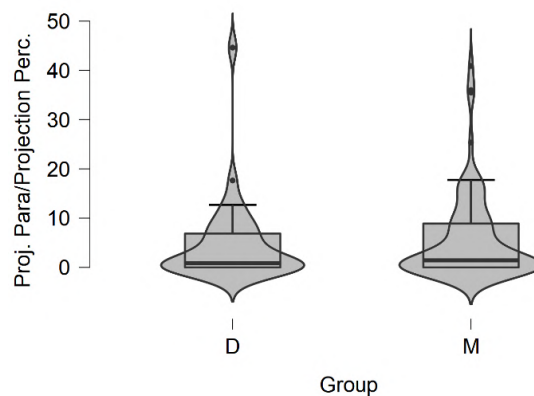


Figure 143 The boxplots comparing the proportion of paratactic projection in projections in Master and PhD sub-corpora

For abnormally distributed data, the Mann-Whitney U test is first referred to. When the alternative hypothesis is group D being smaller than group M, the p-value is 0.386 (see Table 96), and the VS-MPR

value is 1, the alternative hypothesis can be accepted in a sense. But the Hodges-Lehmann Estimate value is quite small (-3.945×10^{-5}), and the value of the effect size (Rank-Biserial correlation) is only -0.034, the difference between the two groups is thus not big.

Independent Samples T-Test

	W	df	p	VS-MP R*	Hodges-L ehmann Estimate	Rank-Biseri al Correlation	SE Rank-Biserial Correlation
Proj. Para/ Projection Perc.	1236.500		0.386	1.000	-3.945×10^{-5}	-0.034	0.121

Note. For the Mann-Whitney test, effect size is given by the rank biserial correlation.

Note. For all tests, the alternative hypothesis specifies that group *D* is less than group *M*.

Note. Mann-Whitney U test.

* Vovk-Sellke Maximum *p* -Ratio: Based on a two-sided *p*-value, the maximum possible odds in favor of H_1 over H_0 equals $1/(-e p \log(p))$ for $p \leq .37$ (Sellke, Bayarri, & Berger, 2001).

Table 96 The Mann-Whitney U test for comparing the proportion of paratactic projection in projections in Master and PhD sub-corpora

In the Bayesian Mann-Whitney U test, the Rhat value is 1.045, which is not 1, i.e., the between- and within-chain variance of the MCMC estimates for the delta parameter is not convergent; a difference thus can be found within two groups. When the alternative hypothesis is that two groups are unequal for the proportion, the possibility of accepting the original hypothesis (group *D* = group *M*) is 4.081 times (See Figure 144, left 1) of rejecting it; when the alternative hypothesis is that group *D* being bigger than group *M*, the chance of accepting the original one is 5.602 times (Figure 144, right 1) of rejecting it, with a medium effect size; and when the alternative hypothesis is group *D* being smaller than group *M*, the probability of rejecting it is 3.209 times (Figure 144, left 2) of accepting it. So, the proportion of choosing paratactic projection in all projections by doctoral students is slightly less than that of Master students.

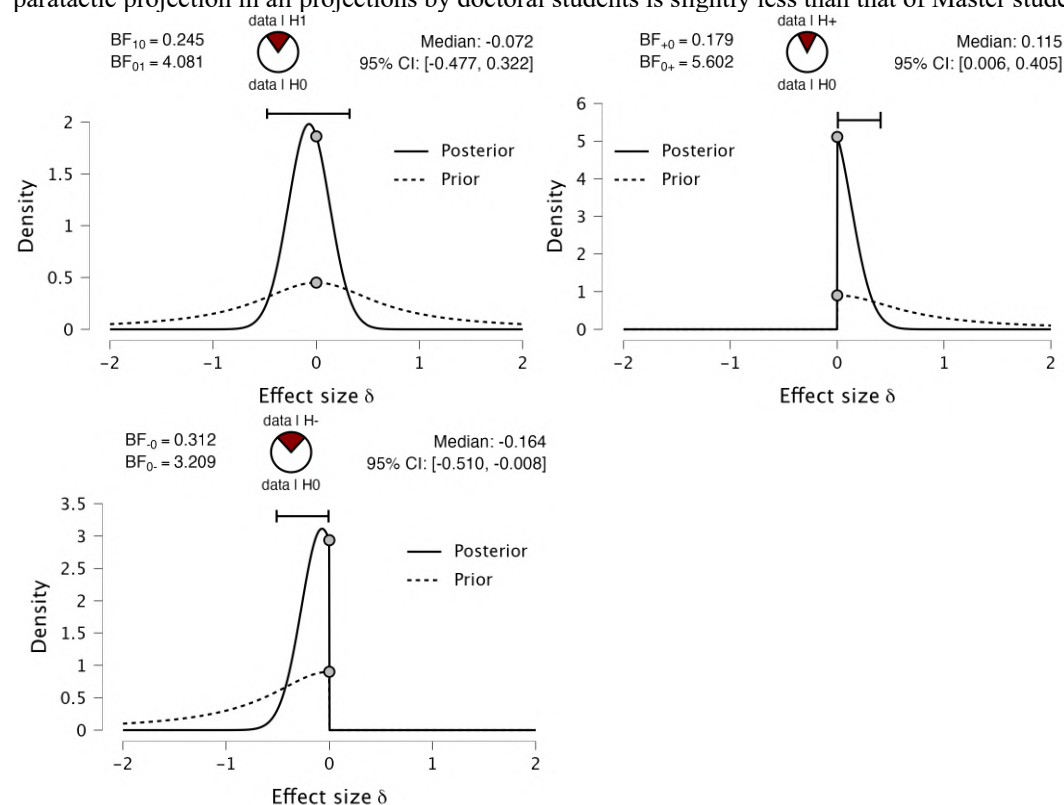


Figure 144 The inferential plots comparing the proportion of paratactic projection in projections in Master and PhD sub-corpora (with different alternative hypotheses)

6.1.3 The Distribution of Hypotactic Projection

As the proportion of paratactic projection in all clause nexuses and projections has been proved to be less, hypotactic projection then takes a large proportion compared to paratactic ones. This section focuses on the number of hypotactic projections, the proportion of hypotactic projection in all clause nexuses, all

hypotactic structures, and projections within two groups to investigate whether students have diversities in choosing the grammatical structure.

In total, 4282 hypotactic projections in doctoral dissertations and 4458 in Master theses can be found that the average of the former is almost 2.5 times of the latter (133.81 in group D and 55.73 in group M), with a range from 66 to 255 in group D and 13 to 154 in group M. (See Table 6.1.3-1 in Appendix 6.1) Data in group D is platykurtic (the kurtosis value is 0.649) and normally distributed (the Shapiro-Wilk value is 0.965, almost 1; the P-value of Shapiro-Wilk is 0.371; the skewness value divided by the standard error of skewness is 1.61, within $-2/+2$; and the kurtosis value divided by the standard error of kurtosis is 0.84, also within $-2/+2$) with almost no skew (the skewness value is 0.666) while in group M is leptokurtic (the kurtosis value is 2.996) and abnormally distributed (the Shapiro-Wilk value is 0.906 but the P-value of Shapiro-Wilk is smaller than 0.001; two divisions are 5.09 and 5.63 respectively, both larger than $+2$) with a right-sided skew (the skewness value is 1.37), which can also be proved by Figure 145 that except for one case in group D and two cases in group M, other data are normally distributed within 50 to 200 in group D and 0 to 130 in group M though are slightly bimodal.

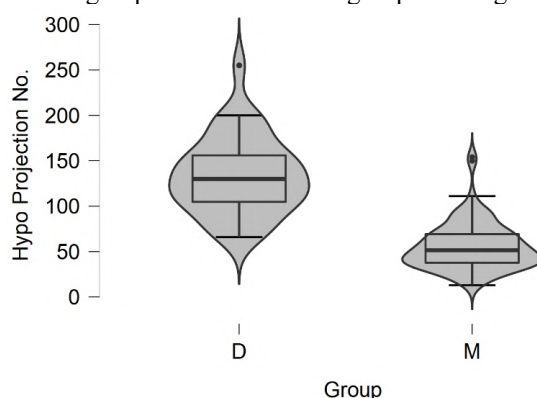


Figure 145 The boxplots comparing the number of hypotactic projections in Master and PhD sub-corpora

6.1.3.1 Projection Hypotaxis/Clause Nexuses

In all clause nexuses, the proportion of hypotactic projection in two groups seems to be the same and is no more than 20% (18.05% in group D and 17.67% in group M). In group D, it ranges from 10.26% to 24.85%, while in group M, it ranges from 6.47% to 35.63%. (See Table 6.1.3.1-1 in Appendix 6.1) Data in group D seems to be platykurtic (the kurtosis value is -0.787), normally distributed (the Shapiro-Wilk value is 0.975; the P-value of Shapiro-Wilk is 0.634; the division of the skewness value by the standard error of skewness is 0.48 and the division of the kurtosis value by the standard error of kurtosis is -0.97 , within $-2/+2$) with a short left-sided tail (the skewness value is -0.197) while in group D is platykurtic (the kurtosis value is 0.584), abnormally distributed (the Shapiro-Wilk value is 0.978; the P-value of Shapiro-Wilk is 0.179; two divisions are 2.12 and 1.10 respectively) with a short right-sided tail (the skewness value is 0.57). As indicated by Figure 146, except for one case in the Master group with the proportion larger than 35%, other data are all normally distributed though slightly bimodal.

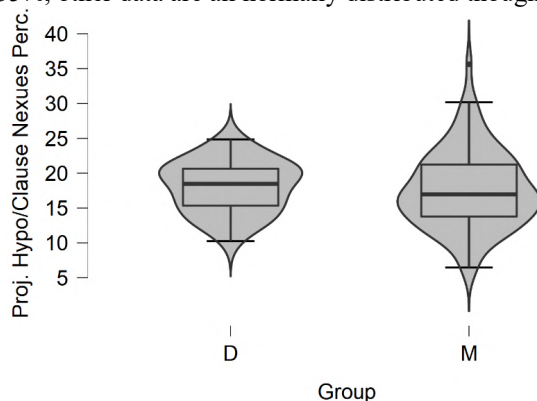


Figure 146 The boxplots comparing the proportion of hypotactic projection in clause nexuses in Master and PhD sub-corpora

For one normally while another abnormally distributed data, both the Welch t-test and Mann-Whitney U test are conducted. (See Table 97) In the former test, the result of statistics and location parameter both indicate that the proportion of hypotactic projection in clause nexuses in group D is larger than that of group M under diverse calculation methods (0.416% and 0.378% respectively); the P-value

0.339 and the SE difference value 0.908 also confirm the variance to some degree; the VS-MPR value is 1.003 (larger than 0.37), indicating that the probability of accepting the alternative hypothesis is 1.003 times of rejecting it. As the effect size value of 0.08 is quite small, more tests should be taken to confirm the statement.

Independent Samples T-Test

	Test	Statistic	df	p	VS-MP R*	Location Parameter	SE Difference	Effect Size
Proj. Hypo/Clause Nexuses Perc.	Welch	0.416	84.404	0.339	1.003	0.378	0.908	0.080
	Mann-W hitney	1396.000		0.228	1.091	0.746		0.091

Note. For all tests, the alternative hypothesis specifies that group *D* is greater than group *M*.

Note. For the Welch t-test, effect size is given by Cohen's *d*. For the Mann-Whitney test, effect size is given by the rank biserial correlation.

Note. For the Welch t-test, location parameter is given by mean difference. For the Mann-Whitney test, location parameter is given by the Hodges-Lehmann estimate.

* Vovk-Sellke Maximum *p*-Ratio: Based on a two-sided *p*-value, the maximum possible odds in favor of H_1 over H_0 equals $1/(-e p \log(p))$ for $p \leq .37$ (Sellke, Bayarri, & Berger, 2001).

Table 97 The independent samples t-test for comparing the proportion of hypotactic projection in clause nexuses in Master and PhD sub-corpora

When the Bayesian Mann-Whitney U test is referred to, the Rhat value is 1.021, meaning that a difference indeed exists between the two groups. And when the alternative hypothesis is that group *D* is unequal to group *M*, the possibility of rejecting it is 4.079 times (Figure 147, left 1) accepting it; when the alternative hypothesis specifies that group *D* is greater than group *M*, the possibility of accepting the original hypothesis (group *D* < group *M*) is only 2.867 times (Figure 147, right 1) of rejecting it; and when the alternative hypothesis is reversed, the probability of accepting the original hypothesis that group *D* having larger proportion than group *M* is 7.066 times (Figure 147, left 2) of rejecting it with a medium effect size. Thus, group *D* tends to have a larger average proportion of hypotactic projection in all clause nexuses than group *M*.

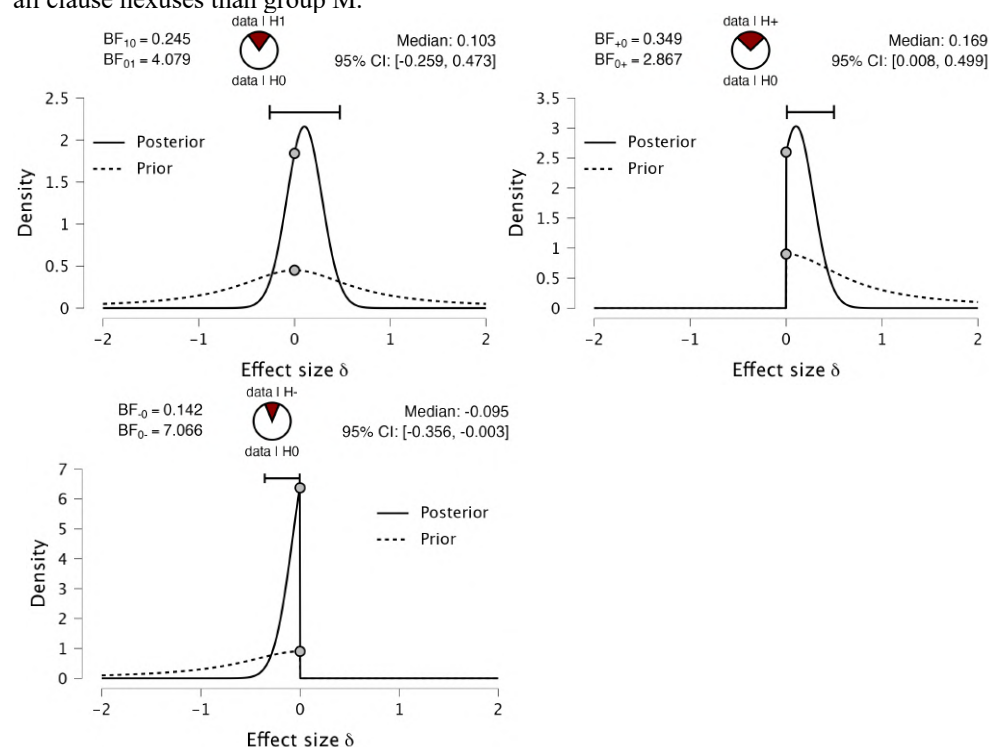


Figure 147 The inferential plots comparing the proportion of hypotactic projection in clause nexuses in Master and PhD sub-corpora (with different alternative hypotheses)

6.1.3.2 Projection Hypotaxis/Hypotaxis

The index hypotactic projection per hypotaxis discusses the choices of hypotactic projections compared

to hypotactic expansions, thus data are either reverse or the same as those described in Section 5.1.3.2: the average proportion in both groups is no bigger than 28% (27.92% in group D and 26.98% in group M), with a range in group D from 16.35% to 37.23% and in group M from 9.28% to 52.25%. (See Table 6.1.3.2-1 in Appendix 6.1) Data in group D are leptokurtic and normally distributed (as have been analyzed in the Section 5.1.3.2) with a shot left-sided tail (the skewness value is -0.214) while in group M are leptokurtic (the kurtosis value is 0.741), abnormally distributed (the P-value of Shapiro-Wilk is only 0.002 ; the division of the skewness value by the standard error of skewness is 2.97 , being larger than $+2$; though Shapiro-Wilk value 0.947 being close to 1 and the kurtosis value divided by the standard error of kurtosis is 1.39) with a right-sided skew (the skewness value is 0.798). In fact, except for five cases with high propositions larger than 40% and one with a very low proportion (9.275%) in group M, other data are still normally distributed. (See Figure 148)

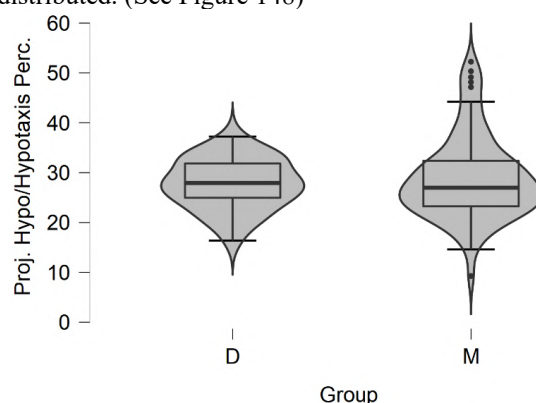


Figure 148 The boxplots comparing the proportion of hypotactic projection in hypotaxis in Master and PhD sub-corpora

Because of the skewness of data distribution in group M, both Welch and Mann-Whitney tests are conducted. (See Table 98) In the Welch test, the statistic value -0.933 and location parameter value -1.21 all indicate the proportion difference within the two groups and group D is smaller than group M; and when the alternative hypothesis is set as this, the p-value 0.177 proves the high possibilities of accepting the alternative hypothesis instead of the original one; the effect size value -0.174 certifies the validity of this test. At the same time, the effect size -0.004 for the Mann-Whitney U test suggests the invalidity of this test.

Independent Samples T-Test

	Test	Statistic	df	p	Location Parameter	SE Difference	Effect Size	SE Effect Size
Proj. Hypo/ Hypotaxis Perc.	Welch	-0.933	92.815	0.177	-1.210	1.297	-0.174	0.210
	Mann-W hitney	1275.000		0.488	-0.102		-0.004	0.121

Note. For all tests, the alternative hypothesis specifies that group D is less than group M.

Note. For the Welch t-test, effect size is given by Cohen's d. For the Mann-Whitney test, effect size is given by the rank biserial correlation.

Note. For the Welch t-test, location parameter is given by mean difference. For the Mann-Whitney test, location parameter is given by the Hodges-Lehmann estimate.

Table 98 The independent samples t-tests for comparing the proportion of hypotactic projection in hypotaxis in Master and PhD sub-corpora

This statement that group D has less proportion than group M on the proportion of hypotactic projection per hypotaxis can also be proved by the Bayesian Mann-Whitney U test as the Rhat value is 1.083 . When the alternative hypothesis is that group D is unequal to group M, the possibility of accepting the original hypothesis is 4.056 times (See Figure 149, left 1) of accepting the alternative one; when the alternative is that group D has a larger proportion than group M, the probability of accepting the original one is 6.25 times (Figure 149, right 1) of rejecting it; and when alternative is group D being smaller than group M on the index, the possibility of accepting the original hypothesis is 3.003 times (See Figure 149, left 2) of accepting the alternative one with a medium effect size. Thus, the proportion of hypotactic projection in all hypotaxis chosen by doctoral students is less than that by Master students, which is contrary to the testing results described in Section 5.1.3.2, though detailed data have changed.

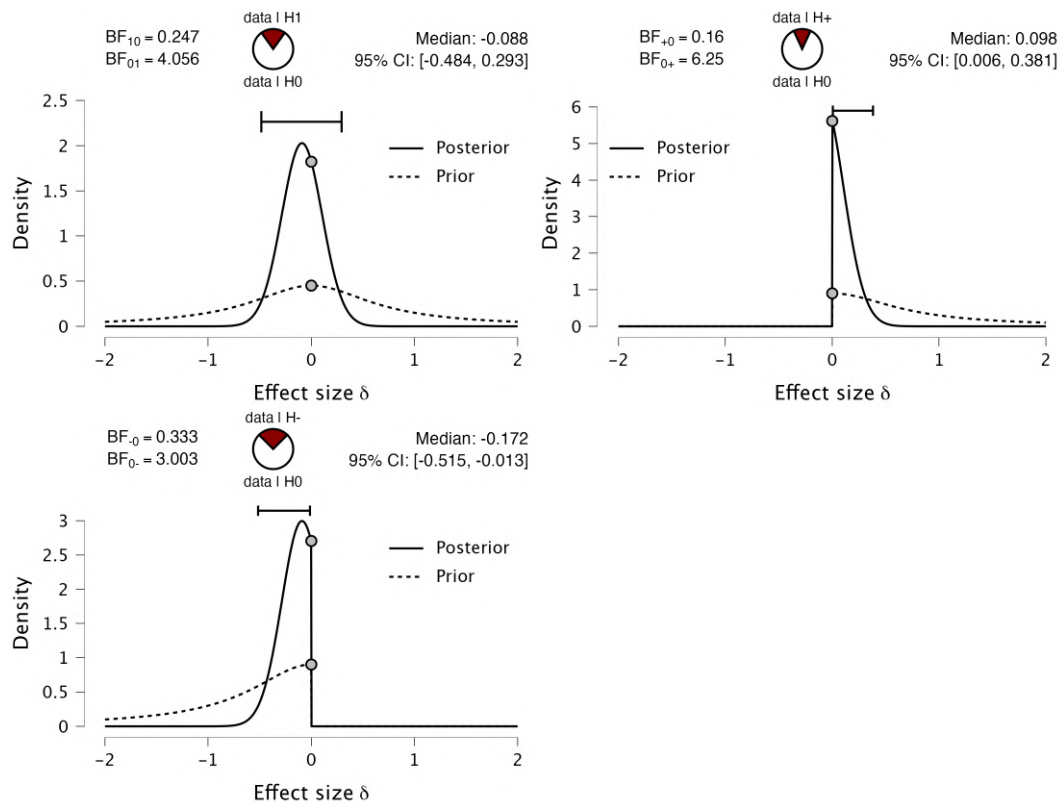


Figure 149 The inferential plots comparing the proportion of hypotactic projection in hypotaxis in Master and PhD sub-corpora with diverse alternative hypotheses

6.1.3.3 Projection Hypotaxis/Projection

The index hypotactic projection per projection explores the choice of hypotactic projections versus paratactic projections; the testing results are without no doubt reverse of those described in Section 6.1.2.3, in that the average proportion is 95.124% in group D and 94.433% in group M. The proportion ranges from 55.376% to 100% in the former group and from 59.091% to 100% in the latter one. (See Table 6.1.3.3-1 in Appendix 6.1) Though the proportion of group D has been proved to be slightly larger than that of group M, data of testing results still have changed. In the Mann-Whitney U-test, the p-value is 0.446 when the alternative hypothesis specifies that group D has a larger proportion than group M, and the VS-MPR value 1 confirms diversities within the two groups. But this difference is quite small, as proved by the Hodges-Lehmann Estimate value 8.222×10^{-6} . The effect size of this test is also small, as demonstrated by the Rank-Biserial Correlation value 0.016. (See Table 99)

Independent Samples T-Test

	W	df	p	VS-MP R*	Hodges-Lehmann Estimate	Rank-Biserial Correlation	SE Rank-Biserial Correlation
Proj. Hypo/ Projection Perc.	1300. 500		0.446	1.000	8.222×10^{-6}	0.016	0.121

Note. For the Mann-Whitney test, effect size is given by the rank biserial correlation.

Note. For all tests, the alternative hypothesis specifies that group D is greater than group M.

Note. Mann-Whitney U test.

* Vovk-Sellke Maximum p -Ratio: Based on a two-sided p-value, the maximum possible odds in favor of H₁ over H₀ equals $1/(-e p \log(p))$ for $p \leq .37$ (Sellke, Bayarri, & Berger, 2001).

Table 99 The Mann-Whitney U test for comparing the proportion of hypotactic projection in projections in Master and PhD sub-corpora

So, the Bayesian Mann-Whitney U test is also conducted, and the Rhat value is calculated as 1.01, suggesting a small difference between the two groups. As demonstrated in Figure 150, when the alternative hypothesis is that group D is unequal to group M, the possibility of rejecting it is 4.52 times (left 1) of accepting it; when the alternative one is that group D has a larger proportion than group M, the possibility of rejecting it is 4.306 times (right 1) of accepting it; and when the alternative hypothesis is set as group D being smaller than group M on the testing index, the probability of accepting the original

hypothesis is 4.76 times (left 2) of accepting the alternative one. By comparing these numbers, a conclusion can be reached that the proportion of hypotactic projection in all projections written by doctoral students is greater than that by Master students, though this difference is extremely small and can even be ignored when the result is compared with other indexes.

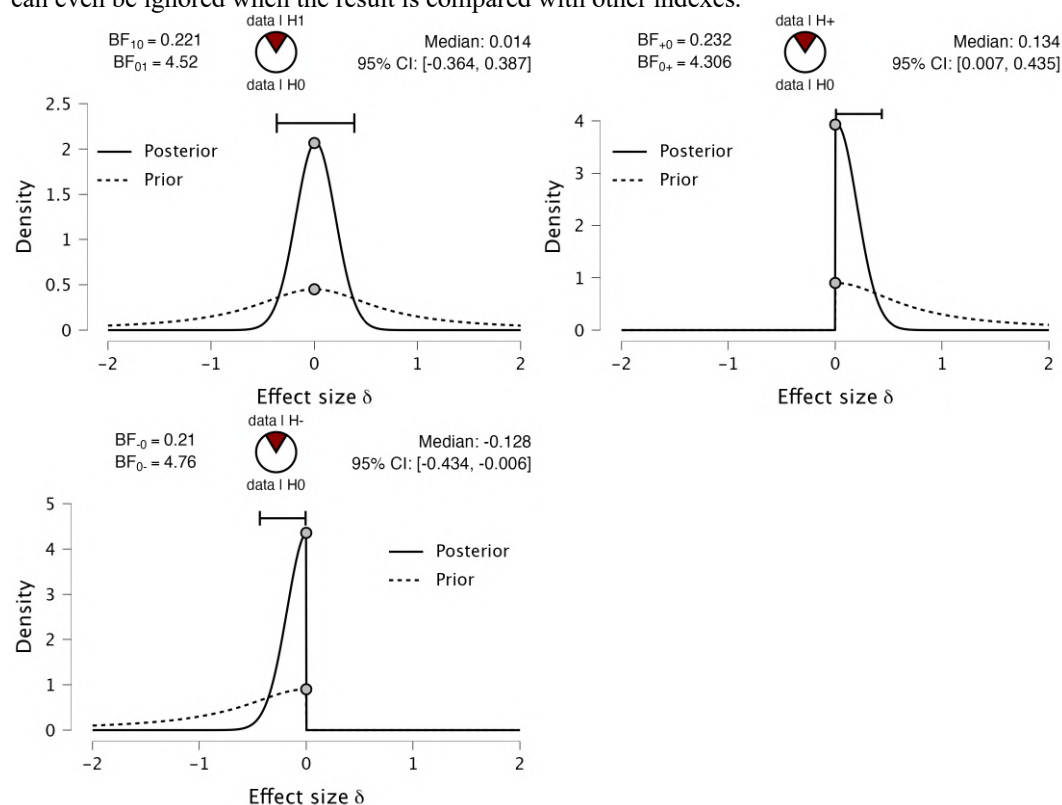


Figure 150 The inferential plots comparing the proportion of hypotactic projection in projections in Master and PhD sub-corpora with diverse alternative hypotheses

A short conclusion for paratactic projection, except for the denominator parataxis that the doctoral group is greater than the Master group, the proportion of the other two indexes (per clause complexes and projection) chosen by doctoral students are less than those by Master students; so does the proportion of hypotactic projection per hypotaxis; though the proportion of paratactic projection in all clause nexuses does not influence the choice of hypotactic projection in all clause nexuses, a reverse result was tested; and though in projection, the choice of paratactic projection by doctoral students is less than that by Master students, the proportion of hypotactic projection per projection, however, have slight difference within two groups.

6.1.4 Projection Composition

In Sections 3.1.2 & 3.1.4, the projection has been described to include locution and the idea that the comparison of projection choice should consider the conditions of choosing these two categories to investigate how the preference of specific markers can influence the frequency of organizational grammatical categories and then make up the proportion of hypotactic or paratactic projection structures in chosen texts.

6.1.4.1 Locution (Descriptive Statistics; Locution/Projection)

In clause nexuses, the total number of locutions chosen by doctoral candidates is 1,415, and by Master students is 1,242. In detailed texts, it can range from 9 to 166 in doctoral dissertations and from zero to 44 in Master theses. (Table 6.1.4.1-1 in Appendix 6.1) The average number of choosing locutions of Group D (44.219) is around 2.8 times of Group M (15.525). When Figure 151 is referred to, except for two cases larger than 100, most of the data would concentrate on 9 to 90 in group D and 0 to 40 in group M.

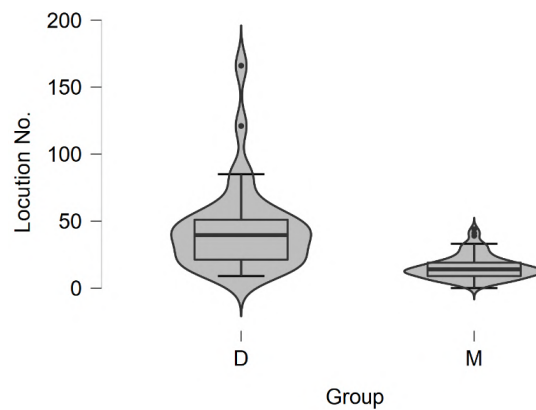


Figure 151 The boxplots of the number of locations in Master and PhD sub-corpora

The proportion of location in all projections can range from 8.65% to 65.05% in doctoral dissertations and zero to 66.67% in Master theses, with the mean in the former group (29.49%) being a little bit larger than the latter one (27.59%). (See Table 6.1.4.1-2 in Appendix 6.1) Data in both groups seem to be platykurtic (the kurtosis value in Group D is 0.382 and in Group M is 0.514), normally distributed (the Shapiro-Wilk value in group D is 0.954 and in group M is 0.979; the P-value of Shapiro-Wilk in the former is 0.191 and in the latter is 0.204; the division of the skewness by the standard of skewness in Group D is 1.61 and in Group M is 1.98, within $-2/+2$; the division of kurtosis value by the standard error of kurtosis in group D is 0.47 and in group M is 0.97, also within $-2/+2$) with a right-sided skewness as the skewness value in group D is 0.666 and in group M is 0.532. In fact, when Figure 152 is referred to, the proportion in group D mainly ranges from 8.65% to around 58% except for one that is higher than 60%, and data are normally distributed, though a little bit bimodal; and the proportion in group M are also bimodal except two cases with high proportions larger than 58%.

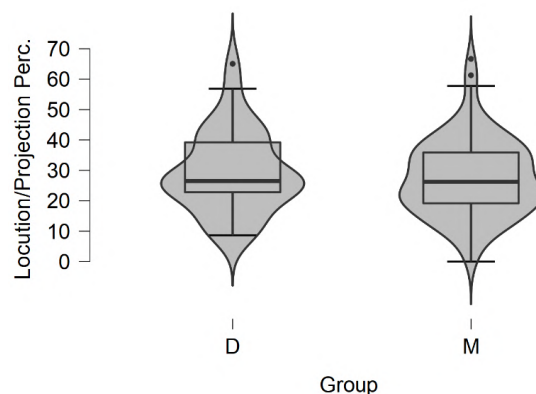


Figure 152 The boxplots comparing the proportion of location per projection in Master and PhD sub-corpora

Independent Samples T-Test

	Test	Statistic	df	p	VS-M PR*	Location Parameter	SE Difference	Effect Size	SE Effect Size
Location /Projection Perc.	Student	0.696	110	0.244	1.069	1.893	2.721	0.146	0.210
	Mann-Whitney	1388.500		0.243	1.070	1.788		0.085	0.121

Note. For all tests, the alternative hypothesis specifies that group D is greater than group M.

Note. For the Student t-test, effect size is given by Cohen's d. For the Mann-Whitney test, effect size is given by the rank biserial correlation.

Note. For the Student t-test, location parameter is given by mean difference. For the Mann-Whitney test, location parameter is given by the Hodges-Lehmann estimate.

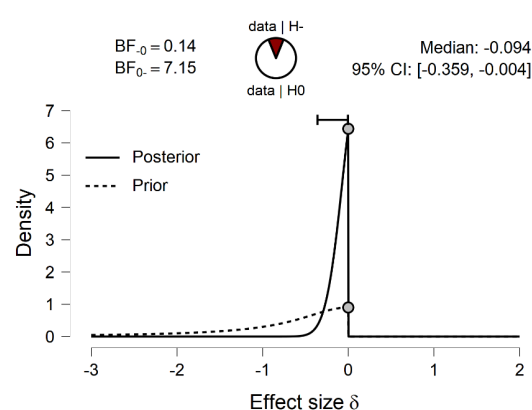
* Vovk-Sellke Maximum p -Ratio: Based on a two-sided p -value, the maximum possible odds in favor of H_1 over H_0 equals $1/(-e p \log(p))$ for $p \leq .37$ (Sellke, Bayarri, & Berger, 2001).

Table 100 The independent samples t-tests for comparing the proportion of location in projections in Master and PhD sub-corpora

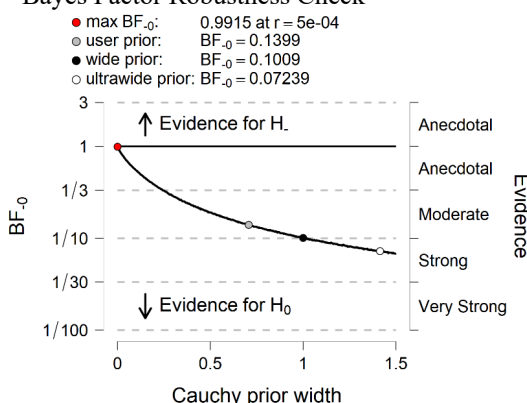
For normally distributed data, both the student t-test and the Mann-Whitney U test are conducted. From the former test in Table 100, the statistic value 0.969 and the location parameter value 1.893 both indicate differences within two groups under diverse calculation methods; the p-value 0.244, though is not smaller than 0.05, still suggests the possibility of accepting the alternative hypothesis; the VS-MPR value 1.069 is larger than 0.37, diversity hence exists. But the effect size for the student t-test is 0.146; further tests are thus necessary for comparison.

In Bayesian independent samples t-test, when the alternative hypothesis specifies that the proportion in group D is less than that in group M, the possibility of accepting the original one is 7.15 times (see Figure 153) of accepting the alternative hypothesis (when the alternative is group D being larger than group M, the probability is 2.482; and when the alternative is group D being unequal to group M, the chance is 3.685, because of space limit, are not presented here), with a medium effect size. The evidence for the alternative hypothesis, as suggested by the Bayes factor robustness check, is moderate to strong when the Bayes factor is 7.15, and the number of cases for moderate to strong evidence should at least be 90 for testing, as indicated by the sequential analysis plot.

Prior and Posterior



Bayes Factor Robustness Check



Sequential Analysis

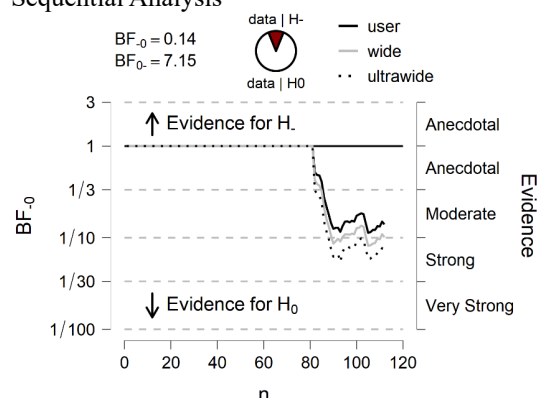


Figure 153 The inferential plots comparing the proportion of locution in projections in Master and PhD sub-corpora

So, the proportion of locution in all projections chosen by doctoral candidates is certified to be higher than that by Master students.

6.1.4.2 Idea (Descriptive Statistics; Idea/Projection)

The total number of ideas chosen by doctoral students is 3,136, and by Master's students, 3,448 that, the average number in the former group (98) is 2.28 times of the latter (43.1). In some doctoral dissertations, only 37 are chosen, while in some, the number can reach up to 155; in the Master thesis, the choice of ideas ranges from 6 to 132. (See Table 6.1.4.2-1 in Appendix 6.1) Data in group D are platykurtic (the kurtosis value is -0.708) and normally distributed (the p-value of Shapiro-Wilk is 0.676) and most of the number concentrate on 40 to 150; while in group M, except for three cases with the number being larger than 90, other are leptokurtic (the kurtosis value is 3.685) and also normally distributed (See Figure 154).

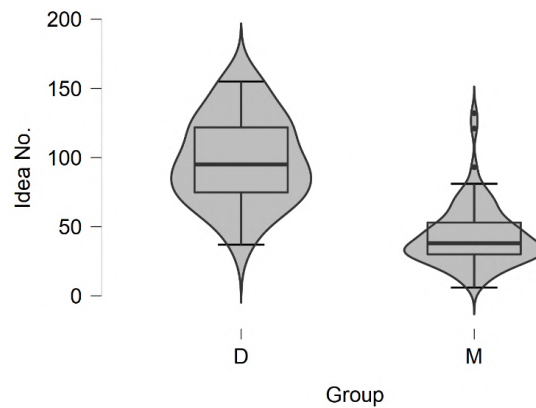


Figure 154 The boxplots of the number of ideas in Master and PhD sub-corpora

The proportion of the idea in projections is the very reverse of that described in Section 6.1.4.1 that the mean proportion in group D is 70.51% and in group M is 72.41%, and the proportion ranges from 34.95% to 91.35% in the former group and 33.33% to even 100% in Master group. (See Table 6.1.4.2-2 in Appendix 6.1) Data in both groups are platykurtic and normally distributed with a short left-sided tail, which can be deduced from statistics in the table and can also be proved by descriptions in location per projection part before. From Figure 155, it is clear that except for one case in group D and two cases in group M lower than 40%, other data are around 50% to 100% and are slightly bimodally distributed.

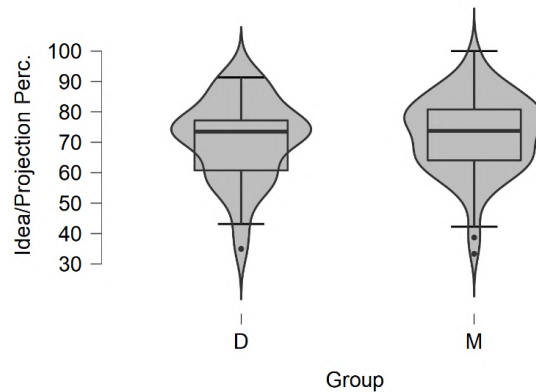


Figure 155 The boxplots comparing the proportion of ideas in projections in Master and PhD sub-corpora Tests for comparison within two groups can refer to Section 6.1.4.1 that as the proportion of locution per projection in group D is larger than that in group M, the condition of idea/projection proportion is on the contrary, i.e., fewer ideas (in projections) are chosen by doctoral candidates in writing dissertations.

6.1.5 Summary of Projection Comparison

This section discusses projection quantitatively and compares its choice within doctoral dissertations and Master theses from three aspects: the proportion of projection in all clause nexuses; projection related to taxis, i.e., the proportion of paratactic or hypotactic projection in all clause nexuses, hypotaxes, and projections respectively; and the composition of projection, i.e., the proportion of locution and idea. In clause nexuses, the proportion of projection and hypotactic projection chosen by doctoral candidates is higher than that by Master students, as well as the proportion of locution in projections; opposite testing results are received in all other testing results.

6.2 Locution (Quantitatively Comparison)

The number of locution structures and the proportion of locution in all projections have been described and compared in former sections; this section will then start from the proportion of locution in all clause nexuses, followed by paratactic or hypotactic locution proportions when the denominator is clause nexus, parataxis or hypotaxis, paratactic or hypotactic projection, and locution as the choice of locution can be intersected with taxis.

6.2.1 General Distribution of Locution (Locution/Clause Nexuses)

The average proportion of locution in all clause nexuses in two groups is no bigger than 6% (5.55% in group D and 4.96% in group M), though the proportion can be 13.61% in some doctoral dissertations or 13.06% in some Master theses. (See Table 6.2.1-1 in Appendix 6.2) Data seem to be leptokurtic (the kurtosis value is 1.156) and normally distributed (the Shapiro-Wilk value is 0.901; the P-value of Shapiro-Wilk is 0.006; the division of skewness value by the standard error of skewness is 2.86, larger than +2; the division of the kurtosis value by the standard error of kurtosis is 1.43, within $-2/+2$) in group D with a right-sided skew (the skewness value is 1.184); and platykurtic (the kurtosis value is 0.732), normally distributed (the Shapiro-Wilk value is 0.961; the P-value of Shapiro-Wilk is 0.017; two divisions are 2.91 and 1.38 respectively) in group M with a short right-sided tail (the skewness value is 0.783). However, Figure 156 clearly shows that except for two proportions larger than 12% in group D and two cases higher than 10% in group M, other data are normally distributed in two groups, and the average proportion in group D is lower than that in group M.

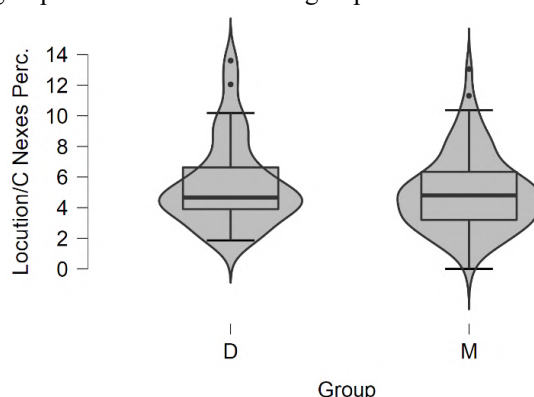


Figure 156 The boxplots comparing the proportion of locution in clause nexuses in Master and PhD sub-corpora

For controversial data distribution, the independent samples t-test method for comparing data in two groups will be the Mann-Whitney U test that when the alternative hypothesis is set as group D being greater than group M, the p-value is 0.215 (not far away from 0.05), the maximum possible odds in favor of the alternative hypothesis than the original one has reached up to 1.113 (the VS-MPR value, higher than 0.37), both the Hodges-Lehmann estimate (0.431) and the rank-biserial correlation value (0.096) indicate differences within two group due to various calculation methods, and group D indeed has larger proportion than group M. (See Table 101)

Independent Samples T-Test

	W	df	p	VS-MPR*	Hodges-Lehmann Estimate	Rank-Biserial Correlation	SE Rank-Biserial Correlation
Locution/C Nexuses Perc.	1403.000		0.215	1.113	0.431	0.096	0.121

Note. For the Mann-Whitney test, effect size is given by the rank biserial correlation.

Note. For all tests, the alternative hypothesis specifies that group D is greater than group M.

Note. Mann-Whitney U test.

* Vovk-Sellke Maximum p -Ratio: Based on a two-sided p -value, the maximum possible odds in favor of H_1 over H_0 equals $1/(-e p \log(p))$ for $p \leq .37$ (Sellke, Bayarri, & Berger, 2001).

Table 101 The Mann-Whitney U test for comparing the proportion of locution in clause nexuses in Master and PhD sub-corpora

The same testing result can also be received from the Bayesian Mann-Whitney U test that when the alternative hypothesis is that the location of group D is smaller than the location of group M, the proportion of accepting the original hypothesis (group D > group M) is 8.87 times (see Figure 157) of accepting the alternative one (other two hypotheses are 1.78 times when the alternative is that group D > group M, and 2.965 times when the alternative is that two groups are equal) with the Rhat value being 1.038 (difference exists). So, doctoral candidates have a higher average proportion of locution in all clause nexuses than Master students. As locution involves hypotactic or paratactic structures, the choice of which affects the higher proportion of locution in group D is to be probed into.

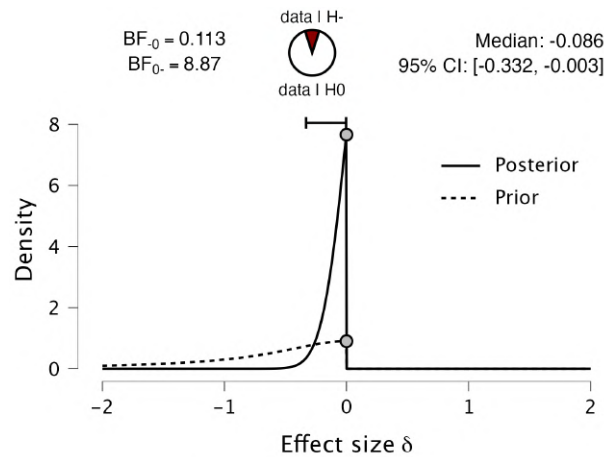


Figure 157 The inferential plot comparing the proportion of location in all clause nexuses in Master and PhD sub-corpora (with the alternative hypothesis group D < group M)

6.2.2 The Distribution of Paratactic Locution

The total number of paratactic locutions in some doctoral dissertations can be 72, and in some master's theses can be only 18, with a mean of the former group 7.125 and the latter just 2.4. (See Table 6.2.2-1 in Appendix 6.2) Data are leptokurtic and abnormally distributed with a long right-sided tail in two groups as the kurtosis values are 14.572 in group D and 5.256 in group M, the Shapiro-Wilk value is 0.551 in group D and 0.673 in group M (both are far away from 1), the P-value of Shapiro-Wilk in both groups are smaller than 0.001, the divisions of the skewness value by the standard error of skewness are 8.5 in group D and 8.35 in group M (both are larger than +2), the divisions of the kurtosis value by the standard error of kurtosis are 18.01 and 9.88 respectively (also larger than +2), and the skewness value in group D is 3.519, and in group M is 2.245. Figure 158 presents data distribution more clearly that except for one doctoral dissertation with the number of paratactic locutions being higher than 60 and another one being higher than 30, other data are bi-modally distributed and concentrate within 0 to 20; and the number of paratactic locutions in most Master theses are zero.

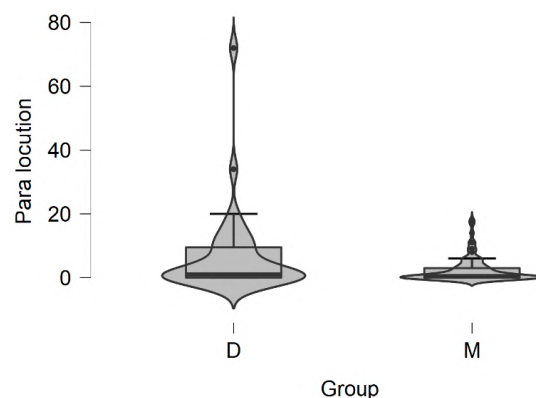


Figure 158 The boxplots of the number of paratactic locutions in Master and PhD sub-corpora

6.2.2.1 Locution Parataxis/Clause Nexuses

The index paratactic locution per clause nexuses is hard to compare as the number of paratactic locutions is too small, and data are skewed and distributed in two groups. From Table 6.2.2.1-1 (see Appendix 6.2), it is obvious that the average proportion of paratactic locution in all clause nexuses in the two groups is extremely small as it is only 0.783% in group D and 0.803% in group M, though the proportion can be 7.17% in some doctoral dissertations and 6.55% in some Master theses. Data in the two groups are both leptokurtic (the kurtosis value in Group D is 13.814 and in Group M is 5.929) and abnormal (the Shapiro-Wilk value in group D is 0.597 and in group M is 0.664, both far away from 1; the P-value of Shapiro-Wilk in two groups are both smaller than 0.001; the divisions of the skewness value by the standard error of skewness in two groups are 8.03 and 8.68 respectively, larger than +2; the divisions of the kurtosis value by the standard error of kurtosis are 17.08 and 11.14 correspondingly, also larger than +2) distributed with a right-sided skew (the skewness value in Group D is 3.325 and in Group M is 2.336).

In fact, as suggested by Figure 159, except for one case with a proportion higher than 6% in the doctoral group and 5 cases with a proportion larger than 3% in group M, most of the data are zero or nearly zero.

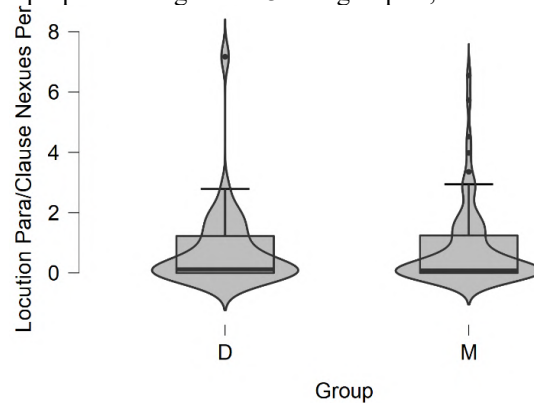


Figure 159 The boxplots comparing the proportion of paratactic locution in clause nexuses in Master and PhD sub-corpora

For abnormally distributed data, the Mann-Whitney U test is conducted firstly for comparison: when the alternative hypothesis is that group D has a larger proportion than group M, the p -value is 0.412, and the VS-MPR value is one though the maximum value should be smaller than 0.37 when favoring the alternative hypothesis; the Hodges-Lehmann estimate value (6.343×10^{-5}) is extremely small, as well as the rank-biserial correlation value 0.026. So, the difference between the two groups almost does not exist. (See Table 102)

Independent Samples T-Test

	W	df	p	VS-MP R*	Hodges-L ehmann Estimate	Rank-Biseri al Correlation	SE Rank-Biser ial Correlation
Locution Para/Clause Nexuses Perc.	1313. 000		0.412	1.000	6.343×10^{-5}	0.026	0.121

Note. For the Mann-Whitney test, effect size is given by the rank biserial correlation.

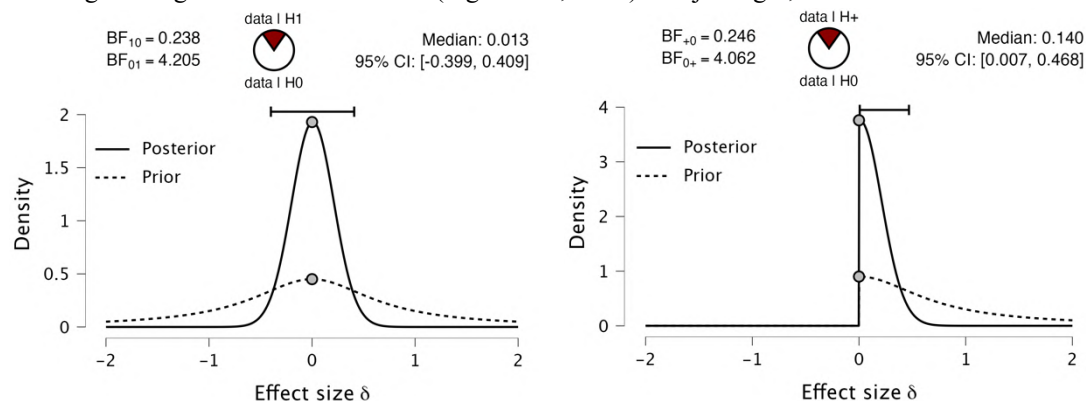
Note. For all tests, the alternative hypothesis specifies that group D is greater than group M.

Note. Mann-Whitney U test.

* Vovk-Sellke Maximum p -Ratio: Based on a two-sided p -value, the maximum possible odds in favor of H_1 over H_0 equals $1/(-e p \log(p))$ for $p \leq .37$ (Sellke, Bayarri, & Berger, 2001).

Table 102 The Mann-Whitney U test for comparing the proportion of paratactic locution in clause nexuses in Master and PhD sub-corpora

The same testing result can also be received from the Bayesian Mann-Whitney U test (the R_{hat} value is 1.064) that when the alternative hypothesis is two groups having various proportions, the possibility of accepting the original hypothesis (two groups are equal) is 4.205 times (Figure 160, left 1) of rejecting it; when the alternative hypothesis is that group D has larger proportion than group M, the probability of accepting the original one is 4.062 times (Figure 160, right 1) of accepting the alternative one; and when the alternative hypothesis is group D having smaller average proportion than group M, the probability of favoring the original one is 4.358 times (Figure 160, left 2) of rejecting it, with a medium effect size.



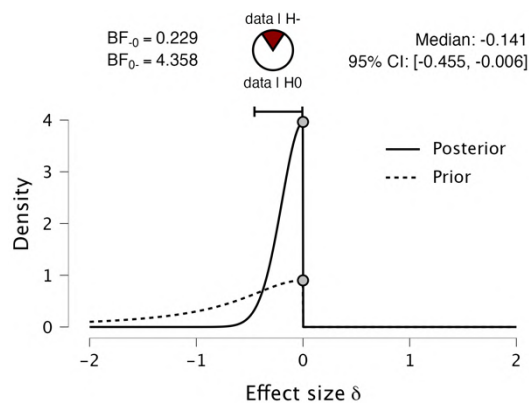


Figure 160 The inferential plots comparing the proportion of paratactic locution in clause nexuses in Master and PhD sub-corpora (with diverse alternative hypotheses)

By comparing the possibilities, a conclusion can be drawn: a slight difference exists between the two groups, though the average proportion of paratactic locution in all nexuses in doctoral dissertations is a little bit larger than that in Master's theses.

6.2.2.2 Locution Parataxis/Parataxis

The index paratactic locution per parataxis excluded all hypotactic structures, and the proportion of paratactic locution in all paratactic structures is also not high because of the small number of paratactic locations. Though in some doctoral dissertations, it can reach up to 19.25%, and in some Master theses, it can be 15.94%, the average proportion in group D is only 2.22%, and in group M is 2.07%. (See Table 6.2.2.2-1 in Appendix 6.2) Data are also leptokurtic (the kurtosis value in Group D is 11.214, and in Group M is 5.188), abnormally (the Shapiro-Wilk value in Group D is 0.618, and in Group M is 0.653; the P-value of Shapiro-Wilk in both groups are smaller than 0.01; the divisions of the skewness value by the standard error of skewness are 7.32 in Group D and 8.42 in Group M; the divisions of the kurtosis value by the standard error of kurtosis are 13.86 in group D and 9.75 in group M) distributed with a short right-sided tail (the skewness value in Group D is 3.031, and in Group M is 2.266) in two groups.

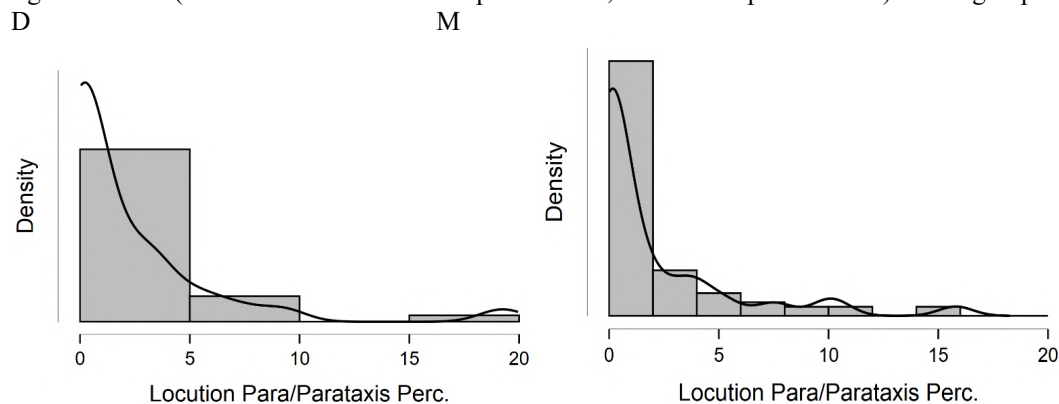


Figure 161 The distribution plots comparing the paratactic locution in parataxis in Master and PhD sub-corpora

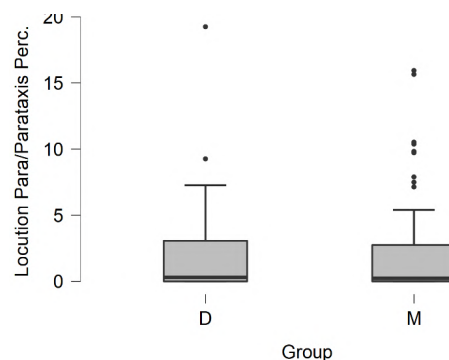


Figure 162 The boxplots comparing the paratactic locution in parataxis in Master and PhD sub-corpora From Figure 161 & Figure 162, data distribution in the two groups is quite clear except in two cases with the proportion being larger than 9% in group D and 7 cases, with high proportions ranging from 6% to 17%

in group M, other data mainly concentrate in zero to 4% and most of them are zero.

Though data are abnormally distributed in two groups, they can still be compared. In the Mann-Whitney U test (See Table 103), when the alternative hypothesis specifies that group D has a greater proportion than group M, the p-value is 0.352, and the VS-MPR value is 1.001, which means that the alternative hypothesis is highly favored. Differences within the two groups can be presented by the Hodges-Lehmann estimate value (6.580×10^{-5}) and with the rank-biserial correlation value of 0.044, slight variety can be found.

Independent Samples T-Test

	W	df	p	VS-MP R*	Hodges-L ehmann Estimate	Rank-Biseri al Correlation	SE Rank-Biseri al Correlation
Locution Para/Parataxis Perc.	1336.000		0.352	1.001	6.580×10^{-5}	0.044	0.121

Note. For the Mann-Whitney test, effect size is given by the rank biserial correlation.

Note. For all tests, the alternative hypothesis specifies that group *D* is greater than group *M*.

Note. Mann-Whitney U test.

* Vovk-Sellke Maximum *p* -Ratio: Based on a two-sided *p*-value, the maximum possible odds in favor of H_1 over H_0 equals $1/(-e p \log(p))$ for $p \leq .37$ (Sellke, Bayarri, & Berger, 2001).

Table 103 The Mann-Whitney U test for comparing the paratactic locution in parataxis in Master and PhD sub-corpora

When the Bayesian Mann-Whitney U test is referred to, the alternative hypothesis above is supported as the possibility of accepting the original hypothesis (group *D* > group *M*) is 6.518 times (Figure 163, left 2) of accepting the alternative one; and other hypotheses are 4.16 times (Figure 163, left 1. The alternative hypothesis is two groups having unequal proportions) or 3.057 times (Figure 163, right 1. The alternative hypothesis is that group *D* is greater than group *M*). Thus, doctoral candidates choose a higher proportion of paratactic locution in all paratactic structures than Master students, though this outcome may be affected by peculiarities, especially found in the doctoral group.

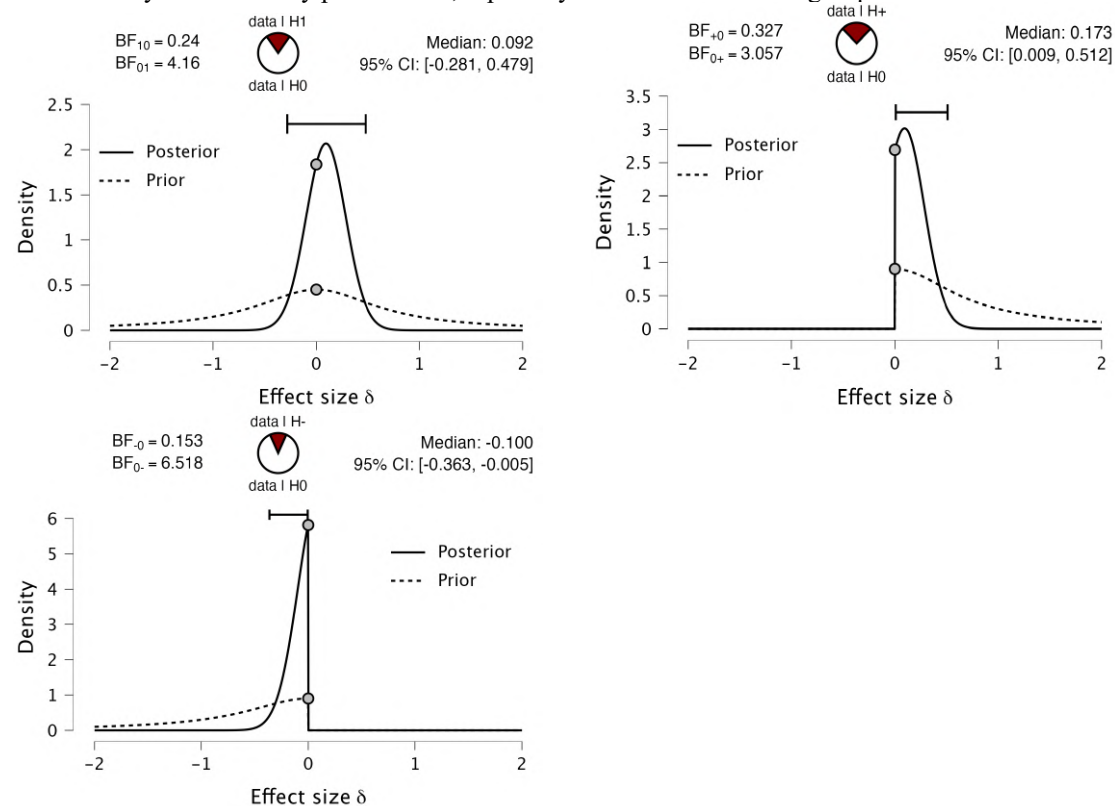


Figure 163 The inferential plots comparing the paratactic locution in parataxis in Master and PhD sub-corpora (with diverse alternative hypotheses)

6.2.2.3 Locution Parataxis/Projection Parataxis ³⁰

The index paratactic locution per paratactic projection excluded paratactic idea; the average proportion hence can be high (91.31% in group D and 88.48% in group M), and the proportion in the two groups differs greatly from each other (50% to 100% in group D and zero to 100% in group M). (See Table 6.2.2.3-1 in Appendix 6.2) Data are platykurtic (the kurtosis value is 1.784) and abnormally (the Shapiro-Wilk value is 0.635, the P-value of Shapiro-Wilk is smaller than 0.001; the division of the skewness value by the standard error of skewness is -4.12, smaller than -2; the division of the kurtosis value by the standard error of kurtosis is 2.21, larger than +2) distributed in group D with a short left-sided tail (the skewness value is -1.717); and in group M are leptokurtic (the kurtosis value is 5.143), abnormally (the Shapiro-Wilk value is 0.526; the P-value of Shapiro-Wilk is smaller than 0.001; two divisions are -9.05 and 9.67 respectively, not within -2/+2) distributed with a long left-sided tail (the skewness value is -2.434).

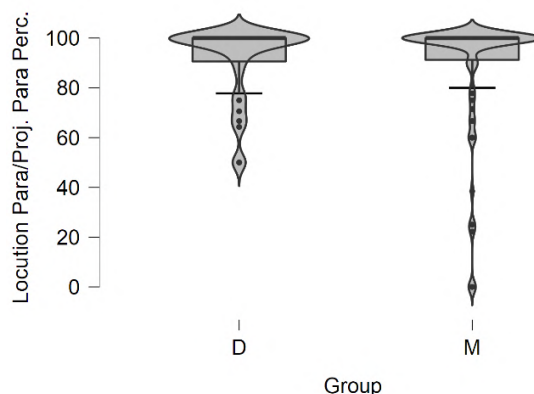


Figure 164 The boxplots comparing the proportion of paratactic locution in paratactic projection in Master and PhD sub-corpora

Data distributions are more clearly demonstrated in Figure 164 that except 5 cases with the proportion being smaller than 80 in group D, others are around 100%; and except 9 cases with the proportion randomly scattered within 0 to 80% in group M, others concentrate in 80% to 100%.

The first test for comparing data in two groups is, of course, the Mann-Whitney U test because of their abnormally distributed data. As suggested by Table 104, when the alternative hypothesis is that group D has a smaller average proportion than group M, the p-value is 0.3, and the VS-MPR value of 1.019 proves a difference between the two groups. However, the difference is quite small, as indicated by the Hodges-Lehmann estimate value (-3.009×10^{-6}) and the rank-biserial correlation value -0.051.

Independent Samples T-Test

	W	df	p	VS-M PR*	Hodges-L ehmann Estimate	Rank-Biseri al Correlation	SE Rank-Biserial Correlation
Locution Para/Proj. Para Perc.	1214.500		0.300	1.019	-3.009×10^{-6}	-0.051	0.121

Note. For the Mann-Whitney test, effect size is given by the rank biserial correlation.

Note. For all tests, the alternative hypothesis specifies that group D is less than group M.

Note. Mann-Whitney U test.

* Vovk-Sellke Maximum p -Ratio: Based on a two-sided p -value, the maximum possible odds in favor of H_1 over H_0 equals $1/(-e p \log(p))$ for $p \leq .37$ (Sellke, Bayarri, & Berger, 2001).

Table 104 The Mann-Whitney U test for comparing the proportion of paratactic locution in paratactic projections in Master and PhD sub-corpora

The second test for comparison is then the Bayesian independent samples t-test that when the alternative hypothesis is two groups having unequal average proportions, the possibility of supporting the original one is 4.044 times (See Figure 165, left 1) of rejecting it; when the alternative hypothesis is that group D has larger proportion than group M, the multiple is 4.728 (ibid, right 1) for supporting the original one; and when the alternative hypothesis is that group D has smaller proportion than group M, the multiple is 3.533 (ibid, left 2). So, the proportion of paratactic locution in paratactic projections chosen by doctoral candidates is slightly smaller than that chosen by Master students.

³⁰ When processing the data, 0 to 0 needs to be calculated as 100 percent of 0 instead of 0 percent of 0.

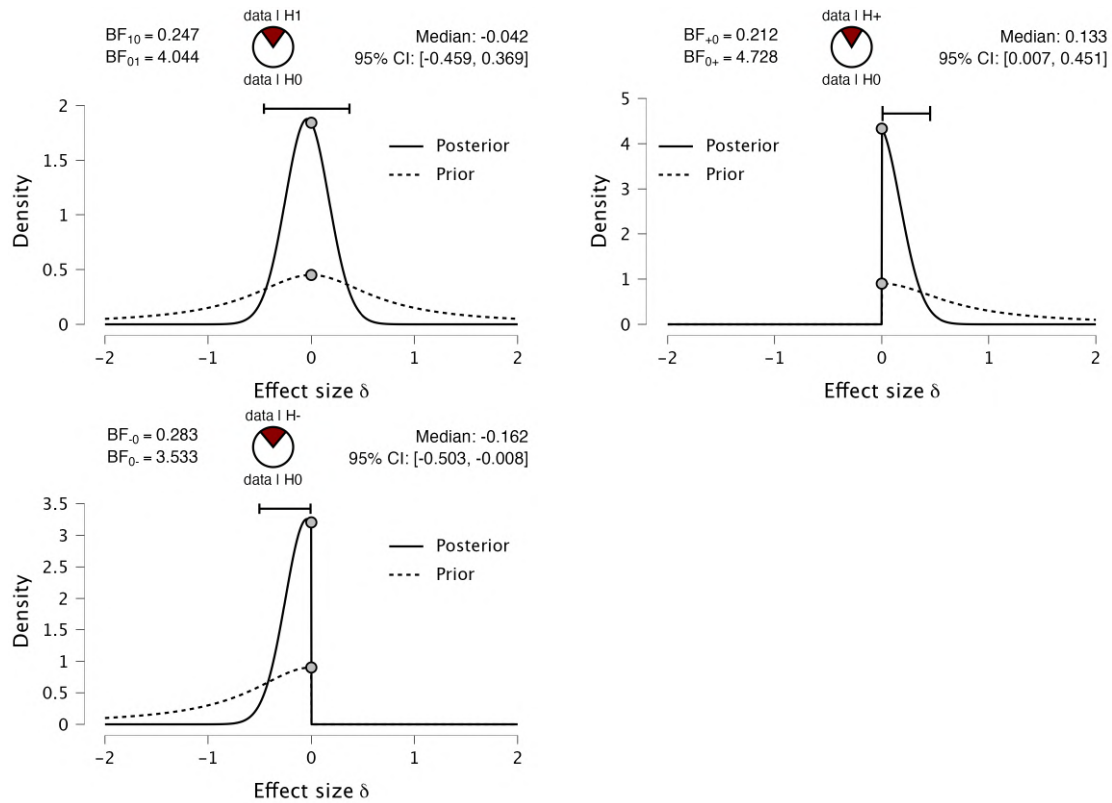


Figure 165 The inferential plots comparing the proportion of paratactic location in paratactic projections in Master and PhD sub-corpora

6.2.2.4 Locution Parataxis/Locution

The index paratactic location per locution demonstrates the choice of paratactic or hypotactic locations. In Table 6.2.2.4-1 (see Appendix 6.2), the average proportion of the two groups is around 10% (10.51% in group D and 12.53% in group M), though the proportion in some doctoral dissertations can be 59.50% and in some master's theses can be as high as 88.89%. Data in the two groups are both leptokurtic (the kurtosis value in group D is 3.967 and in group M is 4.127), abnormally (the Shapiro-Wilk value in group D is 0.776 and 0.732; the p-value of Shapiro-Wilk are both smaller than 0.001; two divisions of the skewness value by the standard error of skewness are 4.13 and 7.04 respectively, both larger than +2; and divisions of the kurtosis value by the standard error of kurtosis are 4.90 and 7.76 respectively, also both larger than +2) distributed with a right-sided tail (the skewness value in group D is 1.713, in group M is 1.894). In fact, data in group D are bi-modally distributed except for one case with a high proportion (59.50%), and in group M are mostly scattered within zero to 50% except for three cases with high proportions. (See Figure 166)

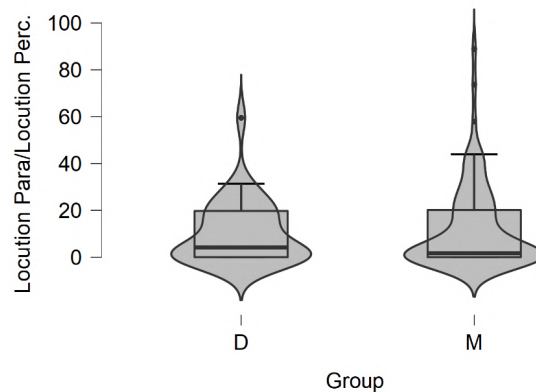


Figure 166 The boxplots comparing the proportion of paratactic location in locutions in Master and PhD sub-corpora

In the Mann-Whitney U test, when the alternative hypothesis is that group D has a larger average proportion than group M, the p-value is 0.48, and the VS-MPR value is 1. A difference is hard to tell because the Hodges-Lehmann estimate value (1.487×10^{-5}) and the rank-biserial correlation value 0.006

are both small. (See Table 105)

Independent Samples T-Test

	W	df	p	VS-MP R*	Hodges-L ehmann Estimate	Rank-Biseri al Correlation	SE Rank-Biseri al Correlation
Location Para/Locution Perc.	1288.000		0.480	1.000	1.487×10^{-5}	0.006	0.121

Note. For the Mann-Whitney test, effect size is given by the rank biserial correlation.

Note. For all tests, the alternative hypothesis specifies that group *D* is greater than group *M*.

Note. Mann-Whitney U test.

* Vovk-Sellke Maximum *p* -Ratio: Based on a two-sided *p*-value, the maximum possible odds in favor of *H*₁ over *H*₀ equals $1/(-e p \log(p))$ for $p \leq .37$ (Sellke, Bayarri, & Berger, 2001).

Table 105 The Mann-Whitney U test for comparing the proportion of paratactic locution in locutions in Master and PhD sub-corpora

Again, the Bayesian independent samples t-test is referred to for further comparison: the support for the original hypothesis is 4.427 times (Figure 167, left 1) of rejecting it when the alternative hypothesis is that two groups have diverse proportions; 4.456 times (ibid, right 1) with the Rhat value 1.022 and a medium effect size when the alternative is group *D* being greater than group *M*; and 4.399 times (ibid, left 2) when the alternative hypothesis is that group *D* has a smaller average proportion than group *M*. Thus, the proportion of paratactic locution in all locutions chosen by doctoral candidates has no big difference when compared to those chosen by Master students, though the former might be slightly smaller than the latter group. This results from scattered data in group *M* and three cases with high proportions, which, if cleaned, might result in a reverse testing outcome.

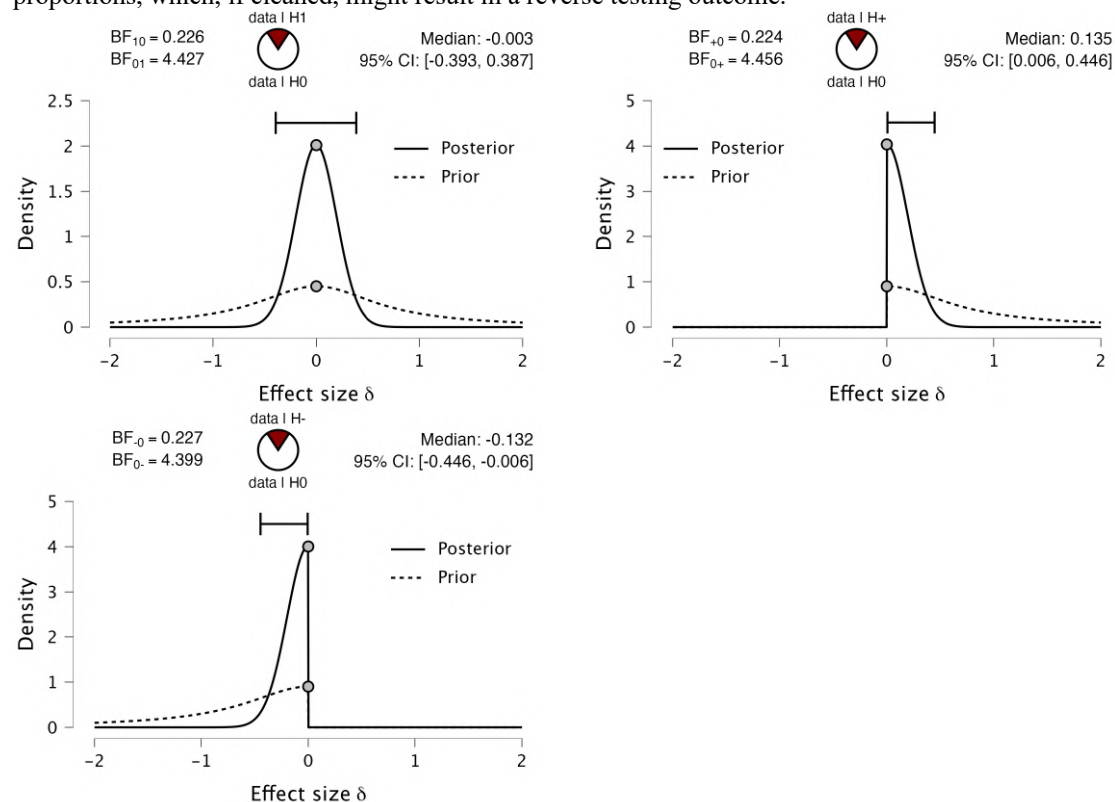


Figure 167 The inferential plots comparing the proportion of paratactic locution in locutions in Master and PhD sub-corpora

So, for paratactic locution, except for in paratactic structures, the average proportion in the doctoral group is apparently larger than that in the Master group, average proportions of paratactic locution in other structures, i.e., clause nexuses, locutions, and even paratactic projections have no big difference within two groups though the latter two structures are more favored by Master students. As peculiarities always exist, testing outcomes are not perfect. But these cases with strange proportions cannot be deleted as they authentically reflect the language choices of students' thesis writing.

6.2.3 The Distribution of Hypotactic Locution

The number of hypotactic locutions used in the two groups is also very small in that it can range from 9 to 132 with an average number of 37.09 and a total number of 1,187 in the doctoral group and zero to 35 with an average of 13.13 and a sum of 1,050 in the Master group. (See Table 6.2.3-1 in Appendix 6.2) Data in group D are leptokurtic (the kurtosis value is 7.394), abnormally (the Shapiro-Wilk value is 0.822, the P-value of Shapiro-Wilk is smaller than 0.001; the division of the skewness value by the standard error of skewness is 5.21; the division of the kurtosis value by the standard error of kurtosis is 9.14) distributed with a right-sided skew and in group M are platykurtic (the kurtosis value is 0.389), abnormally (the Shapiro-Wilk value is 0.952; the p-value of Shapiro-Wilk is 0.005; two divisions are 2.98 and 0.73) distributed with a short right-sided tail (the skewness value is 0.801). Figure 168 shows more clearly that except for one case from group D with the number 132, other data are normally distributed and concentrated in 9 to 100; and data in group M seem to be normally distributed except for one case with the number 35.

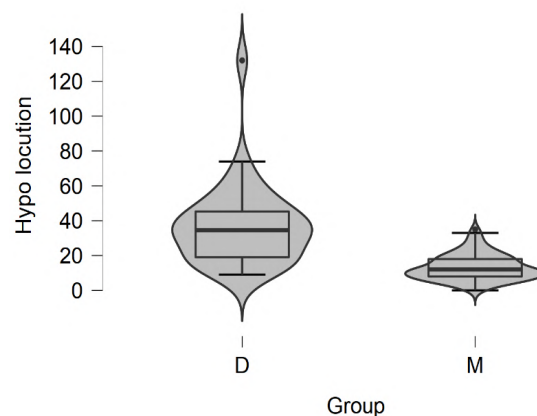


Figure 168 The boxplots comparing the number of hypotactic locutions in Master and PhD sub-corpora

6.2.3.1 Locution Hypotaxis/Clause Nexuses

When the proportion of hypotactic locution is discussed, the denominator can be clause nexuses, hypotaxis, hypotactic projection, and locution, in which the proportion of the first one is the smallest as the number of clause nexuses is the highest among four denominators. From Table 6.2.3.1-1 (see Appendix 6.2), the average proportion of group D (4.765%) can be a little bit larger than that of group M (4.159%), though in some doctoral dissertations, the proportion can be 10.82% and in some Master theses it can be 8.50%. Data in the doctoral group are slightly leptokurtic (the kurtosis value is 0.95), abnormally (the Shapiro-Wilk value is 0.914; the p-value of Shapiro-Wilk is 0.014; the division of the skewness value by the standard error of skewness is 2.49, larger than +2; and the kurtosis value divided by the standard error of kurtosis is 1.17, within $-2/+2$) distributed with a right-sided tail (the skewness value is 1.030) and in Master group are platykurtic (the kurtosis value is -0.288), and normally (the Shapiro-Wilk value is 0.987; the p-value of Shapiro-Wilk is 0.616; two divisions are 0.88 and -0.54 respectively, both within $-2/+2$) distributed with a short right-sided tail (the skewness value is 0.237). When the boxplots (See Figure 169) are referred to, except for one case with a proportion higher than 8% and one higher than 10%, other data are tri-modally distributed in group D, and they concentrate in 2% to 8%.

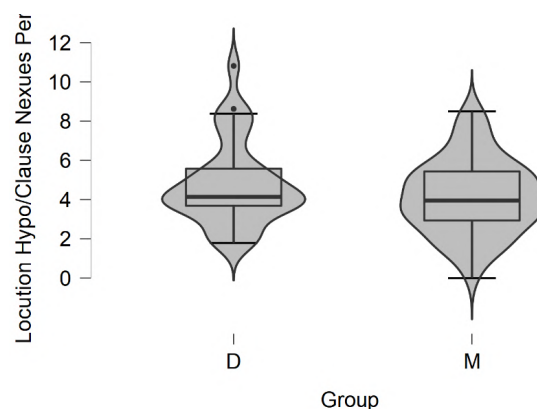


Figure 169 The boxplots comparing the proportion of hypotactic locution in clause nexuses in Master and PhD sub-corpora

For one normally and one abnormally distributed data comparison, both the Welch t-test and Mann-Whitney U test are adopted. In Table 106, the p-value of the former test is only 0.084 when the alternative hypothesis specifies that group D is greater than group M, the VS-MPR value 1.768 confirms diversities within two groups, the value of location parameter (0.606) and SE difference (0.433) both indicate differences quantitatively in statistical methods, and this test has a medium effect size as the value is 0.299.

Independent Samples T-Test

	Test	Statistic	df	p	VS-MPR [*]	Location Parameter	SE Difference	Effect Size	SE Effect Size
Locution Hypo/Clause Nexuses Perc.	Welch	1.398	51.871	0.084	1.768	0.606	0.433	0.299	0.212
	Mann-W hitney	1456.000		0.129	1.392	0.422		0.137	0.121

Note. For all tests, the alternative hypothesis specifies that group *D* is greater than group *M*.

Note. For the Welch t-test, effect size is given by Cohen's *d*. For the Mann-Whitney test, effect size is given by the rank biserial correlation.

Note. For the Welch t-test, location parameter is given by mean difference. For the Mann-Whitney test, location parameter is given by the Hodges-Lehmann estimate.

^{*} Vovk-Sellke Maximum *p* -Ratio: Based on a two-sided *p*-value, the maximum possible odds in favor of *H*₁ over *H*₀ equals $1/(-e p \log(p))$ for $p \leq .37$ (Sellke, Bayarri, & Berger, 2001).

Table 106 The independent samples t-test for comparing the proportion of hypotactic locution in clause nexuses in Master and PhD sub-corpora

The alternative hypothesis above can also be proved by the Bayesian Mann-Whitney U test that the possibility of accepting the original hypothesis (group *D* > group *M*) is 10.825 times rejecting it with the *R*hat value being 1.08 and small effect size, as suggested by Figure 170, and other multiples are 2.088 (the alternative hypothesis is two groups being unequal) and 1.155 (the alternative hypothesis is group *D* > group *M*) respectively. Thus, the average proportion of hypotactic locution in all clause nexuses of group *D* is obviously higher than that of group *M*.

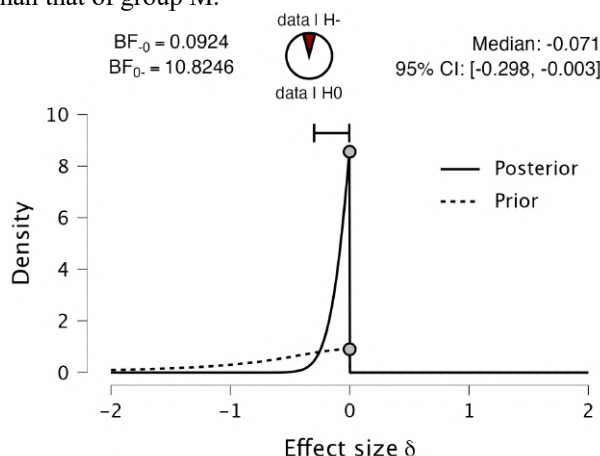


Figure 170 The inferential plot comparing the proportion of hypotactic locution in clause nexuses in Master and PhD sub-corpora (with the alternative hypothesis group *D* < group *M*)

6.2.3.2 Locution Hypotaxis/Hypotaxis

The proportion of hypotactic locution in all hypotactic structures excluded hypotactic expansions as well as hypotactic ideas, thus, it won't be high enough. From Table 6.2.3.2-1 (see Appendix 6.2), the average proportion can be around 7%, though it in group *D* (7.339%) is larger than in group *M* (6.853%) and the proportion ranges from 2.34% to 15.48% in the former group and zero to 13.75% in the latter one. Data seem to be platykurtic (the kurtosis value is 0.408), normally (the Shapiro-Wilk value is 0.938, nearly to 1; the p-value of Shapiro-Wilk is 0.068; the division of skewness by the standard error of skewness is 1.93, within $-2/+2$; and the division of the kurtosis value by the standard error of kurtosis is 0.50, also within $-2/+2$) distributed in group *D* with a right-sided tail (the skewness value is 0.798) and platykurtic (the kurtosis value is -0.484), normally (the Shapiro-Wilk value is 0.989; the p-value of Shapiro-Wilk is 0.74; two divisions are 0.51 and -0.91 , both are within $-2/+2$) distributed in group *M* with a short right-sided tail (the skewness value is 0.137). In fact, data in group *D* are scattered between 2% to 10%,

except three cases within 10% to 15% and one case being higher than 15%, are thus abnormally distributed, and data in group M, though slightly bimodal, are normally distributed. (See Figure 171)

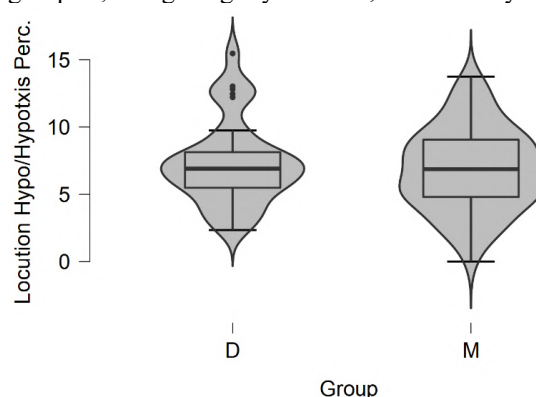


Figure 171 The boxplots comparing the proportion of hypotactic locution in hypotaxis in Master and PhD sub-corpora

So, the best independent samples t-test for one abnormally and another normally distributed data comparison can be the Mann-Whitney U test that when the alternative hypothesis is set as group D having a larger average proportion than group M, the p-value is 0.339, the VS-MPR value is 1.003, and referencing methods for difference testing, i.e. both the Hodges-Lehmann estimate and rank-biserial correlation have small values (0.345 and 0.051 respectively). (See Table 107) So, the alternative hypothesis can be accepted in a sense, though the difference between the two groups might not be obvious.

Independent Samples T-Test

	W	df	p	VS-MP R*	Hodges-L ehmann Estimate	Rank-Biseri al Correlation	SE Rank-Biseri al Correlation
Locution Hypo/Hypotaxis Perc.	1345.000		0.339	1.003	0.345	0.051	0.121

Note. For the Mann-Whitney test, effect size is given by the rank biserial correlation.

Note. For all tests, the alternative hypothesis specifies that group D is greater than group M.

Note. Mann-Whitney U test.

* Vovk-Sellke Maximum p -Ratio: Based on a two-sided p -value, the maximum possible odds in favor of H_1 over H_0 equals $1/(-e p \log(p))$ for $p \leq .37$ (Sellke, Bayarri, & Berger, 2001).

Table 107 The Mann-Whitney U test for comparing the proportion of hypotactic locution in hypotaxis in Master and PhD sub-corpora

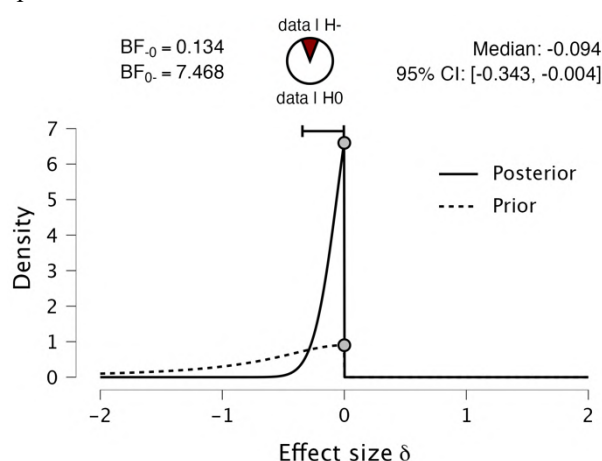


Figure 172 The inferential plot comparing the proportion of hypotactic locution in hypotaxis in Master and PhD sub-corpora (with the alternative hypothesis group D < group M)

Hence, the Bayesian Mann-Whitney U test is conducted: when the alternative hypothesis is that group D has a smaller average proportion than group M, the possibility of accepting the original hypothesis (group D > group M) is 7.468 times (See Figure 172) of rejecting it, with the Rhat value 1.032 and a

medium effect size, other multiples are 3.673 (the alternative hypothesis is two groups being unequal) and 2.435 (the alternative hypothesis is group D > group M) separately. A conclusion can be drawn here that group D has a larger average proportion than group M on the index hypotactic location per hypotaxis.

6.2.3.3 Locution Hypotaxis/Projection Hypotaxis

The proportion of hypotactic locution in hypotactic projections investigates the choice of hypotactic locations or hypotactic ideas as they together constitute hypotactic projection structures. From Table 6.2.3.3-1 (see Appendix 6.2), though the Master group has a wider range of the proportion (from zero to 64.71%) than the doctoral group (from 8.65% to 51.77%), the average mean of the latter group (27.18%), however, is larger than that of the former (24.42%). Data in group D seem to be platykurtic, right-sided skew, and normally distributed as its kurtosis value is -0.356 (not far away from 0), skewness value is 0.335, Shapiro-Wilk value is 0.965 (not far away from 1), the p-value of Shapiro-Wilk is 0.367, the value of skewness by the standard error of skewness is 0.81 (within $-2/+2$), and the value of kurtosis by the standard error of kurtosis is -0.44 (also within $-2/+2$); and in group M are leptokurtic (the kurtosis value is 1.374), right-sided skew (the skewness value is 0.687), and abnormally (the Shapiro-Wilk value is 0.972, the p-value of Shapiro-Wilk is however 0.078, two divisions are 2.55 and 2.58 correspondingly, being larger than $+2$) distributed. Figure 173 proves the judgment that data in group M scatters within zero to 50% except for one case, with its proportion being 64.706%, and in group D are even tri-modal.

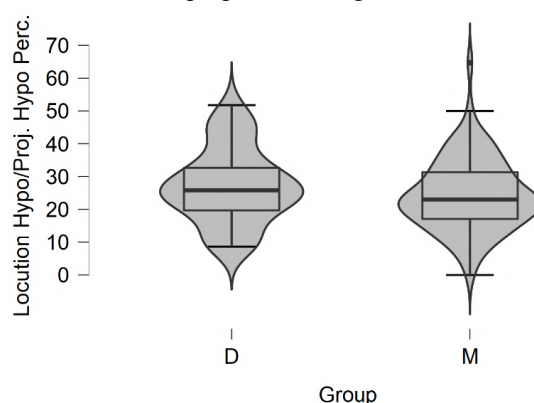


Figure 173 The boxplots comparing the proportion of hypotactic locution in hypotactic projections in Master and PhD sub-corpora

For abnormally distributed data in two groups, the Mann-Whitney U test is chosen: when the alternative hypothesis is set as group D having a larger average proportion than group M, the p-value is 0.105, with the VS-MPR value being 1.558, the alternative hypothesis can thus be accepted. The Hodges-Lehmann estimate value indicates that a difference between the two groups can be 2.946%, with a small to medium effect size, as suggested by the rank-biserial correlation value of 0.153. (See Table 108)

Independent Samples T-Test

	W	df	p	VS-MPR *	Hodges-L ehmann Estimate	Rank-Biserial Correlation	SE Rank-Biseria l Correlation
Locution Hypo/Proj. Hypo Perc.	1475.500		0.105	1.558	2.946	0.153	0.121

Note. For the Mann-Whitney test, effect size is given by the rank biserial correlation.

Note. For all tests, the alternative hypothesis specifies that group D is greater than group M.

Note. Mann-Whitney U test.

* Vovk-Sellke Maximum p -Ratio: Based on a two-sided p -value, the maximum possible odds in favor of H_1 over H_0 equals $1/(-e p \log(p))$ for $p \leq .37$ (Sellke, Bayarri, & Berger, 2001).

Table 108 The Mann-Whitney U test for comparing the proportion of hypotactic locution in hypotactic projections in Master and PhD sub-corpora

The same result can be received from the Bayesian Mann-Whitney U test that the probability of favoring the original hypothesis (group D > group M) is 9.08 times rejecting it, with the Rhat value 1.03 and a small to medium testing effect size. The other two multiples are 3.145 (the original hypothesis is two groups having the same proportion) and 1.902 (the original hypothesis is group D < group M). (See Figure 174)

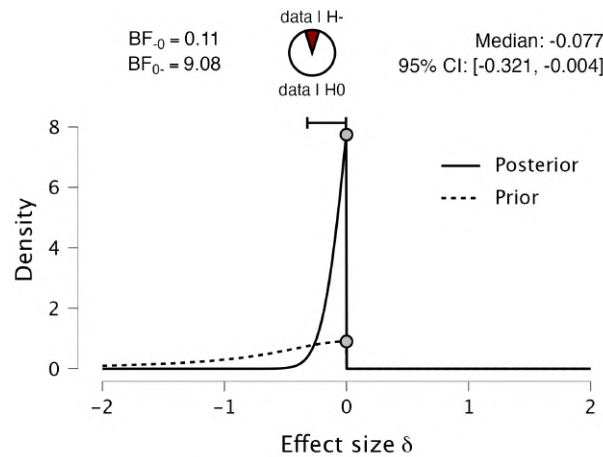


Figure 174 The inferential plot comparing the proportion of hypotactic location in hypotactic projections in Master and PhD sub-corpora (with the alternative hypothesis group D < group M)

6.2.3.4 Locution Hypotaxis/Locution

The proportions of hypotactic location in locution in two groups are the reverse condition of that described in Section 6.2.2.4 that the average proportion in group D can be as high as 89.49% and in group M reaches up to 86.22%. Data are leptokurtic and abnormally distributed (as has been proved before) with left-sided tails, as the skewness values are -1.713 in group D and -2.113 in group M. (See Table 6.2.3.4 in Appendix 6.2) The Bayesian Mann-Whitney U test, however, shows diverse multiples for proportions comparison in two groups: when the alternative hypothesis is that two groups have unequal proportions, the possibility of accepting the original hypothesis is 4.16 times (Figure 175, left 1) of rejecting it; when the alternative hypothesis is that group D has larger average proportion than group M, the probability of accepting the original hypothesis is 3.15 times (ibid, right 1) of rejecting it; and when the alternative hypothesis being group D has smaller proportions than group M, the multiple of favoring the original hypothesis is 6.122 (ibid, left 2) compared to the alternative one with the Rhat value 1.018 and a medium effect size. Thus, group D tends to have a higher proportion of hypotactic location in all locutions compared to group M.

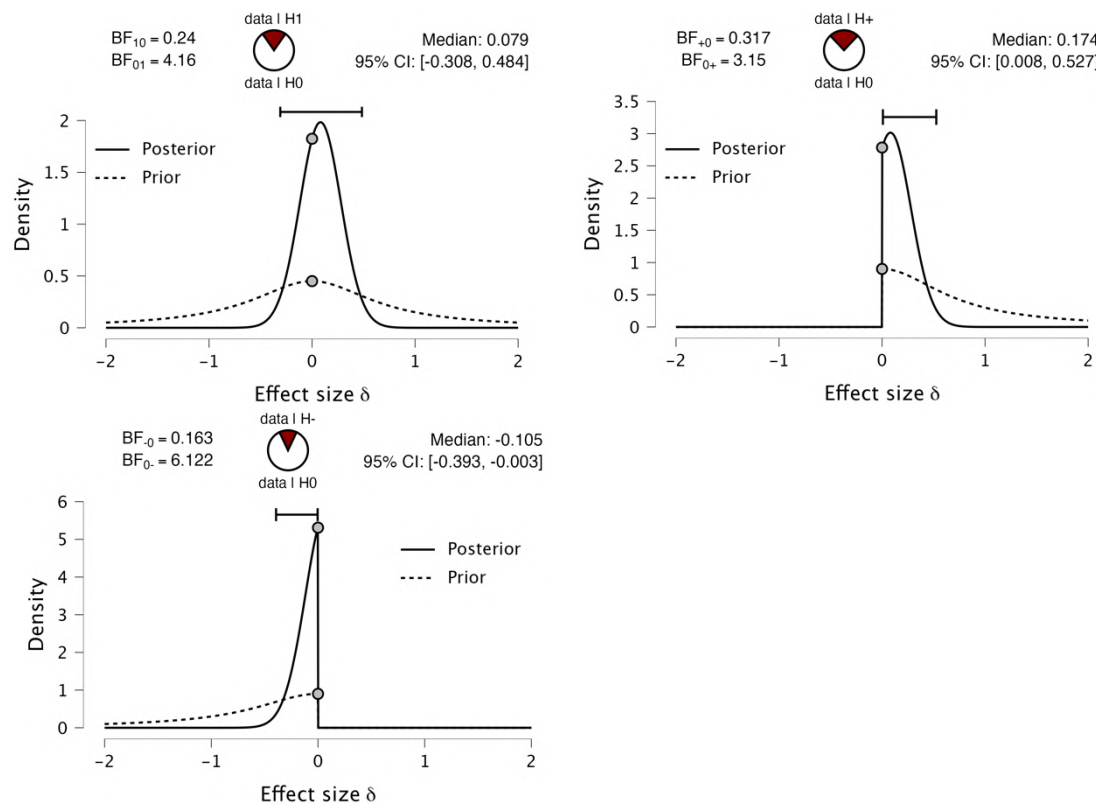


Figure 175 The inferential plots comparing the proportion of hypotactic location in locutions in Master and PhD sub-corpora (with diverse alternative hypotheses)

The proportions of hypotactic locution in the doctoral group have been demonstrated to be higher than those in the Master group no matter what the denominator is (clause nexuses, hypotaxis, hypotactic projections, locutions). One probable reason comes from a higher number of hypotactic locutions chosen by doctoral candidates, while the number can be zero in some Master theses, or the overuse of hypotactic ideas in hypotactic projections by Master students, and another can be a smaller number of hypotactic structures but more paratactic ones that are adopted in Master theses. More reasons need to be synthesized and considered after Ideas are quantitatively compared within two groups.

6.2.4 Summary of Locution Comparison

This section started with the average proportion comparison of locutions in all clause nexuses within two groups; then compared the usage of paratactic locution, the proportion of it with the denominator being clause nexus, parataxis, paratactic projection and locution in two groups; the usage of hypotactic locution and the proportion of hypotactic locution in all clause nexuses, in hypotactic structures, in hypotactic projections and all locutions chosen by doctoral candidates and Master students from our corpus. Except for the proportion of paratactic locution in all clause nexuses, in paratactic projections, and in locutions, the two groups have no big difference (though the first one turns out to be group D > group M and the latter two are group D < group M), the average proportions of other indexes are all higher in the doctoral group. A preliminary conclusion can be doctoral candidates' favoring of locutions, no matter whether hypotactic or paratactic ones, reflect their respect for references that they would rather present authors' original words than paraphrase them, which can change meanings during the transmitting approaches.

6.3 Locution (Qualitatively Analysis)

6.3.1 General Information of Locution

The division of projection in Section 3.1.4 has been described as quoting or reporting and has demonstrated detailed differences between two concepts by giving corresponding examples chosen from our corpus, but this division primarily distinguishes the two by mode, i.e., by two tactic interdependency relations of parataxis and hypotaxis. In explaining the influence of specific lexis on clause relations, the primary change lies in the difference of speech function that except for linking a statement, i.e., a proposition as a clause, some lexis also led to clauses with a speech function, i.e., a proposal. Hence, for lexical discussion, locution will be seen as propositions or proposals following a verbal lexis.

6.3.2 Locution - Proposal

The number of proposals after a projecting verb in two groups is quite small as in Master theses; there are only 136 proposals, while in doctoral dissertations, only 109 are discovered. (See Table 6.3.1-1 in Appendix 6.3) Most of which can be ascribed to *suggest* that 44 are applied in doctoral dissertations and 43 in Master theses, followed by *require* (only 2 in group D but 24 in group M), *argue* (19 in group D and 11 in group M) and *propose* (11 in group D and 10 in the latter). Some lexes are chosen in two groups, such as *advocate*, *agree*, *maintain*, *point out*, *tell*, and *warn*, some are unique in Master's theses like: *announce*, *ask*, *demand*, *insist*, *joke*, *remind*, *reproach*, *say*, and *state*, while some can only be found in doctoral dissertations: *admit*, *affirm*, *claim*, *conclude*, *contend*, *recommend*, and *stress*. So, seeing from the variety of these lexes, the result that 19 kinds are chosen in group M and 17 in group D indicates a slight distinction between the two groups, though the subject in the former greatly outnumbers that in the latter.

In qualitatively analyzing locution-proposals, at least one case would be picked out from each verb, i.e., if only one case for the corresponding projecting verb is found in the Master or doctoral group, then it must be included, but if more than 1 case for any marker can be found in both groups, then one for each would be considered. Hence, for proposals in the locution category, all together 35 instantiations are taken into account. Based on the analyzing framework from SFG and Appraisal Theory intersection described in Section 3.2.3, kinds of types with their numbers for these cases are clearly classified, in which type 0000 (four cases) does exist for objectively choosing verbs in interpersonally projecting clauses by the writer, 7 cases (6 belong to type +000 and 1 for type 00+0) indicate that the writer almost takes the impartial stand in projecting proposals, then 9 instantiations (4 for type +0+0, 3 for ++00, one +-00, and one -0+0) demonstrate writers' moderate subjectivity in adopting these projecting verbs into academic writing, and 15 left ones (6 are in type +++++, 3 for ++++0, 2 for ++0+ and --+0, and only one for type +-+0 and --++) are subjectively interfered by writers with intense emotions that not only authors of the projecting clauses are strongly approved for or negate, but also inclined readers for these projections are closely attached to and required for agreement by these writers.

6.3.2.1 Objectivity (One Type)

“unknown” - “neutral” - “keep” - “present” (type 0000)

The objectivity class identified as type 0000 is special in it can only be discovered in locution-proposals instead of other categories as projected words are either quoted or directed reported in the clause complex without being mixed up with the writer's admission or permission, proposals projected usually have no connection with writers' affirmation or negation of the content being described in compared to propositions, the distance within the writer and readers is not close when no agreement is required from them, and the aim for most writers is to show the proposal without changing its context, attitude, mood, voice or even a word from reported subjects. The unmarkedness in all indicators is thus signed as the type 0000, which can be instantiated by verbs such as *contend*, *state*, and *tell* etc., in a clause complex.

E.g., 6.3.1 ||| The first maxim of cooperative principle is the quantity maxim, || which **states*** || that a person's contribution to social intercourse should provide sufficient information, || but the information should not be too much, and over what in need. ||| ---- CMAC14PR_4_421

In presenting the content of the quantity maxim in cooperative principle in E.g., 6.3.1, the writer hypotactically reported it by choosing the locutionary verb *state*, which does not encompass any emotional coloring, and the contents of the reported clauses are not affirmed or denied by the writer, i.e., the writer just describes the essence of the maxim, though a proposal is adopted, which comes from the maxim itself instead of the writer's intention. As to readers, this kind of description does not require anyone to accept it but only presents what *a person's contribution to social intercourse should* do and shows the writer's keeping a distance from readers in just introducing the maxim. So, the projection is an objective demonstration of the quantity maxim.

E.g., 6.3.2 ||| The participants were **told*** || that they would be asked to give their opinions about a range of topics. ||| ---- CPhD08_0012_725

By choosing the locutionary verb *tell* in a passive voice in E.g., 6.3.2, the request for the participants that they would be asked for *their opinions about a range of topics* is impartially presented that the writer had not permitted it to happen, nor did the writer refuse it; so does the content of the request that being different from a proposition which can be confirmed or denied by the writer, the proposal does not have any truth value, so cannot be judged by the writer. If the participants refer to readers of this thesis and the requirement is an activity that will be conducted on them in the future, the tense of the clause then would be inconsistent, and the writer would become teachers that ask for opinions, the projecting clause with the verb *tell* is undoubtedly redundant. Hence, readers are excluded from the activity and have no connection to the whole statement, i.e., this proposal is totally objectively presented by the writer without any involvement of the writer, the author, and readers.

6.3.2.2 Impartiality (Two Types)

1. “admit” - “neutral” - “keep” - “present” (type +000)

The class impartiality for a locution-proposal category means that the writer either supports the author or the subject of the clause complex grammatically or semantically or draws the distance with readers closer to emotions, can therefore be realized as type +000 or 00+0, in which the former is demonstrated by verbs such as *announce*, *propose*, *remind*, *suggest*, etc. in chosen instantiations. Though type 0+00, 000+, -000, and 000- are theoretically possible, they might not be highly realized, especially in locutionary proposals.

E.g., 6.3.3 ||| Gardner, MacIntyre, and Lysynchuk (1990) **propose*** || that language programs in general should be evaluated in terms of non-linguistic outcomes (including willingness to communicate, a principal non-linguistic outcome), in addition to the evaluation of linguistic performance. ||| ---- CPhD13_0003_59

In E.g., 6.3.3, no matter whether the writer regards the proposal *language programs be evaluated in terms of non-linguistic outcomes* as a rational one or not, the proposal by Gardner, MacIntyre, and Lysynchuk is selected as an important addition and stated by the verb *propose* in the writer's thesis with a confirmation. Thus, without considering the relationship between the writer and readers, the first indicator, i.e., admission or distance, can at least be positively marked. On this basis, the proposal is then presented to readers as a projection.

E.g., 6.3.4 ||| Compared with the traditional view [[that learning is the accumulation of facts or the development of skills]], || constructivism **suggests*** || that individuals are engaged right from birth in building up personal understanding from their experiences. ||| ---- CMAC09ME_24_209

When the writer sees the view *learning is the accumulation of facts or ...* as a traditional one and compares the point from constructivism with it, the proposing activity of the latter is firstly admitted by the writer; but the comparison does not indicate acceptance of the contents described in the proposed clause, nor has the writer asked for readers' acknowledgment for the content. So, the projection in E.g. 6.3.4 only has one indicator marked as positive and is ascribed to type +000.

2. “unknown” - “neutral” - “close” - “present” (type 00+0)

Another type of class impartiality in locution-proposals is 00+0 in the writer’s attitude toward authors or the subject being projected is obscure, but to readers is intimate.

E.g., 6.3.5 ||| The American philosopher Saul Steinberg had even **said***: || “trying to define humor is also a definition of humor.” ||| ---- CMAC14PR_4_95

In E.g., 6.3.5, the writer seems to go along with the American philosopher Saul Steinberg by using the past perfect tense *had said*, but the choice of *even* then makes the writer’s attitude doubtful and whether the writer confirms the projected clause remains unknown, the first two indicators to authors are thus unmarked. But the writer tries to tell readers the saying and endeavors to close the gap between them, though simply presenting the projection in an objective status, the whole clause complex is marked as 00+0 type.

6.3.2.3 Moderate Subjectivity (Four Types)

1. “admit” - “neutral” - “close” - “present” (type +0+0)

The class moderate subjectivity indicates that two of the four indicators are marked. Though theoretically, eight possible types should be discovered in each category, only four, i.e., +0+0, ++00, +-00, and -0+0, are concluded from the locution-proposal category up till now. Among these, the first one is usually realized by verbs such as: *advocate*, *ask*, *joke*, *maintain*, etc.

E.g., 6.3.6 ||| Many researchers **advocate*** || that teachers should focus on content more than form (Faigley & Witte, 1981; Sommers, 1980). ||| ---- CPhD16_0001_1753

Whether the writer supports the proposal *teachers should focus on content more than form* or not in E.g., 6.3.6, he/she at least admits the activity of *advocate* by so many researchers that the proposal is projected, references of corresponding scholars are listed in the bracket at the end of the clause complex and then attempts to objectively present the proposal to readers in a locutionary tone. So, the clause complex is marked as +0+0 type that the writer affirms scholars on the one hand and amiably transfers these opinions to readers on the other hand.

E.g., 6.3.7 ||| Rassaei (2014) further **maintains*** || that corrective feedback research should be more on social interaction relationship and on the way [[how to turn corrective feedback to individual learners || rather than on testifying the cognitive processing of learning corrective feedback could results in]]. ||| ---- CMAC17_02_34

The same goes E.g., 6.3.7, as the writer won’t present Rassaei’s proposal to readers and even adopts the locutionary verb *maintain* with *further* before it if he/she disapproves of it. In explaining the research method of corrective feedback to readers, the writer excerpts two clauses rather than a few words from the author’s viewpoints, which suggests his/her aim of letting readers know better about the object being discussed.

2. “admit” - “affirm” - keep” - “present” (type ++00)

Besides the verbs *demand* and *require*, type ++00 can also be realized by *advocate* that the writer in this type normally affirms the proposal from authors or subjects while showing no interest in readers’ approval or relations with them.

E.g., 6.3.8 ||| It is because of the not enough information Cameron provides || that leads to the violation of sub-maxim of quantity [[which **demands*** || that the speaker should make the contribution as informative as required in the conversation]]. ||| ---- CMAC15_01_353

In E.g., 6.3.8, the writer firstly admits the demand by the sub-maxim of quantity that *the speaker should make the contribution as informative as required in the conversation* and then conveys his/her confirmation on the content of the proposal, based on which the reason of violating the sub-maxim of quantity is tenable and thus demonstrated, i.e., *Cameron does not provide enough information*. So, the writer partly interferes in projection and meaning transferring to readers no matter if they want to know or approve of it or the other way around.

E.g., 6.3.9 ||| And nobody missed the first class of the semester || and the teacher **required*** || that all the students in both of the experimental and control class should not be absent in the week of 18, || so that all the students took part in the tests with no one missing. ||| ---- CMAC12LI_6_503

When the teacher is not the writer but just a research conductor, the demonstration of the proposal by the writer is just to show how experiments are performed and what the requirement from the teacher is that leads to the success of the research, so a distance within the writer and readers is kept, and the proposal is objectively presented to readers. If the writer does not go along with the teacher’s requirement, he/she will not depict the consequence *so that all the students took part in the tests with no one missing* afterward. When the teacher is the writer him(herself), the obvious is that the teacher’s decisions are all accepted by the writer, and the experiment is noted down as a result. Hence, the clause complex can be marked as “admit-affirm-keep-present,” i.e., the ++00 type.

3. “admit” - “deny” - “keep” - “present” (type +-00)

The type +-00 is quite special in that it temporarily appears in the locution-proposal category rather than in others.

E.g., 6.3.10 ||| Although she **agrees*** || that the source text is no longer the most important factor but still one of the sources of information taken by the translator, || she doesn’t think || source-culture conventions must be replaced by target-culture convention in every translation. ||| ---- CMAC07CU_35_91

The projected clause *the source text is no longer the most important factor* in E.g., 6.3.10 seems to be a proposition, but in fact, is a proposal that should be rewritten as *the source text should no longer be regarded as the most important factor*. If the writer confirms the content of the proposal, the conditional relation marked by *although* would not be chosen in the clause complex, so the result for the second indicator is “deny” rather than “affirm.” Normally, the first indicator is consistent with the second one, but the subject’s attitude toward the proposal, i.e., *she agrees*, is indeed admitted by the writer because the verb *agree* would not appear in the projecting part otherwise. So, the clause complex is marked as a +-00 type, especially when the writer keeps a distance from readers, shows the contents of the proposal impartially to them, but takes an orthodox attitude toward the subject or authors.

4. “distance” - “neutral” - “close” - “present” (type -0+0)

E.g., 6.3.11 ||| Kahn **argued*** || that the syllable has to be represented as an autosegmental unit || rather than being contained within syllable boundaries. ||| ---- CMAC12PH_1_220

The proposal *that the syllable has to be ...* in E.g. 6.3.11 is not hard to identify the second indicator that the writer neither affirms the content nor shows negative evaluation to readers about his/her attitude. As no requirement for ascertainment from readers is suggested by the writer, the fourth indicator is also unmarked and signed as *present*. But in adopting the verb *argue*, a distance can be found between the writer and Kahn as the verb must be changed if the writer admits Kahn’s activity or manner of putting forward the proposal, the distance between readers, however, is narrowed down as the verb can demonstrate the writer’s intention to communicate with them. Hence, a moderate subjectivity is pervaded in the projection and is therefore marked as type -0+0.

6.3.2.4 Intense Subjectivity (Six Types)

1. “admit” - “affirm” - “close” - “present” (type +++0)

Intense subjectivity indicates that the writer brings strong emotions to a clause complex that, in the process of projecting, the proposal is not objectively presented to readers in fact. Except for the types ++++ and --++ in all indicators are marked, still, ten kinds of types can be marked theoretically. But for locution-proposals, only four of them have been discovered, in which the type +++0 can be realized by verbs such as *admit*, *point out*, *warn*, etc.

E.g., 6.3.12 ||| For example, Knickerbocker (2007, p. 64) **admitted*** || that the result in her study should be interpreted with caution, || because she just investigated a small sample of 22 participants. ||| ---- CPhD10_0009_453

In revealing the writer’s attitude to the interpretation of Knickerbocker’s study result in E.g., 6.3.12, no matter the reason in the last clause *because she just investigated ...* is provided by the author or the writer him/herself, the distance within the writer and readers is close as the choice of *admit* displays a conversational property. And by referencing Knickerbocker’s words with the verb *admit*, the writer, on the one hand, manifests his/her permission to the viewpoint semantically and conveys the confirmation of the contents of the proposal on the other hand that *the result in her study indeed should be interpreted with caution*, otherwise the explanation afterward is needless. So, the clause complex is marked as a +++0 type that, without the writer’s involvement, the reference would not be presented to readers, and readers would not be influenced by this proposal.

E.g., 6.3.13 ||| When Agnes condemns Steerforth’s influence over David || and **warned*** him || “you have made a dangerous friend”, (DC, p. 427) || David is still obliged by his gullibility || to accept her warning. ||| ---- CMAC10LIT_7_33

For the literary description of Agnes’ behavior on David in E.g., 6.3.13, if the verb *warn* is adopted by the author and the writer just reports it from his/her angel, then the writer indeed considers Steerforth’s influence over David is negative and approves Agnes’ warning on him as well as the saying that David indeed *have made a dangerous friend*; if the verb is just chosen by the writer, then the first indicator must be “admit” and the second turns out to be “affirm” as the description of David’s reaction in the last clause is *still obliged by his gullibility*, indirectly reflecting the writer’s agreement on the paratactically quoting and conveying an expectation on David that he should not make such a dangerous friend. So, the first two indicators are both marked that the writer supports Agnes, no matter the grammatical or semantic aspect or the content of the proposal. The same goes for the distance to readers as a more detailed description can suggest a stronger intention of closing the gap with them by the writer. As a result, this clause complex belongs to a +++0 type.

2. “admit” - “affirm” - “keep” - “seek for agreement” (type ++0+)

E.g., 6.3.14 ||| This **requires*** || that, while using language, we should not only pay attention to its accuracy || but also make corresponding variations in lexicogrammar according to the change of any element in field of discourse, tenor of discourse and mode of discourse. ||| ---- CPhD10_0015_107

The verb *require*, though seems to be impartial in E.g., 6.3.14 with its subject *this*, the adoption of *we* by the writer not only demonstrates his/her confirmation of the requirement, the acceptance of the proposal that *should not only pay attention to...but also...*, but also asks readers to take the same attitude toward the proposal, i.e., to act the same as the writer does, an intense subjectivity is thus added to the verb *require* in the projection. But the distance between the writer and readers still exists as the pronoun *we* refers more to the writer him/herself than taking readers into account. Hence, this clause complex is a ++0+ type.

E.g., 6.3.15 ||| This theory **stresses*** || that writing instruction should be supplemented with teachers, teaching materials, sample writings and so forth, || and should decompose writing tasks by stages, || so as for learners to break away from the auxiliary “scaffolding” and write independently. ---- CPhD10_0015_105

Another example of the ++0+ type comes from the verb *stress* in E.g., 6.3.15, the verb *stress* itself does not possess any emotion, but when the subject is *this theory* written by the writer, it indicates that the stress comes from the writer and the proposal is his/her expectation on readers that *writing instruction should be supplemented with teachers*, ... So, the activity of *stress* and the proposal must be approved of by the writer, and then the writer seeks for agreement from readers, and at the same time keeps certain distance with them. That means the writer provides a solution for readers on writing instructions that detailed practice just relies on their choice.

3. “distance” - “deny” - “close” - “present” (type --+0)

E.g., 6.3.16 ||| Eleven students **claim*** || that sentence 1 and 2 should precede sentence 9 (amounting to 27.5%), || which suggests || that they still follow a particular-general pattern or an inductive thinking modal. ||| ---- CPhD05_0003_227

In showing the survey result to readers in E.g., 6.3.16, the writer keeps apart from the subject, i.e., *eleven students*, as the choice of the projecting verb *claim* fully demonstrates the affiliation of the following proposal that it is closely tied up with these eleven students instead of the writer. The first indicator is hence marked as “distance.” If the writer affirms the proposal, then the suggestion *that they still follow a...pattern or a...modal* in the fourth clause would not be accompanied by a negative attitude. The whole description of the claim, though, inclines toward seeking repudiation from readers by the writer, still just presents a phenomenon with the writer’s subjective understanding through establishing a close relationship with readers. The clause complex is, therefore, a --+0 type in the intense subjectivity class.

E.g., 6.3.17 ||| The poet **reproaches*** his friend || that he should not leave the poet in the storm without warning, “And make me travel forth without my cloak”. ||| ---- CMAC18_03_613

In closely demonstrating the activity of the poet to readers, the choice of the projecting verb *reproach* not only reflects the writer’s negative emotion toward the poet that a distance is thus kept within them but also shows the writer’s disapproval of the proposal from the poet, the clause complex can therefore be signed as --+0 type. But if the verb *reproach* is chosen by the author who depicts the poet, then the clause complex is another story, i.e., another projection that should be added reference information and belongs to 00+0 type in the impartiality class.

4. “admit” - “deny” - “close” - “present” (type +-+0)

E.g., 6.3.18 ||| Despite these advantages, however, it should be **warned*** || that this technique can only provide some visual evidence for the general trend of language change in certain constructions, || and whether such trend conforms to the existing theories and hypotheses still requires a further linguistic interpretation. ||| ---- CPhD19_0002_837

The type +-+0 in the intense subjectivity class is demonstrated in E.g., 6.3.18, in which the writer first confirms the following proposal as a warning and then thinks that this warning should be noticed by readers. Hence, the writer, on the one hand, supports the subject *it* and, on the other hand, intends to make a connection with readers and tell them about the necessity of this warning. But *this technique* seems to be a disavowal to the writer as it *can only provide some visual evidence for ...*, The writer, as a result, does not ask for agreement or denial from readers but just presents the warning as it can be. So, without the writer’s intense attitude, whether *this technique* works or not remains unknown to readers.

5. “admit” - “affirm” - “close” - “seek for agreement” (type ++++)

The ++++ type in intense subjectivity class suggests that all indicators are marked according to the semantics, context, grammar, and contents of the projection that the writer shows certain emotions to the author or the subject of the projecting clause and to the readers he/she inclines to at the same time. Verbs for this type of projecting locution-proposals can be *affirm*, *insist*, *maintain*, *point out*, *recommend*, etc.

E.g., 6.3.19 ||| Since it is concerned with speech acts, || it is based on the work of Austin and Searle, [[who **pointed out*** || that language can be used to do things || as well as say things]]. ||| ---- CMAC05PR_24_373

The verbal phrase *point out* just signifies the putting forward of a viewpoint from someone, but when the subject refers to scholars such as Austin and Searle, and the writer explains the foundation of his/her work, i.e., on speech acts based on their viewpoints, an intense emotion then is quite obvious as not only the activity of putting forward the proposal is approved, but also the content of the proposal that *language can be used to do things as well as say things* is supported by the writer. The aim of depicting the work foundation by referring to Austin and Searle's words is to seek readers' agreement on the proposal as well so that further study based on the viewpoint is meaningful and can be accepted by them. So, the writer displays the interpersonal relations to both authors and readers in the projecting clause complex, which is certainly marked as a ++++ type.

6. "distance" - "deny" - "close" - "seek for agreement" (type --++)

E.g., 6.3.20 ||| Different interpretations can be found with *qi-cheng-zhuan-he*, || though scholars **agree*** || that the pattern is a universal discourse pattern even to poetry writing. ||| ---- CPhD05_0003_2075

The type --++ is slightly different from ++++, though both are fully marked of their indicators that the writer in the former conveys a negative attitude toward the author or the subject. As indicated by E.g., 6.3.20, other scholars agree that *the pattern should be regarded as a universal discourse pattern...* except the writer because the enhancement: conditional marker *though* is chosen. So, the writer not only keeps a distance from these scholars who have the claim but also expresses disagreement with the proposal from these scholars. Besides, in describing *different interpretations* with *qi-cheng-zhuan-he* to readers, the writer hopes that they can take part in discovering more interpretations in addition to the proposal of seeing it as *a universal discourse pattern even to poetry writing*."

A summary for locution-proposal part is not on the interpersonal inclination of representative projecting verbal markers that they played in clause complexes as well as their types concluded from the framework based on SFG & Appraisal Theory, but on some general information of these projections: 1. projecting verbs usually followed by modal verbs in the projected proposals that a request for or an intention to someone rather than facts or depictions on something would generally be projected though sometimes modal verbs are omitted by writers and hard to identify; 2. not all projections are excerpted from the literature review part though the proportion can be high enough, and the proportion of projections in each thesis part can be another research project that deserves further investigation; 3. some verbs in the Master group can be infrequent that they would not appear in doctoral group at all as these usually come from literary works and the selection of them are discussions of plots by writers, verbs adopted in clause complexes from topics such as syntax and semantics in linguistics, applied linguistics, methodological studies, theoretical exploration, etc. are common words; and 4. the kind of verbs in doctoral group is less than that in Master group, which closely correlates to the number of writers that as different student can perform diverse writing behaviors, the higher number of subjects indicates the higher possibility of varieties in choosing these verbs.

6.3.3 Locution - Proposition

The number of propositions projected by locutionary verbs in clause complexes in two groups can be as high as one thousand when all together, 938 have been discovered in the Master group, and 1,094 are found in the doctoral group. (See Table 6.3.1-2 in Appendix 6.3) The most frequent verb can be *argue* (268 in doctoral group while only 85 in Master group), *say* (188 in Master group while 118 in doctoral group), and *point out* (145 in group D and 126 in group M), indicating that doctoral students are more critical in projecting clauses from authors at the same time try to be objective in directly transferring authors' words to readers; some verbs are also welcomed by students in both groups: *claim* (69 in group M and 96 in group D), *state* (62 vs. 53), *explain* (45 vs. 60), and *conclude* (56 vs. 39), suggesting that except for directly quoting statements, explanation and conclusion from authors and writers themselves are also of significance; still some are common verbs: *propose*, *tell*, *report*, *suggest*, *ask*, and *admit*; then are some key markers with not obvious frequencies: *assert*, *emphasize*, *maintain*, *stress*, *put forward*, and *summarize* that persuasion or advises are given by writers when some projections are reinforced; and rare words but are chosen in both groups: *discuss*, *declare*, *insist*, *convince*, *put*, *reply*, *convey*, *express*, *complain*, *deny*, *answer*, *interpret*, *acknowledge*, and *add*. Verbs like *point*, *announce*, *blame*, *clarify*, *affirm*, *disagree*, *shout*, *mention*, and *remark* only appear in the Master group due to the description of literary works in some cases while *contend*, *quote*, *recognize*, and *talk about* only come out in doctoral theses. Though the total number of projecting verbs in the Master group is not as high as that in the doctoral one, the variety of lexes in the former group indeed outnumbers 42 versus 37.

The qualitative analysis of instantiations from locution-propositions resembles that from

locution-proposals in Section 6.3.2, and in total, 81 cases are selected from 46 lexes listed above for indicators annotation. The result manifests that more than one-third belong to the moderate subjectivity class in which half are +0+0 type, then ++00 type accounts for 9, and -0+0 has 6; another one-third are marked as intense subjectivity, 11 of which are fully marked (9 are ++++ type, one for +-++ and --++ type), and 17 are marked as the third class (11 in which are ++0+ type, 3 for +++0 type, 2 for type --+0 and one for -0++); around a quarter (20 exactly) belongs to +000 type and 3 for -000 type. The reason that type 0000 is missing in locution-propositions partly originates from writers' stance that they more or less mix emotions or attitudes with their expressions, leaving scarcely possibility for objectivity in clause complexes writing, and partly comes from annotation that the more viewpoints conveyed in the clause complex, the more probable writers are for being understood as impartiality rather than objectivity.

6.3.3.1 Impartiality (Two Types)

1. "admit" - "neutral" - "keep" - "present" (type +000)

Type +000 in locution-propositions is the most welcomed as compared to that in other categories or to other types in the same grammatical category that the writer usually shows his/her admission or permission to the projecting verb, which is realized by a list of words in annotated instantiations: *acknowledge, answer, convey, declare, deny, disagree, discuss, emphasize, express, interpret, report, reply, say, state, suggest, summarize, tell*, etc.

E.g., 6.3.21 ||| Duan Han-wu (1999) **disagrees*** || that there is consensus on the scope of antonyms in effect || and he adopts a different classification of his own morphologically and semantically, [[in which AA are classified according to the semantic meaning]]. ||| ---- CMAC10SE_18_109

Though the locutionary projecting verb in E.g., 6.3.21 is *disagree*, which entails a negative mood toward the author Duan Han-wu, the writer supports the activity of the author's disagreement in describing the contents of the proposition except for showing attitude on them, as the behavior depicted in the third clause, i.e., *he adopts a different classification of ...* is based on the disagreement. The first indicator is thus marked as "admit", but not "distance", and the whole clause complex is +000 type when no emotions to readers are mingled.

E.g., 6.3.22 ||| Harter S. (1983), a self-esteem researcher at the present age, **said*** || "Self-esteem has never been defined clearly as if it had been treated as a conclusion, but in fact everyone differs sharply from others on the understanding of the concept." ||| ---- CPhD08_0001_948

The paratactic quotation in E.g., 6.3.22 projected by the verb *said*, in fact, is objective. Thus, the whole clause complex can be regarded as 0000 type. But as the writer chooses viewpoints from Harter S. instead of other scholars, at least an affirmative attitude toward the activity of *saying* is expressed by the writer. Then, whether the proposition is supported or not, whether the writer cares about the relationship with readers or not, and whether an agreement is required or not remain obscure.

E.g., 6.3.23 ||| He **summarized*** || that entertain is more employed than other engagement resources based on data. ||| ---- CMAC13CU_21_334

By selecting the verb *summarize* in the past tense in E.g., 6.3.23, the writer firstly considers the proposition *entertain is more employed than ...* as a summary and then confirms the activity from the subject (he) that the attitude to the projected clause is unknown and emotions to readers are hidden. The clause complex is +000 type, though the distance between the writer and readers is not too far away in the process of presenting.

2. "distance" - "neutral" - "keep" - "present" (type -000)

E.g., 6.3.24 ||| Extrapolating from the work of Philips (1984, among others) on reticence, || McCroskey and Richmond (1990) **maintain*** || that people [[who perceive themselves to be poor communicators]] are apt to be less willing to communicate. ||| ---- CPhD13_0003_1449

The distance between the writer and the subject in E.g., 6.3.24 is demonstrated twofold: as the statement *that people ... are apt to be less willing to communicate* by McCroskey and Richmond is *extrapolated from the work of Philips on reticence*, the reliability on the statement and two authors is hence reduced, and the writer seems to exclude him/herself from the statement; and as the choice of *maintain* stresses the behavior of speakers that the proposition is ascribed to McCroskey and Richmond, little correlation to the writer's stance can therefore be claimed. Though the writer keeps away from authors, no inclination to approach readers has emerged, and the clause complex, as a result, is -000 type.

E.g., 6.3.25 ||| When the participants worked on the questionnaire, || they were not **told*** || that this was a questionnaire under construction. ||| ---- CPhD08_0012_641

In E.g., 6.3.25, factors that the passive voice, past tense, and negativity of the locutionary verb *tell* and the choice of the demonstrative pronoun *they* for *the participants* by the writer codetermines the first indicator as "distance." In an effort to ensure the veracity of the questionnaire results, the writer maintains a certain distance from the participants (in the case of a questionnaire designed by the writer) or strives to present an objective account of the questionnaire process (in the absence of any direct

involvement by the writer) by stating that the participants were not informed of certain details. For the former condition, the clause complex is ++0 type, and for the latter is -000.

6.3.3.2 Moderate Subjectivity (Three Types)

1. “admit” - “neutral” - “close” - “present” (type +0+0)

The type +0+0 in locution-propositions is also higher than that in other grammatical categories in displaying the writer’s interpersonal relationship with authors/subjects and readers in projecting clauses while showing objectivity to the contents of the propositions. Verbs such as *argue*, *assert*, *claim*, *complain*, *explain*, *express*, *mention*, *point*, *point out*, *propose*, *quote*, *remark*, *say*, *shout*, etc. can lead to clauses in this type.

E.g., 6.3.26 ||| He **asserts*** || that language is like the majority of complex systems [[that exist in nature || and that empirically exhibit hierarchical structure]]. ||| ---- CPhD08_0012_297

E.g., 6.3.26 seems to have no big difference from E.g., 6.3.23, though the tense changes slightly, but in revealing the writer’s support for the subject (he), the option of the locutionary projecting verb *assert* on the one hand clarifies the belonging of the assertion and on the other hand, makes the distance with readers closer that the writer inclines to judge with them together about the authenticity of the proposition except for presenting the contents.

E.g., 6.3.27 ||| For the grammar has too many rules to remember, the students have **expressed*** || that they often forgot these grammar rules || if they were not often used. ||| ---- CMAC14WR_04_619

In demonstrating students’ expressions or even complaints to readers in E.g., 6.3.27, the writer obviously is not in the same group as those students who easily forget grammar rules but has a closer relationship with readers as he/she wants to discuss with them and then find a solution to the problem. Hence, the writer not only admits the statement from students but also delivers wishes to readers. As a phenomenon from personal experiences has no truth value, the writer is hard to say confirms the contents of the expression.

E.g., 6.3.28 ||| To illustrate this point, || they **quoted*** a famous Chomskyan example (ibid. p. 150): || Children with English speaking parents readily adopt that “The man is here.” becomes a question by reversal of noun and verb “Is the man here?” ||| ---- CPhD10_0009_244

When *they* as a pronoun is chosen to refer to the subject of a projection, the distance between the writer and readers becomes closer as *they* as a pronoun is the referential word to both sides from the same angle in pointing to the subject. Then, support for the subject comes from the verb *quote*, that the aim of quoting a proposition from Chomsky is *to illustrate the point*. Though the projected words are a famous example, the writer does not express his/her agreement with it but just paratactically presents it as a quotation to inclined readers.

2. “admit” - “affirm” - “keep” - “present” (type ++00)

Type ++00 in moderate subjectivity class for the locution-proposition category can be realized by verbs such as *conclude*, *contend*, *convey*, *insist*, *interpret*, *point out*, *propose*, *put*, *state*, etc.

E.g., 6.3.29 ||| So the author **concludes*** || that though there is obvious field cognitive style differences among students with high English level, || the “field” difference doesn’t have much influence on their choose or application of reading strategies. ||| ---- CMAC17_04_467

When the author refers to the writer in E.g., 6.3.29, the conclusion is made by the writer. Hence, it is admitted, and the contents of the proposition being concluded must be accepted, the former two indicators are marked as “admit” and “affirm.” Then, the writer hopes that readers can agree with the statement that *the ‘filed’ difference doesn’t have much influence on ...* though a distance is kept between them. The clause complex is a ++0+ type. But when the author is a dependent subject, the writer just presents the conclusion to readers, though the activity of making a summary is supported and the contents being described in the projected clauses are confirmed. The clause complex, therefore, is ++00 type that the writer shows no requirement on readers.

E.g., 6.3.30 ||| Wang Jianrong (2009) **interprets*** in the thesis “An Analysis of the Kite Image in The Kite Runner” || that the kite image has narrative functions || and plays an important part in constructing the themes of the novel. ||| ---- CMAC15_05_206

As the interpretation that *the kite image* can have *narrative functions* and *play an important part in constructing the themes of the novel* by Wang Jianrong in E.g., 6.3.30 is confirmed or accepted by the writer, the proposition is demonstrated in the thesis, and the author’s understanding on the kite image is admitted. Though a distance exists between the writer and readers, the presence of the proposition enables readers to know the significance of the object described in projected clauses.

E.g., 6.3.31 ||| Just as Young (1996) **puts*** it || if the teacher acts more like a friend helping students to learn and less like an authority figure [[making them perform]], || he can cultivate a surrounding with low classroom anxiety. ||| ---- CPhD08_0001_124

The paratactic projection of the quotation in E.g., 6.3.31 by the verb *put* shows Young’s opinions to

readers originally, and the choice of *just as* not only suggests the writer's confirmation of the contents of the proposition that when *the teacher acts more like a friend*, he can decrease the anxiety in the classroom, but also shows support for Young in putting forward the thought. Instead of requiring identification from readers, the reference by the writer more resembles a suggestion to him/herself to act like a friend in the classroom.

3. "distance" - "neutral" - "close" - "present" (type -0+0)

Type -0+0 simply reflects the writer's interaction with authors or the subject of the locution and with planned readers. Usually, verbs such as *ask*, *convince*, *report*, *stress*, etc. lead to propositions of this type. E.g., 6.3.32 ||| However, one low proficiency students **reported***, || "if we needn't take the English exams, I would not learn it at all". ||| ---- CMAC17_01_449

In showing students' viewpoints on the English exams to readers in E.g., 6.3.32, the writer paratactically quotes a student's words and describes the student as having low proficiency, which seems to degrade the ability of the student and expresses the writer's disagreement on the report. Though the content of the proposition, i.e., the student would not learn English at all if they did not take the exam, has not been approved or disapproved by the writer, his/her negative attitude toward the subject (the student) is obvious. So, the clause complex is -0+0 type that by transferring the complaint to readers, the intention of asking for help from readers is uncovered.

E.g., 6.3.33 ||| The participants **stressed*** || that anxiety decreases || when teachers make the class interesting and fun. ||| ---- CPhD08_0001_1536

The same type comes from E.g., 6.3.33 when the writer regards the participants' claim as an emphasis and chooses the verb *stress* to project their claims, the distance within them is apparent no matter whether the writer agrees with the description in the proposition or not. By projecting the participants' viewpoints, the writer hopes to discuss with readers to either expect their responses on how to *make the class interesting and fun* or find a solution to the problem of decreasing participants' anxiety during the class.

6.3.3.3 Intense Subjectivity (Seven Types)

1. "admit" - "affirm" - "keep" - "seek for agreement" (type ++0+)

Type ++0+ is important in the intense subjectivity class when locution-propositions are investigated. Except for the frequency and popularity of verbs such as: *affirm*, *claim*, *clarify*, *emphasize*, *insist*, *maintain*, *put*, *put forward*, *recognize*, *remark*, etc., the writer's intention of seeking for agreement from readers but at the same time keeping distance with them makes the projecting verb strange but valuable. E.g., 6.3.34 ||| Zhang Jian-li (1995) has **clarified*** || that antonyms refer to pairs of adjectives varying in qualities in most cases and not an umbrella term for the opposition of meaning, || which we will follow in the thesis, || and hence in this way we don't employ 'adjective antonym', but just the term 'antonym' for short. ||| ---- CMAC10SE_18_139

Though the verb *clarify* is used to project a proposition on the definition of *antonyms* from Zhang Jian-li in E.g., 6.3.34, the foundation of the discussion in the following part of the thesis would not be built if the writer does not approve for the proposition and attempt to convince readers of the confirmation of the projected contents. If readers hold an opposite opinion on the annotation of the term *antonym*, further discussions by the writer would be meaningless, and understanding from readers would be unnecessary. But if readers insist on seeing *antonyms* as an *umbrella term for the opposition of meaning*, the writer still follows the clarification from Zhang Jian-li. So, support for the author is apparent, especially after his clarification has been claimed to be rational by the writer.

E.g., 6.3.35 ||| As John F. Kennedy **put*** it, || "Our most basic common link is that we all inhabit this planet", || we are beginning to realize || that "ancient prejudices and other barriers to intercultural understanding can fall quickly to the wayside when survival itself is at stake." ||| ---- CMAC07CU_35_143

Though the choice of *we* in the third clause in E.g., 6.3.35 seems to be a sign of drawing readers together, the distance between the writer and readers, in fact, still exists as only the writer realizes the viewpoint *that ancient prejudices and ... can fall quickly...* and perceives it to be reasonable, but not readers though the writer indeed seeks for agreement from them after having admitted the quotation from John F. Kennedy, reckoned the contents by directly projecting it, and then paratactically presented the proposition to readers through the verbal mark *put*.

E.g., 6.3.36 ||| Selinker (1972) **remarks*** || that fossilization is one of the unique properties of interlanguage and perhaps the most crucial fact, || which any adequate theory of second language learning will have to explain. ||| ---- CPhD10_0018_256

By mentioning the whole projection from Selinker, whether the third clause is the original words or ideas from the author or not, the writer emphasizes the status of fossilization in interlanguage that it must be explained by *any adequate theory of second language learning*. As the dissertation is on second

language learning, *fossilization* is presented by projecting the proposition with the verb remark to readers, suggesting support for Selinker on the remark and confirmation of the statement on *fossilization* from the writer, conveying the expectation of recognition from readers and at the same time maintaining a seemingly objective relation with them.

2. “admit” - “affirm” - “close” - “present” (type +++0)

Sometimes, writers would speak for readers without expecting their responses to project locution-propositions by verbs like: *acknowledge*, *admit*, *complain*, etc., to convey their own attitudes, emphasize what they believe, and provide more information to readers.

E.g., 6.3.37 ||| In response to the aforementioned questions, five out of eight students **acknowledged*** || LPBL or PBL would help promote their language proficiency especially in listening and speaking. ||| ---- CPhD11_0007_1479

In E.g., 6.3.37, the writer does not have any requirement for readers to judge whether *LPBP* or *PBL* would really promote students' language proficiency or not, nor does the writer intend to ask readers to discuss the effect of LPBP or PBL on them, but just to tell readers about the condition of students' response to former questions through supporting for their acknowledgment and accepting their statement. So, the clause complex is +++0 type.

E.g., 6.3.38 ||| When the majority of the students **complain*** || that they do not go further in the university || the teacher's deep knowledge is needed most. ||| ---- CMAC09ME_24_687

The writer in E.g., 6.3.38 seems to keep him/herself away from the subject of the clause complex, i.e., many of the students, in fact, accept the complaint from students and recognize that *they do not go further in the university*, and based on which, one solution to the problem is provided for readers, that is, to rely on deep knowledge. The aim of the projection is to discuss with readers the solution to the complaint instead of presenting it; the clause complex, therefore, is marked as +++0.

3. “distance” - “deny” - “close” - “present” (type --+0)

E.g., 6.3.39 ||| On the contrary, he **blames*** her wife Edna || that she does not pay much attention to her children. ||| ---- CMAC18_04_904

No matter the verb *blame*, in E.g., 6.3.39 is from the literary work that the writer is describing or not, the demonstration of the plot *he blames his wife Edna* after *on the contrary* to readers suggests the writer's distance to the subject and disagreement on the reason of the blame on the one hand, and intention to persuade readers to hold the same attitude toward the subject on the other hand though the latter is not of necessity. The clause complex can thus be marked as --+0 type.

E.g., 6.3.40 ||| With power of words, they **convince*** women || that they are pure, unselfish, charming and above all mindless. ||| ---- CMAC06LIT_31_33

If women are the same as E.g., 6.3.40 describes that *they are pure, unselfish, charming and above all mindless*, then they would not be *convinced* by the people who know the *power of words*. So, in the writer's eye, the choice of the verb *convince* also conveys his/her dissatisfaction towards the subject of the projecting clause as well as disagreement with the statement of the proposition. The pronoun of the subject *they*, shortens the gap between the writer and readers as by standing together, the writer attempts to unite with readers and seek their opposition to the statement, though they usually would not take any action. The clause complex is, hence, another case for the --+0 type.

4. “distance” - “neutral” - “close” - “seek for agreement” (type -0++)

Type -0++ temporarily only appears in locution-propositions and is thus a special type, especially when the writer focuses more on interaction with readers rather than authors or the subject of the projecting clause and keeps away from the latter.

E.g., 6.3.41 ||| The dialogue happens || when Helen (H) comes to comfort Jane (J) || after Mr. Brocklehurst **announces*** || that Jane is a liar in front of all the students and teachers. ||| ---- CMAC10PR_19_574

According to the writer in E.g., 6.3.41, the announcement from Mr. Brocklehurst brings harm to Jane. Hence, Jane needs comfort from someone (Helen). The writer does not regard the announcement as a positive or suitable behavior whether he/she agrees with its content or not, and at the same time, describes the proposition to readers for their disapproval. Otherwise, the dialogue of comfort between Helen and Jane would not be reasonable and necessary when being presented to readers. This clause complex is absolute -0++ type.

5. “admit” - “affirm” - “close” - “seek for agreement” (type ++++)

Type ++++ can be locution-propositions after verbs such as: *add*, *argue*, *assert*, *conclude*, *discuss*, *summarize*, *talk about*, etc.

E.g., 6.3.42 ||| From the four dimensions of communicative competence, we can **conclude*** || that communicative competency is an integrated language ability [[which enables the learners to use language appropriately in different kinds of tasks || and use language for social action with other people]]. ||| ---- CMAC05ME_21_145

As the conclusion in E.g., 6.3.42 is made by the writer him/herself, it is no wonder that the subject is

admitted, and the content of the proposition is affirmed. The first two indicators are, therefore, positively marked. When *we* as a pronoun is taken as the subject of the projecting clause, the distance between the writer and readers becomes closer; and by stating the conclusion, readers are informed of what exactly the communicative competency is, no matter if they agree with the depiction from the writer or not though they indeed are hoped to accept the conclusion.

E.g., 6.3.43 ||| As we have **talked about***, || interlanguage is a language system [[which is intermediate between the learner's mother tongue]] and the second language [[that he is learning]]]. ||| ---- CPhD10_0018_182

Another case (E.g., 6.3.43) for the ++++ type is similar to the example above, though the proposition is paratactically quoted. The pronoun *we* in the projecting clause also draws the distance between the writer and readers close; the verb *talk about* suggests that the writer is communicating with readers about what the interlanguage is; and the perfect tense of the verb *have talked about* indicates that the activity of discussing *interlanguage* has already happened and the contents of the proposition have been proved by the writer to be reasonable. If readers disagree with the connotation of *interlanguage*, i.e., if a consensus has not been reached within them, further communication would then be unnecessary.

6. "admit" - "deny" - "close" - "seek for agreement" (type +-++)

E.g., 6.3.44 ||| Therefore, in his study, the subjects who consistently answer all the sentences correctly (i.e., three out of three) are regarded to be consistent || though he **admits*** || that some misinterpretations may be caused by performance factors. ||| ---- CPhD15_0001_944

The proposition in E.g., 6.3.44 that *some misinterpretation may be caused by performance factors* though has been admitted by the subject *he*, in fact, is disapproved by the writer, and readers are even affected to jointly oppose to the misinterpretations. That is, the writer supports the activity of the subject's recognition but shows disagreement on the content of the proposition and endeavors to unite with readers for objection. Not only the method taken by the subject is negated, but also his research result based on this method is controversial in the writer's opinion. The clause complex, therefore, is annotated as a +-++ type.

7. "distance" - "deny" - "close" - "seek for agreement" (type --++)

E.g., 6.3.45 ||| It seems that Jay should have **answered*** || that it is not an appropriate gift for Manny, || and that would be a normal and right answer. ||| ---- CMAC15_01_593

Apparently, in E.g., 6.3.45, the writer is not satisfied with the answer from Jay and provides another version for him that *it is not an appropriate gift for Manny*, the first two indicators are thus both negative as the writer keeps a distance with Jay and denies his answer in writing as *should have answered*. The uncertainty conveyed by *it seems that* seems to present the writer's interaction with readers and intention of seeking advice or other answers from them. The clause complex is hereby marked as --++ type.

In a word, students strive to objectively transmit locutions from referenced authors or subjects to readers in the process of projecting propositions, either with the aim of presenting what they have known or mastered, laying a foundation for their research for further discussions, reaching to a consensus by informing readers of the contents, or with the intention of seeking for answers, solutions or responses, though moderate or intense subjectivity is inevitable.

6.3.4 Summary of Locution Marker Choices

In projecting locutions, if proposals from authors or the subject of the clause are not objectively projected, then they would be more or less understood as the writer's proposal as well; and if propositions are mingled with emotions, either kind of attitudes are expressed to the subject by the writer or responses are required from readers and their reflection on the writer's activity is demanded. As chosen verbs with their tense, voice, mood, modality, and transitivity for projecting locutions directly have influences on the expression and construal of meaning in communication, writers' attitudes or views on selected locutions can be mixed on the one hand, their attitudes toward authors and expectations from readers can be presented on the other hand, which verb is the most suitable for the context of the clause complex and how can the projected verb be fully taken advantage of is therefore of great significance.

6.4 Idea (Quantitatively Comparison)

Another important category in projection is the idea that the content of a mental clause instead of a verbal one is represented in a clause complex. Likely, the comparison for ideas will firstly be the proportion in all clause nexuses within two groups, then conditions for paratactic ideas in four parts: clause nexus, parataxis, paratactic projection, and idea, and lastly, conditions of hypotactic ideas with the denominator being clause nexus, hypotaxis, hypotactic projection, and idea. Though reverse statistics and testing results compared to the locution part can be received from some indexes, detailed discussions are still

necessary.

6.4.1 General Distribution of Idea (Idea/Clause Nexuses)

For the proportion of ideas in clause nexuses, the means in the two groups are very close (13.43% in group D and 13.64% in group M) though the proportion in group M (3.77% to 30.65%) has a broader range than group D (6.47% to 20.26%). From Table 6.4.1-1 (see Appendix 6.4), it is obvious that data in group D are platykurtic (the kurtosis value is -0.928), normally (the Shapiro-Wilk value is 0.962 , the p -value of Shapiro-Wilk is 0.316 ; the division of the skewness value by the standard error of skewness is 0.34 , within $-2/+2$; and the division of the kurtosis value by the standard error of kurtosis is -1.15 , also within $-2/+2$) distributed with a very short right-sided tail (the skewness value is 0.14), and in group M are also platykurtic (the kurtosis value is 0.735), but abnormally (the Shapiro-Wilk value is 0.955 ; the P -value of Shapiro-Wilk is only 0.007 ; two divisions are 2.95 and 1.38 correspondingly) distributed with a right-sided skew (the skewness value is 0.794). In fact, the distribution of data in group D is bimodal, and in group M is normal except in one case with a high proportion of over 30%. And most of the proportion concentrate in 10% to 15% in both groups. (See Figure 176)

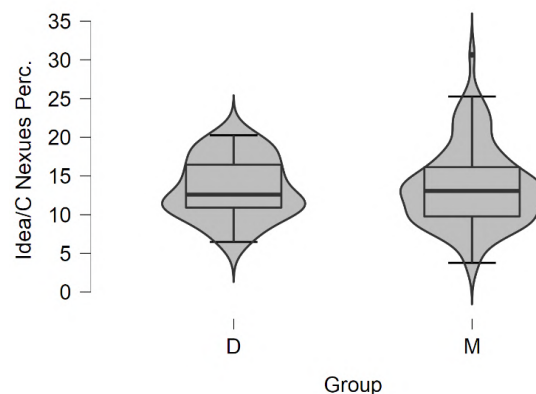


Figure 176 The boxplots comparing the proportion of ideas in clause nexuses in Master and PhD sub-corpora

The Mann-Whitney test is first conducted to compare abnormally distributed data. As shown in Table 109, when the alternative hypothesis comes to be group D having larger proportions than group M, the p -value is 0.436 , and the VS-MPR value is 1 , a difference is thus hard to tell within the two groups. The Hodges-Lehmann estimate value is 0.16 , meaning that the variance is very small. As the effect size of this test by the rank-biserial correlation value is only 0.02 , more tests are necessary.

Independent Samples T-Test

	W	df	p	VS-M PR*	Hodges-Le hmann Estimate	Rank-Biseri al Correlation	SE Rank-Biseri al Correlation
Idea/C Nexuses Perc.	1305.500		0.436	1.000	0.160	0.020	0.121

Note. For the Mann-Whitney test, effect size is given by the rank biserial correlation.

Note. For all tests, the alternative hypothesis specifies that group D is greater than group M.

Note. Mann-Whitney U test.

* Vovk-Sellke Maximum p -Ratio: Based on a two-sided p -value, the maximum possible odds in favor of H_1 over H_0 equals $1/(-e p \log(p))$ for $p \leq .37$ (Sellke, Bayarri, & Berger, 2001).

Table 109 The Mann-Whitney U test for comparing the proportion of ideas in clause nexuses in Master and PhD sub-corpora

In the Bayesian Mann-Whitney test, when the alternative hypothesis is two groups having unequal average proportions, the possibility of accepting the original hypothesis is 4.583 times (See Figure 177, left 1) of accepting the alternative one; when the alternative hypothesis indicates that group D has a larger average proportion than group M, the probability of supporting for the original one is 4.605 times (ibid, right 1) of accepting the alternative hypothesis, with the Rhat value 1.03 and a medium effect size; and when the alternative hypothesis is set as group D being smaller than group M, the multiple is 4.562 (ibid, left 2) for favoring of the original hypothesis. By comparing these multiples, the same result can be reached that the difference between the two groups is not obvious, though doctoral candidates may

choose fewer ideas in their clause nexuses as compared to Master students.

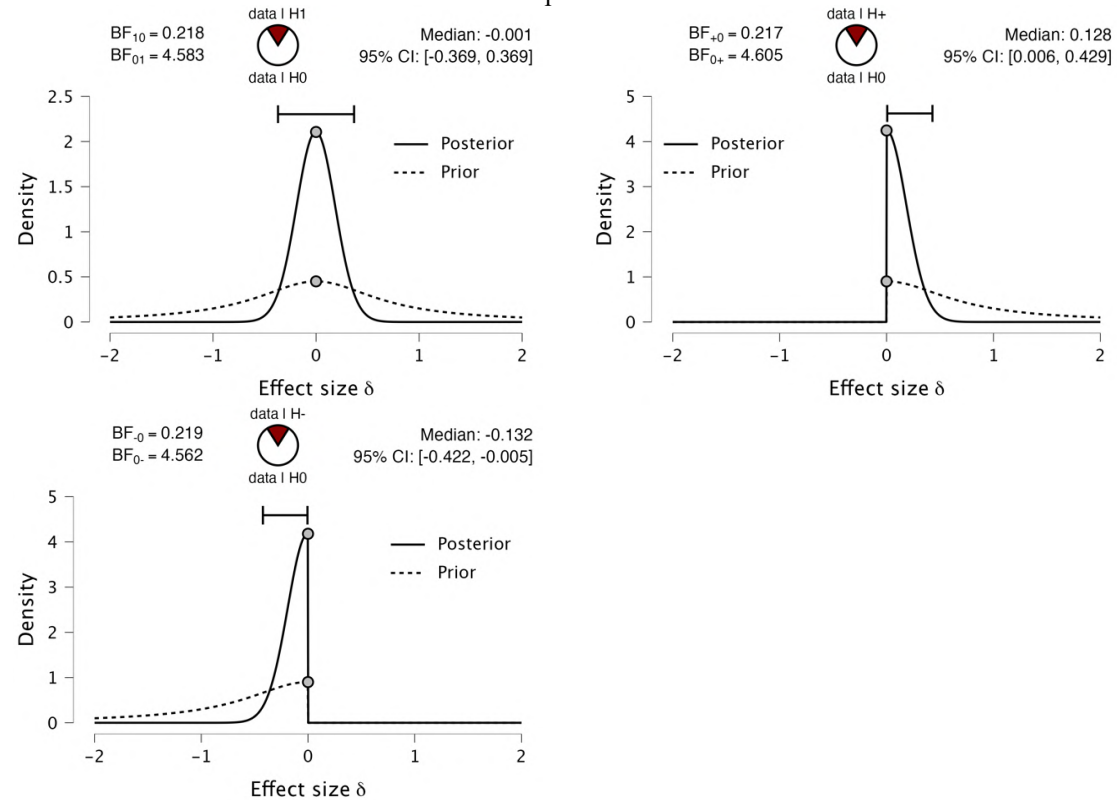


Figure 177 The inferential plots comparing the proportion of ideas in all clause nexuses in Master and PhD sub-corpora (with diverse alternative hypotheses)

6.4.2 The Distribution of Paratactic Idea

The total number of paratactic ideas in both groups is extremely small, i.e., only 41 in the doctoral group and 50 in the Master group, which means that the average number chosen by students in the former group is only 1.28 and in the latter group is only 0.625. (See Table 6.4.2 in Appendix 6.4) Data in two groups are, of course, abnormally distributed due to small numbers. From the boxplots (Figure 178), chosen numbers can be seen randomly scattered in two groups: most of them are 0, except one being 3, one being 5, one being 8, and another being 11 in group D; and some have 1 or 2 paratactic ideas, two cases have 3, one has 4, one has 7 and another has 8 in group M.

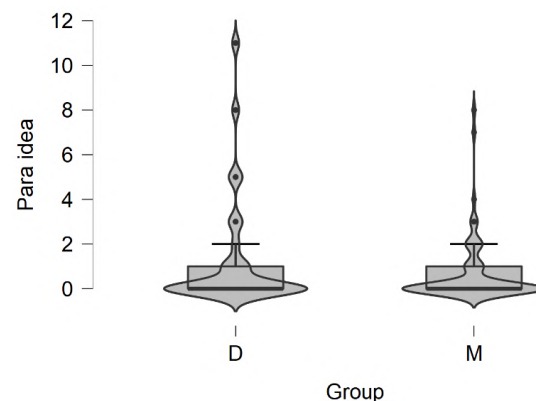


Figure 178 The boxplots of the number of paratactic ideas in Master and PhD sub-corpora

6.4.2.1 Idea Parataxis/Clause Nexuses

Due to the extremely small number of paratactic ideas with distributions that are abnormal and skewed in both groups, tests for data distribution on paratactic ideas part are left out no matter what the denominator is. The independent samples t-tests would not have enough effect size, i.e., they are not convincing for small data comparison, only the Bayesian Mann-Whitney U test with diverse alternative hypotheses is

thus conducted for comparison in the following four indexes.

For the index paratactic idea per clause nexuses, the average proportion in group D is only 0.148%, and in group M, it is 0.191%, meaning that only 1 or 2 paratactic ideas can be found in 1,000 clause nexuses. For such a small number and proportion, the possibility of accepting the original hypothesis is 4.405 times (Figure 179, left 1) of rejecting it in the Bayesian Mann-Whitney U test when the alternative hypothesis is group D > group M, with a Rhat value 1.016 and a medium effect size; other two multiples of supporting for the original hypothesis are 4.089 (ibid, right 1; the alternative hypothesis is group D < group M) and 4.242 (ibid, left 2; the alternative hypothesis is two groups being unequal) correspondingly. So, group D has a smaller average proportion of paratactic ideas in all clause nexuses than group M, though this difference is not obvious.

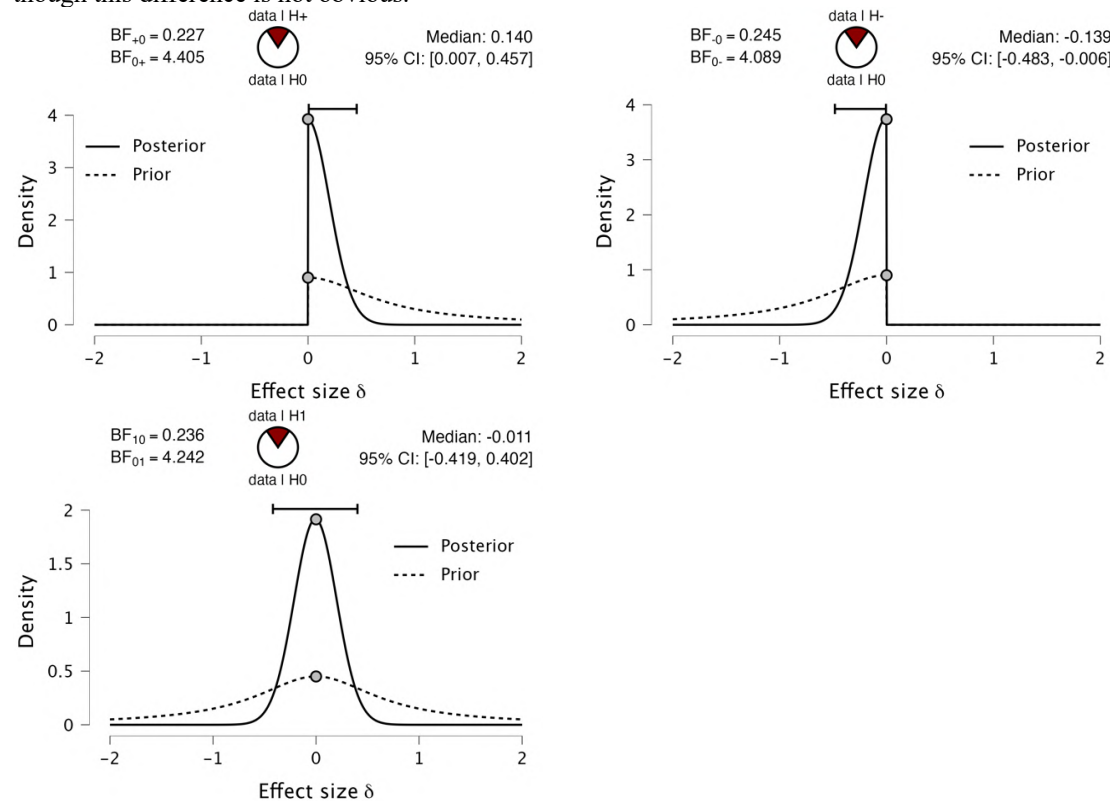
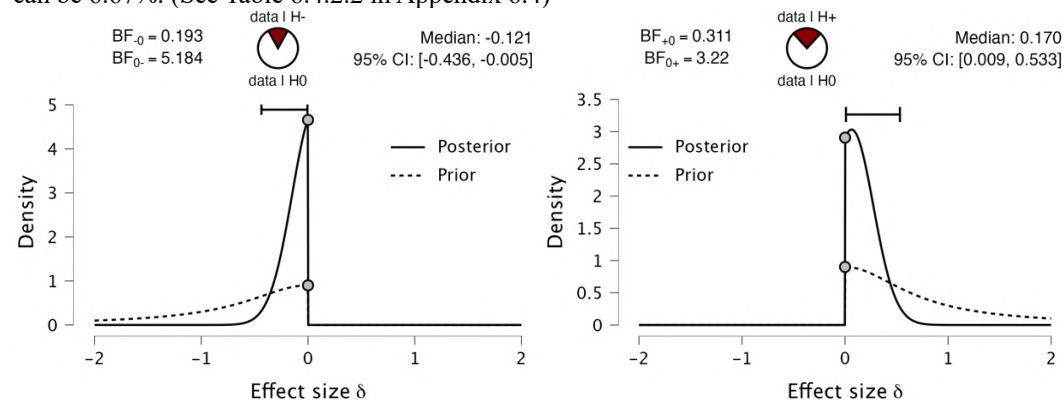


Figure 179 The inferential plots comparing the proportion of paratactic ideas in clause nexuses in Master and PhD sub-corpora (with diverse alternative hypotheses)

6.4.2.2 Idea Parataxis/Parataxis

The average proportion of paratactic ideas in paratactic structures becomes higher as hypotactic structures are all excluded, but is still lower than 1% in both groups (0.425% in group D and 0.456% in group M), though the proportion in some doctoral dissertations can be 3.64% and in some master's theses can be 6.67%. (See Table 6.4.2.2 in Appendix 6.4)



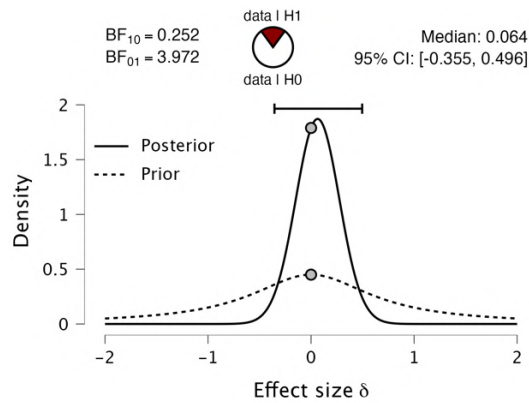


Figure 180 The inferential plots comparing the proportion of paratactic ideas in parataxis in Master and PhD sub-corpora (with diverse alternative hypotheses)

In the Bayesian test, the multiple of accepting the original hypothesis (group D > group M) is 5.184 (Figure 180, left 1) compared to the alternative one, with the Rhat value 1.009 and a medium effect size; the multiple of accepting the original hypothesis (group D < group M) is 3.22 (ibid, right 1) compared to the alternative one; and the possibility of supporting for the original hypothesis that two groups have the same average proportion is 3.972 times (ibid, left 2) of rejecting it. Hence, for the proportion of paratactic ideas in all paratactic structures, the doctoral group has a higher average value than the Master group.

6.4.2.3 Idea Parataxis/Projection Parataxis

As the proportion of paratactic locutions in paratactic projection can be around 90%, as suggested by Section 6.2.2.3, the index paratactic idea per paratactic projection then has a reverse condition as compared to the former. So, the average proportion of paratactic ideas in paratactic projections in group D is only 8.69%, and in group M is only 11.52%. (See Table 6.4.2.3 in Appendix 6.4)

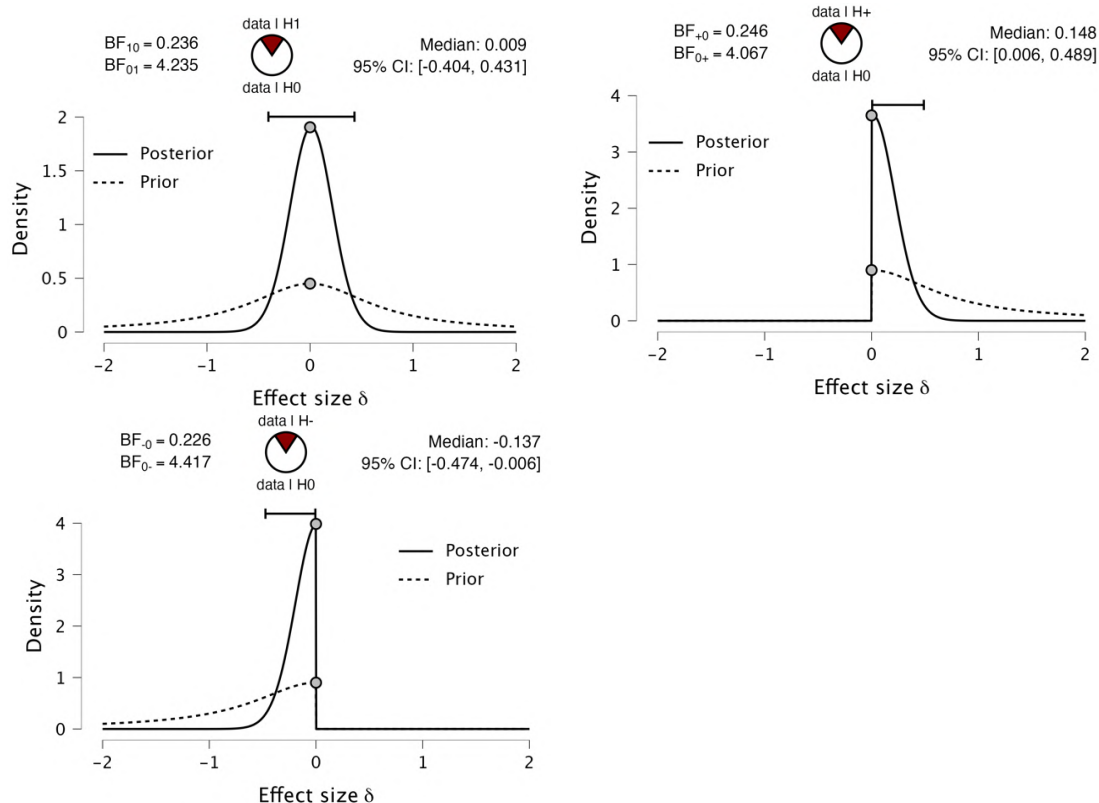


Figure 181 The inferential plots comparing the proportion of paratactic ideas in paratactic projections in Master and PhD sub-corpora (with diverse alternative hypotheses)

Though the average proportion of this index in group D is slightly higher than that in group M, as can be deduced from testing results in the paratactic locution/paratactic projection part, multiples of accepting the original hypothesis are, however, presented as diverse numbers in the Bayesian Mann-Whitney U test: it can be 4.235 (Figure 181, left 1) when the alternative hypothesis is two groups having unequal average proportions or can be 4.067 (ibid, right 1) when the alternative one is group D being greater than group M

or is 4.417 (ibid, left 2) when the alternative hypothesis is that group D has a smaller average proportion than group M, with the calculating value 1.004 and a medium effect size. Thus, doctoral candidates still have higher average proportions than Master students on the index paratactic idea per paratactic projections, though this diversity is not so apparent.

6.4.2.4 Idea Parataxis/ Idea

The index paratactic idea per idea excluded hypotactic ideas that only 1.44% in group D and 1.62% in group M are paratactic ideas, though 16.92% in some doctoral dissertations or 18.18% in some Master theses can be found. (See Table 6.4.2.4 in Appendix 6.4) Data are leptokurtic (the kurtosis value in group D is 13.913 and in group M is 7.676) and abnormally distributed with a right-sided tail (the skewness value in group D is 3.557 and in group M is 2.666) in both groups.

In the Bayesian Mann-Whitney U test, when the alternative hypothesis is two groups having unequal average proportions, the possibility of accepting the original one is 4.17 times (Figure 182, left 1) of accepting the alternative one; when the alternative is group D < group M, the multiple of supporting the original one is 4.017 (ibid, right 1) compared to the alternative hypothesis; and when the alternative hypothesis is set as group D < group M, the probability of accepting the original one is 4.327 times (ibid, left 2) of rejecting it, with the Rhat value 1.007 and a medium effect size. Therefore, the doctoral group also has a larger average proportion than the Master group on the index paratactic idea per idea, though this difference can almost be ignored because of the small number of paratactic ideas in the two groups.

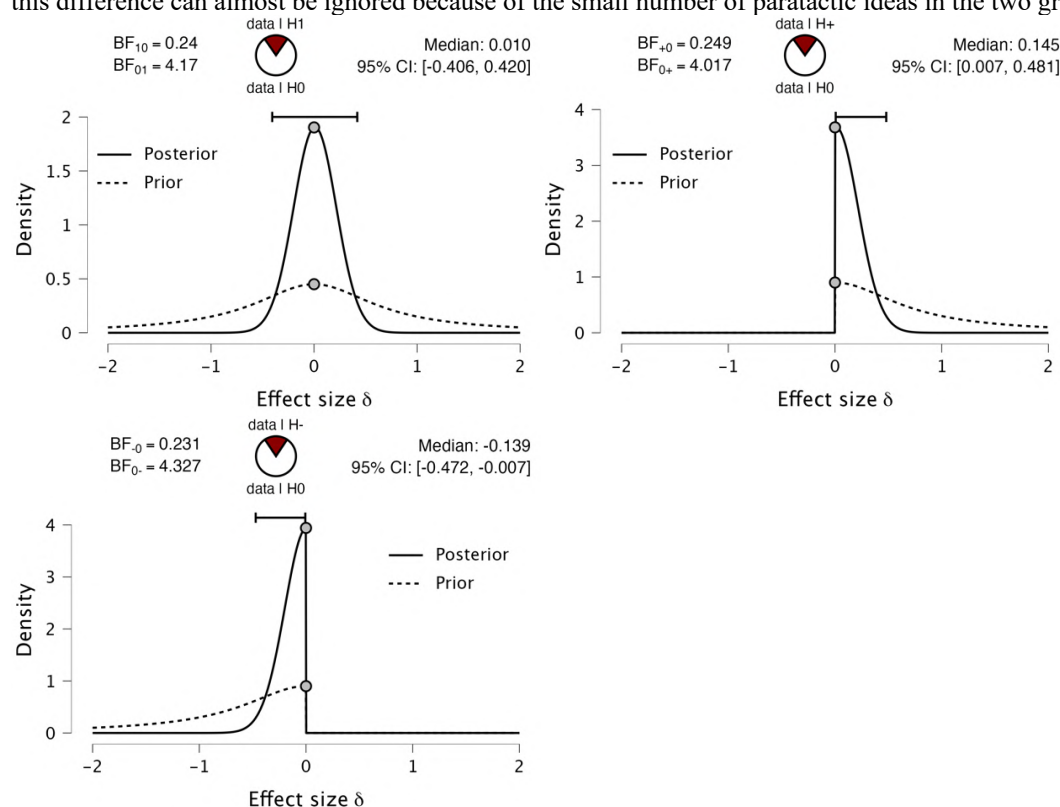


Figure 182 The inferential plots comparing the proportion of paratactic idea in ideas in Master and PhD sub-corpora (with diverse alternative hypotheses)

Hence, for paratactic ideas, except the index paratactic idea per parataxis that doctoral candidates have higher average proportions than Master group, no obvious differences can be seen on other three indexes, i.e., when the dominator is clause nexuses, paratactic projection or idea, though the average proportion of paratactic ideas per clause nexuses are more favored in Master group while the proportion with the denominator being the latter two are more preferred in doctoral group. One reason can be the higher number of clause nexuses and smaller number of paratactic structures in doctoral dissertations; one comes from an extremely small number of paratactic ideas in both groups that varieties are hard to investigate, and data are thus not convincing; one originates from doctoral students' preference for paratactic ideas rather than paratactic locutions; and still one is the favor of hypotactic ideas by Master students, which will be demonstrated in the next section.

6.4.3 The Distribution of Hypotactic Idea

The condition of the hypotactic idea is better as the number of which can be counted and compared by usual methods. Likewise, except the denominator being clause nexuses or hypotaxis, the condition of hypotactic idea per hypotactic projection has a reverse result compared to hypotactic location per hypotactic projection, and the index hypotactic idea per idea also has a reverse testing result as to paratactic idea per idea.

In total, the number of hypotactic ideas found in Master theses (3,398) is higher than that discovered in doctoral candidates (3,095), but the mean number in the doctoral group (96.72) is around 2.28 times that in the Master group (42.48). (See Table 6.4.3-1 in Appendix 6.4) Data in group D is platykurtic (the kurtosis value is -0.777), normally (the Shapiro-Wilk value is 0.973 ; the P-value of Shapiro-Wilk is 0.598 ; the division of the skewness value by the standard error of skewness is 0.35 , within $-2/+2$; and the division of the kurtosis value by the standard error of kurtosis is -0.96 , also within $-2/+2$) distributed with a quite short right-sided skew (the skewness value is 0.145) while in group M is leptokurtic (the kurtosis value is 3.892), abnormally (the Shapiro-Wilk value is 0.892 ; the P-value of Shapiro-Wilk is smaller than 0.001 ; two divisions are 5.79 and 7.32 respectively, both larger than $+2$) distributed with a right-sided skew (the skewness value is 1.558) as suggested by the table.

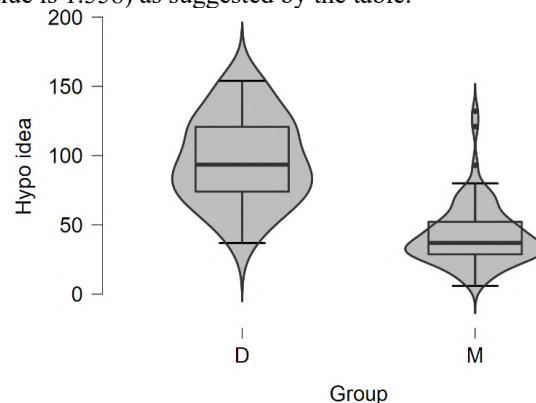


Figure 183 The boxplots comparing the number of hypotactic ideas in Master and PhD sub-corpora

From Figure 183, data in group D are normally distributed, and most of them concentrate in 50 to 150; while in group M are normally distributed except for three cases with high numbers being larger than 90 and left are smaller than 50.

6.4.3.1 Idea Hypotaxis/Clause Nexuses

The average proportion of hypotactic ideas in all clause nexuses in two groups can be around 13% (13.28% in group D and 13.45% in group M), and the proportion in group D ranges from 5.38% to 20.26% and in group M ranges from 3.77% to 30.65%. (See Table 6.4.3.1-1 in Appendix 6.4)

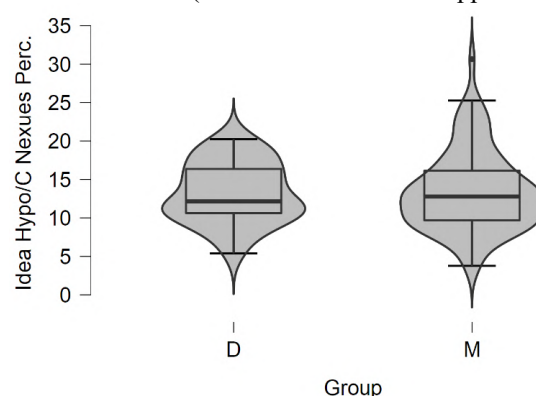


Figure 184 The boxplots comparing the proportion of hypotactic ideas in clause nexuses in Master and PhD sub-corpora

Data in the doctoral group are platykurtic (the kurtosis value is -0.831), normally (the Shapiro-Wilk value is 0.968 ; the P-value of Shapiro-Wilk is 0.458 ; the division of the skewness value by the standard error of skewness is 0.19 , within $-2/+2$; and the division of the kurtosis value by the standard error of kurtosis is -1.03 , also within $-2/+2$) distributed with almost no skew (the skewness value is 0.078) while in Master group are platykurtic (the kurtosis value is 0.758), abnormally (the Shapiro-Wilk value is 0.953 ; the P-value of Shapiro-Wilk is 0.005 ; two divisions are 3.06 and 1.42 correspondingly) distributed with a

short right-sided skew (the skewness value is 0.824). When boxplots are referred to, data in group D are slightly bi-modal, most of which concentrate in 5% to 20%, and that in group M are normally distributed except in one case with its proportion being as high as 30.65, others concentrate in 5% to 15%. (See Figure 184)

For one normally and one abnormally distributed data comparison, both the Welch t-test and Mann-Whitney U test are adopted. In Table 110, the p-value of the former test is 0.575, however, with the VS-MPR value being 1. As the Welch test has an extremely small effect size (-0.037) and the difference between the two groups is only 0.17%, as indicated by the location parameter, the Mann-Whitney U test result is consulted. But, the effect size value of 0.034 and the location parameter value of 0.282 also cannot provide enough information for comparison, though the VS-MPR value is also 1 and the p-value is 0.392.

Independent Samples T-Test

	Test	Statistic	df	p	VS-M PR*	Location Parameter	SE Difference	Effect Size	SE Effect Size
Idea Hypo/C Nexuses Perc.	Welch	-0.190	75. 686	0.575	1.000	-0.171	0.902	-0.037	0.209
	Mann-W hitney	1323.000		0.392	1.000	0.282		0.034	0.121

Note. For all tests, the alternative hypothesis specifies that group D is greater than group M.

Note. For the Welch t-test, effect size is given by Cohen's d. For the Mann-Whitney test, effect size is given by the rank biserial correlation.

Note. For the Welch t-test, location parameter is given by mean difference. For the Mann-Whitney test, location parameter is given by the Hodges-Lehmann estimate.

* Vovk-Sellke Maximum p -Ratio: Based on a two-sided p -value, the maximum possible odds in favor of H_1 over H_0 equals $1/(-e p \log(p))$ for $p \leq .37$ (Sellke, Bayarri, & Berger, 2001).

Table 110 The independent samples t-test for comparing the proportion of hypotactic idea in clause nexuses in Master and PhD sub-corpora

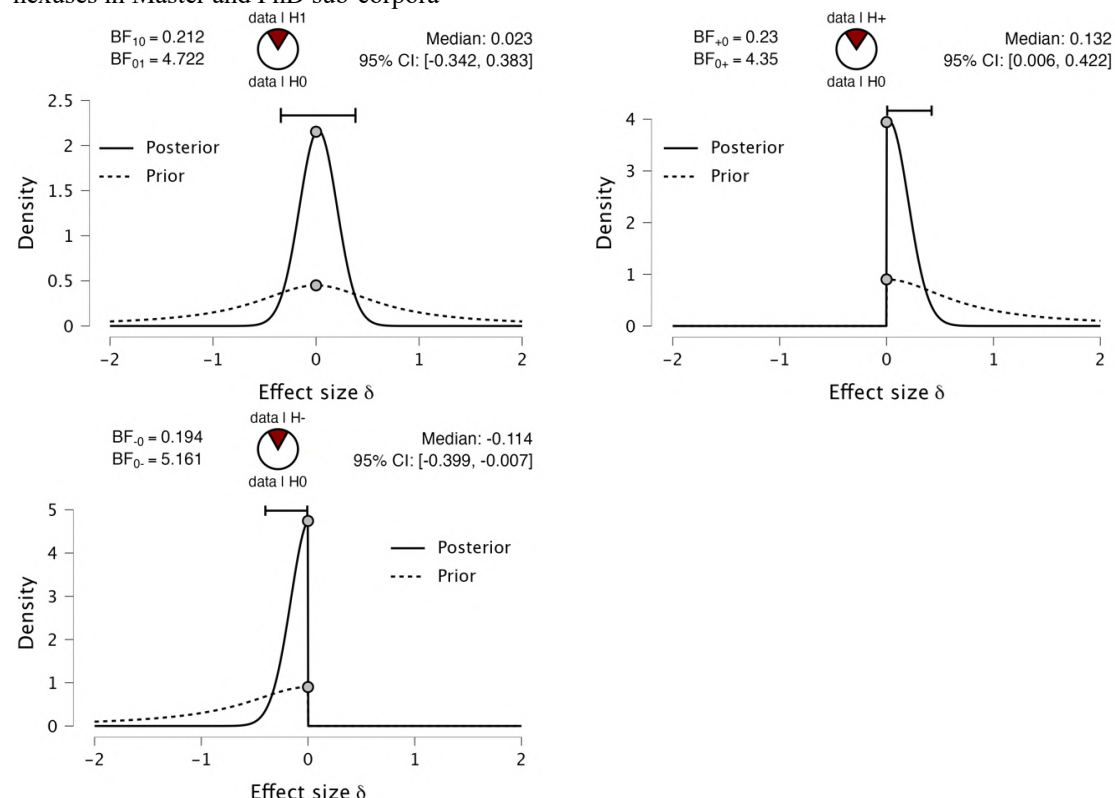


Figure 185 The inferential plot comparing the proportion of hypotactic ideas in clause nexuses in Master and PhD sub-corpora (with diverse alternative hypotheses)

The Bayesian Mann-Whitney U test is thus quite necessary: when the alternative hypothesis is two groups having diverse average proportions, the possibility of accepting the original one is 4.722 times (Figure 185, left 1) of supporting the alternative one; when the alternative hypothesis is group D being greater than group M, the multiple is 4.35 (ibid, right 1) of supporting the original hypothesis; and when

the alternative hypothesis is set as group D being smaller than group M, the probability of accepting the original hypothesis is 5.161 times (ibid, left 2) of accepting the alternative one, with a Rhat value 1.011 and a medium effect size. Thus, doctoral candidates have a higher average proportion than Master students on the index hypotactic idea per clause nexuses.

6.4.3.2 Idea Hypotaxis/Hypotaxis

The average proportion of hypotactic ideas in hypotactic structures is around 20% in both groups (20.36% in group D and 21.98% in group M), which is quite reasonable as five categories are included in hypotactic structures. The proportion of this index can range from 8.57% to 33.82% in doctoral dissertations and from 7.5% to 48.30% in Master's theses. (See Table 6.4.3.2-1 in Appendix 6.4)

Data in group D are platykurtic (the kurtosis value is -0.153), normally (the Shapiro-Wilk value is 0.988 , very close to 1; the p-value of Shapiro-Wilk is 0.966 ; the division of skewness by the standard error of skewness is 0.45 , within $-2/+2$; and the division of the kurtosis value by the standard error of kurtosis is -0.189 , also within $-2/+2$) distributed with a very short right-sided tail (the skewness value is 0.186); while in group M are leptokurtic (the kurtosis value is 1.276), abnormally (the Shapiro-Wilk value is 0.929 ; the p-value of Shapiro-Wilk is smaller than 0.001 ; two divisions are 3.83 and 2.40 , both larger than $+2$) distributed with a right-sided skew (the skewness value is 1.03), which corresponds to distributions described above. Seen from Figure 186, the statement above is clearer that except in three cases with high proportions larger than 35%, other data are still normally distributed and concentrated within 10% to 30%.

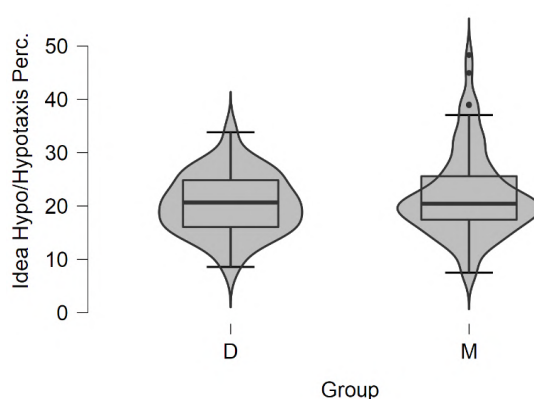


Figure 186 The boxplots comparing the proportion of hypotactic ideas in hypotaxis in Master and PhD sub-corpora

Independent Samples T-Test

	Test	Statistic	df	p	VS-MP R*	Location Parameter	SE Difference	Effect Size	SE Effect Size
Idea Hypo/Hypot axis Perc.	Welch	-1.211	81.4 99	0.115	1.481	-1.619	1.337	-0.234	0.211
	Mann-W hitney	1193.500		0.290	1.025	-0.695		-0.068	0.121

Note. For all tests, the alternative hypothesis specifies that group D is less than group M.

Note. For the Welch t-test, effect size is given by Cohen's d. For the Mann-Whitney test, effect size is given by the rank biserial correlation.

Note. For the Welch t-test, location parameter is given by mean difference. For the Mann-Whitney test, location parameter is given by the Hodges-Lehmann estimate.

* Vovk-Sellke Maximum p -Ratio: Based on a two-sided p -value, the maximum possible odds in favor of H_1 over H_0 equals $1/(-e p \log(p))$ for $p \leq .37$ (Sellke, Bayarri, & Berger, 2001).

Table 111 The Mann-Whitney U test for comparing the proportion of hypotactic idea in hypotaxis in Master and PhD sub-corpora

The independent samples t-test for one normally and another abnormally distributed data comparison can be the Welch's t-test and the Mann-Whitney U test. From Table 111, the p-value is 0.115 in the former test when the alternative hypothesis is that group D has a smaller average proportion than group M, with the VS-MPR value being 1.481, a difference thus does exist. The difference degree can be suggested by the location parameter value -1.619 , i.e., group D is 1.619% smaller than group M. The effect size value -0.234 indicates that this test has a small to medium effect size. As the effect size of the latter test is only

-0.068, attention will not be paid to it.

For this obvious testing results, multiples for accepting the original hypothesis in the Bayesian Mann-Whitney U test do not differ too much from each other: 4.054 (Figure 187, left 1) when the alternative hypothesis is two groups being unequal; 6.631 (ibid, right 1) when the alternative hypothesis is that group D has a larger average proportion than group M with a Rhat value 1.031 and a medium effect size; and 2.919 (ibid, left 2) when the alternative hypothesis is group D < group M. Therefore, for the index hypotactic idea in hypotaxis, doctoral dissertations have a smaller average proportion than Master theses. This might result from a higher amount of parataxis in the former group.

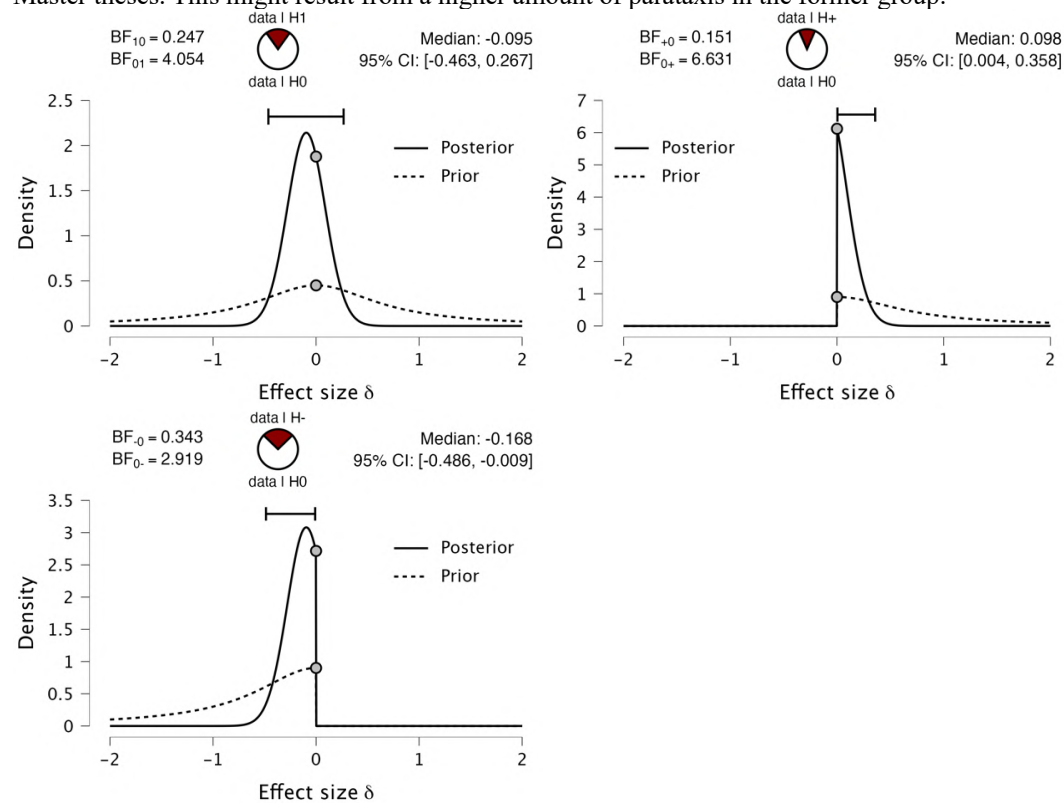
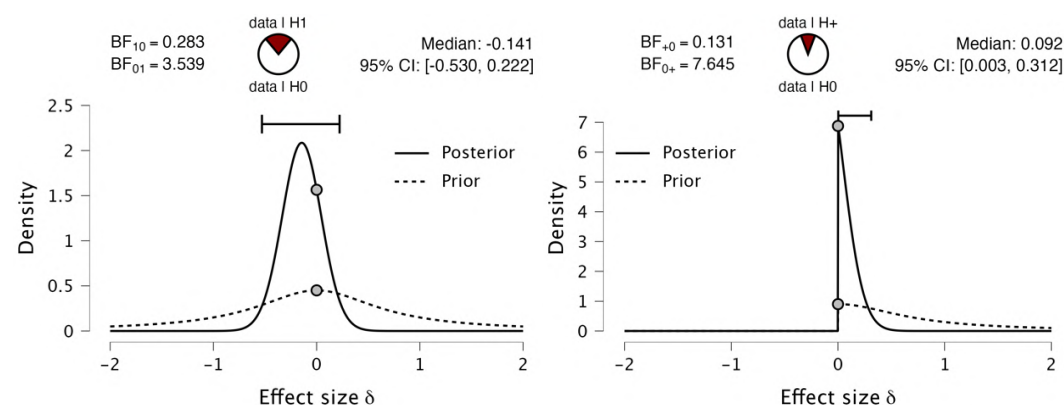


Figure 187 The inferential plots comparing the proportion of hypotactic ideas in hypotaxis in Master and PhD sub-corpora (with diverse alternative hypotheses)

6.4.3.3 Idea Hypotaxis/Projection Hypotaxis

The condition of hypotactic idea per hypotactic projections is just the reverse condition of hypotactic location in hypotactic projections that the average proportion of the former is thus more than 70% in two groups (72.82% in group D and 75.3% in group M) and the proportion in group D ranges from 48.24% to 91.35% and in group M from 35.29% to 100%. (See Table 6.4.3.3-1 in Appendix 6.4)

Data distributions and the independent samples t-test can be referred to in Section 6.2.3.3. Though the result of comparing this index can be inferred from the hypotactic location per hypotactic projections section that the doctoral group must have a smaller average proportion than the Master group, multiples for accepting the original hypothesis in the Bayesian Mann-Whitney U test are still diversified: 3.539 (Figure 188, left 1) when the alternative hypothesis is two groups being unequal; 7.645 (ibid, right 1) when the alternative hypothesis is that group D has a larger average proportion than group M with a Rhat value 1.062 and a small to medium effect size; and 2.302 (ibid, left 2) when the alternative hypothesis is group D < group M.



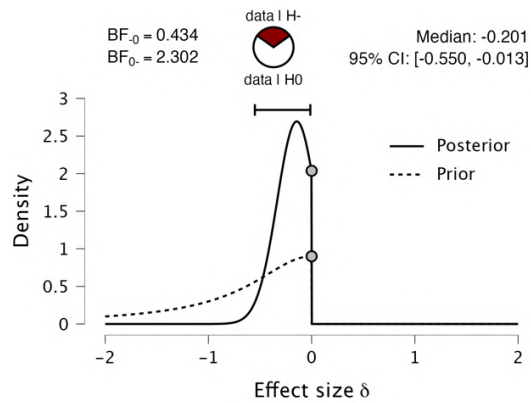


Figure 188 The inferential plots comparing the proportion of hypotactic ideas in hypotactic projections in Master and PhD sub-corpora (with diverse alternative hypotheses)

6.4.3.4 Idea Hypotaxis/ Idea

The proportion, average proportion, and data distribution of the index hypotactic idea per idea can be inferred from Section 6.4.2.4, in which conditions of paratactic ideas per idea have been described clearly. (See also Table 6.4.3.4-1 in Appendix 6.4) Though the result of comparing hypotactic ideas per idea is obvious that the doctoral group must have a smaller average proportion than the Master group, multiples for accepting the original hypothesis in the Bayesian Mann-Whitney U test are also diversified from those tested in Section 6.4.2.4: 4.264 (Figure 189, left 1) when the alternative hypothesis is two groups being unequal; 4.51 (ibid, right 1) when the alternative hypothesis is that group D has a larger average proportion than group M with a Rhat value 1.024 and a medium effect size; and 4.045 (ibid, left 2) when the alternative hypothesis is group D < group M. But this difference is extremely small because of the scattered data in both groups that have been mentioned before.

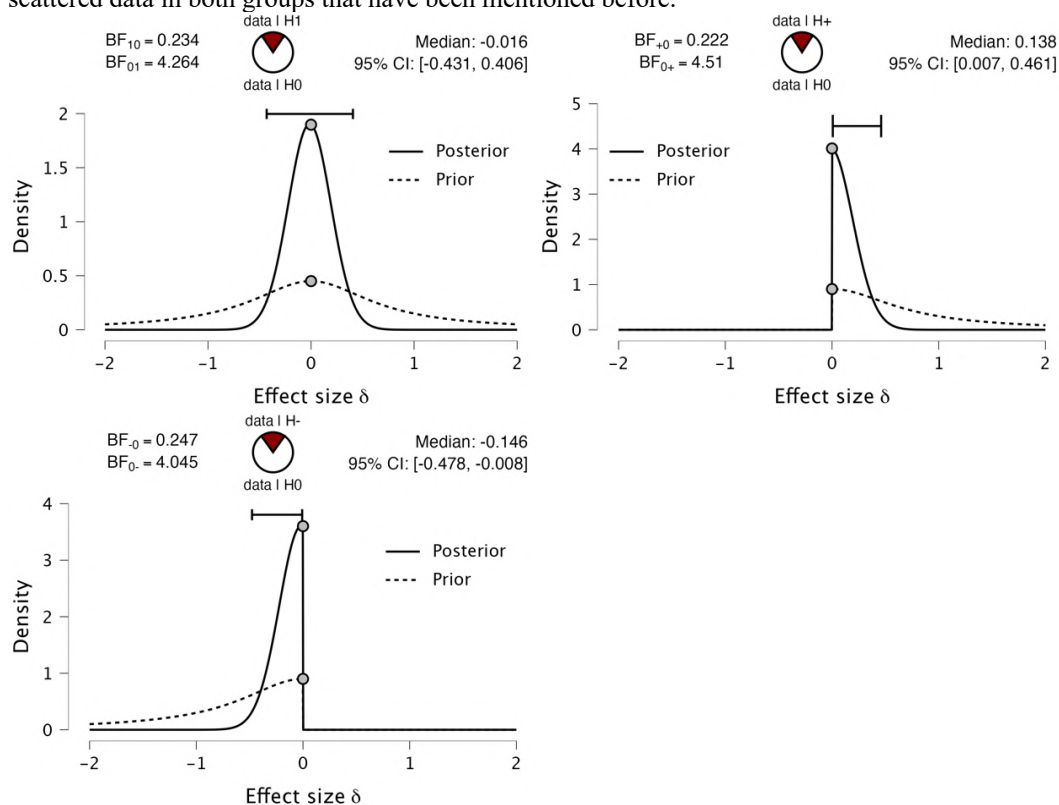


Figure 189 The inferential plots comparing the proportion of hypotactic ideas in ideas in Master and PhD sub-corpora (with diverse alternative hypotheses)

So, for comparing hypotactic ideas chosen in two groups, except for the denominator clause nexuses, other indexes have consistent results that doctoral candidates have a smaller average proportion than Master students, though the proportion of hypotactic ideas per idea is slightly diversified. One explanation for these results is the higher number of hypotactic structures and hypotactic projections in the doctoral group; another can be the lower number or scattered distribution of paratactic ideas in the Master group.

6.4.4 Summary of Idea Comparison

This section quantitatively compared the choice of ideas in doctoral dissertations and Master theses from the proportion of ideas in clause nexuses; paratactic ideas in clause nexuses, paratactic structures, paratactic projections, and ideas; and hypotactic ideas in clause nexuses, hypotactic structures, hypotactic projections, and ideas, to answer the question how two groups differ from each other, where do these differences come from, how to explain diversities or the sameness within two groups in choosing grammatical structures, why do students prefer to this kind instead of another kind, whether the result is meaningful, and how can the test be efficient for proving diversities within data. Though the number of paratactic ideas in both groups is so small that testing results must have been influenced, variances do not have to be in existence.

6.5 Idea (Qualitatively Analysis)

6.5.1 General Information of Idea

The qualitative analysis of ideas will follow the structure in the locution part that by different speech functions, ideas are divided into proposals and propositions as well. Besides demonstrating lexical choices with corresponding frequencies in Master and doctoral groups, instantiations chosen from each lexis are still annotated according to the framework of integrating SFG and Appraisal Theory to investigate students' interference in clause complex writing for meaning-conveying and information transmitting. By explaining principles for annotating one or two cases in each type, whether students are objectively or negatively projecting ideas and whether they interact with authors and inclined readers or not will be obviously revealed.

6.5.2 Idea - Proposal

The number and kind of proposals in projected ideas are even less than that in projected locutions. As suggested by Table 6.5.1-1 (see Appendix 6.5), together, 62 proposals are found in all Master theses, and only 35 have been discovered in doctoral dissertations. Except for the verb *hope*, more than half (37 exactly) in the Master group belong to it, and 14 are adopted in the doctoral group, other verbs such as *expect* (8 in group M and 5 in group D), *believe*, *propose*, *mean* and *wish* are also welcomed by students from both groups. Some verbs, however, are only used by Master students: *suggest*, *think*, *assume*, *decide*, *sense*, *turn out*, and *understand*, and some are only chosen by doctoral candidates: *show*, *concern*, *discover*, *hint*, *imply*, and *regard*. Hence, in total, 19 kinds of verbal lexes, 13 appear in Master theses and 12 in doctoral dissertations.

From 25 annotated instantiations on the subjectivity-objectivity cline, 13 belong to the moderate subjectivity class in which 6 are ++00 type, 5 are -0+0 type and 2 are +0+0 type that has never come out before; 8 are intense subjectivity class, 3 of them are all marked, 2 for +++0 and ++0+ types, one for --+= type; leaving four for the impartiality class (3 for +000 and one 0+00 type).

6.5.2.1 Impartiality (Two Types)

1. "admit" - "neutral" - "keep" - "present" (type +000)

The impartiality class for idea-proposals can either be +000 type or 0+00 type as both are concerned about the writer's support for either semantical or grammatical expressions or contents of the proposal, in which the former can be realized by verbs like: *believe*, *expect*, and *sense*.

E.g., 6.5.1 ||| Many humanistic teachers **believe*** || there should be no lesson plans or standard curricula || and that grades should be de-emphasized or abolished completely. || ---- CPhD08_0001_556

For objectively present ideas from *many humanistic teachers*, the verb *believe* is chosen in E.g., 6.5.1, and two proposals are hypotactically reported with the modal auxiliary *should*. The writer's support for the contents of proposals cannot be judged, but at least these thoughts are allowed to exist. As no sign for closing the gap between the writer and readers turns up, no requirement for readers can be interpreted, and the clause complex is annotated as +000 type.

E.g., 6.5.2 ||| Nevertheless, David is still lost in profound self-reproach for his failure to cherish this delicate blossom, || **sensing*** || he should take the inescapable responsibility for Dora's death, || because he realizes || his criticism has hurt the innocent feeling of fragile Dora. || ---- CMAC10LIT_7_309

In describing David's feelings for Dora in E.g., 6.5.2, the selection of the projecting verb *sense* followed by a proposal indicates David's inner activity and regret for Dora and, at the same time, reflects the writer's attitude that what David thinks at this moment should all be transmitted to readers, i.e., they have the right to know more about David and the main reason for Dora's death. Hence, the writer

supports the burst from David, though the description of the proposal remains unknown. And the whole clause complex is an objective present to readers without asking for their acceptance or slander of the subject's thoughts.

2. “unknown” - “affirm” - “keep” - “present” (type 0+00)

E.g., 6.5.3 ||| A high score **means*** || the country can be classified as collective; || a lower score is associated with cultures [[that promote individualism]]. ||| ---- CMAC13CU_27_613

In E.g., 6.5.3, the writer cannot be judged to approve for a *high score* or not, but one thing is for sure, i.e., the proposal *the country can be classified as collective* is confirmed no matter whether it is proposed by the writer or other researchers. If it is the latter condition, and if the writer has a different opinion on the proposal, then the contrast between a high score and a lower score described in the third clause would be invalid. As the interaction with readers in the clause complex cannot be estimated, the instantiation is marked as 0+00 type.

6.5.2.2 Moderate Subjectivity (Three Types)

1. “admit” - “affirm” - “keep” - “present” (type ++00)

Type ++00 is common in that it temporarily is the most frequent one in idea-proposals compared to other types as verbs such as *concern*, *discover*, *mean*, *propose*, *show*, *understand*, etc. can all back up to projections in this type.

E.g., 6.5.4 ||| Though claiming || that lexis seems more important than grammar, || linguists do not **mean*** || that teachers only need to teach lexis, || and should exclude grammar from classrooms. ||| ---- CPhD05_0003_1827

In E.g., 6.5.4, teachers' opinion that they *only need to teach lexis and should exclude grammar from classroom* is quite conflict with that of linguists, thus, the writer emphasizes this point by indicating *linguists do not mean...* after an enhancement: conditional clause and presents it to readers. Although the admission of linguists' thinking and the confirmation of the contents of their proposals have been clarified, the writer has no interaction with readers or has any requirement on them.

E.g., 6.5.5 ||| This suggests || that a majority of students realize || what their weakness is || and **understand*** || that they should make up for their weakness after class || in order to learn English well. ||| ---- CMAC11ME_16_349

The statement that students know their weaknesses and have the consciousness of making up for the weakness in E.g., 6.5.5 is apparently a viewpoint from the writer as the obscure pronoun *this* at the beginning of the clause complex, in fact, indicates a reflection of the writer. So, the activity *understand* is of high probability proven by the writer, and the proposal, though it is an introspection from students, is also supported. The clause complex is marked as ++00 type as a result.

2. “distance” - “neutral” - “close” - “present” (type -0+0)

Type -0+0 in the moderate subjectivity class is also common for idea-proposals when it is realized by verbs like *expect*, *hope*, *think*, *turn out*, *wish*, etc. Writers of this type keep their distance from the authors on one side and shorten the distance with readers on the other but convey nothing about the content of proposals.

E.g., 6.5.6 ||| The last element of foreign language anxiety, fear of negative evaluation, relates to anxiety observed, || when one **expects*** || that others would evaluate him or her negatively. ||| ---- CMAC10SP_11_461

The distance between the writer and readers in E.g., 6.5.6, in fact, is not as close as we have imagined as the conversational atmosphere is not obvious when the whole clause complex is inspected. But the writer keeps so far away from the subject of the projecting clause that the indefinite pronoun *one* is adopted in writing, making the distance to readers thus one bit closer. The case can be marked as -000 or -0+0 type from diverse angles in interpretation.

E.g., 6.5.7 ||| Despite the satisfaction of most students, some students, especially the students of higher writing proficiency, **wished*** || that the teacher could give them more suggestions to help them write idiomatic English essays. ||| ---- CPhD16_0001_1545

The wish in E.g., 6.5.7 coming from students is seen as dissatisfaction as opposed to the satisfaction that the writer might think these students as captious in hoping for *more suggestions to help them write idiomatic English essays*. Hence, by drawing a closer distance from readers, the writer intends to make it clear whether the requirement from some students is reasonable or not. The clause complex is marked as -0+0 type when the proposition has not been proved.

3. “admit” - “neutral” - “close” - “present” (type +0+0)

Type +0+0 in idea-proposals is the least compared to that in other grammatical categories.

E.g., 6.5.8 ||| Constructivists **believe*** || that different things even should be learnt in very different methods || while they were provided with [[what seem to be very similar experiences learned]]. ||| ---- CMAC12ME_5_285

In E.g., 6.5.8, the proposal *that different things even should be learnt ...* is a belief by constructivists, and the writer confirms this by selecting the verb *believe*, the first indicator is hereby marked. When the subject turns as *they* in the third clause to refer to these constructivists, the distance between readers becomes closer as the communicating attribute is manifested, the third indicator is also positively marked. However, the content of the proposal has not been affirmed, and the whole clause complex is presented to readers to inform them of ideas from constructivists, the other two indicators are unmarked.

E.g., 6.5.9 ||| David at last **decides***: || “better to be naturally Dora than anything else in the world”. (DC: P765) || since he can’t adapt Dora to himself, || he resolves to adapt himself to Dora. || ---- CMAC10LIT_7_299

By demonstrating David’s activity and explaining reasons for the activity in paratactically quoting his decision in E.g., 6.5.9, the writer endeavors to attract attention from readers. The name *David* and the pronoun *he* indicates a far distance between the writer and the subject, but the distance between readers is consequently bridged. As the decision is important, the description here is helpful for readers to get acquainted with David’s experience. So, the clause complex is +0+0 type.

6.5.2.3 Intense Subjectivity (Four Types)

1. “admit” - “affirm” - “close” - “present” (type +++0)

E.g., 6.5.10 ||| Grene actually **hinted*** || that the appreciation of the early plays should take into account the social context. ||| ---- CPhD17_0002_273

The clause complex in E.g., 6.5.10 either informs readers of the writer’s understanding of the importance of *taking the social context into account for the appreciation of the early plays* or demonstrates the writer’s acceptance of Grene’s proposals and then his/her proposal to readers that *the appreciation of the early plays by taking the social context into account* is useful and should be regarded. No matter what the condition is, one thing is for sure: the writer admits Grene’s hint and supports what he proposed, and at the same time advises readers to consider the proposal. Hence, the instantiation is +++0 type.

E.g., 6.5.11 ||| These studies **suggest*** || that metaphor should be understood from various aspects, || which directs us to investigate the cognitive mechanism of metaphor. ||| ---- CMAC20_05_335

The suggestion in E.g., 6.5.11 seems to have no relation to the writer as the subject of the projecting clause is *these studies*, in fact, originates from the writer’s interpretation that the proposal *metaphor should be understood from various aspects* is made by the writer him/herself. Thus, the first two indicators are marked as “admit” and “affirm.” Then, according to the writer, the suggestion *directs us to investigate...*, obviously indicating that he/she is standing together with readers that in further investigations, readers are asked to consider more about the *cognitive mechanism of metaphor* through reading the thesis written by the writer.

2. “admit” - “affirm” - “keep” - “seek for agreement” (type ++0+)

E.g., 6.5.12 ||| Based on the analysis of the above examples violating the Maxim of Relation, || this author **assumes*** || that in the translating process, the translator should make all his contributions relevant to the discussed topic, || as well as to the target readers’ background, || which requires || that the translator should follow the Maxim of Relation to achieve the communicative goal. ||| ---- CMAC09PR_28_515

When the subject of the projecting clause in E.g., 6.5.12 *this author* refers to the writer him/herself, the activity of putting forward the idea is, of course, admitted, and the proposals are highly proven to be effective by the writer. When the subject *this author* means another scholar, the idea is firstly important and should be noticed, proposals based on *the analysis of above examples* are then accepted and described to readers. No matter what the word denotes, the former two indicators on the subjectivity-objectivity cline are surely marked. Though repeated proposals are provided for readers and the intention for their acceptance of these proposals is overwhelming, the writer keeps a certain distance from them when *the translator* rather than other pronouns is adopted to refer to inclined readers. The clause complex is thus marked as ++0+ type.

E.g., 6.5.13 ||| It is thus **proposed*** || that the interdisciplinary researches of cognitive linguistics, Gestalt psychology and cognitive translation can be adopted as a new perspective for the present research, [[on whose basis studies on theoretical construction of cognitive translation strategies and translators’ cognitive psychological processing]] get a manipulated theoretical frame. ||| ---- CPhD18_0002_100

The proposal in E.g., 6.5.13 is put forward by the writer; thus, the content of the proposal must be supported, and the first two indicators are positively marked. The aim of selecting the verb *propose* is not only to inform readers of the interdisciplinary research method but also to expect a response from them, i.e., readers hoped to accept the proposal and then apply it in further research. Hence, the clause complex is ++0+ type with the writer’s intense interference.

3. “distance” - “deny” - “keep” - “seek for agreement” (type --0+)

E.g., 6.5.14 ||| Opposed to this view is the much longer tradition of narrative inquiry [[which **regards*** ||

the natural flow of discourse, the “parole”, should be the core of the study instead of the written or fixed form of language || as only the natural flow of language can be the manifestation of mind]]. ||| ---- CPhD20_0002_252

If the writer in E.g., 6.5.14 agrees with the idea, then the description *the much longer tradition* would not be used. Hence, the writer keeps distance from the subject of the projecting clause *traditional narrative inquiry* and refers it to *which* on the one hand, and shows negative attitudes toward the proposal by this inquiry that *the natural flow of discourse* should not be *the core of the study* but *the written or fixed form of language* on the other hand. At the same time, the writer hopes that readers can be influenced and then jointly fight for this regard though a distance within them is manifested. The clause complex is therefore marked as --0+ type.

4. “admit” - “affirm” - “close” - “seek for agreement” (type +++)

Type ++++ for idea-proposals is not the most frequent one, but it still plays an important part, especially when verbs like *hope*, *imply*, *wish*, etc., can be adopted in fulfilling projections.

E.g., 6.5.15 ||| So much has been discussed in terms of language nature as well as social and cultural factors [[which affect L2 learning]], || that I **hope*** || that we can have a better idea of [[what L2 writing is]] || and that we can find an appropriate way to learn and teach L2 writing in a maximized effective way. ||| ---- CPhD05_0003_1887

E.g., 6.5.15 is a typical instantiation for the +++++ type as the writer straightforwardly conveys his/her expectations to readers by writing as *I hope*. And the selection of *we* as a pronoun clearly shortens the distance with readers that they are hoping to accept the writer’s proposals, i.e., to *have a better idea of what L2 writing is and find a way to learn and teach it ...*

E.g., 6.5.16 ||| In the further study, the writer **wishes*** || that more rigorous researchers could deeply explore the characteristics of evidentiality adopted in academic discourse. ||| ---- CMAC17_05_644

The same condition is in E.g., 6.5.16, that the subject of the projecting clause is just the writer him/herself; the proposal to readers that they *could deeply explore the characteristics of evidentiality adopted in academic discourse* is confirmed by the writer. Though the object is addressed as *more rigorous researchers*, the distance with readers is not quite far away as this wish is directly conveyed to them, and their acceptance of the proposal is highly expected. Hence, the clause complex is +++++ type rather than ++0+ type.

So, for proposals in mental projections, writers would usually transfer authors’ expectations to readers or convey their own hopes to them, though they understand that this effort might be in vain. But as they have the freedom of communicating through thesis writing, the opportunity would be taken full advantage of to inform readers of what they have mastered, to affect readers and change their conventional thinking, to expect responses from readers, and to mutually fulfill their aims.

6.5.3 Idea - Proposition

The number of propositions projected by mental verbs in clause complexes in two groups can be the highest as all together, 2,323 are discovered in the Master group, and 1,989 are found in the doctoral group. (See Table 6.5.1-2) The most frequent verbs can be *show* (346 in the doctoral group and 331 in the Master group), *find* (265 and 242 correspondingly), *indicate* (253 in group D and 225 in group M), and *believe* (211 in Master’s while only 116 in doctoral) that they almost take up for half of all propositions. A slight difference in the frequency within the two groups in the former three verbs indicates that students endeavor to be objective in projecting mental propositions, while more choice on *believe* means Master students’ preference for personal ideas no matter from authors or themselves. Some verbs are also welcomed by students in both groups: *mean* (196 in group M and 138 in group D), *think* (198 vs. 97), *see* (168 vs. 114), *know* (172 vs. 72) and *suggest* (70 vs. 164), suggesting that except for reporting ideas from other scholars, students would usually mingle with their interpretations and understanding toward referenced viewpoints. Still, some verbs are common: *prove*, *assume*, *consider*, *understand*, *realize*, *demonstrate*, and *imply*; and some are key markers with not obvious frequencies: *find out*, *determine*, *investigate*, *discover*, *reflect*, *predict* and *feel* that the subjectivity of authors or the subject of the projecting clauses is fully emphasized, leaving fewer rooms for writers’ involvement; then are rare words but chosen in both groups: *afraid*, *identify*, *judge*, *suppose*, *decide*, and *remember*. Verbs like *guess*, *intend*, *deem*, and *wonder* only appear in the Master group, while *expect* only comes out in doctoral theses. In all 34 kinds of lexes, the variety in the Master group is 33, and in the doctoral group is 30, which differs not too much compared with their frequencies.

After annotating 58 instantiations from mental propositions in the subjectivity-objectivity cline framework, it is astonishing that most of them (altogether 34) belong to the moderate subjectivity class, in which 23 are ++00 type, 7 are +0+0 type, 2 for -0+0 type, and one each for special types that have never been found before: 0++0 and -+00; then 15 are marked as impartiality that type +000 accounts for

7, and three types that have temporarily all been discovered: type 00+0 (also 5), -000 (2 cases) and 0+00 (only 1); left 9 for the intense subjectivity class, in which 3 are fully marked, 5 for +++0 type and only one for type +-+0. The objectivity class (marked as 0000 type) is still missing in this category; one reason comes from annotating custom as the activity of referencing for scholars' ideas has already confirmed writers' admission of their existence sometimes, another is the fact that students indeed interfere with those propositions reported from other scholars though they struggle to avoid it.

6.5.3.1 Impartiality (Four Types)

1. "admit" - "neutral" - "keep" - "present" (type +000)

The familiar and common type +000 in mental propositions can be realized by verbs like: *afraid*, *believe*, *discover*, *find out*, *investigate*, etc.

E.g., 6.5.17 ||| Teachers are **afraid*** || that students will develop fossilized incorrect grammar without correction. ||| ---- CPhD16_0001_129

In E.g., 6.5.17, no matter whether the teacher really feels afraid of the phenomenon that *students will develop fossilized incorrect grammar without correction* or it is the writer who considers that teachers should feel afraid of the assumption, an attitude toward the occurrence of the thought is expressed by the writer. Hence, the first indicator is positively marked. But as no other sign for the content of the proposition or readers can be analyzed, the clause complex is only +000 type.

E.g., 6.5.18 ||| She **believes*** || although the domestic researches on conversational repair are still at initial stages, || this field has been developed quickly in recent years. ||| ---- CMAC20_02_119

Regardless of the writer's opinion on whether the *domestic research on conversational repair* in recent years has developed quickly or not, the subject of the projecting clause *she* confirms the proposition, at least as the verb *believe* is chosen to describe her attitude. Then, no other information, such as the writer's intention for readers to agree on this point or conversational atmosphere, is provided, the clause complex is +000 type.

2. "unknown" - "neutral" - "close" - "present" (type 00+0)

Besides verbal proposals, type 00+0 can also be marked for mental propositions. Verbs for this type usually are *afraid*, *decide*, *know*, *realize*, *think*, etc.

E.g., 6.5.19 ||| She **decided*** || she would do something for herself at that very moment. ||| ---- CPhD17_0004_1554

As the writer is describing the subject's (she) decision to readers in e.g., 6.5.19, the distance within them is shortened. But the writer does not show any admission to the decision, though it has already happened judging from the past tense of the verb, nor does he/she convey any approval for the proposition, so the requirement from readers is unnecessary. The clause complex is simply 00+0 type.

E.g., 6.5.20 ||| Mitchell **realizes*** || that he has made a mistake, || since he should have not talked about the Propocia. ||| ---- CMAC15_01_297

The same goes for E.g., 6.5.20 when the writer just presents Mitchell's realization to readers without making clear whether he/she supports the activity of realization or the content described in the projected proposition as the realization has no connection to the writer, nor does it have any influence on the writer or readers.

3. "distance" - "neutral" - "keep" - "present" (type -000)

E.g., 6.5.21 ||| The reporter-speaker intends for the hearer to **believe*** || that only the content of the reporting originates from the reported speaker. ||| ---- CPhD06_0004_387

Being different from E.g., 6.5.18, in E.g., 6.5.21, the verb *believe* makes the clause complex -000 type as the writer does not close the gap between him/her and the subject of the projecting clause *the reporter-speaker* especially when *intend* is selected to describe the subject's activity, nor does the writer stand in line with the object, i.e., the hearer as whether the latter believes the proposition or not has no relation to the writer. Hence, the whole projection is a presentation to readers of the reporter-speaker's intention, though a line is clearly drawn between the writer and the subject.

E.g., 6.5.22 ||| The researcher's attention was raised || and it was **supposed*** || that the differences between Level 1 and the other two levels were caused by the fact [[that they had more classes associated with computer because of their majors]]. ||| ---- CPhD14_0002_1229

When the researcher in E.g., 6.5.22 is not the writer but another scholar, the writer then keeps a distance from the subject of the projecting clause as the original pronoun *it* before the projecting verb *suppose* refers to the researcher from whom the supposition comes; and when the researcher is the writer him/herself, the choice of *it was supposed that* also indicates a distance that for objectively reflecting ideas, the writer excluded him/herself on purpose by changing the subject of reporting a proposition to an impersonal pronoun. If it is the former condition, the clause complex is -000 type; otherwise, it turns into -++0 as the writer confirms the supposition and attempts to inform readers of it.

4. “unknown” - “affirm” - “keep” - “present” (type 0+00)

E.g., 6.5.23 ||| <In regard to the organization of L1 mental lexicon,> it is **supposed*** || that words are stored in the mind in the form of a semantic network. ||| ---- CMAC12MO_3_181

In presenting a widely believed viewpoint to readers, the choice of *it is supposed that* excludes the function of the subject but emphasizes the contents of the proposition, as it might be deemed by the writer that who originally has the idea is not important, but the idea itself matters. As the real personal subject is unknown, the relationship between the writer and the subject remains insignificant. And by demonstrating the proposition to readers, the viewpoint at least is confirmed no matter what readers think and no matter they accept it or not. The clause complex is thus 0+00 type.

6.5.3.2 Moderate Subjectivity (Five Types)

1. “admit” - “affirm” - “keep” - “present” (type ++00)

Type ++00 in mental propositions is the most frequent one when considering not only other types but also other grammatical categories, and verbs for this type are also various: *consider, determine, discover, expect, find, find out, indicate, judge, mean, prove, realize, reflect, see, show, suggest, think, understand*, etc.

E.g., 6.5.24 ||| As a prediction, we would **expect*** || that one’s self-perceived communicative competence, though confounded with variables like apprehension and confidence, is rooted in his or her real language proficiency. ||| ---- CPhD13_0003_1353

Though the subject of the projecting clause *we* in E.g., 6.5.24 seems to close the distance with readers, in fact, excludes readers as the expectation comes from the writer but not others, hence, not only the activity of expecting is admitted, but also the content of the proposition is confirmed that *one’s self-perceived communicative competence is rooted in his or her real language proficiency*. The clause complex is ++00 type, and the writer does not have any expectation from readers to back up the proposition.

E.g., 6.5.25 ||| In opposition to Widdowson’s view, many linguists argue || that SFL provides tools for the text analyst, the teacher and the student to **understand*** || that meaning is based on choices || and those choices are available in the language (Martin, 1985/89; Martin, Christie & Rothery, 1994; Berry, 1995, 1996). ||| ---- CPhD09_0002_247

If the proposition *meaning is based on choices and those choices are available in the language* is not a fact or confirmed by most scholars or does not deserve to be known by the subject *the text analyst, the teacher and the student*, then the projecting verb would not be chosen as *understand* in E.g., 6.5.25. The proposition is described as a tool provided by SFL that has been endorsed by so many linguists listed in the bracket, on the one hand, the writer agrees with the proposition, and on the other hand, the writer considers the understanding of the viewpoint to be necessary, the first two indicators are thus positively marked. But readers are not advised on accepting the statement, nor do they directly be communicated to, the clause complex is therefore only ++00 type.

E.g., 6.5.26 ||| In the first case, it is the relation of cause and effect, || which indicates || that the matter has been resolved || and, in the second case, the sentence forms a conditional relationship, || which **suggests*** || that the matter has not been settled yet. ---- CMAC06SY_17_287

The projecting verb *suggest* in E.g., 6.5.26 means “indicate” that the viewpoint obviously comes from the writer instead of others though the subject of the projection is *which*, denoting the conditional relationship within a sentence. Therefore, the writer not only approves the content of the proposition but also conveys the ideas by choosing a demonstrative pronoun. As no sign of communicating with readers has been demonstrated, the clause complex is also ++00 type.

2. “admit” - “neutral” - “close” - “present” (type +0+0)

Verbs for projecting mental propositions marked as +0+0 type are *assume, consider, find, identify, imply, wonder*, etc.

E.g., 6.5.27 ||| He **assumes*** || that binding principles - more specifically Binding Principle A - apply cyclically on the basis of the information contained at the level of the syntactic phase. ||| ---- CPhD15_0001_260

In E.g., 6.5.27, the subject’s assumption must be admitted by the writer, otherwise, the viewpoint would not be reported in the thesis. However, the content of the proposition has not been supported during the process of being unrevealed to readers. So, the writer does not convey his agreement on the statement, nor does he/she ask readers for acceptance. The clause complex is just to present binding principles to other scholars. The clause complex is therefore marked as +0+0.

E.g., 6.5.28 ||| He modified Trubetzkoy’s phonemic theory || and made a study of it in a totally different way, || because he **considered*** || that it is the construction of the whole distinctive feature system [[that determines the position of the marked and unmarked (Jakobson, 1983, cited in Croft, 1990/2009)]]. ||| ---- CMAC10SY_9_111

In E.g., 6.5.28, it is hard to say that the writer confirms the content of the projected clause *it is the construction of the ...* though Croft and Jakobson in the bracket both endorse for it, at least admits the activity of *his* consideration as the thoughts can explain the subject's modification of Trubetzkoy's phonetic theory. And in revealing these reasons to readers, the adoption of *he*, in fact, draws the relationship with readers closer. So, the clause complex is another case of +0+0 type.

3. "distance" - "neutral" - "close" - "present" (type -0+0)

E.g., 6.5.29 ||| Thus, he **assumes*** || that waiting for the teacher to ask him by the name is still a wise and polite way. ||| ---- CMAC10SP_11_225

Though the same verb *assume* is chosen in E.g., 6.5.29, the clause complex is another type as *still* in the projected clause seems to convey the writer's negative attitude toward the subject that *his* viewpoint is outdated. Hence, the first indicator is marked as "distance" rather than "admit." The distance to readers tends to be marked as "close," although the writer's inclination to communicate with readers about *his* consideration is not obvious.

E.g., 6.5.30 ||| There was also another wide range of definition, || which **deems*** || that humor is of the same grade with intelligent, comedy, joke, etc. ||| ---- CMAC14PR_4_101

In describing the essence of *humor* to readers in E.g., 6.5.30, the relative pronoun *which* being as the subject of the projecting clause indicates the writer's maintaining a distance away from it as who has the ideas is unknown to not only readers but also to the writer though one thing for sure is that the definition has a wide range. But whether the statement is confirmed or supported by the writer remains blurry. The clause complex is marked as a -0+0 type.

4. "unknown" - "affirm" - "close" - "present" (type 0++0)

E.g., 6.5.31 ||| Amir blames himself badly, || for he betrayed Hassan || and planned to frame Hassan, || so in Clause 7 he said || he "was a liar, a cheat, and a thief", || which **intends*** || that he keeps the dirty secret, || deceives his father || and conceals the truth. ||| ---- CMAC15_05_590

Though the subject of the projecting clause in E.g., 6.5.31 is also the relative pronoun *which*, the idea originates from the writer him/herself as the proposition is the writer's interpretation of Amir's locution. It is hard to say that the writer supports the subject as he/she is not sure of the interpretation or just selects *which* on purpose to make the proposition objective. So, the first indicator is unmarked, but the second is "affirm." The distance between the writer and readers is not far, for the whole clause complex is an explanation of Amir's self-blame to them. Therefore, the case is a unique 0++0 type.

5. "distance" - "affirm" - "keep" - "present" (type -+00)

E.g., 6.5.32 ||| As it is well **known*** || that operationalizing and quantifying one factor is not an easy job, || a time span based on the joint influence of all the factors is hardly anything more than arbitrary. ||| ---- CPhD08_0012_265

Another unique type -+00 comes from the case when the subject of the projecting clause is *it*, and the projecting verb is chosen as *know* with its passive voice in E.g., 6.5.32 that an extensive consensus is reported by the writer, the distance is kept, but the content of the projected clause is confirmed, the two former indicators are thus marked as "distance" and "affirm." But as a present rather than an exchange or a discussion about some problems, the distance is kept from readers; the latter two indicators are unmarked.

6.5.3.3 Intense Subjectivity (Three Types)

1. "admit" - "affirm" - "close" - "present" (type +++0)

Type +++0 in intense subjectivity class for mental propositions can be realized by verbs like: *demonstrate, feel, imply, prove, remember*, etc.

E.g., 6.5.33 ||| Any change of semantic field in relation to the terms will be understood as marked, || and **imply*** || there has been a quite different event frame added in. ||| ---- CPhD17_0001_1259

The implication in E.g., 6.5.33 is an interpretation of the *change of semantic field in relation to the terms* by the writer, so not only the activity of *imply* is admitted, but also the description of the implication is approved by the writer. This explanation aims to inform readers of the additional event frame regardless of their acceptance or rejection. So, except for the last indicator, the other three are all positively marked.

E.g., 6.5.34 ||| Many researches have **proved*** || that it is a creative and effective teaching approach. ||| ---- CMAC07RE_19_717

In E.g., 6.5.34, the present perfect tense of the projecting verb *prove* suggests the writer's agreement on the projection *it is a creative and effective teaching approach*, otherwise, positive adjectives *creative* and *effective*, the endorsement from *many researches*, and the tense would all be changed. And the support for the research can be interpreted from the verb *prove*. As the writer not only wants to present the viewpoint to readers but also intends to receive feedback on the statement, the distance between them is hence closed. The clause complex is +++0 type.

2. “admit” - “deny” - “close” - “present” (type +-+0)

E.g., 6.5.35 ||| Those with low motivation may **feel*** || that there is no need to revise their writing. ||| ---- CPhD16_0001_199

The writer in E.g., 6.5.35 seems to have a distance with the subject of the projecting clause, though *those* is used to refer to it. From the dissertation, it is known that the subject is students, and the writer admits the existence of their feelings for *revising their writing*, though the choice *may* as a modal verb indicates disapproval from the writer on the content of their viewpoint. The clause complex seems to be a complaint to readers or just a declaration of the situation that the writer encounters. No matter which conditions it is, the distance between the writer and readers is not so far. Thus, the case is marked as +-+0 type.

3. “admit” - “affirm” “close” - “seek for agreement” (type ++++)

E.g., 6.5.36 ||| So we should **understand*** || that cultural knowledge and cultural awareness are the two important aspects [[which compose communicative ability || and is an important teaching content to reaching the teaching aim]]. ||| ---- CMAC08CU_33_221

The idea in E.g., 6.5.36 in the first place is admitted by the writer as the activity of *understand* is put forward by the writer *de facto* and then is confirmed of its contents as the viewpoint comes from the writer’s cognition that *cultural knowledge and cultural awareness are the two important aspects*. The selection of *we* as a pronoun draws the relation with readers closer, and the modal auxiliary *we should* indicates that not only the writer should know the viewpoint, but also readers should be aware of it and better confirm the statement. That is, the writer’s intention of persuading readers of his/her thoughts is revealed in the clause complex, which is then marked as ++++ type.

Relatively speaking, students intend to be objective in projecting mental propositions, regardless of whether the source is other scholars, authors, or their ideas. The former are original thoughts that are presented to readers in either an indirect or direct manner and which may be based on the student’s own interpretations of the authors’ activities, speeches, or viewpoints. The latter are perspectives, cognitions, or reflections that are conveyed by the students with a certain degree of personal emotion or attitude. It is not typically expected that readers will agree or accept the presented mental propositions; however, some students do utilize these propositions to persuade readers to align with their own perspectives.

6.5.4 Summary of Idea Marker Choices

In mental projections, except for the number of lexical choices, the kind for mental propositions is almost three times that for mental proposals, and the proportion of corresponding frequency for these lexes in the latter takes up more than ninety percent in both Master theses and doctoral dissertations, one obvious difference when considering about the subjectivity-objectivity annotation for instantiations lies in the attitude toward inclined readers that students have a higher tendency of persuading readers of accepting their perspectives or viewpoints in projecting proposals than in reporting propositions. One explanation can be the adoption of modal auxiliaries in proposals while being changed into normal verbs with diverse tenses or voices in propositions; another can be the characteristics of some verbs that are transitive with direct or indirect objects in projecting proposals, specifying objects that are influenced, despite most of the verbs projecting propositions are intransitive, excluding possibilities of designating people that would be impressed on. In addition, objects in mental proposals are usually turned into personal emotions, objective realities, phenomena, etc., in propositions, thus leaving less room for creating a communicating atmosphere with readers.

6.6 Conclusion for Projection Conditions

By adopting the same data calculating and comparison methods just as those in Chapter Five on the investigation of categories from hypotactic and paratactic projection to Quoting Locution and Reporting Locution, and to Quoting Idea and Reporting Idea; from Locution and Idea to propositions and proposals; from grammatical choices to lexical (projecting keywords) instances in Chapter Six, the purpose of horizontally and vertically digging out how different Master’s and doctoral students are in directly and indirectly referencing and citing other scholars’ words or thoughts is realized.

In general, the first Section 6.1 comparatively probes into some necessary indexes within Master theses and doctoral dissertations: the average proportion of projection in all clause nexuses; paratactic projections in clause nexuses, paratactic structures, and all projections; hypotactic projection in all clause nexuses, hypotactic structures and all projections; and the composition of projection, i.e., the number and the average proportion of locution and idea in all projecting structures that though hypotactic projection has a higher ratio in all clause nexuses in doctoral dissertations compared to Master theses, the lower proportion of ideas in all projections in the former group is an interesting phenomenon.

The quantitative analysis of the locution part in Section 6.2 then compared the choice of paratactic locutions from denominators being clause nexuses, paratactic structures, paratactic projections, and locutions to calculate the average proportions in each group, as well as hypotactic locutions when denominators are changed to clause nexuses, hypotactic structures, hypotactic projections and again locutions after conducting hypothesis tests within two groups on the index the average proportion of locution in all clause nexuses. Except for the vague result of the proportion of paratactic locutions because of few cases discovered in the corpus, though it might be slightly higher in the doctoral group than in the Master's, the ratio of locutions and hypotactic locutions are both greater in the theses written by doctoral candidates.

Section 6.3 pays attention to verbal lexes by dividing them into markers that project proposals or propositions according to the speech functions they perform in a clause complex and presenting each verb chosen in theses with their frequencies, the condition of which verbs are usually seen in each group, which markers are both common in two groups with high frequencies, which are common but have varieties of their preferences in two groups, and which verbs are unique or peculiar in each group has been demonstrated. By analyzing lexical choices of these projecting verbs with modal auxiliaries, tense, voice, transitivity, and categorizing instantiations into diverse types on the subjectivity-objectivity cline, the interpersonal metafunction of projection-locution in thesis writing has been revealed that students, on the one hand, endeavors to show their affirmation or negation to authors or the subject of projecting clauses, on the other hand, attempts to seek for agreement from readers on accepting referenced locutions. The biggest difference between verbal proposals and propositions lies in the degree of objectivity in presenting locutions to readers, as writers in the former structure can absolutely exclude themselves from projected words without interfering with any emotion or attitude, while in the latter, they cannot.

For the quantitative analysis of the idea part in Section 6.4, though the number of paratactic ideas in most theses is zero, which would contaminate calculating statistics, the Bayesian independent samples t-test has also been conducted for a precise result and as a supplement to means comparison. Through the statistics of grammatical indexes on the proportion of ideas in all clause nexuses; the selection of paratactic ideas in clause nexuses, paratactic structures, paratactic projections, and ideas; and the proportion of hypotactic ideas in all clause nexuses, in hypotactic structures, hypotactic projections and again ideas in two groups, a confirm statement has been reached to that doctoral candidates are in favor of paratactic ideas while Master students have a preference for hypotactic ideas, and two groups are almost the same in adopting ideas in all clause nexuses.

Besides listing kinds of lexes with corresponding frequencies in projecting mental proposals or propositions in clause complexes in Section 6.5, the annotation of chosen instantiations and the categorization of their types on the subjectivity-objectivity cline has also been uncovered that in presenting ideas to readers, students sometimes intermingle their own thoughts with original ones that they referred to, perplexing readers' differentiation between writers' interpretation and authors' expression; and in projecting mental proposals rather than propositions, the interaction with readers is more demanding as they are either asked to realize authors' thoughts or required to fulfill writers' expectations.

For a detailed description of annotation types from altogether 199 projection instantiations: in the impartiality class, the number of +000 type ("admit"-neutral-"keep"-present") is the highest (20 exactly) in locution-propositions while in other categories are no more than 10; type 00+0 ("unknown"-neutral-"close"-present), 0+00 ("unknown"-affirm-"keep"-present), and -000 ("distance"-neutral-"keep"-present) are all found in idea-propositions but appears respectively in locution-proposals, idea-proposals, and locution-propositions. In the moderate subjectivity class, type +0+0 ("admit"-neutral-"close"-present) comes out 15 times in locution-propositions while no more than 10 times in other categories; type ++00 ("admit"-affirm-"present"-keep) accounts for 23 in idea-propositions but only few in other categories; type -0+0 ("distance"-neutral-"close"-present) can be discovered in all four; but type +-00 ("admit"-deny-"keep"-present) can only be found in locution-proposals, and types 0++0 ("unknown"-affirm-"close"-present) and -+00 ("distance"-affirm-"keep"-present) are only annotated in idea-propositions; and the kind of types in idea-propositions can be the highest (five types with 34 cases). In the intense subjectivity class, type +++0 ("admit"-affirm-"close"-present) spreads over all categories; type ++0+ ("admit"-affirm-"keep"-seek for agreement) appears for 11 times in locution-propositions but disappears in idea-propositions; type --+0 ("distance"-deny-"close"-present) happens to be twice in locution-proposals and locution-propositions; type +-+0 ("admit"-deny-"close"-present) arises once in locution-proposals and idea-propositions; type --0+ ("distance"-deny-"keep"-seek for agreement) only emerges in idea-proposals and type -0++ ("distance"-neutral-"close"-seek for agreement) only in locution-propositions; and the kind of types in locution-propositions can be the highest (four kinds with 17 cases). For all marked types ("admit"-affirm-"close"-seek for agreement) in the intense

subjectivity class, type +-++ (“admit”-“deny”-“close”-“seek for agreement”) and --++ (“distance”-“deny”-“close”-“seek for agreement”) are only marked in locution-propositions. Above all, the objectivity class (type 0000, “unknown”-“neutral”-“keep”-“present”) is only marked in locution-proposals. Though more types should be discovered theoretically, and the frequency of each type can be more thoroughly demonstrated when a larger quantity of instantiations from all projecting verbs are annotated instead of just selecting one or two cases in each verb from Master and doctoral theses for comparison, the statistics above at least suggest the generality of some grammatical-lexical characteristics in students’ projecting locutions or thoughts to perform the interpersonal metafunction of language. As the qualitative analyzing framework is individually constructed, and the annotation for each clause complex depends highly on personal interpretation and cognition of meaning construing, the reliability and validity of research results need consideration in further research, and the framework can be modified according to diverse genres or research purposes. Anyhow, the consistent annotation of instantiations and comparisons of the universality and individuality of grammatical-lexical choices based on this consistency is of great significance.

An interesting and confusing projecting verb can be *suggest* as in locution-proposals, it means “advise” or “propose”; in locution-proposition, it means “express,” and in idea-proposition, especially in written texts, it is “convey”; while in mental-proposals or mental propositions, it can also have the meaning of “indicate,” hence in annotating or interpreting instantiations, besides making clear of the grammatical category of projected clauses, the meaning of *suggest* also depends on the context and the intention the writer has in transmitting information or viewpoints to readers.

As Martin and White (2005, p. 95) point out, the nature of projection, as a resource within the Appraisal system of engagement, is “dialogically directed towards aligning the addressee into a community of shared value and beliefs,” the investigation of projection chosen in thesis writing can thus not only indicate how writers strive to enter the academic circle when they show their attitude toward referenced locutions or ideas but also suggest how they affect readers by presenting knowledge they have learned, expressing their expectation on readers and communicate with them for a solution.

the former slightly exceeds the latter, with the selection of elaboration being only 15.34%.

In all hypotactic structures, more than half are enhancement type (50.28%), followed by idea (22.00%), elaboration (16.20%), locution (6.80%), and extension (4.72%); in paratactic ones, extension (79.65%) is the most favored by students, then is enhancement (11.21%), elaboration (6.69%), locution (1.94%), and idea (0.51%). By comparing the proportions, it is obvious that paratactic extension should be the most common selection, then followed by hypotactic enhancement, hypotactic idea, hypotactic elaboration, paratactic enhancement, etc.

To examine the intersection of taxis in each logico-semantic type, except the category extension that paratactic extension (91.52%) is far more frequent than hypotactic extension (8.48%), hypotactic structures are all predominated over paratactic ones in other categories, i.e., elaboration (79.11% vs. 20.89%), locution (84.54% vs. 15.46%), enhancement (87.53% vs. 12.47%), and idea that the proportion of hypotactic idea in all ideas even reaches up to 98.55%, whereas that of paratactic idea is only 1.45%.

7.1.1.2 Category Distribution of CC System in Doctoral Dissertations

The category distribution of clause complex system in doctoral dissertations differs not too much from that in Master's theses when Figure 191 is referred to. Though the proportion of parataxis (35.23%) is over one-third, it is still not high enough compared to that of hypotaxis (64.77%). The proportion of expansion (81.13%) is more than four times that of projection (18.87%) instead of being the same as the latter. In projection, the selection of idea (68.91%) is more than two times that of locution (31.09%); and in the expansion category, enhancement (46.07) is the most favored, then is the extension (39.69%) that both these two types have the proportion being larger than one third, and elaboration (14.24%) is the least choice for students.

Clause Complex	Hypotaxis		15623	64.77%				
	Parataxis		8499	35.23%				
	Complex	Idea	3136	68.91%	1.31%	41	98.69%	19.81%
								3095
								0.48%
								2.68%
								7.60%
								16.11%
								83.89%
								1187
								31.09%
								1415
24122	Projection	Elaboration	2786	14.24%	21.00%	585	79.00%	14.09%
								2201
								6.88%
								80.35%
								6.01%
								9.60%
								52.49%
								9017
								46.07%
								9017
24122	Expansion	Extension	7768	39.69%	87.91%	6829	12.09%	6.01%
								939
								80.35%
								9.60%
								52.49%
								9017
								46.07%
								9017
								46.07%
								9017

Figure 191 The number with corresponding frequency for taxis and logico-semantic types in the clause complex system (PhD sub-corpus)

When hypotactic structures are investigated, enhancement is the most common category with its percentage of 52.49%, followed by idea accounting for 19.81%, and elaboration for 14.09%, whereas locution only takes up 7.6% and extension only 6.01%. In parataxis, the proportion of extension (80.35%) takes the leading position that other categories are only a quarter of it, in which enhancement is 9.6%, elaboration accounts for 6.88%, locution falls to 2.68% while the idea is only 0.48%. Proportions suggest that paratactic extension is the most favorite category by doctoral candidates, then is hypotactic enhancement with their proportions being both higher than 50%; hypotactic idea and hypotactic elaboration are also welcomed; left categories are no more than 10% in which paratactic idea is the least one chosen for doctoral dissertations.

To compare taxis choices in each logico-semantic type, it is amazing that hypotactic idea should take up to 98.69% while paratactic idea is only 1.31%; the proportion of hypotactic enhancement (90.95%) is also larger than 90%, leaving paratactic enhancement to 9.60%; extension has an opposite condition that paratactic extension accounts for as high as 87.91% but paratactic extension for 12.09%; then is the proportion of hypotactic locution (83.89%), with paratactic locution being 16.11%; the last one can be hypotactic elaboration (79%) when paratactic elaboration reaches up to 21%.

Comparison M vs. D

The biggest difference in category distribution in the clause complex system in the Master's and doctoral groups lies in the choice of paratactic extension and hypotactic enhancement, the former of which is

more preferred by Master's students as the number is the highest in the corresponding group while the latter is more welcomed by doctoral candidates when the number 8,201 is compared to 6,829, though the proportion of these two categories in no matter each logico-semantic type or in respective taxis has no obvious diversity within two groups. When the number of taxis intersected with logico-semantic types are compared in descending order in each group, paratactic enhancement is the fifth in the Master's group, followed by hypotactic locution and hypotactic extension, whereas is the seventh in the doctoral group, the former of which is hypotactic locution and hypotactic extension. Hence, except for paratactic extension, hypotactic categories are all highly opted in organizing clauses instead of parataxis by doctoral candidates. Broadly, the diversity in choosing ideas or locution in the Master's group is larger than that in the doctoral group, while the difference between hypotaxis or parataxis and enhancement or extension is more distinct in doctoral dissertations.

It is no wonder that paratactic elaboration, paratactic locution, and paratactic idea are the least three categories for choice because of less paradigmatic or lexical selection of the language essentially in function realization and meaning transmitting on the one hand, and students' familiarity and preference for hypotactic structures on the other. In two groups, enhancement is the most favored, followed by extension, idea, elaboration, and locution when each category is ordered by its number separately. The proportion of expansion or projection in two groups changes slightly, and so does the proportion of taxis intersected with idea, locution, and elaboration.

7.1.1.3 Category Distribution of CC System in the Corpus

To combine calculation results from Master's theses with those from doctoral dissertations, the category distribution of clause complex system in the whole corpus can thus be fulfilled. (See Figure 192) Of all 49,445 clause nexuses, 62.84% are hypotaxis while 37.16% are parataxis. The higher possibility of choosing these nexuses goes to expansion (81.31%) rather than projection (18.69%), i.e., no more than one nexus is the projection in each of the five nexuses.

Clause Complex	Hypotaxis					
	31071	62.84%				
	Parataxis					
	18374	37.16%				
	Projection	9241	18.69%	Idea		0.50%
				6584		71.25%
				1.38%		91
				98.62%		6493
				2.29%		7.20%
				15.81%		420
				84.19%		2237
				28.75%		2657
				6.78%		1246
				79.06%		4704
Expansion	40204	81.31%		Elaboration		15.14%
				5950		14.80%
				20.94%		1246
				79.97%		14694
				89.81%		10.19%
				40.70%		16362
				5.37%		1668
				10.46%		1923
				89.25%		17892
				44.50%		15969

Figure 192 The number with corresponding frequency for taxis and logico-semantic types in the clause complex system in Master and PhD sub-corpora

In projections, less than one clause nexus in three nexuses turns out to be locution (the proportion is 28.75%) when the idea takes up as high as 71.25%. In expansion, enhancement (44.50%) and extension (40.70%) are both larger than one-third, and the former is more usual, leaving only 14.8% to elaboration. To calculate the proportion of these five categories, enhancement (36.19%) is the most welcomed structure, with its probability being larger than one-third, the same goes for extension as its proportion is 33.09%, then followed by idea (13.32%), elaboration (12.03%) and the last choice is locution, which accounts for only 5.37%.

In hypotaxis, it can be foreseen that more than half of the structure are enhancements (the detailed percentage is 51.4%), then is the idea with its proportion of 20.89%, elaboration ranks third when it accounts for 15.14%, locution the fourth for 7.2%, and the last choice is the extension for only 5.37%. In parataxis, extension predominates with its percentage of 79.97%, enhancement the second for 10.46%, then elaboration (6.78%), locution (2.29%), and idea (0.50%) gradually. From the number of these ten types, hypotactic enhancement (accounts for 32.297% in all clause nexuses) is the most welcomed, paratactic extension goes after it with its proportion 29.718%, then is hypotactic idea (13.132%), hypotactic elaboration (9.514%), hypotactic locution (4.524%), paratactic enhancement (3.889%),

hypotactic extension (3.373%), paratactic elaboration (2.520%), paratactic locution (0.849%), and paratactic idea (0.184%) is the last one. Thus, more than 60% of clause nexuses belong to hypotactic enhancement or paratactic extension, and paratactic locution or paratactic idea scarcely exists as compared to other types in the system.

When taxis and logico-semantic types are intersected, idea is greatly divided: hypotactic idea even takes up 98.62% while paratactic idea for only 1.38%, then is paratactic extension for 89.81% while hypotactic extension for 10.19%, hypotactic enhancement accounts for 89.25% while paratactic enhancement for 10.75%, the last but two belongs to locution that the ratio of hypotactic locution is 84.19% while for paratactic locution is 15.81%, and least difference exists within hypotactic elaboration (79.06%) and paratactic elaboration (20.94%).

The reason for quantitative differences, such as the diversity of hypotactic enhancement versus hypotactic extension in hypotaxis, can correlate to qualitative quality when the former is more delicate and highly elaborated than the latter in the system network, which manifests “the general principle governing the relationship between instantiation in text and systemic elaboration discussed in relation to transitivity system” (Matthiessen, 2002).

7.1.1.4 Comparison with Wu's Outcome

To compare the category distribution above with that of Wu's, the system outcome received by Wu is demonstrated here: (see Figure 193)

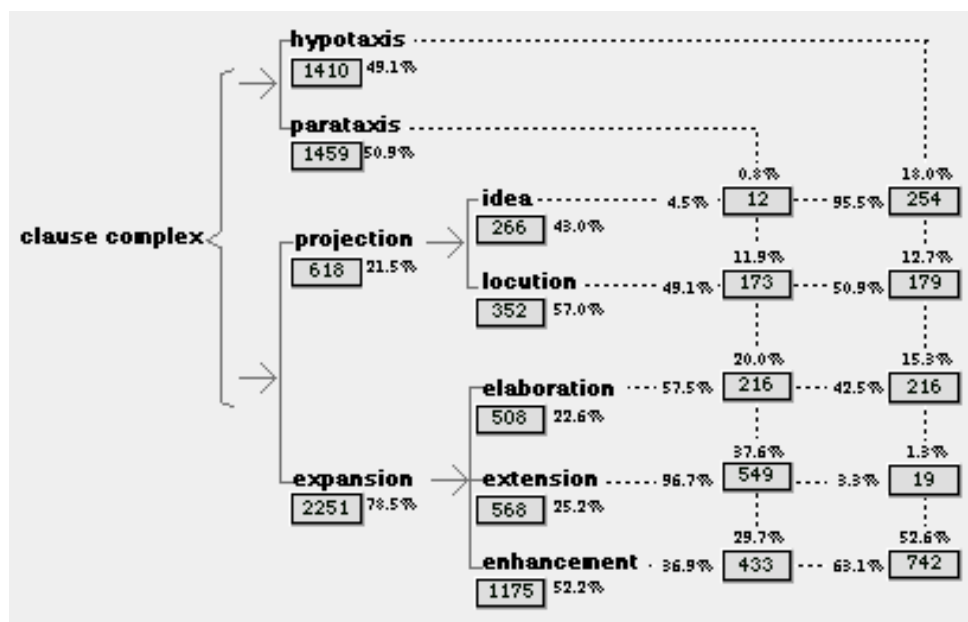


Figure 193 The number with corresponding frequency for taxis and logico-semantic types in the clause complex system (From Wu, 2000, p. 255)

From together 2,869 clause nexuses, frequencies, and probabilities in Wu's version show differences but also similarities: in general, hypotaxis (49.1%) and parataxis (50.9%) are “fairly well balanced” (Wu, 2000, p. 255), which is an ideal result but a discrepancy to the corpus' outcome above. This may indicate the particularity of the academic writing genre that texts as a medium for thesis writing certainly will result in a higher possibility of hypotactic structures and more complex clause nexuses being compared to synthetic forms or spoken genres only. An interesting result can be the division of expansion and projection that the percentage 78.5% versus 21.5% differs not too much from our corpus' outcome, which seems to be attributed to the sophistication of the grammatical system and its realization.

Within the projection category, a reverse result is presented that locution (57%) should be more frequently used than idea (43%)! This would be highly impossible for the thesis writing genre as locution mainly comes from directly quoted clause complexes, which may be avoided application by students when they intend to objectively evaluate scholars' viewpoints and reveal their abilities for various reporting expressions. In the expansion category, the percentage of enhancement (52.2%) that accounts for more than half of the total occurrences and roughly even divided proportion of elaboration (22.6%) and extension (25.2%) provides another possibility of category distribution in the clause complex system and proves the excessive application of paratactic extension (especially marked by “and”) by students to some extent in our corpus. The least number of ideas compared to the other four categories in Wu's corpus differs greatly from the corpus' outcome above that idea structure has a higher proportion than elaboration, clearly demonstrating the characteristics of thesis writing genre when projection cannot be

refrained from because of the literature review contexture.

In hypotaxis, the order of each category has no change from our corpus' result though hypotactic enhancement (52.6%), hypotactic elaboration (15.3%), and hypotactic locution (12.7%) slightly increase while hypotactic idea (18.0%) and hypotactic extension (1.3%) decrease. The same goes for the order of categories in parataxis when extension is the highest, followed by enhancement, elaboration, locution, and idea. But the proportions of them are quite different as extension decreases from 79.97% to only 37.6%, enhancement, however, increases from 10.46% to 29.7%, elaboration also increases from 6.78% to 20%, locution from 2.29% to 11.9%, and idea rises from 0.5% to 0.8% insignificantly. Thus, paratactic extension in Wu's corpus gives room for paratactic enhancement, paratactic elaboration, and paratactic locution, which reveals an almost perfect distribution of these categories. This also results in a differentiation when ten sub-categories are considered in the whole system: the proportion of hypotactic enhancement (25.863%), paratactic extension (19.136%), and paratactic enhancement (15.092%) together take up more than 60% of all clause nexuses; paratactic enhancement takes the third position, and paratactic locution the fourth with its proportion being 10.178%; paratactic locution (6.03%) is the seventh while hypotactic extension (0.662%) and paratactic idea (0.418%) are least applied in Wu's corpus; the proportion of paratactic extension in the whole system decreases as compared to our corpus' outcome (from 29.718% to 19.136%) while that of other paratactic structures all increase; and the proportion of hypotactic locution in the system increases in Wu's from ours' (4.524% to 6.239%) while that of other hypotactic structures all decrease.

When the distribution of taxis from the perspective of logico-semantic types is investigated, the biggest difference lies in extension that paratactic extension should account for 96.7% whereas paratactic extension for only 3.3%, then idea that hypotactic idea takes up 95.5% while paratactic idea for only 4.5%. Though the extension is not so sharply contrasted in our corpus, the condition of the ideas is almost the same, as the hypotactic ideas can be more than 90% in all idea structures. In enhancement, paratactic enhancement (36.9%) is more than one-third, which is not the same as our corpus, though hypotactic enhancement (63.1%) still has a higher percentage. In elaboration, paratactic elaboration (57.5%) is more than hypotactic elaboration (42.5%), and their percentage difference is only 15%, which greatly differs from our results. Another strange point comes from locution that hypotactic locution (50.9%) and paratactic locution (49.1%) are almost balanced, though the former is a little bit higher than the latter, which is not the same as that from our corpus when these two are contrasted.

In a word, the higher possibility of choosing paratactic enhancement (obvious markers can be temporal: later relation by *and then...* or *then*), paratactic elaboration (exposition or exemplification relation), and paratactic locution (directly referencing) in Wu's result certifies the particularity of thesis writing genre in our corpus that paratactic enhancement can be more common in the literary genre to emphasize sequential order of events, paratactic elaboration is more welcomed in the expository or scientific genre to describe characteristics of objects with examples, and paratactic locution is often seen in speeches or advertisement to appeal to listeners or audiences. The thesis writing genre, however, is not a necessity or strong requirement for these grammatical structures to perform the functions mentioned above. But interesting is the frequency order of sub-categories in hypotaxis or parataxis that this similarity may be a generality in all genres.

7.1.1.5 Category Range of CC System in the Corpus

When category distribution in the clause complex system in two groups is comprehensively inspected and when Section 7.1.2.2 is related to exploring the range of the possibility of choosing each category in texts despite personal writing behavior or preference, Figure 194 is completed and presented below.

In taxis, hypotaxis can range from 44.28% to 81.88%, while parataxis ranges from only 18.12% to no more than 56% (55.72% exactly). And at least 60% of clause nexuses are expansions (64.27% to 92.04%), while projections cannot exceed 36% (the highest is 35.63%), though they must exist in texts (the lowest proportion is 7.64%).

In Projection, the proportion of idea (33.33% to 100%) or locution (0% to 66.67%) has a broad scope according to preference and demand that the former can take up all projections while the latter can disappear in some texts. The same goes for the division of expansion, as elaboration can be 4.73% to 28%, enhancement can range from 25% to 63.93%, and extension has a wider range: 19.45% to 64.35%. Hence, whether extension or enhancement is more welcomed depends on fixed text, but this only affects the least frequency of elaboration in some expansions.

In hypotaxis, the proportion of enhancement can reach up to 78.31% from 29.2%, followed by idea from 7.5% to 48.3%, and elaboration from 2.12% to 35.1%. In some texts, locution or extension can disappear as the former ranges from 0% to 15.48% and the latter from 0% to 14.79%. In expansion, more than half can be the extension as it ranges from 57.02% to 95.12%, then is the enhancement from 1.22% to 25.88%. When most of the paratactic structures are extensions, and few are enhancements, elaboration

(0% to 22.13%), locution (0% to 19.25%), and idea (0% to 6.67%) can be zero. Thus, hypotactic enhancement and paratactic extension will always play a major role in clause nexuses, whereas paratactic idea is always minor as it is never higher than 6.67%.

Clause Complex		Hypotaxis					
		-	44.28%~81.88%				
		Parataxis					
		-	18.12%~55.72%				
		Idea		0%~18.18%	0%~6.67%	7.5%~48.3%	
		-	33.33%~100%		-	81.82%~100%	
		Projection		0%~19.25%	0%~15.48%		
		-	7.64%~35.63%	Locution	0%~88.89%	-	0%~100%
		-	0%~66.67%		-	-	
		Elaboration		0%~75%	0%~22.13%	2.12%~35.10%	
		-	4.73%~28%		-	25%~100%	
		Expansion		57.02%~95.12%	0%~14.79%		
Extension		65.38%~100%	-	0%~34.62%			
-	64.37%~92.04%		-	19.45%~64.35%			
Enhancement		1.22%~25.88%	29.2%~78.31%				
-	25%~63.93%		-	-			

Figure 194 The range of frequencies for taxis and logico-semantic types in the clause complex system (from Master and PhD sub-corpora)

To discuss logico-semantic types with taxis, the percentage will be horizontally compared in the system: in idea, paratactic idea can only as high as 18.18% while hypotactic idea ranges from 81.82% to 100%; then is enhancement that paratactic enhancement always exists as the lowest can be 1.12%, but will not be higher than 33.33%, with hypotactic enhancement ranging from 66.67% to 98.88%; for extension, at least 65.38% are paratactic extension and this even can reach up to 100% that hypotactic extension be totally ignored; the division of elaboration is rather blurry as hypotactic elaboration can take up from 25% to 100% and paratactic elaboration from 0% to 75%, meaning that hypotactic elaboration always exists and can be more than one fourth; the condition of locution is the most strange one as hypotactic locution can be zero, so does paratactic locution, but another condition can be the fact that all locutions are hypotactic locutions, leaving paratactic locution also to zero though in some texts it can be as high as 88.89%.

Hence, in different texts, no matter which genre the text belongs to, the choice of clause nexus relationship by writers at least includes the division of hypotactic enhancement versus paratactic enhancement, high probability of hypotactic idea and paratactic extension, and hypotactic elaboration to some extent, leading to a high proportion of expansion relations and a part of projections in clause complexes no matter they are paratactic or hypotactic ones.

7.1.2 Grammatical Comparison

The comparison of clause complex in two groups from the grammatical perspective involves altogether 73 indexes that the Bayes Factor values, the comparison of range versus mean results, and the comparison of mean and Bayes Factor values for these indexes will all be taken into account to reveal how different two groups are and where do these differences come from.

7.1.2.1 BF Value Comparison

From the perspective of Bayes Factor values in all 73 indexes, one index (extension hypo/extension) is confirmed that group D > group M as the value is 32.284; 19 indexes are obvious as values are larger than 10, in which group D > group M in 12 indexes (expansion hypo/clause nexuses, 15.44; enhancement/expansion, 14.07; extension hypo/clause nexuses, 13.185; expansion hypo/expansion, 13.13; hypotaxis/clause nexuses, 12.79; extension hypo/hypotaxis, 12.765; words/clauses, 12.5; enhancement hypo/clause nexuses, 12.37; enhancement hypo/enhancement, 12.37; enhancement/clause nexuses, 11.59; locution hypo/clause nexuses, 10.825; extension hypo/expansion hypo, 10.606) and group M > group D in 7 indexes (extension para/extension, 14.682; extension para/clause nexuses, 14.531; expansion para/clause nexuses, 13.18; expansion para/expansion, 13.13; parataxis/clause nexuses, 12.79; enhancement para/enhancement, 11.50; extension/expansion, 10.39); 6 indexes suggest that differences within two groups are not so obvious as their BF value differences are 1 to 2 (the higher

value minus the lower one), in which group D is slightly greater than group M in two indexes (elaboration hypo/elaboration and idea para/parataxis) while smaller in four indexes (enhancement para/parataxis, enhancement para/expansion para, projection para/clause nexuses, and locution para/projection para); 13 indexes indicate that differences within two groups are almost the same as the higher BF value minus the lower one is smaller than 1, in which group D has slightly larger tendency value than group M in 10 indexes (expansion hypo/hypotaxis, elaboration para/expansion para, extension para/parataxis, extension para/expansion para, projection para/parataxis, projection hypo/projection, locution para/clause nexuses, idea para/projection para, idea para/idea, idea hypo/clause nexuses) and group M > group D in 3 indexes (expansion para/parataxis, idea para/clause nexuses, idea hypo/idea); three indexes (elaboration para/parataxis, locution para/locution, and idea/clause nexuses) show no difference as BF value differences within two groups are smaller than 0.1; and other indexes present that either group D has larger average proportion than group M or has smaller one, with their Bayes Factor values usually within 5-10. Detailed distribution of these indexes is demonstrated in Table 112, and their values and comparing results are listed in Table 113.

Category	D > M	D < M	D = M	All
Confirm	1			1
Obvious (>10)	12	7		19
9-10	1	3		4
8-9	3	3		6
7-8	4	4		8
6-7	4	7		11
5-6		2		2
Not Obvious (BF 1-2)	2	4		6
Almost the same (BF <1)	10	3		13
Equal (BF <0.1)			3	3
ALL	40	33	3	73

Table 112 Summarized categories of grammatical indexes for comparing clause complexes in Master and PhD sub-corpora (according to Bayes Factor values)

Section	Parameters	Object		Object	BF ₊₀ / BF ₀₊ (D vs. M)	Notes 2
4.3						
	Words/Clauses	D	>	M	12.5	
	Chosen CC/1/2Clauses	D	<	M	6.21	
	Clause Nexuses/CC	D	<	M	8.8	
	Clause Nexuses L1 Percentage	D	>	M	8.3	
	Clause Nexuses L2 Percentage	D	<	M	7.3	
	Clause Nexuses L3 Percentage	D	<	M	9.7	
	Clause Nexuses L4 Percentage	D	>	M	6.6	
	Parataxis/Clause Nexuses	D	<	M	12.79	
	Hypotaxis/Clause Nexuses	D	>	M	12.79	
5.1						
	Expansion/Clause Nexuses	D	<	M	6.654	
	Expansion Para/Clause Nexuses	D	<	M	13.18	
	Expansion Para/Parataxis	D	<	M	>4.289, <5.154, =4.682	Almost the same
	Expansion Para/Expansion	D	<	M	13.13	
	Expansion Hypo/Clause Nexuses	D	>	M	15.44	

	Expansion Hypo/Hypotaxis	D	>	M	>5.277, <4.294, =4.735	Almost the same
	Expansion Hypo/Expansion	D	>	M	13.13	
	Elaboration/Expansion	D	<	M	6.477	
	Extension/Expansion	D	<	M	10.39	
	Enhancement/Expansion	D	>	M	14.07	
5.2						
	Elaboration/Clause Nexuses	D	<	M	7.257	
	Elaboration Para/Clause Nexuses	D	<	M	6.957	
	Elaboration Para/Parataxis	D	=	M	>4.842, <4.76, =4.801	
	Elaboration Para/Expansion Para	D	>	M	>5.492, <4.749, =5.093	Almost the same
	Elaboration Para/Elaboration	D	<	M	<5.56, >3.243, =4.097	
	Elaboration Hypo/Clause Nexuses	D	<	M	6.556	
	Elaboration Hypo/Hypotaxis	D	<	M	8.261	
	Elaboration Hypo/Expansion Hypo	D	<	M	9.626	
	Elaboration Hypo/Elaboration	D	>	M	>5.739, <3.824, =4.59	Not Obvious
5.4						
	Extension/Clause Nexuses	D	<	M	8.884	
	Extension Para/Clause Nexuses	D	<	M	14.531	
	Extension Para/Parataxis	D	>	M	>4.595, <4.43, =4.511	Almost the same
	Extension Para/Expansion Para	D	>	M	>5.014, <4.35, =4.661	Almost the same
	Extension Para/Extension	D	<	M	14.682	
	Extension Hypo/Clause Nexuses	D	>	M	13.185	
	Extension Hypo/Hypotaxis	D	>	M	12.765	
	Extension Hypo/Expansion Hypo	D	>	M	10.606	
	Extension Hypo/Extension	D	>	M	32.284	Confirm
5.6						
	Enhancement/Clause Nexuses	D	>	M	11.59	
	Enhancement Para/Clause Nexuses	D	<	M	9.05	
	Enhancement Para/Parataxis	D	<	M	>3.33, <5.11, =4.03	Not Obvious
	Enhancement Para/Expansion Para	D	<	M	>3.86, <5.59, =4.57	Not Obvious
	Enhancement Para/Enhancement	D	<	M	11.50	
	Enhancement Hypo/Clause Nexuses	D	>	M	12.37	
	Enhancement Hypo/Hypotaxis	D	>	M	8.85	
	Enhancement Hypo/Expansion Hypo	D	>	M	7.09	
	Enhancement Hypo/Enhancement	D	>	M	12.37	
6.1						
	Projection/Clause Nexuses	D	>	M	6.698	
	Projection Para/Clause Nexuses	D	<	M	>3.893, <5.047, =4.395	Not Obvious
	Projection Para/Parataxis	D	>	M	>4.793, <4.492, =4.637	Almost the same

	Projection Para/Projection	D	<	M	>3.209, <5.602, =4.081	
	Projection Hypo/Clause Nexuses	D	>	M	7.066	
	Projection Hypo/Hypotaxis	D	<	M	6.25	
	Projection Hypo/Projection	D	>	M	>4.76, <4.306, =4.52	Almost the same
	Locution/Projection	D	>	M	7.15	
	Idea/Projection	D	<	M	7.15	
6.2						
	Locution/Clause Nexuses	D	>	M	8.87	
	Locution Para/Clause Nexuses	D	>	M	>4.358, <4.062, =4.205	Almost the same
	Locution Para/Parataxis	D	>	M	6.518	
	Locution Para/Projection Para	D	<	M	>3.533, <4.728, =4.044	Not Obvious
	Locution Para/Locution	D	=	M	>4.399, <4.456, =4.427	
	Locution Hypo/Clause Nexuses	D	>	M	10.825	
	Locution Hypo/Hypotaxis	D	>	M	7.468	
	Locution Hypo/Projection Hypo	D	>	M	9.08	
	Locution Hypo/Locution	D	>	M	6.122	
6.4						
	Idea/Clause Nexuses	D	=	M	>4.562, <4.605, =4.583	
	Idea Para/Clause Nexuses	D	<	M	>4.089, <4.405, =4.242	Almost the same
	Idea Para/Parataxis	D	>	M	>5.184, <3.22< =3.972	Not Obvious
	Idea Para/Projection Para	D	>	M	>4.417, <4.067, =4.235	Almost the same
	Idea Para/Idea	D	>	M	>4.327, <4.017, =4.17	Almost the same
	Idea Hypo/Clause Nexuses	D	>	M	>5.161, <4.35, =4.722	Almost the same
	Idea Hypo/Hypotaxis	D	<	M	6.631	
	Idea Hypo/Projection Hypo	D	<	M	7.645	
	Idea Hypo/Idea	D	<	M	>4.045, <4.51, =4.264	Almost the same

Table 113 The list of comparing results of clause complex indexes in Master and PhD sub-corpora (with BF values and key categories)

No matter whether categories like “not obvious,” “almost the same,” and “equal” are taken into account or not, doctoral dissertations can be proved to have developed from Master’s theses on the aspect of clause complexes choices as the former group has larger average proportions in more indexes of grammatical categories than the latter one as indicated by tables above.

7.1.2.2 Comparison within Range, Mean, and BF Values

From Table 114, except that two indexes (words/clauses and clause nexuses/CC) are presented as numbers, other indexes are all presented as proportions. The first interesting phenomenon is that a narrower range of proportions for these indexes in a group does not mean a smaller average proportion of the index of this group, and vice versa. Indexes with a narrower range but a larger average proportion in the doctoral group can be the proportion of clause nexuses on Level 1, the proportion of hypotactic expansion, hypotactic extension, enhancement, hypotactic enhancement, projection or hypotactic projection in all clause nexuses, paratactic extension in paratactic structures, hypotactic expansion in hypotactic structures, etc., and with a broader range but a smaller average proportion can be paratactic projection or paratactic locution in clause nexuses, paratactic expansion in paratactic structures and paratactic projection in projections. Note 1 presents the comparison result of the mean in two groups and note 2 is the Bayesian independent sample t-test result for indexes.

Section	Parameters	Range		Mean		Note 1	Note 2
		D	M	D	M	D vs. M	
4.3	Words/Clauses ³¹	18.96 - 32.53	18 - 30.94	25.68	23.1	>	
4.3	Chosen CC/1/2Clauses	42.06 - 66.57	39.18 - 79.35	56.05	57.06	<	
4.3	Clause Nexuses/CC ³²	1.32 - 1.76	1.25 - 2.02	1.52	1.56	<	
4.3	Clause Nexuses L1 Percentage	56.84 - 75.52	49.49 - 80.09	66.17	64.88	>	
4.3	Clause Nexuses L2 Percentage	21.85 - 29.12	16.53 - 33.69	25.34	25.83	<	
4.3	Clause Nexuses L3 Percentage	2.13 - 11.42	2.36 - 13.70	6.64	7.34	<	
4.3	Clause Nexuses L4 Percentage	0.14 - 3.72	0 - 4.77	1.52	1.55	<	>
4.3	Parataxis/Clause Nexuses	18.12 - 46.20	24.39 - 55.72	34.89	39.01	<	
4.3	Hypotaxis/Clause Nexuses	53.80 - 81.88	44.28 - 75.61	65.11	60.99	>	
5.1	Expansion/Clause Nexuses	75.16 - 87.41	64.37 - 92.04	81.02	81.38	<	
5.1	Expansion Para/Clause Nexuses	18.12 - 45.42	23.18 - 54.25	33.96	38.01	<	
5.1	Expansion Hypo/Clause Nexuses	34.26 - 59.70	25.17 - 64.01	47.07	43.38	>	
5.2	Elaboration/Clause Nexuses	6.16 - 21.32	4.23 - 24.71	11.69	12.43	<	
5.2	Elaboration Para/Clause Nexuses	0.36 - 5.13	0 - 9.91	2.31	2.57	<	
5.2	Elaboration Hypo/Clause Nexuses	4.18 - 20.90	1.06 - 21.47	9.37	9.86	<	
5.3	Extension/Clause Nexuses	15.14 - 45.08	22.09 - 48.88	31.97	34.19	<	
5.3	Extension Para/Clause Nexuses	13.01 - 40.77	16.59 - 45.85	28.11	31.2	<	
5.3	Extension Hypo/Clause Nexuses	1.21 - 7.55	0 - 9.86	3.86	2.98	>	
5.4	Enhancement/Clause Nexuses	26.30 - 49.05	18.62 - 57.26	37.37	34.77	>	
5.4	Enhancement Para/Clause Nexuses	1.39 - 6.75	0.38 - 12.44	3.54	4.23	<	
5.4	Enhancement Hypo/Clause Nexuses	23.24 - 45.87	16.21 - 53.99	33.83	30.53	>	
6.1	Projection/Clause Nexuses	12.59 - 24.85	7.64 - 35.63	18.98	18.61	>	
6.1	Projection Para/Clause Nexuses	0 - 8.27	0 - 6.55	0.93	0.99	<	
6.1	Projection Hypo/Clause Nexuses	10.26 - 24.85	6.47 - 35.63	18.05	17.67	>	
6.2	Locution/Clause Nexuses	1.86 - 13.61	0 - 13.06	5.55	4.96	>	
6.2	Locution Para/Clause Nexuses	0 - 7.17	0 - 6.55	0.78	0.8	<	>
6.2	Locution Hypo/Clause Nexuses	1.79 - 10.82	0 - 8.50	4.77	4.16	>	
6.3	Idea/Clause Nexuses	6.47 - 20.26	3.77 - 30.65	13.43	13.64	<	=
6.3	Idea Para/Clause Nexuses	0 - 1.10	0 - 2.59	0.15	0.19	<	
6.3	Idea Hypo/Clause Nexuses	5.38 - 20.26	3.77 - 30.65	13.28	13.45	<	>
5.1	Expansion Para/Parataxis	77.87 - 100	84.06 - 100	97.36	97.47	<	
5.2	Elaboration Para/Parataxis	1.05 - 14.33	0 - 22.13	6.62	6.63	<	=
5.3	Extension Para/Parataxis	57.22 - 91.38	57.02 - 95.12	80.1	80.01	>	
5.4	Enhancement Para/Parataxis	4.13 - 25.88	1.22 - 24.76	10.65	10.82	<	
6.1	Projection Para/Parataxis	0 - 22.19	0 - 15.94	2.64	2.53	>	
6.2	Locution Para/Parataxis	0 - 19.25	0 - 15.94	2.22	2.07	>	
6.3	Idea Para/Parataxis	0 - 3.64	0 - 6.67	0.43	0.46	<	>
5.1	Expansion Hypo/Hypotaxis	62.77 - 83.65	47.75 - 90.73	72.3	71.15	>	

³¹ Number Count.

³² Number Count

5.2	Elaboration Hypo/Hypotaxis	6.67 - 25.52	2.12 - 35.10	14.34	15.91	<	
5.3	Extension Hypo/Hypotaxis	2 - 12.94	0 - 14.79	6	4.85	>	
5.4	Enhancement Hypo/Hypotaxis	42.58 - 63.69	29.20 - 78.31	51.97	50.4	>	
6.1	Projection Hypo/Hypotaxis	16.35 - 37.23	9.28 - 52.25	27.7	28.91	<	
6.2	Locution Hypo/Hypotaxis	2.34 - 15.48	0 - 13.75	7.34	6.85	>	
6.3	Idea Hypo/Hypotaxis	8.57 - 33.82	7.5 - 48.30	20.36	21.98	<	
5.1	Expansion Para/Expansion	23.29 - 57.00	28.33 - 66.20	41.82	46.69	<	
5.1	Expansion Hypo/Expansion	43.00 - 76.71	33.80 - 71.67	58.18	53.31	>	
5.1	Elaboration/Expansion	7.40 - 27.40	4.73 - 28.00	14.41	15.29	<	
5.1	Extension/Expansion	19.45 - 56.17	24.66 - 64.35	39.39	42.03	<	
5.1	Enhancement/Expansion	33.00 - 59.29	25.00 - 63.93	46.21	42.67	>	
5.2	Elaboration Para/Expansion Para	1.08 - 14.43	0 - 23.01	6.85	6.82	>	
5.3	Extension Para/Expansion Para	71.77 - 93.53	59.29 - 95.12	82.25	82.09	>	
5.4	Enhancement Para/Expansion Para	4.28 - 25.88	1.22 - 27.10	10.9	11.1	<	
5.2	Elaboration Hypo/Expansion Hypo	8.94 - 35.00	2.61 - 43.53	19.8	22.41	<	
5.3	Extension Hypo/Expansion Hypo	3.03 - 16.49	0 - 24.71	8.26	6.88	>	
5.4	Enhancement Hypo/Expansion Hypo	61.43 - 84.45	47.06 - 96.73	71.95	70.71	>	
5.2	Elaboration Para/Elaboration	2 - 39.45	0 - 75.00	19.66	22.66	<	
5.2	Elaboration Hypo/Elaboration	60.55 - 98.00	25.00 - 100.00	80.36	77.34	>	
5.3	Extension Para/Extension	80.19 - 96.46	65.38 - 100.00	87.55	90.98	<	
5.3	Extension Hypo/Extension	3.54 - 19.81	0 - 34.62	12.46	8.02	>	
5.4	Enhancement Para/Enhancement	3.26 - 17.30	1.12 - 33.33	9.53	12.37	<	
5.4	Enhancement Hypo/Enhancement	82.70 - 96.74	66.67 - 98.88	90.47	87.63	>	
6.1	Projection Para/Projection	0 - 44.62	0 - 40.91	4.88	5.75	<	
6.1	Projection Hypo/Projection	55.38 - 100.00	59.09 - 100	95.12	94.43	>	
6.1	Locution/Projection	8.65 - 65.05	0 - 66.67	29.49	27.59	>	
6.1	Idea/Projection	34.95 - 91.35	33.33 - 100	70.51	72.41	<	
6.2	Locution Para/Projection Para	50 - 100	0 - 100	91.31	88.48	>	<
6.3	Idea Para/Projection Para	0 - 50	0 - 100	8.69	11.52	<	>
6.2	Locution Hypo/Projection Hypo	8.65 - 51.77	0 - 64.71	27.18	24.42	>	
6.3	Idea Hypo/Projection Hypo	48.24 - 91.35	35.29 - 100	72.82	75.3	<	
6.2	Locution Para/Locution	0 - 59.50	0 - 88.89	10.51	12.53	<	
6.2	Locution Hypo/Locution	40.50 - 100	0 - 100	89.49	86.22	>	
6.3	Idea Para/Idea	0 - 16.92	0 - 18.18	1.44	1.62	<	>
6.3	Idea Hypo/Idea	83.08 - 100	81.82 - 100	98.56	98.38	>	<

Table 114 The list of comparing results of clause complex indexes in Master and PhD sub-corpora (with ranges and means)

When the Bayesian testing result for comparison is different from the mean comparison result marked in note 1, a new result would be labeled. In fact, except for the index hypotactic idea per clause nexuses that

group D indeed has a larger average proportion than group M, almost no difference can be found in other indexes (the proportion of clause nexuses on level 4, paratactic locution per clause nexuses, idea per clause nexuses, paratactic elaboration per parataxis, paratactic idea per parataxis, paratactic locution per paratactic projection, paratactic idea per paratactic projection, paratactic idea per idea, and hypotactic idea per idea) and should thus be marked as “equals to.”

For the relevance within indexes, some are reverse that the comparison result of one index can cause the opposite result of another, such as the index *the proportion of clause nexuses on each level*: the result “group D > group M” on level 1 comes from the result that group D has smaller average proportion on other levels, i.e., on level 2-4 (though slight difference exists within two groups on Level 4), hence the relevance is written as “clause nexuses L1 percentage (group D > group M) ← the percentage of clause nexuses on Level 2-4 (group D < group M),” in which “←” is interpreted as *comes from*. Still, some indexes remain consistent that their relevance can be described as follows in detail:

For the comparison result that group D > group M:

Hypotaxis/Clause Nexuses ← Hypotactic Expansion/Hypotaxis ← Hypotactic Extension / Hypotactic Expansion & Hypotactic Enhancement / Hypotactic Expansion;³³

Projection/Clause Nexuses ← Locution/Projection ← Hypotactic Locution/Locution;

Projection/Clause Nexuses ← Hypotactic Projection/Projection ← Hypotactic Locution/ Hypotactic Projection;

Hypotactic Expansion/Clause Nexuses ← Hypotactic Extension/Hypotactic Expansion & Hypotactic Enhancement/ Hypotactic Expansion;

Enhancement/Clause Nexuses ← Hypotactic Enhancement/Enhancement;

Locution/Clause Nexuses ← Hypotactic Locution/Locution;

Hypotactic Projection/Clause Nexuses ← Hypotactic Locution/Hypotactic Projection;

Paratactic Projection/Parataxis ← Paratactic Locution/Paratactic Projection;

Hypotactic Expansion/Expansion ← Hypotactic Extension/ Hypotactic Expansion & Hypotactic Enhancement/ Hypotactic Expansion;

Enhancement/Expansion ← Hypotactic Enhancement/Enhancement.

For the comparison result that group D < group M:

Parataxis/Clause Nexuses ← Paratactic Expansion/Parataxis ← Paratactic Enhancement/Paratactic Expansion;

Expansion/Clause Nexuses ← Elaboration & Extension/Expansion ← Paratactic Elaboration/Elaboration & Paratactic Extension/Extension;

Expansion/ Clause Nexuses ← Paratactic Expansion/Expansion ← Paratactic Enhancement/ Paratactic Expansion;

Paratactic Expansion/Clause Nexuses ← Paratactic Enhancement/ Paratactic Expansion;

Paratactic Projection/Clause Nexuses ← Paratactic Idea/ Paratactic Projection;

Elaboration/Clause Nexuses ← Paratactic Elaboration/Elaboration;

Extension/Clause Nexuses ← Paratactic Extension/Extension;

Idea/Clause Nexuses ← Paratactic Idea/Idea;

Hypotactic Projection/Hypotaxis ← Hypotactic Idea/Hypotactic Projection;

Idea/Projection ← Paratactic Idea/Idea.

By comparing these descriptions, it is obvious that preference for hypotactic enhancement, hypotactic extension, and hypotactic locution by doctoral students is the reason for those indexes, as group D has a higher average proportion than group M, while the favor for paratactic enhancement, paratactic idea, paratactic extension, and paratactic elaboration by Master students cause to higher average proportions of some other indexes in group M.

As to clause nexuses, it is the higher number of hypotactic extensions, hypotactic enhancements, and hypotactic locutions chosen by doctoral students that causes to the higher number of hypotactic expansions and hypotactic projections, and then to hypotactic structures in group D, and the higher number of hypotactic locutions can result to the higher number of hypotactic projections or locutions and then to projections in group D; while it is the preference of paratactic elaboration, paratactic extension, paratactic enhancement, paratactic locution and paratactic idea by Master’s students that result to the higher number of paratactic expansion and paratactic projection and then to parataxis in group M, the favor of expansion by Master’s students comes from their higher choice of paratactic expansion and paratactic elaboration, paratactic extension, and paratactic enhancement in detail on the one hand, and

³³ The description can be interpreted as: the higher average proportion of hypotactic structures in all clause nexuses in group D than that in group M comes from the higher average proportion of hypotactic expansion and originates from the higher average proportion of hypotactic extension and hypotactic enhancement choices.

originates from the higher number of elaborations and extensions, and of paratactic elaborations, hypotactic elaborations, paratactic extensions, and paratactic enhancements in detail on the other. In parataxis, it is the higher number of paratactic locutions that causes more paratactic projections in the doctoral group, while the preference for paratactic enhancement and paratactic elaboration can cause a higher number of paratactic extensions in the Master's group. In hypotaxis, more hypotactic expansion structures come from the higher number of hypotactic extensions and hypotactic enhancements in group D as compared to group M, while the higher number of hypotactic ideas can cause more hypotactic projections in the Master group. In expansion, the doctoral group has more hypotactic expansions because of the higher number of hypotactic extensions and hypotactic enhancements, while the less paratactic expansions are due to the lower number of paratactic enhancements compared to the Master's group.

Hence, a short conclusion can be drawn here that doctoral students prefer structures such as hypotactic enhancement, hypotactic extension, hypotactic locution, and paratactic locution, while Master's students are more in favor of paratactic enhancement, paratactic elaboration, paratactic extension, paratactic idea, and hypotactic idea.

7.1.3 Lexical Comparison

As grammatical categories of the clause complex are judged by specific words or phrases, i.e., each lexis marks a kind of relationship within two clauses in a clause nexus, varieties with corresponding frequencies of lexis chosen by students will be presented and illustrated in detail in following sections from five sub-categories, i.e., elaboration, extension, enhancement, locution, and idea. Though markers are not sorted as paratactic or hypotactic, most can only denote one "taxis". The total number of main markers in each sub-category is smaller than that is calculated in related grammatical categories as some clause nexuses are implicit ones that no obvious markers can be found for understanding their relations on the one hand, and some markers are not calculated, marked or considered due to annotation workload on the other.

7.1.3.1 Annotating Markers in Elaboration

Most elaborations are clarification relations rather than exposition or exemplification ones that clause nexuses marked by hypotactic lexis *which* and *it is that* are outstanding (1,365 and 430 respectively in Master's theses and 1,015 and 371 in doctoral dissertations) as indicated by Table 115.

Elaboration 25							
M		D		M		D	
No.	Lexis	Lexis	No.	No.	Lexis	Lexis	No.
Exposition (paratactic) 15				Exemplification 7			
99	colon	colon	144	26	such as	such as	22
88	that is	that is	71	20	for example	for example	14
47	that is to say	i.e.	33	16	examples/instance	for instance	6
26	i.e.	dash	26	5	e.g.	e.g.	4
13	dash	that is to say	24	5	for instance	like	2
12	in other words	in other words	17	5	take ... as examples	examples/instance	2
12	namely	namely	15	4	like	take ... as examples	1
5	semicolon	semicolon	7	Clarification 3			
1	in another saying	in particular	1	1365	Clarification: which	Clarification: which	1015
1	or rather	put it another way	1	430	It is that	It is that	371
1	put it another way			276	Clarification (others)	Clarification (others)	328
1	simply put						
1	specifically						
1	to some extent						

Table 115 Main markers for annotating the elaboration relations in Master and PhD sub-corpora
Still, some clarification clause nexuses are implicit that no markers are chosen. In exposition relations

(paratactic elaboration), except for punctuation marks such as colon, dash, and semicolon, phrases like that is, that is to say, i.e., in other words and the word namely are most frequently applied in. For exemplification, phrases such as, for example, for instance, and the abbreviation e.g., are common in both groups. In paratactic exposition relation, markers such as in another saying, or rather, simply put, specifically, and to some extent appear once in Master's theses but are not found in PhD dissertations.

7.1.3.2 Annotating Markers in Extension

Most of the extension structures are positive additions marked by *and* (at least one-third in two groups) and adversative additions by *while* (more in the doctoral group) and *but* (more in the Master's group) as indicated by Table 116. Then the alternative relation by *or*, positive addition by *not only...but also...*, and adversative addition by *whereas* are outstanding, though the number of the last one is not obvious in the Master's group. For variation: replacive relations, Master's students prefer *instead of* and *not...but...* while doctoral students are more in favor of *not...but...* and *rather than*. Both groups would choose *except* to connect clauses with variation: subtractive relations, though *but* is chosen in the Master's group and *except that* appears often in the doctoral group. The negative additive relation is only presented by the word *nor* and the phrase *neither...nor...* in both groups, and the number for this relation is quite small. The high frequency of parataxis in extension relation in two groups originates from the choice of *and*, positive additive relation, *but*, *or*, and *not only...but also...* Master's students would choose *but* to indicate variation: replacive relation within clauses, but doctoral candidates would not.

Extension 29							
M		D		M		D	
No.	Lexis	Lexis	No.	No.	Lexis	Lexis	No.
Addition 13				Variation 12			
Addition: Additive 6				Variation: Replacive 8			
3368	and	and	2503	59	instead of	not... but...	53
575	additive	additive	533	51	not... but...	rather than	51
137	not only...but also...	not only...but also...	127	37	rather than	instead of	27
25	as well as	as well as	32	13	instead	rather	20
7	apart from	besides	14	11	but	instead	12
7	besides	apart from	8	9	rather	but not	4
Addition: Negative 2				6	not	not	4
3	nor	nor	7	4	but not		
3	neither...nor...	neither...nor...	4	Variation: Subtractive 4			
Addition: Adversative 5				8	except	except	6
655	but	while	503	5	but	except that	5
430	while	but	475	5	except for	only	1
46	whereas	whereas	152	1	other than	other than	1
45	without	without	24				
16	(and) yet	(and) yet	19	Alternation 4			
				204	or	or	272
				5	either...or...	either...or...	20
				1	if not	if not	2
				1	or else	or else	1

Table 116 Main markers for annotating the extension relations in Master and PhD sub-corpora

7.1.3.3 Annotating Markers in Enhancement

The most common relation in enhancement is cause: purpose marked by the preposition *to* (more than 1,000 in both groups) and then temporal: same time by *when* (more than 1,000 in the Master's group). Relations such as manner: means by the preposition *by* (more than 500 in two groups), cause: reason by *because* (larger than 400 in two groups), and positive condition by *if* (also larger than 400) are also outstanding in both groups, followed by relations like cause: purpose by *in order to* (larger than 300 in the Master's group), cause: reason by *since* (larger than 300 in the doctoral group), manner: means by *as*,

(also larger than 300 in the doctoral group) and cause: reason by *so* (more than 300 in the Master's group). Some relations, i.e., cause: reason by *as*, condition: concession by *although*, manner: comparison by *as*, (larger than 200 in both groups in these three kinds) and manner: means by (*and*) *thus* (larger than 200 in the doctoral group), and condition: concessive by *though* (also larger than 200 in the doctoral group) are welcomed by students. (See Table 117)

Enhancement 94							
M		D		M		D	
No.	Lexis	Lexis	No.	No.	Lexis	Lexis	No.
Temporal 19				Cause: Reason (Continue)			
Temporal: Same time 9				22	considering	through	9
1004	when	when	802	14	through	because of	8
86	in	in	168	7	because of	considering	7
72	while	while	122	3	and so	and so	6
20	once	as	30	2	as a result	as a result	3
19	(and) meanwhile	once	19	2	seeing that	seeing that	2
18	as	(and) meanwhile	11			in case of	1
18	in the process of	whenever	10	Cause: Result 3			
4	whenever	in the process of	9	131	so that	so that	164
3	as soon as	as soon as	4	32	to	to	6
Temporal: Earlier 2				3	such that	such that	3
99	before	before	89	Cause: Purpose 8			
16	until	until	29	1097	to	to	1004
Temporal: Later 8				396	in order to	in order to	298
198	and then	and then	125	106	so as to	for	144
186	after	after	83	84	for	so as to	60
41	then	then	21	59	so that	so that	3
31	since	(and) finally	18	7	in order that	in order that	2
22	(and) finally	since	17	2	with the aim of	with the aim of	1
2	afterwards	afterwards	1	1	for the sake of		
1	(and) ultimately	(and) ultimately	1	Condition 31			
1	whereafter			Condition: Positive 9			
Spatial 5				411	if	if	424
Spatial: Same place 5					then...	then...	
32	where	as far as	64	26	as long as	if...then...	51
21	as far as	where	59	25	if...then...	as long as	22
8	thereby	thereby	16	3	(and) in that case	provided that	5
3	whereby	in so far as	6	1	assuming that	in case	2
1	wherever	whereby	3	1	in case	(and) in that case	2
Manner 12				1	provided that	assuming that	1
Manner: Means 5				1	supposing that		
577	by	by	859	Condition: Negative 7			
183	as	as	332	20	without	unless	22
144	(and) thus	(and) thus	226	19	no matter	no matter	15
5	via	thereby	15	13	otherwise	without	13
4	thereby	via	3	6	unless	otherwise	6
		in that way	1	2	but for	if not...then...	2
Manner: Comparison 7				1	or else	or else	1

208	as	as	285	1	if not...then...		
25	(and) thus	(and) thus	27	Condition: Concessive 15			
22	like	like	12	244	although	although	267
9	(and) so	as if	4	128	though	though	217
8	as if	as though	4	56	even though	even though	46
		as much as	3	37	however	however	36
		(and) so	2	20	even if	despite	27
Cause 27				20	on the one hand...on the other hand...	even if	24
Cause: Reason 16				20	without	without	13
401	because	because	476	12	despite	on the contrary	5
322	so	since	315	6	on the contrary	yet	5
245	as	as	288	5	in spite of	in spite of	4
219	since	for	130	3	nevertheless	on the one hand...on the other hand...	2
103	for	so	126	2	while	after all	1
71	therefore	in that	112	1	after all	but	
29	and therefore	therefore	73	1	yet		
28	in that	and therefore	49		but		
26	with	with	41				

Table 117 Main markers for annotating the enhancement relations in Master and PhD sub-corpora

The choice of enhancement is thus to present purposes of research, emphasize the importance of time, describe means for conducting research, provide explanations for research results or methods, and prescribe limits to the statement of facts, though sometimes concessions and comparisons would be made by students. One point to notice is the frequent choice of the preposition *as* because it can perform four functions in connecting clauses: manner: means, cause: reason, manner: comparison, and temporal: same time, thus, the differentiation is more confusing and misleading. For five categories of enhancement, Master's students prefer causal (3,412 in total), temporal (1,841), manner (1,185), and then conditional (1,086) and spatial (65) relations while doctoral candidates tend to choose causal (3,331), manner (1,773), temporal (1,559), and then conditional (1,213) and spatial (148) relations in theses writing. In detail, besides *when*, the word *in* and *while* are also common in functioning as temporal: same time markers within clauses; *and then* and *after* are both common in two groups in connecting two clauses with temporal: later relations; only *before* and *until* are chosen for conveying temporal: earlier relations; Master's students use more *where* than *as far as* in spatial: same place clause nexuses while doctoral students prefer to *as far as* rather than *where* in the same structure; except *as*, *(and) thus* can also be found in denoting a manner: comparison relation in addition to a manner: means relation though the number of which is not so obvious in the former relation; doctoral students choose more cause: reason clause nexuses (1,646 in doctoral group and 1,494 in Master's group) that *because*, *since*, *as*, *for*, *so*, and *in that* are usual markers though the former five are also welcomed by Master's students but with diverse frequencies (*because*, *so*, *as*, *since* and *for* in order), whereas Master's students use more cause: purpose relations (1,752 in Master's group and 1,512 in doctoral group) in enhancement: causal categories, connecting two clauses with *in order to*, *so as to* and *for* in addition to the preposition *to*; for cause: result relation, *so that* would often be discovered in two groups; compared to positive conditions, students have preference for concessions in conditional clause nexuses that doctoral candidates (555 and 647 respectively) use more than Master's students (470 and 507 correspondingly) in these two structures though they (59 in total) have slightly less negative conditions than the latter (62); without considering the application of *then...*, phrases *as long as* and *if...then...* are usual in addition to *if* in connecting positive conditional clause nexuses; and not to mention *but*, markers such as *even though*, *even if*, *however*, and *without* would sometimes be seen except for *although* and *though* in two groups in transmitting concessive meanings within clauses. For peculiarities, *in that way* (Manner: means), *as though* and *as much as* (manner: comparison), and *in case of* (cause: reason) can be found in doctoral dissertations, while *whereafter* (temporal: later), *for the sake of* (cause: purpose), *supposing that*

(condition: positive), *but for* (condition: negative), and *nevertheless* and *while* (condition: concessive) are chosen in some master theses.

7.1.3.4 Annotating Markers in Locution

For markers leading to proposals in locution relation, *suggest*, *say*, *require*, *propose*, *argue* and *advocate* is the top ones with high frequencies in Master theses, and *suggest*, *argue* and *propose* are important in PhD dissertations that *suggest*, *say*, *propose* and *argue* are common in both groups, as well as *require* and *advocate* though the latter two are more frequent in Master's theses. Some markers are unique in Master's theses such as: *ask*, *state*, *advise*, *announce*, *define*, *demand*, *insist*, *joke*, *mention*, *offer*, *persuade*, *remind* and *reproach*, still some can only be found in PhD dissertations: *require*, *stress*, *admit*, *conclude*, *decide* and *deny*. (See Table 118)

Locution 69							
M		D		M		D	
No.	Lexis	Lexis	No.	No.	Lexis	Lexis	No.
Proposal 22				Proposition (Continue)			
43	suggest	suggest	44	22	suggest	propose	28
25	say	argue	19	20	admit	tell	27
24	require	propose	19	15	stress	assert	20
15	propose	say	6	14	propose	admit	16
11	argue	claim	5	12	put forward	ask	14
10	advocate	agree	4	11	emphasize	emphasize	14
4	ask	recommend	4	9	assert	declare	8
4	state	advocate	3	9	discuss	convince	7
3	agree	point out	3	9	report	discuss	7
2	point out	require	2	8	point	insist	7
1	advise	stress	2	8	reply	put	7
1	announce	admit	1	7	insist	contend	6
1	define	conclude	1	7	summarize	complain	5
1	demand	decide	1	6	convey	quote	5
1	insist	deny	1	6	convince	stress	5
1	joke	warn	1	6	declare	recognize	4
1	mention			6	express	summarize	4
1	offer			4	announce	acknowledge	3
1	persuade			4	blame	convey	3
1	remind			4	clarify	deny	3
1	reproach			4	put	express	3
1	warn			3	affirm	interpret	3
Proposition 42				3	answer	maintain	3
				3	maintain	answer	2
167	say	argue	268	2	deny	put forward	2
118	point out	point out	145	2	disagree	reply	2
85	argue	say	113	2	interpret	add	1
69	claim	claim	96	2	shout	affirm	1
62	state	explain	60	1	acknowledge	call	1
56	conclude	state	53	1	add	talk about	1
55	tell	report	44	1	complain		
45	explain	conclude	39	1	recognize		
30	ask	suggest	38	1	remark		

Table 118 Main markers for annotating the locution relations in Master and PhD sub-corpora

For markers in connecting propositions in locution relation, Master's students would often choose a list of words: *say* (167 altogether), *point out* (118), *argue*, *claim*, *state*, *conclude*, *tell*, *explain*, *ask*, *suggest*, *admit*, *stress*, *propose*, *put forward*, *emphasize*, etc., and doctoral candidates have preference for markers such as: *argue* (268), *point out* (145), *say* (113), *claim*, *explain*, *state*, *report*, *conclude*, *suggest*, *propose*, *tell*, *assert*, *admit*, *ask*, and *emphasize*, in which some markers have equal importance in two groups: *say*, *point out*, *argue*, *claim*, *state*, *conclude*, *explain*, *suggest*, *admit*, *emphasize*, *discuss*, etc. Still, some markers are more favored in Master's theses, like *tell*, *ask*, *stress*, and *put forward*, while *propose* and *assert* are more welcomed in PhD dissertations. Words like *report*, *point*, *announce*, *blame*, *clarify*, *disagree*, *shout*, and *remark* can only be found in Master's theses, and *contend*, *quote*, *call* and *talk about* only appear in PhD dissertations. The overuse of *argue* by doctoral candidates is the main reason for a higher proportion of locutions in their dissertations that critical thinking can be presented when referring to scholars' sayings.

7.1.3.5 Annotating Markers in Idea

Idea 66							
M		D		M		D	
No.	Lexis	Lexis	No.	No.	Lexis	Lexis	No.
Proposal 16				Proposition (Continue)			
4	suggest	contend	4	43	find out	understand	31
3	think	show	4	40	hope	explore	28
2	assume	believe	3	32	assume	investigate	28
2	hope	hope	2	31	imply	consider	25
2	propose	propose	2	20	investigate	realize	23
1	believe	assume	1	20	reflect	discover	22
1	decide	concern	1	18	feel	predict	20
1	expect	demonstrate	1	16	determine	reflect	13
1	feel	discover	1	16	explore	hope	12
1	flout	hint	1	10	discover	maintain	12
1	follow	imply	1	10	maintain	feel	11
1	mean	mean	1	8	afraid	find out	11
1	sense	regard	1	8	expect	expect	8
1	turn out	wish	1	8	predict	identify	7
1	understand			7	check	propose	6
1	wish			7	judge	suppose	6
Proposition 44				6	propose	check	5
331	show	show	346	5	doubt	decide	5
242	find	find	265	4	decide	elucidate	4
225	indicate	indicate	253	4	identify	ascertain	3
211	believe	suggest	164	4	wonder	wonder	3
198	think	mean	138	3	summarize	afraid	2
196	mean	believe	116	3	suppose	judge	2
172	know	see	114	2	elucidate	summarize	2
168	see	think	97	2	guess	doubt	1
72	prove	know	72	2	intend	remember	1
70	suggest	assume	57	1	attempt		
56	consider	imply	44	1	deem		
52	realize	demonstrate	42	1	forget		
50	understand	determine	39	1	remember		
48	demonstrate	prove	31				

Table 119 Main markers for annotating the idea relations in Master and PhD sub-corpora

Markers for proposals in idea relation are not usual, though *assume*, *hope*, *propose*, *believe*, *mean*, and *wish* are both chosen by Master's and doctoral students (50 in total in Master group and 41 in doctoral group). (See Table 119) Master's students have their way of connecting idea clauses: *suggest*, *think*, *decide*, *expect*, *feel*, *flout*, *follow*, *sense*, *turn out*, and *understand* are chosen in their theses, so do doctoral candidates: *contend*, *show*, *concern*, *demonstrate*, *discover*, *hint*, *imply*, and *regard* are only selected in doctoral dissertations. The condition for propositions in idea relation is more complex: Master's students have a higher tendency to choose words such as: *show* (331), *find* (242), *indicate* (225), *believe* (211), *think* (198), *mean* (196), *know* (172), *see* (168), *prove*, *suggest*, *consider*, *realize* and *understand*, and doctoral candidates tend to choose *show* (346), *find* (265), *indicate* (253), *suggest* (164), *mean* (138), *believe* (116), *see* (114), *think*, *know* and *assume* though some words are common in both groups: *show* (larger than 300 in both groups), *find* (larger than 200), *indicate* (more than 200), *mean*, *understand*, *demonstrate*, *investigate*, *reflect*, *feel*, *maintain*, etc. Some words are more welcomed in Master's theses: *believe*, *think*, *know*, *see*, *prove*, *suggest*, *realize*, *find out*, *hope*, while some are more favored by doctoral candidates such as *consider*, *assume*, *imply*, *determine*, *explore*, *discover*, etc. Markers like: *guess*, *intend*, *attempt*, *deem*, and *forget* are unique in Master's theses, and *ascertain* is only used in doctoral dissertations. The obvious is that the preference for *believe*, *think*, *mean*, *know*, and *see* by Master's students is the main reason for their higher proportion of hypotactic ideas and idea structures compared to PhD candidates.

In general, *suggest* is the only marker for connecting clauses with not only the locution relation but also the idea that it can lead to clauses as no matter propositions or proposals because of its dual meanings, i.e., meaning either providing suggestions for someone or indicating something to people.

7.1.4 Realization Analysis

Section 7.1.2 has demonstrated that doctoral students prefer hypotactic enhancement, hypotactic extension, hypotactic locution, and paratactic locution. In comparison, Master's students tend to choose paratactic extension, hypotactic idea, paratactic enhancement, and paratactic elaboration relations in their theses. Because clause relations are realized by specific words or phrases, the frequency choice of these hypotactic or paratactic markers is thus connected with corresponding grammatical stratification in comparing two groups to investigate their lexicogrammatical differences and explain known grammatical differentiations. In Master's theses, the reason for the higher number of paratactic extensions lies in their more choice of addition: additive relation marked by *and* (865 more than the doctoral group) and addition: adversative relation realized by *but* (180); the difference of hypotactic ideas comes from their preference for words like: *know* (100 more than doctoral candidates), *think* (99), *believe* (95), *mean* (58), *see* (54), *prove* (41), *find out* (32), *consider* (31), *realize* (29), *hope* (28) and *understand* (19); the higher proportion of paratactic enhancements can be ascribed to the application of temporal: later relation marked by *and then* (73 more than doctoral dissertations) and *then* (20), temporal: same time relation by *(and) meanwhile* (8), and manner: comparison by *(and) so* (7); and the more paratactic elaborations can be attributed to more selection of exposition relation signed by *that is to say* (23 more) and *that is* (17). As to PhD dissertations, the condition of hypotactic enhancement relation is complex that functions such as manner: means indicated by the preposition *by* (282 more) and *as* (149), cause: reason by *since* (96), *in that* (84), *because* (75), and *as* (43), temporal: same time by *in* (82) and *while* (50), condition: concessive by *though* (89), *although* (23) and *despite* (15), manner: comparison by *as* (77), spatial: same place by *as far as* (43) and *where* (27), cause: purpose by *for* (60), cause: result by *so that* (33), and condition: positive by *if...then...* (26) all result to a higher proportion of this structure; more hypotactic extensions originate from the use of addition: adversative function marked by *whereas* and variation: replacive by *rather than* (14 more); hypotactic locution is because of the appearance of more markers like: *argue*, *point out*, *claim*, *report*, *suggest*, and *assert*; and the higher proportion of paratactic locutions is due to their favor of *explain* and *propose*.

7.1.5 Summary of CC System Comparison

In short, differences and similarities can be discovered not only in the systemic distribution of grammatical categories but also in grammatical indexes and annotating markers that realize these clause nexus relations by comparing quantitative data of clause complexes from Master's theses and doctoral dissertations that differences outnumber similarities and are more outstanding. The reason for lexical comparison without taxis resides in noting and calculating deficiencies of SysFan and energy or time-consuming by manual recording. This direction is also meaningful for revealing selective differences within students from diverse educational levels in meaning construing and realization in a particular genre.

7.2 Four Kinds within AW & Error Analysis

This section comes from the process of clause complexes annotation that the clause relation category of extension: addition within thesis written by Master's students and PhD candidates can be confused as a large number of this kind are unrelated or even wrong clause nexuses on the one hand and left relations are inevitably misused by different learners because of their cognition and language abilities on the other. Hence, four kinds of additions are advocated here, and the implicit one is mentioned and introduced in detail in Section 7.3, leaving the other three for analysis in the first part. Then, the location of these four kinds and their divisions are demonstrated in the second part. To expand four layers in addition to the whole system, the location in any category of clause complex is described in Section 7.2.3. The last part focuses on types of wrong clause complexes with descriptions and explanations of cases written by students to present difficulties for annotation and the importance of mastering basic knowledge of clause complex categories in thesis writing.

7.2.1 Four Kinds of Additions within Academic Writing

During the process of corpus annotation, it is found that both Master's students and PhD candidates are not very clear about logico-semantic relations in that they would sometimes produce ambiguous or wrong clause complexes, which brings confusion for understanding and categorization for not only teachers and students but also researchers. Taking the word *and* for example, four kinds of relations can be discovered when it is used to denote extension: positive additive relation in clause connection: unrelated addition, unequal addition, implicit addition, and wrongly addition.

7.2.1.1 Unrelated

When two clauses are paratactically adjoined together without any causal or temporal relationship, and they are not even so related, or they are talking about diverse objects or having different themes, this kind of additional relation is then an unrelated addition. For example:

E.g., 7.2.1 ||| [1:] Foreign language study is a continuous process, || [+21:] **and*** learners should improve the four abilities step by step || [+2+2:] **and*** make use of them every day. ||| ---- CMAC08CU_33_377

E.g., 7.2.2 ||| The utterance can be a warning || when it is uttered by a mother to her young child [[who is afraid of dog || **and*** is walking towards a dog]]. ||| ---- CPhD17_0002_1083

The first *and* in E.g., 7.2.1 seems to paratactically connect two clauses for an additional relationship, the theme of the first clause is *foreign language study*, in fact, while that of the second clause is *learners*, meaning that the clause theme has been changed by choice of *and*. Logically, the second clause should be concerned with either characteristic of *foreign language study* or specification of *continuous process* when *and* is chosen for the positive: additive relation just as the second *and* does within the second and the third clause. This kind of relation, if the conjunctive *and* has to be used for clause connection by the writer, is then an unrelated addition, which can be changed to separate clauses with a full stop in between; otherwise, the relation *and* would be understood as a cause: reason relation that the former clause demonstrates the reason of the content described in the latter one. In E.g., 7.2.2, the clause *who is afraid of dog* and the following clause *who is walking towards a dog* actually express separate meanings, though they share the same subject, *the young child* as the former depicts the status of the child with the tense of the verb being simple present while the latter describes the child's activity and the tense of the verb is present continuous. Normally, the clause relation of these two clauses can be a positive addition or an adversative one if they have to be connected. However, the writer's intention, as the example shown, is quite different in that *and* is chosen to indicate the child's activity happening at the same time with the status *being afraid of dog*, thus, the clause relation turns to be a paratactic enhancement: temporal: same time relation. Two examples above are typical that the clause relation indicated by the conjunctive *and* is an unrelated addition when it is regarded as an extension: addition, though semantically it can be construed as other relations such as causal or temporal.

7.2.1.2 Unequal

In an additional clause complex, one clause can sometimes be embedded into another one semantically by using the conjunctive *and*, that is, the contents of two clauses are of unequal status, or the subject of one clause belongs to the taxonomy of that in another, then the relation in between is unequal addition, which is another marginal clause relation except for unrelated one within the sphere of the additional system. For example:

E.g., 7.2.3 ||| [1:] We can always see the picture of human body in the book of anatomy, || [+2:] **and*** the motor image given by "ribs" is "something resembling a rib in shape or function". ||| ---- CMAC06MO_18_449

E.g., 7.2.4 ||| Mao (Ibid) therefore argues || that the usage of tag question is [[to convey interpersonal meanings]] || **and*** the interpretation of the meanings is indispensable to contextual factors, for instance, speaker's identity or speaker's power. ||| ----CPhD17_0002_1068

The main object described in the former clause in E.g., 7.2.3 is *the picture of human body*, while the theme of the latter clause then turns into *the motor image of ribs*, which is a component of the human body. From the writer's perspective, the phrase *in which* can be more suitable for clause connection than the word *and* as the motor image is for sure included in *the book of anatomy*. But as *and* is chosen by the writer, the clause relation is then categorized into unequal addition. The same goes E.g., 7.2.4 that the clause before *and* discusses the purpose of *tag question*, i.e., *to convey interpersonal meanings*, while the clause after *and* focuses on *the interpretation of meanings*, meaning that the latter clause is embedded in the former semantically when they are adjoined by *and*. The agnation of the clause complex can be: *the usage of ...is to ... meanings, the interpretation of which is ...* that *of which* is a sign for clarification relation. In this condition, the writer's consistency in choosing the conjunctive *and* then provides just an unequal addition relation for two clauses.

7.2.1.3 Wrongly

If two categories discussed above still belong to additional relation, "wrongly addition" is then questionable as it indicates the logico-semantic relation within two clauses that should not be simply addition but is suggested by conjunctive such as *and*, *or*, *but*, etc. when the writer insists on choosing them. For example:

E.g., 7.2.5 ||| [1:] One group majors in primary English education in the three years, || [+2:] **and*** the other groups do not. ||| ---- CMAC08CU_33_499

If the writer in E.g., 7.2.5 intends to present the whole condition of the major in two groups, then the conjunctive *and* seems to be acceptable; but the clause complex is only on the major in separate groups that the contrast is what the writer emphasizes, the choice of *and* thus becomes incompatible and should be changed into the conjunction *but* or the word *while* for an extension: adversative clause relation.

E.g., 7.2.6 ||| Comparatively speaking, English natives' thinking pattern is linear, || **and*** Chinese natives' is circular. || In text organization, linear structure mirrors linear thinking pattern || **and*** circular structure mirrors circular thinking pattern. ||| ---- CPhD10_0015_2181

Again, the conjunctive *and* in the former clause complex E.g., 7.2.6 is wrongly used as the writer presumes to compare the thinking pattern of English natives with Chinese natives, can thus be better replaced by the adversative marker *but*; *and* in the latter clause complex is slightly different because the writer might choose it to demonstrate the mirroring condition of various structures for diverse thinking patterns in text organization in both English and Chinese languages instead of only comparing the linear structure with the circular structure, so *and* for denoting an extension: positive addition relation in the second clause complex is reasonable.

E.g., 7.2.7 ||| He pointed out, || students have to get experience in a real context— || have a continuous activity [[interested in activity itself]]; || producing a real problem as thinking irritants in internal scenarios; possession of the knowledge information || **and*** engage in the necessary observations || to deal with the problem; || must be expanding the methods [[of solving the problems step by step]]; || having a opportunity [[to test ideas by application]], || **and*** have a clear sense. ||| ---- CMAC12ME_5_311

The second type of wrongly used addition is ascribed to inconsistency that the verb before and after the conjunctive *and* belongs to different tenses or voice, such as *having an opportunity* and *have a clear sense* in the last clause nexus in E.g., 7.2.7. If the comma within two clauses is deleted, then the clause *to test ideas...* and the clause *have a clear sense* are connected by *and* for a right positive additive relation and are collectively embedded in the noun *opportunity*, which is in the other way around totally wrong in choosing the tense of two verbs *having* and *have* as the example shows. Another inconsistency is more outrageous that the clause after *and* begins with the verb *engage*, as presented by the third line of the example, while before it, is absent of a verb at all and is replaced by a noun phrase *possession of the knowledge*. The writer's intention of choosing *and* in between is to paratactically and positively combine two structures, but the way of construing them is peculiar.

E.g., 7.2.8 ||| The LA group, when speaking, more or less failed to reduce vowels without stress, || **and that is***, they made little distinction in duration between stressed and unstressed syllables, || with each syllable fully articulated. ||| ---- CMAC11PH_09_350

Normally, the saying *and that is* exists and is usually followed by a relative clause with the beginning type *WH-* for implying a paratactic clarification relation is, however, redundantly used for a paratactic clarification relation as suggested by E.g., 7.2.8 that the phrase *that is* is enough for the

indication of clause relation, the word *and* in fact should be deleted. Then, this kind is not a positive additive relation within two clauses.

7.2.1.4 Implicit

For an implicit addition relation, the clause nexus does not have explicit markers as other relations but is still paratactically organized and categorized into an extension: addition either semantically or structurally, just like what E.g., 7.2.9 has presented:

E.g., 7.2.9 ||| Manny [[who does not think himself to be a kid]] seems to be more mature than his peers, ||* he cares about international issues, || likes to read those books beyond a 10 year old boy's understanding, ||* pays more attention to readings. ||| ---- CMAC15_01_681

The first comma can best be changed into a colon to demonstrate the clarification relation within the first and the second clauses, as the writer is explaining *Manny seems to be more mature than his peers*. The following three clauses are paratactically and additively connected for this clarification; hence, the relation between the third and the fourth clause is a positive additive one, though the conjunctive *and* has not been written.

E.g., 7.2.10 ||| It is unacceptable for Jay, || **and*** he keeps on judging this unsuitable thing with his own language in the satire style. ||| ---- CMAC15_01_571

E.g., 7.2.11 ||| Cameron finds his fat pants [[being stashed in the bush by Mitchell]], || **and*** he is so angry about it. ||| ---- CMAC15_01_741

E.g., 7.2.12 ||| Because of the use of the phrase "turn around" [[which is a pun]], the sub-maxim of manner is violated in this conversation, || **and*** the humorous effects are produced. ||| ---- CMAC15_01_857

E.g., 7.2.13 ||| A text has texture, || **and*** this is [[what distinguishes it from something [[that is not a text]]]]. ||| ---- CPhD05_0003_407

Some cases with *and* for clause connection seem to be additive relations are, however, proved to be other implicit kinds, as shown in E.g., 7.2.10-13 above, though the writer chooses *and* to join two clauses together in E.g., 7.2.10, the relation is not an addition but paratactic enhancement: cause: reason or temporal: later time as the agnation of the clause complex can be *It is unacceptable for Jay, so he keeps on judging ...* or *It is unacceptable for Jay, then he keeps on judging...*, in which the former is more likely. The clause complex in E.g., 7.2.11 is also an implicit paratactic enhancement: cause: reason relation instead of being a positive: additive one as the writer is conveying that *Cameron is so angry as he finds his fat pants ...* In E.g., 7.2.12, it is through the use of the pun phrase that *the humorous effects are produced, and* for connecting two clauses can thus actually be written as *and thus* for meaning a paratactic enhancement: manner: means relation that the contents in the later clause would not be possible unless the means are taken as have been described in the former clause. The word *and* in E.g., 7.2.13 is an implicit sign for paratactic elaboration: clarification relation because the second clause specifies the meaning conveyed in the former clause and turns the whole clause complex to be *A text has texture, to be more specific, this is what ...* Usually, examples like these kinds are not considered as wrong but as other implicit categories because clause complexes appear to be grammatically right, though are logically strange and acquire for readers' comprehension and judgment on clause relations.

7.2.2 The Location of Four Additions

To intuitively grasp the connotation and denotation of the four kinds of additions discussed above, Figure 195 is depicted and provided for comprehending the construction of the "addition" system: being analogous with earth layers, the inner core belongs to explicit addition, the outer core represents implicit addition, the mantle refers to unequal addition, the crust can be unrelated addition, and the troposphere is wrongly addition. The circular or ring area indicates the proportion of each kind, which can, though, be stretched out or drawn back according to the practical application in each text or by diverse users, approximately assumed to be in such appearance. The part included in the dotted line is the inner circle, i.e., explicit and implicit additional relations are common and total additions; the part between the dotted line and the real line is marginal as unequal addition, though belongs to the additional relation, is rare in academic writing. As an unrelated addition can either be regarded as an addition or as other relations based on its context, it is located on the crust unquestionably. Wrong addition is the case that needs modification and thus should not be included in the system of addition, that is, within the red real line as suggested by the figure.

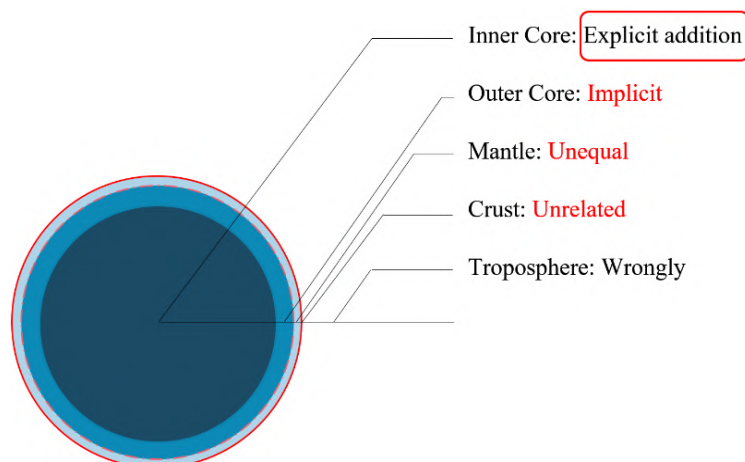


Figure 195 The location of four kinds of additions with their general proportion

7.2.3 The Location of Four Layers in any Category of CC

To extend the location of four layers in the “addition” system to other categories, the inner core can be any logico-semantic category with explicit markers, such as temporal enhancement, conditional enhancement, exemplification in elaboration, locution in projection, etc., as prescribed in Figure 196. Other parts are not changed when compared with the location of four kinds (implicit, unequal, unrelated, and wrongly) in “addition” system that unequal and unrelated layers are still marginal. In any category, the proportion of each layer can also stretch out or drawback due to the characteristics and the essence of the category itself, but as the proportion of implicit ones would be shown in Section 7.3, explicitness (any category) usually takes up the most part in all layers. And though slight changes can happen because of personal choices of these layers, the generalization remains almost the same. Hence, the figure presenting the location with proportions follows that revealed in Figure 195.

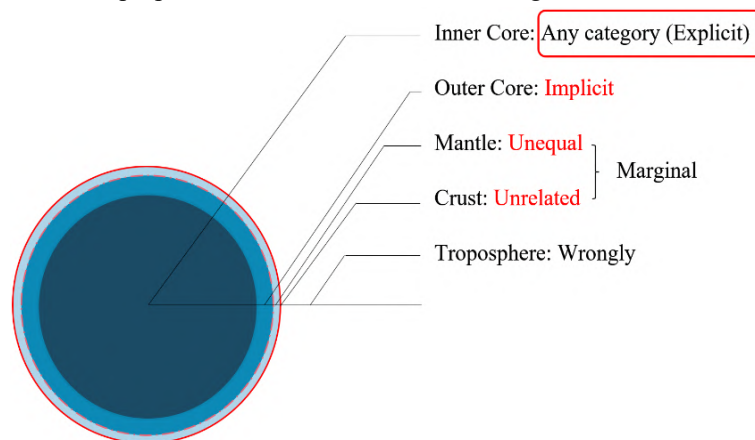


Figure 196 The location of four layers in any logico-semantic category with their general proportion

7.2.4 Annotation for Wrong CCs

In the process of annotating chosen data in thesis writing, most of the clause nexuses are right and complete for identification, while a portion of clause complexes are not correctly organized for some common and typical reasons: adding, missing, substituting, inconsistent, not-splitting, punctuation mistakes, and peculiar usages overall.

7.2.4.1 Adding

“Adding” is easy to understand and find as some words are redundant for clause connection and meaning conveying:

E.g., 7.2.14 ||| The attitude [[which a word assigns]] is not an indexical meaning to the here-and-now scene, || **but instead***, it is indexical knowledge [[that might be invisible in a representation of associative scene, such as trends of thought, bodily changes, expressive episodes, action tendencies, and shifts in attention]]. ||| ---- CPhD17_0001_867

Essentially, the saying *but instead* used in E.g., 7.2.14 does not exist and should be separated into *but* and *instead*, the former of which refers to an extension: addition adversative relation within clauses and the latter indicates the paratactic extension: variation: replacive relation. If the word *instead* is deleted, the clause complex can be *the attitude ... is not an indexical meaning, but an indexical knowledge ...*, in which *it is* in the original clause needs to be omitted. In this condition, the clause connection is variation: replacive relation as marked by the phrase *not... but...* And if *but* is deleted, the original clause complex still works in that *it* after the comma refers to the subject of the first clause, *the attitude*, leaving the clause relation also a variation: replacive one. So, *but* is redundant obviously. E.g., 7.2.15 || Dewey advocated || that teaching method should be paid attention, || that is, ignoring teachers [[how to teach]], || **but rather than*** guiding students [[how to learn]]. || --- CMAC12ME_5_579

The word *but* in the clause *but rather than guiding ...* in E.g., 7.2.15 is also needless as the phrase *rather than* is already enough for indicating a variation: replacive relation. And if the *but* has to be chosen instead of *rather than*, the clause should be written as: *ignoring teachers..., but not guiding students...* that *but not* is acceptable when the clause relation needs to be replacive.

E.g., 7.2.16 || A natural conclusion can be arrived at [[that none of the translated versions has presented the sociosemiotic meanings and functions perfectly, || **though*** in appropriateness does exist in respective versions, || **yet*** the defects do not outweigh the merits]]. || --- CMAC11PR_10_571

The repeated use of words for suggesting hypotactic enhancement: condition: concessive clause relation is not strange as this kind of error can be a language transfer from the Chinese phrase *虽然...但是...* (suiran ... danshi..., *although... but...*) that two words are often combined in writing Chinese sentences for the concessive relation. In English, however, one word is sufficient for indicating a clause relation, just like the word *though* in E.g., 7.2.16 shows. This *though* is a must for joining the first and the third clause semantically due to its location, i.e., at the beginning of the second clause. The word *yet* is then not necessary and needs modification.

7.2.4.2 Missing

“Missing” is the opposite of “adding” as it means the situation when required words for clause connection are, however, absent or forgotten. Usually, mistakes can be the missing conjunctions or conjunctive, prepositions, pronouns, or other elements in a clause nexus:

E.g., 7.2.17 || But Jay puts some additional explanations for Joe’s throwing up, from his understanding, || the reason Joe throws up is [[he needs time to digest this idea]]. || --- CMAC15_01_569

Semantically, the second clause in E.g., 7.2.17 provides explanations for the theme of the first clause; and functionally, their clause relation is a clarification that the latter is more precise and gives more information to support the point of the former. But no marker for clarification relation such as *i.e.*, *to be precise*, *viz.*, etc. is shown up in the clause nexus.

E.g., 7.2.18 || The fourth is image-G accretion strategy, || motivated by supplement of semantic slots of default value, || can be vindicated by the non-correspondence type. || --- CPhD18_0002_1584

Being a non-finite clause, the absence of *which is* for clarification in the second clause in E.g., 7.2.18 is feasible, but the clause relation within the second and the third is a paratactic extension: positive additive one that no marker such as *and* is chosen. The meaning might be changed according to the reader’s comprehension as the gap can be construed as a condition: concessive or an adversative one when a marker is missing. A reason for this phenomenon can be a direct translation from Chinese sentences that, normally, no conjunctive appears in a Chinese clause nexus.

E.g., 7.2.19 || Langacker (2004) considers || that image is the ability of human kind to perceive a certain scene from different perspectives, || forming different images, || comprehending and commanding this scene and various things || by taking different visual angles, || choosing different attention and domains, || giving salience to different aspects. || --- CMAC06MO_18_335

E.g., 7.2.19 seems to be another case for the missing of an extension: positive addition marker that no conjunctive can be found in a list of paratactic clauses, especially between the last and the last second one. In addition, readers might be confused because the last two clauses are non-finite whether the information is additional to the clause *taking different visual angles* or to the clause *comprehending and commanding this scene ...* Another problem comes from the second clause followed by a comma: the relationship within it and following clauses are obscure as it can be an exemplification (following clauses are exemplified by *different perspectives*), an addition (following clauses are additive for *the ability to perceive...*), or even a clarification when other clauses are explained by the whole second clause. In fact, the writer is explaining the clause *forming different images* and then revealing a list of activities that happened temporally. The relation within the second and the third clause is thus a clarification, and that between the last and the last second one is a paratactic enhancement: temporal: later time type that can be signed by a conjunctive such as *and then*.

E.g., 7.2.20 || Teaching skills portfolio is prepared and carried out by the guide teacher as a routine part of teaching and learning, || **and*** it is especially related to [[what has been taught]], || contents is in harmony with the information from the **assessment** is used diagnostically, || it focuses on the individual learner's specific strengths and weakness, needs, etc. || ----CMAC12ME_5_599

In all four clauses in E.g., 7.2.20, the first clause nexus is paratactically additively connected as the marker *and* is apparent, the third clause, however, seems to explain the contents of the second (*contents* are connected to *what has been taught*), and the fourth clause then seems to clarify the object that has been mentioned in the third one (the theme *it* refers to the *assessment*), leaving alone the absence of *that* for defining *the assessment* and into which the clause *is used diagnostically* can be embedded. So, when conjunctive is missing in clause complexes, the clause relation would be really hard to identify, from which ambiguities in readers' comprehension are then produced.

E.g., 7.2.21 || The personification employed by Phil flouts the sub-maxim of quality [[that speaker does not say || what he or she believes to be false]], || and it is Phil's answer [[that makes the audience laugh]], || the humorous effects are achieved by this kind of flouting. || ---- CMAC15_01_539

The meaning conveyed in E.g., 7.2.21 is clear that *the humorous effects are achieved by the kind of flouting* that is described in former clauses, so the clause relation of the last one and former clauses is a paratactic enhancement: manner: means and can be achieved by the addition of the marker *thus* after *the humorous effects are*.

E.g., 7.2.22 || The textbook has gone through several editions since its publication, || short context reading remaining to be a major form. || ---- CPhD14_0002_491

When the writer keeps the non-finite form of the verb, i.e., uses *remaining* in the second clause in E.g., 7.2.22, a preposition such as *with* for denoting an adverbial is then missing, and the clause turns into a phrase in this condition; but when the verb *remaining* is changed to *remains*, the clause nexus can be construed as *The textbook has gone ..., but short context ...* that the clause relation is a paratactic extension: adversative one and the writer's intention for emphasizing the *short context reading* is foregrounded. Which modification is more convincing depends on the context of this clause complex.

E.g., 7.2.23 || The author of the present dissertation has searched all issues (from 1992 to 2016) of the international journal Language and Literature [[sponsored by PALA]], || no articles study Shaw's plays. || ----CPhD17_0002_353

The clause complex in E.g., 7.2.23 is obviously absent of connectors for suggesting the hypotactic enhancement: cause: result relation as it can be understood as *the author ... has searched all issues..., with the result that no articles study Shaw's plays*. Or, a paratactic extension: addition: adversative relation marked by *but* can be acceptable as the result described in the second clause is unexpected. But no matter which word is added to the gap within two clauses, the result that *no articles study Shaw's plays* is what the writer really intends to emphasize.

E.g., 7.2.24 || Students should realize || that some major factors [[that influence them in constructing their desired writer identity]], || which contributes to improving their writings. || ---- CMAC20_04_505

According to the writer from E.g., 7.2.24, the realization of major influencing factors of writing can contribute to its improvement, but the clause as *influence them in constructing ...* is embedded in the subject *major factors*, causing the absence of Rheme in the clause *some major factors ...*. This can be a typical Chinglish in clause combining especially when students have not mastered the component divisions of clauses and clause complexes but have to apply them in academic writing.

E.g., 7.2.25 || Instead of receiving the automatic flow of information || we are asking the speaker/writer to tell us || what we want to know, || which categorizes the sentence that follows, || hence the exposure of the clause relation. || ---- CPhD05_0003_861

Except for the absence again of the Rheme in the last clause *hence the exposure of ...* in E.g., 7.2.25, the connection within the third and the fourth clause is also missing that the relation is a paratactic extension: positive addition one, thus the conjunctive *and* can be inserted.

E.g., 7.2.26 || Johns (1984) even points out || that even "some Chinese teachers of English have never learned how to use lexical cohesion or discourse ties to achieve coherence in texts"; || thus how they can teach students to write coherent texts. || ---- CPhD05_0003_103

If the last clause in E.g., 7.2.26 performs an interrogative function, then it should be written as *how can they teach students ...?* with the question mark in the end; however, the word sequence as shown in the clause suggests an absence of the Rheme undoubtedly that *how they can ...* can only be a Theme and then be embedded into a complete clause such as *is questionable*. Otherwise, the last clause is incomplete and strange that it should not be considered a clause.

7.2.4.3 Substituting

A clause nexus with its connection for clause relation having been replaced by another word (or other words) can also be seen as a wrongly written structure that its modification relates unquestionably to substitution and again selection:

E.g., 7.2.27 ||| At the same time, the no-guaranteed time leads to the teachers and classmates' evaluation [[to be a matter of form, with unable [[to evaluate their learning attitudes and processes]]]]. ||| ----CMAC12ME_5_7

Before discussing, one point is to be made clear that in E.g., 7.2.27, the structure *to be a ... with unable ...* is embedded in the noun *evaluation*, and the same with the phrase *to evaluate their ...* as being embedded in the word *unable*. So, if *unable* remains in the structure, the preposition *with* should either be added with a finite verb *being* after it or be replaced by a conjunctive *while* for indicating an addition: adversative relation as the latter of which is just what the writer intends to reinforce; but if *with* remains unchanged, then the adjective *unable* needs to be substituted by its noun form *inability* for a complete, grammatical compatible expression.

7.2.4.4 Inconsistent

The inconsistent clause complex often refers to the inconsistency of word forms or structures before and after the conjunctive *and* when it is used for an additive clause connection, or just of grammatical elements such as Subject, tense, or mood of verbs, the part of speech, etc. that are applied in the same place in two clauses in a clause nexus. The modification of this kind often involves the latter part instead of the former as it comes after either temporally or sequentially:

E.g., 7.2.28 ||| Generally, Well is regarded as a marker [[of ending || what has been talked about (an old topic) || **and*** switches to a new topic]]. ||| ---- CPhD17_0002_1046

When the clause *switches to a new topic* is paratactically connected to the clause *ending what has been ...* by the extension: positive additive marker *and* in E.g., 7.2.28, and both are characteristics depicted for the word *marker* and then are embedded in it due to the preposition *of* after, they are in the same status and thus should have the same form for their verbs, i.e., *switches* should be changed into a non-finite verb *switching*, in order to be in consistency with the verb *ending* in the former clause.

E.g., 7.2.29 ||| But, [[if one always follows the commands of others || **and*** who only obediently obey the commands || or always indirectly or timidly request others]] will be thought as a wimp || **and*** is inferior or less powerful than others. ||| ----CPhD17_0002_1138

Clauses are strangely mixed in E.g., 7.2.29 that the subject before and after two *and* are both inconsistent: according to the writer, a kind of person *will be thought as a wimp* and will be thought as being *inferior or less powerful than others*, in which *and* is just a connection for positive additive relation, the latter clause loses the information of *being considered by other people*, the fact being described is then distorted; the clause complex before *will be thought as a wimp* has multiple subjects and clause relations that if the hypotactic enhancement: condition relation marked by *if* remains, the whole clause needs to be followed by *... will be thought as a wimp*, and the clause after the first *and* should be revised as *only obediently obeys the commands* for consistency; if the whole constructure is to be embedded in the clause *will be ...*, then *if* should be deleted, and the clause complex turns to be *the person who always follows..., and who only obediently obeys ... or always ... requests others*. So, the positive additive conjunctive *and* seems to be easy to apply, but it is the word mostly wrongly used by students.

E.g., 7.2.30 ||| Kaplan's study has been criticized as ethnocentric and privileging English (Matalene 1985), || **examining** only L2 products (Mohan & Lo 1985), **dismissive** of differences among related languages labeled as "Oriental" (Hinds 1983), || **and implying*** a negative context of L1 transfer (Raimes 1991). ||| ---- CPhD20_0002_548

On account of the clause relation in the last three clauses in E.g., 7.2.30, it is obvious that the finite verb *being* before the clause *dismissive of differences...* is absent because of the clauses' positive additive relation that the verb at the beginning of each clause should be consistent on account of their form. This type of error can also be considered as a missing one, but more importantly, the writer has no awareness of maintaining consistency for clauses that have the same status in parataxis.

E.g., 7.2.31 ||| In humanistic view, human beings should be consider as "whole-persons", || rather than **analyzing*** each element such as behavior, cognition and feelings separately. ||| ---- CMAC07RE_19_281

The subject of the former clause in E.g., 7.2.31 is human beings, while the latter is humanistic researchers that the performers of analyzing are these researchers (though they are human beings biologically): if two clauses are hypotactically connected by the non-finite marker *rather than* for an extension: addition: replacive relation, and have the same subject, the voice, and tense of the verb in them then should usually be consistent, i.e., the clause nexus can be revised as *human beings should be considered as ..., rather than being analyzed of each element ...* or as *they should consider human beings*

as ..., rather than analyzing each element ...

7.2.4.5 Not-splitting

“Not-splitting” means those clause nexuses that should have been separated apart but are not because of writers’ inefficiency of logico-semantic relations within clauses:

E.g., 7.2.32 ||| In the traditional micro-teaching classroom, teaching skills training are mainly based on the explanation and evaluation of teacher, || students train their teaching skills with the guidance of teacher’s instruction and demand, || the atmosphere is always nervous and oppressive, || students’ training is lack of creativity and actively. ||| CMAC12ME_5_551

In E.g., 7.2.32, though clauses can be connected according to reader’s comprehension that the first clause nexus is a positive additive one, the second is paratactic enhancement: cause: reason marked by *so*, and the third again is an additive one, the first and the second clause, in fact, should better be separated into independent clauses as the former focuses on *the explanation and evaluation of teacher* while the latter emphasizes *teachers’ instruction and demand* for teaching skills training. The same goes for the third and the fourth clauses as one is about *the atmosphere* but the other is more on characteristics of students’ training. These four clauses discuss people or staff in the same context, so when writers are incapable of organizing them into a logical sequence with suitable relations in academic writing, mistakes would be easily made as the case does.

7.2.4.6 Punctuation Mistakes

Mistakes of punctuation marks within clauses are common and easy to identify, especially when a comma should have been written as a colon in exposition, clarification, or instantiation relations:

E.g., 7.2.33 ||| It has been mentioned || that the whole tourism translation can be divided into two different processes, || **one is** the communication between the writer of the source text and the translator, || **the other*** is the communication between the translator and the target readers. ||| ---- CMAC09PR_28_567

The second clause in E.g., 7.2.33 clearly demonstrates two processes in tourism translation as suggested by following two clauses with their beginning *one is...*, *the other is...*, the punctuation mark thus should better be a colon, but not a comma.

E.g., 7.2.34 ||| However, some researchers are in disagreement with this theory’s claim, || they don’t deem || that a sense of superiority is the central element in humor. ||| ---- CMAC14PR_4_161

Another case is implicit that the clause relation can only be judged by their meanings, as shown in E.g., 7.2.34: the second clause *they don’t deem that...* exposes the disagreement with the mentioned theory and supports the claim that the former clause has made. A marker for connecting two clauses can be *i.e.*, *to be precise*, *that is*, etc., for the clarification relation, or just a colon.

E.g., 7.2.35 ||| There is a classic instance || as we mentioned in the metaphorical chapter, || based on the experience of the people’s bodily movements, || the body itself can be thought as a container, || and then the CONTAINER schema comes into being. ---- CMAC11SE_19_997

From the third clause in E.g., 7.2.35, the writer begins to provide an example for the statement in former clauses, i.e., the latter three clauses are all about *a classic instance in the metaphorical chapter*, thus the first comma needs to be changed into a colon for a paratactic elaboration: exemplification clause relation within.

7.2.4.7 Wrong Usage

Except for the six kinds of mistakes discussed above, there are some strange usages of conjunctive or prepositions in clause connection in the annotated corpus:

E.g., 7.2.36 ||| Instead of receiving the automatic flow of information || we are asking the speaker/writer to tell us || what we want to know, || which categorizes the sentence that follows, **hence*** the exposure of the clause relation. ||| ---- CPhD05_0003_861

The word *hence* at the beginning of the last clause in E.g., 7.2.36 apparently marks a paratactic enhancement: cause: reason relation that former clauses explain *the exposure of clause relation*. But as the word *exposure* is a noun, the finite verb (*V+ing*, any verb semantically logical) that disappears in this clause should be supplemented. Or, *exposure* can be replaced by its verb form, i.e., *exposes* for the integrality of the clause.

E.g., 7.2.37 ||| Despite all the contributions and the following-up studies inspired by Labovian model, the lack in the discussion of both form and meaning of the narrative structure compels us to explore the ways formal and conceptual structural forms as mechanisms **through which** [[**to*** narrate and to represent conventional constructions of a culture]]. ||| ---- CPhD20_0002_422

When the structure *through which* is used, the following words should at least be a complete clause to present the status after the means is adopted, the clause relation within two clauses is a paratactic enhancement: manner: means one as indicated by the preposition *through*, and the pronoun

which denotes to methods that are described in the former clause (or clauses). In E.g., 7.2.37, the structure after *through which* should thus either be changed to phrases like *to narrate and to represent...* being embedded into a clause, or be added with a subject and written as: *We can narrate and represent ...* for the completion of a clause.

E.g., 7.2.38 ||| Whereas **with** implicit topics or implied themes in a Chinese discourse, **no specified goals to attain**, || **with*** wrong discourse markers, the addressee/reader would have less contextual effect to help him/her to process the meaning, || thus the degree of relevance is low || and will call for the greatest processing effort. ---- CPhD05_0003_1737

Another case of regarding a noun phrase as a clause in writing comes from E.g., 7.2.38 that *no specified goals to attain* is incomplete as the infinitive verb *to attain* is embedded in the noun *goals*, should better be modified as *no specified goals can be attained* for complete meaning conveying at one hand and keep consistent with structures of the latter clause on the other.

7.2.5 Summary of Four Additions and Error Analysis

This section introduces four kinds of additions within academic writing that are categorized from the annotation of clause complexes: unrelated, unequal, wrongly, and implicit, explains each of them with examples from the ChAcE corpus, illustrates the location of these four kinds with the analogy of earth layers, expands them to any category within the clause complex system to prove the feasibility of the location of four layers, and at last discusses types of wrong clause complexes written by Chinese Master's and PhD students as presented by the corpus: adding, missing, substituting, inconsistent, non-splitting, punctuation mistakes, and wrong usages. Except for laying the foundation of implicit relations for further analysis, this section also presents students' understanding and mastery of logico-semantic relations. The level of education might not be the primary factor for language inability, but the transfer of mother tongue and personal preference for some specific structures.

7.3 Implicit & Explicit Clause Complex

This section will be published in the following journal article:

The Implicit Clause Complexes in Thesis Writing: from the *Chinese Academic English* (ChAcE) Corpus

7.4 Common Recursion Patterns and Clause Order

7.4.1 Common Patterns of Clause Complex (Nesting)

This section will be published in the following journal article:

Recursion Patterns of the Clause Complex in Thesis Writing: From the *Chinese Academic English* (ChAcE) Corpus

7.4.2 Clause Order in a Clause Nexus

Another dimension for the discussion of a clause complex concerns the clause order in a clause nexus, especially with explicit markers for clause relations, whether the dependent clause is before or after the independent one represents the emphasis of writers' meaning construction and then is helpful for readers' comprehension and absorption.

Relations	Categories	Linkage	Logico-semantic Markers	Order
Elaboration	Exposition	Parataxis	in other words, in another word, that is, i.e., namely, that is to say, to be specific, or rather, put it another way, simply put, to some extent	$p^{\wedge}q$
		Hypotaxis	/	/
	Clarification	Parataxis	in fact, in reality, in practice, actually, indeed, at least, at lowest, at highest	$p^{\wedge}q$
		Hypotaxis	which, who, whose, where, when	$p^{\wedge}q$
	Exemplification	Parataxis	for example, e.g., for instance, such as, like, in particular, take...as examples	$p^{\wedge}q$
		Hypotaxis	/	/
Extension	Addition	Parataxis	and, both...and..., not only...but also..., moreover, on the one hand...on the other hand..., nor, neither...nor..., but, conversely	$p^{\wedge}q$
		Hypotaxis	while, whereas, besides, apart from, as well as, without	$p^{\wedge}q/q^{\wedge}p$
	Variation	Parataxis	but not, not...but, only, but, except, instead, on the contrary	$p^{\wedge}q$
		Hypotaxis	instead of, rather than, except that, except for, other than, but (for the fact) that	$p^{\wedge}q$
	Alternation	Parataxis	or, alternatively, either...or., or else, elsewhere	$p^{\wedge}q$
		Hypotaxis	if...not (...then)	$p^{\wedge}q$
Enhancement	Temporal	Parataxis	meanwhile, while, when, at the same time, later, subsequently, and, and then, then, (and) afterwards, whereafter, and finally, and ultimately, next, later on, earlier, previously, (and/but) before that, (and/but) first	$p^{\wedge}q$
		Hypotaxis	as, while, in, in the course of, in the process of, when, as soon as, the moment, on, whenever, every time, after, since, once, before, till, until	$p^{\wedge}q/q^{\wedge}p$
	Spatial	Parataxis	and there, at the same place	$p^{\wedge}q$
		Hypotaxis	as far as, in so far as, where, wherever, whereby, thereby, everywhere	$p^{\wedge}q/q^{\wedge}p$
	Manner	Parataxis	and in that way, (and) thus, (and) similarly, (and) so, comparatively, likely	$p^{\wedge}q$
		Hypotaxis	by, thereby, by means of, via, as, as if, like, the way, just as much as, as though, through	$p^{\wedge}q/q^{\wedge}p$
	Cause: reason	Parataxis	(and)so, (and) therefore, for, thus, hence, as a result, since	$p^{\wedge}q$
		Hypotaxis	because, as, since, in case, seeing that, considering, with, through, by, at, as a result, because of, in case of, for, in that, in as much as, thanks to, due to	$p^{\wedge}q/q^{\wedge}p$
	Cause: result	Parataxis	/	/
		Hypotaxis	so that, such that, to	$p^{\wedge}q$
	Cause: purpose	Parataxis	/	/
		Hypotaxis	in order that, so that, in order to, to, so as to, for (the sake of), with the aim of, for fear of	$p^{\wedge}q/q^{\wedge}p$
	Conditional	Parataxis	if...then..., (and) then, and in that case, or else, or, otherwise	$p^{\wedge}q$
		Hypotaxis	if, provided that, as long as, in the event of, unless, but for, without, in case, supposing/assuming that, no matter	$p^{\wedge}q/q^{\wedge}p$
	Concessive	Parataxis	but, (and) yet, still, but nevertheless, however, on the contrary, after all,	$p^{\wedge}q$

		Hypotaxis	even if, even though, although, while, though, despite, in spite of, without	p^q/q^p
Projection	Locution	Parataxis	Verbal Verbs	p^q/q^p
		Hypotaxis	Verbs , that, whether, if, when, where, who, why, how	p^q/q^p
	Idea	Parataxis	Mental Verbs	p^q/q^p
		Hypotaxis	Verbs , that, whether, if, when, where, who, why, how	p^q/q^p

Table 120 The clause order in a clause nexus with markers in the clause complex categories

Table 120 generally presents the clause order in all expansion clause nexus relations with particular logico-semantic markers, in which all paratactic categories are instantiated with *an initial clause + the secondary clause*, are thus recorded as p^q , except for enhancement: cause: result/purpose relations that no markers can be found grammatically, meaning that paratactic causal clause nexuses for result or purpose function are peculiar in the English language. Conditions for hypotactic expansion categories turn out to be more flexible as excluding investigations of hypotactic exposition or exemplification because of inexistence, hypotactic elaboration: clarification, extension: variation/alternation, and enhancement: cause: result relations are all instantiated with *independent clause + dependent clause* no matter which marker is chosen for connection, hence they are recorded as p^q . If independent clauses do not exist, then dependent clauses also won't appear in a clause nexus or a clause complex, and this kind of sequence within clauses is irreversible. However, other categories such as hypotactic extension: addition, enhancement: temporal/spatial/manner, enhancement: cause: reason/purpose, enhancement: conditional/concessive can all have reversible clauses in clause nexuses depending on chosen markers, and this is the type that can clearly demonstrate writer's intention when a dependent clause is located before an independent one in clause nexus writing. Though for some markers like: *besides* and *apart from* in extension: addition relation, *since* and *until* in enhancement: temporal clause nexus, *as far as* in enhancement: spatial, *because*, *since*, *considering*, *thanks to* and *due to* in cause: reason, *in order to*, *for* in cause: purpose, *as long as*, *unless* and *supposing/assuming that* in enhancement: conditional, and *although* and *despite* in enhancement: concessive relation, a dependent clause usually follows one of them and is located before an independent one, so the marked or worthwhile clause nexus now changes into *an independent clause + a dependent one* in writing. For projections, no matter an idea or a locution from the perspective of projection level, the projected clause in a clause nexus always goes after the projection verb as well as the Theme, meaning that the marked clause nexus here should be *a projected clause + projection verb + the Subject or a projecting clause*, which is unique, rare but worthwhile of analyzing for meaning construal.

When a clause nexus does not have any explicit marker for its clause relation, the agnation or equivalents on meaning or logic of this implicit structure would first be considered; then, the clause order turns out to be consistent with the relatively explicit one described above.

7.5 The Complexity of the CC or the Clause Complexical Complexity

Before putting forward the concept of "clause complexical complexity," the parallel term "syntactic complexity" will be discussed with its definition, brief literature review, measurement indexes, problems, and deficiencies to introduce the intention of combining SFL with complexity calculation; then the theoretical background will be provided from three aspects: 1. syntactic complexity in L2 writing; 2. Views on "syntactic complexity" from a functional-typological perspective; and 3. the probabilistic theory of grammar. The definition of the term will start from the definition of "complexity," then to the term that we focus on in this section, followed by the difference between it and "syntactic complexity." In Section 7.5.4, three main formulas will be described for calculating "clause complexical complexity," and though Formula 2 and Formula 3 are better at mixing grammatical categories with levels, Formula 1 will be applied in comparing chosen texts from our corpus due to missing data of detailed grammatical category choices on each level. Since the concept is only an attempt to provide a perspective for further research of the complexity of clauses or sentences, limitation is inevitable.

7.5.1 Origins

After presenting the calculating statistics of grammatical and lexical categories of the clause complex; describing the multi-function characteristic of some specific markers that they perform in connecting clauses; differentiating implicit clause complexes from explicit ones to explain the problems of automatic annotation by machines and qualitatively present the complexity influenced by external

factors such as language transfer of mother tongue, writing behaviors of students, context, etc., and also an internal factor from the language itself; and listing patterns of clause complex recursion with their corresponding counting numbers from a small corpus, the question of how to compare the complexity of these students' writing comes into notice and deserves further researches. Now, can the claim be that the language, especially on the clause complex level, applied in theses written by doctoral candidates as L2 language learners from China is more complex than that by Master's students of almost the same educational background? How can the mentioned "complexity" be measured within a limited scope? In fact, the key point resides in clause relations. As clause is the kernel grammatical category in SFL, the clause relation marked by specific lexis indicates the choice of writers in performing language functions and realizing their meanings in writing, clause complex is hereby identified as a feasible and useful category for measuring its complexity.

In Section 1.1.4, the parallel term "syntactic complexity" has been mentioned and discussed. As it closely relates to writing proficiency, does it mean that more complex syntax writers have acquired and applied in their writing suggests their higher writing proficiency? How can the proficiency of high or low levels be defined and quantified? On the other way, does it mean that writers of higher proficiency, of course, can but would use more complex sentences in conveying their meanings and transmitting knowledge to readers? One interesting and well-known phenomenon is on nominalization that writers from higher educational levels seem to more frequently replace their complex grammatical structures with nominalized expressions not only because of disciplinary standardized terminology requirements but also their conceptualized thinking patterns. Thus, the measurement of "syntactic complexity" provides researchers with the generalization of language utilization from the structural perspective but not with the capability of language application and realization from the functional view. For the indirect relationship between syntax and writing quality, Beers & Nagy (2007, p. 187) once considered that "the ability to produce complex sentences is probably understood as a necessary but not sufficient condition for writing high quality texts," and sentence quality with its complexity is also "fundamentally independent." Another aspect is on recursion that this kind of complexity on the clause complex contains information of clauses on each recursive level, presents a limitation of choosing clauses due to depth and grammatical specialty of language, and reinforces the repetition and variation of clause nexuses within a sentence, which, however, cannot be realized by syntactic complexity analysis.

Over the past half-century, researchers have used various grammatical indexes to quantify syntactic complexity, attempting to figure out its definite relationships with external factors such as writing proficiencies, grades, ages, etc. Research data mainly comes from written texts (Faigley, 1979; Kameen, 1979; Stewart & Grobe, 1979; Crowhurst, 1980; Flahive & Snow, 1980; Sharma, 1980; Homburg, 1984; Bardovi-Harlig & Bofman, 1989; Kawata, 1992; Casanave, 1994; Ortega, 2003; Ravid, 2005; Beers & Nagy, 2007), and these English written research concentrate on the development through increasing ages, grades, curriculum or educational levels, etc., on the one hand, and the relation within learners' complexity and writing proficiency (quality) or complexity differentiates within groups of various compositional levels on the other. Research by Chinese scholars focus on horizontal analysis of the relationship between syntactic complexity and task type or difficulty levels or between it and the writing quality (Chen & Wu, 1998; Tan & Dong, 2007) and vertical analysis of the development (Qin & Wen, 2007; Ji, X., 2009; Bao, G., 2009) also.

Syntactic structures for calculating and analyzing syntactic complexity usually include Word count (W), Sentence (S), Verb phrase (VP), Clause (C), T-unit (T), Dependent clause (DC), Complex T-unit (CT), Coordinate phrase (CP), and Complex nominal (CN); and syntactic complexity indices encompass categories like the length of production units (mean length of sentence or MLS, mean length of T-unit or MLT, mean length of clause or MLC), amounts of coordination (clause per sentence or C/S, verb phrase per T-unit or V/T, clause per T-unit or C/T), amounts of subordination (dependent clause per clause or DC/C, dependent clause per T-unit or DC/T, T-unit per sentence or T/S), overall sentence complexity, i.e., complex T-unit ratio (CT/T), and phrasal sophistication (coordinate phrase per T-unit or CP/T, coordinate phrase per clause or CP/C, complex nominal per T-unit or CN/T, complex nominal per clause or CN/C). (L2 Syntactical Complexity Analyzer by Xiaofei Lu³⁴)

But in taking these indexes for text analysis, no matter whether manually or automatically, problems or peculiar conclusions can be encountered (Lu, 2010, p. 288): the primary is parsing or identifying errors for the analyzing software, most of which are from attachment level or conjunction scope, greatly influencing research outcome and even outnumbering learner errors (such as determiners or agreement errors) made by students; another is that advanced learners should have more idiomaticity (like collocation) problems than grammatical completeness in writing at the sentence level! This comes from

³⁴ <https://aihaiyang.com/software/lca/>

the confusion of investigating phrasal and grammatical structures at the same time, the latter of which in fact should have a lower frequency than the former in a sentence. And according to Lu (*ibid.*, p. 491), longer clauses and T-units are produced by Chinese students of English due to the increasing application of complex phrasal structures such as coordinate phrases and complex nominals rather than the use of dependent clauses or complex T-units when syntactic development research is conducted. As piles of phrases can be easier mastered by students and then be chosen in sentence writing than grammatical structures, this outcome cannot sufficiently verify language development no matter how broad or large the chosen corpus is.

Thus, the research on syntactic complexity seems to be induced by lists of quantitative, comparative, and rationalistic studies when structural indexes are excessively counted and calculated in text analysis and then are randomly compared, clause relations are blindly investigated, leaving theoretical discussions for laying the foundation of these practices, obviously ignored. As syntactic complexity theory is still in its initial stage, issues to be solved and disputes to be clarified are obvious. (See Section 1.1.4 for deficiencies) Undue or deflected application of the theory can cause controversial or even contrary conclusions when diverse methods are taken, measures are conducted, research objects are chosen, and experiments are designed. One thing to note here is that the proposal of clause complex complexity, though manually consuming, does not deny or overturn efforts that scholars of syntactic complexity studies have put into, nor will the syntactic complexity theory be questioned and challenged. The purpose is to provide a perspective, if it can be ascribed to traditional functional attempts, for further comparison of the complexity of complex clauses.

7.5.2 Theoretical Background for the Complexity of the CC

The theoretical foundation of putting forward the concept of “clause complex complexity” and proposing formula in quantitatively calculating indexes, evaluating and comparing academic theses written by students comes from three directions: 1. Syntactic complexity in L2 writing (Foster & Skehan, 1996; Homburg, 1984; Casanave, 1994; Ishikawa, 1995; Hyland, 2003) and its measurement, such as average words of T-unit (Cooper 1981); passive voice and dependent clauses (Kameen, 1979; Homburg, 1984); average clauses of T-unit coordination index (Bardovi-Harlig & Bofman, 1989; Bardovi-Harlig, 1992); description of syntax (Coombs, 1986), etc.; 2. Views on “syntactic complexity” from a functional-typological perspective (Miestamo et al., 2008; Sampson et al., 2009) and from the property of syntactic recursion (van der Hulst, 2010; Sauerland & Trotzke, 2011; Roeper & Speas, 2014), especially claims on the function that “the need to examine linguistic forms not only in terms of their frequency and distribution, but also in terms of the functions which they perform in the context of specific types of extended discourse” (Berman & Nir-Sagiv, 2004, p. 375) should be noticed, and Biber, with his colleagues (e.g., Biber, Gray, & Poonpon, 2011; Biber et al., 2014), advocated that L2 writing researchers should consider about phrasal and clausal complexities or complexification in sufficient functional details and “differentiate among the structural types and syntactic functions of dependent clauses and phrases” (Biber et al., 2014, p. 26) as the method of only counting on the number of quite a few main structural elements appeared in the context are inadequate to represent the construct of syntactic complexity (Ortega, 2015, p. 91); 3. the probabilistic theory of grammar (Halliday, 1991, 1992a, 1993; Halliday & James, 1993, collected in Halliday, 2005b; Matthiessen, 2007a; Manning, 2003, etc.): though this notion can be dated back to B. Buljakovski in 1847 as he suggested that “the probabilistic method could be used for the comparative study of grammar” (Feng, Z., 2023, p. 49), it was Halliday who confirms the feasibility of analyzing grammatical categories quantitatively in SFG field. In addition, SFL intensely connects the study of syntactic complexity and the sphere of educational practice that it successfully supported the L2 writing development of foreign language students (e.g., Byrnes, Maxim & Norris, 2010) and minority students in school (e.g., O’Dowd, 2012) (Ortega, 2015, p. 91), the reversed background from syntactic complexity on SFL studies is thus also reasonable and powerful.

7.5.3 Definition of Clause Complex Complexity

Before defining the term “the complexity of clause complex” or the “clause complex complexity,” the keyword “complexity” has to be first made clear as it is a new word compared to the term “clause complex,” which has been frequently mentioned and regarded as the essential part of this research, and how to define “complexity” directly originates from the essence that language be treated dynamically or statically as the former comes down to its development in whatever form; determines the scope or aspects of clause complex that can be investigated; and influences measuring algorithm of “clause complex complexity” as will be illustrated in later sections.

Though a profusion of scholars (e.g., Ellis, 2003; Housen & Kuiken, 2009; Hunt, 1964, 1970b;

Loban, 1976; Norris & Ortega, 2009; Ortega, 2003; Bulté & Housen, 2012) on first and second language researches have focused on “complexity,” it is still an ambiguous term for language development (Housen & Kuiken, 2009; Norris & Ortega, 2009; Vyatkina et al., 2015) and entails kinds of connotations and operationalizing methods. In applied linguistics, “complexity” has long been referred to as “the more advanced grammatical structures that students exhibit as they progress in their language proficiencies,” especially “grammatical complexity” in relation to writing development (Biber et al., 2011, p. 6). As language development corresponds to the increase of syntactic maturity (Ortega, 2003, 2015), higher maturity signs more proficient writing (McNamara et al., 2010), or in Foster & Skehan’s (1996, p. 304) words, “complexity connects with change and the opportunities for development and growth in the interlanguage system,” complexity turns to be significant and is usually understood as “the sophistication, variety, diversity, or elaboratedness of grammatical resources exhibited in language production” (Ortega, 2015, p. 86). From the ontological perspective, increased complexity is increased hierarchic levels within an organized system, i.e., a network of nodes (the simplest entities or more abstract, higher level sub-parts of the system) and their relations (or connections) as simple or individual entities, though have no internal organization, can have relations to the system as a whole, to its sub-parts, or even to each other. (Givón, 2009, p. 3-4)

In this section, “complexity” includes not only the syntagmatic or hierarchic part mentioned by H. Simon (1962) but also the paradigmatic dimensions within structures, concerns not only about the sophistication, variety, and diversity of language choices but also about its measuring methods for quantification and comparison. Hence, “clause complex complexity” indicates the “complexity” of “clause complex” that the number, type, depth, and interweaving sophistication of sub-structures chosen in a text would be emphasized no matter which genre the text is or by whom the text is produced. The difference of “clause complex complexity” and “syntactic complexity” can be the research object that the former focuses only on sub-categories of the clause complex with embedding clauses being the marginal but essential structure while the latter has a broader sphere and discusses not only syntactic organizations but also phrasal and lexical choices in the syntax, with embedding structures being the core and fundamental elements in most studies; can be the research purposes that the former investigates complexities of texts regardless of producers, producing environments or contexts, style varieties or registers, etc., though these factors do influence the complexity, while the latter primarily probes into language development of first or second language learners through syntactic complexity indexes; can be the parsing methods that clause complexes can only be annotated manually on temporary as the understanding of logico-semantic relations within clauses have not been trained perfectly as to computers, thus the parsing process is laborious and brings difficulties for verification and validity analysis, but annotation for syntactic complexity structures is time and labor saving when most of these works can be realized by software.

This kind of downgrade analysis, i.e., from clause relations to markers denoting the relation or to significant verbs projecting clauses in clause complex studies, provides a possibility for complexity exploration. However, it leaves room for relation inquiry within lexis and clauses, i.e., ignoring the consideration of embedded clauses as an SFL category.

7.5.4 Three Formulas with Comparison

Now that the importance of clause recursion has been demonstrated in the former section, the level at which each clause resides in organizing a clause complex thus needs consideration in the exploration of complexity measurement. And the number of each grammatical category on the corresponding level is another significant factor for calculation. So, the core practice of measuring the clause complex complexity is comprised of the ratio of category number on each level and the number of whole clause complexes found in the text on the one hand and the weighted level on the other hand.

$$\Delta = \sum_{i=1} \frac{\text{No. of clause nexuses on level } i}{\text{No. of clause complexes}} \times i$$

Figure 197 Formula 1 for measuring the clause complex complexity

As indicated by Formula 1 (Figure 197), a multiple sign is chosen for the ratio of the number of clause nexus on each level per the number of whole clause complexes in a text and relevant level (*i*), in order to increase the weight of the level in functioning recursion, and then addition is realized by the summation sign from level 1 to even level 7 provided that a clause nexus can be found on level 7 in a text. Thus, the absolute numerical value after calculation is, what we call, the clause complex complexity value for comparison within texts. Take CMAC10SE_18 from the corpus, for example, the complexity value, as indicated by numbers in Table 121, is:

$$\Delta = 216/216 \times 1 + 136/216 \times 2 + 57/216 \times 3 + 14/216 \times 4 + 6/216 \times 5 + 1/216 \times 6 + 1/216 \times 7 = 3.51$$

No. of CC	CNs (1-7)	On Level 1	On Level 2	On level 3	On Level 4	On Level 5	On Level 6	On Level 7
216	431	216	136	57	14	6	1	1

Table 121 The number of clause complexes and clause nexuses on each level (CMAC10SE_18)

One thing to note is that the denominator of the formula signifies not the number of clauses but of clause complexes, and the latter, in fact, equals the number of clause nexuses on level 1 because one clause complex including at least two clauses, i.e., a clause nexus combined by them, and this clause nexus must be on level 1 when only two clauses exist.

But clause nexuses on each level can signify either location or idea relations in projection clause complexes and can also be classified into elaboration, extension, or enhancement relations when two clauses are expanded. Formula 1 presumes that the choice of grammatical categories (location, idea, expansion, extension, enhancement) by writers or speakers in organizing clauses in a clause nexus has an equal probability, which, however, is proved to be not. The third core factor that must be considered in measuring complexity thus belongs to the proportion of each category in all clause nexuses. Hence, Formula 1 is broadened into Formula 2 (Figure 198) for better investigating the diversities of choosing each grammatical category within texts:

$$\Delta = \sum_{i=1} \frac{\text{No. of A on level } i}{\text{No. of clause complexes}} \times \frac{\text{No. of A}}{\text{No. of clause nexuses}} \times i + \sum_{i=1} \frac{\text{No. of B on level } i}{\text{No. of clause complexes}} \times \frac{\text{No. of B}}{\text{No. of clause nexuses}} \times i +$$

$$\sum_{i=1} \frac{\text{No. of X on level } i}{\text{No. of clause complexes}} \times \frac{\text{No. of X}}{\text{No. of clause nexuses}} \times i + \sum_{i=1} \frac{\text{No. of Y on level } i}{\text{No. of clause complexes}} \times \frac{\text{No. of Y}}{\text{No. of clause nexuses}} \times i +$$

$$\sum_{i=1} \frac{\text{No. of Z on level } i}{\text{No. of clause complexes}} \times \frac{\text{No. of Z}}{\text{No. of clause nexuses}} \times i$$

Note: A is Locution, B is Idea, X is Elaboration, Y is Extension, and Z is Enhancement

Figure 198 Formula 2 for measuring the clause complex complexity

Except for the division of projection and expansion relations, a clause nexus is also influenced by the status of two clauses, as either paratactic or hypotactic relations can be preferred by writers in different contexts, genres, or meaning expressions. The probability of choosing hypotaxis or parataxis has never been the same, as proved by our description of the corpus before. Therefore, grammatical categories in Formula 2 are further divided into hypotactic or paratactic ones for more delicate calculation, as suggested by Formula 3 below:

$$\Delta = \sum_{i=1} \frac{\text{No. of paraA on level } i}{\text{No. of clause complexes}} \times \frac{\text{No. of paraA}}{\text{No. of clause nexuses}} \times i + \sum_{i=1} \frac{\text{No. of hypoA on level } i}{\text{No. of clause complexes}} \times \frac{\text{No. of hypoA}}{\text{No. of clause nexuses}} \times i +$$

$$\sum_{i=1} \frac{\text{No. of paraB on level } i}{\text{No. of clause complexes}} \times \frac{\text{No. of paraB}}{\text{No. of clause nexuses}} \times i + \sum_{i=1} \frac{\text{No. of hypoB on level } i}{\text{No. of clause complexes}} \times \frac{\text{No. of hypoB}}{\text{No. of clause nexuses}} \times i +$$

$$\sum_{i=1} \frac{\text{No. of paraX on level } i}{\text{No. of clause complexes}} \times \frac{\text{No. of paraX}}{\text{No. of clause nexuses}} \times i + \sum_{i=1} \frac{\text{No. of hypoX on level } i}{\text{No. of clause complexes}} \times \frac{\text{No. of hypoX}}{\text{No. of clause nexuses}} \times i +$$

$$\sum_{i=1} \frac{\text{No. of paraY on level } i}{\text{No. of clause complexes}} \times \frac{\text{No. of paraY}}{\text{No. of clause nexuses}} \times i + \sum_{i=1} \frac{\text{No. of hypoY on level } i}{\text{No. of clause complexes}} \times \frac{\text{No. of hypoY}}{\text{No. of clause nexuses}} \times i +$$

$$\sum_{i=1} \frac{\text{No. of paraZ on level } i}{\text{No. of clause complexes}} \times \frac{\text{No. of paraZ}}{\text{No. of clause nexuses}} \times i + \sum_{i=1} \frac{\text{No. of hypoZ on level } i}{\text{No. of clause complexes}} \times \frac{\text{No. of hypoZ}}{\text{No. of clause nexuses}} \times i$$

Note: A is Locution, B is Idea, X is Elaboration, Y is Extension, and Z is Enhancement

Figure 199 Formula 3 for measuring the clause complex complexity

These two formulas have a disadvantageous quality compared to Formula 1. In instances where a more exacting standard for manual annotation is necessary, it is essential to provide a clear indication of the grammatical category, clause status, and level of each clause nexus. While software such as SysFan can be useful for marking and noting, it has not been designed for the specific purpose of counting and calculating relations at each level.

A comprehensive analysis of the texts from the corpus that elucidates the practice of these three formulas is beyond the scope of this study due to the constraints of space.

7.5.5 Formula Application

Though Formula 3 presents the complexity from both grammatical categories and levels perspectives, the missing data of each category on levels 1 to 7 results in our choice of Formula 1 here in comparing Master's theses and PhD dissertations from the corpus. Table 122 and Table 123 present the value of clause complex complexity calculated by Formula 1 in the chosen corpus (80 MA theses and 32 PhD dissertations), respectively, with the average value of 2.32 in the Master's group and 2.20 in the PhD group. Judging merely from the value, it seems that Master's students apply more complex clause complexes than PhD candidates in writing their theses to convey ideas, make statements, review literature, and describe their methods and data.

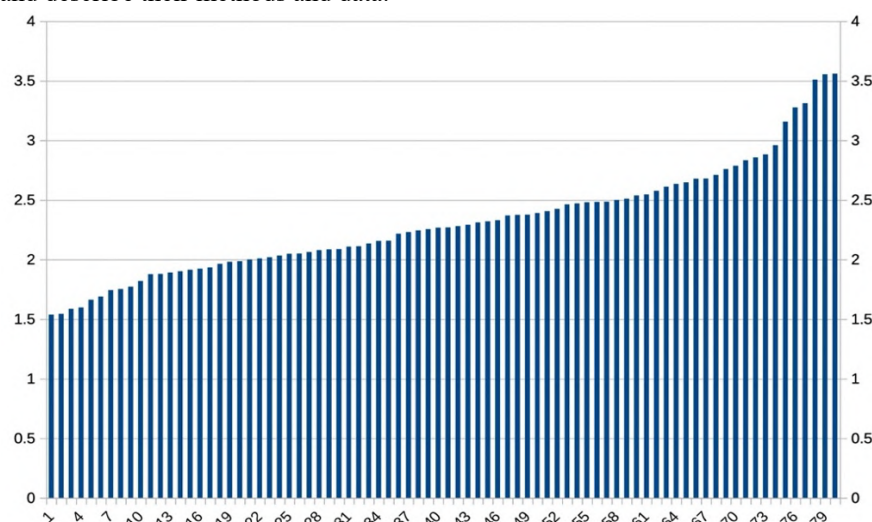


Table 122 The clause complex complexity value by Formula 1 (Master sub-corpus)

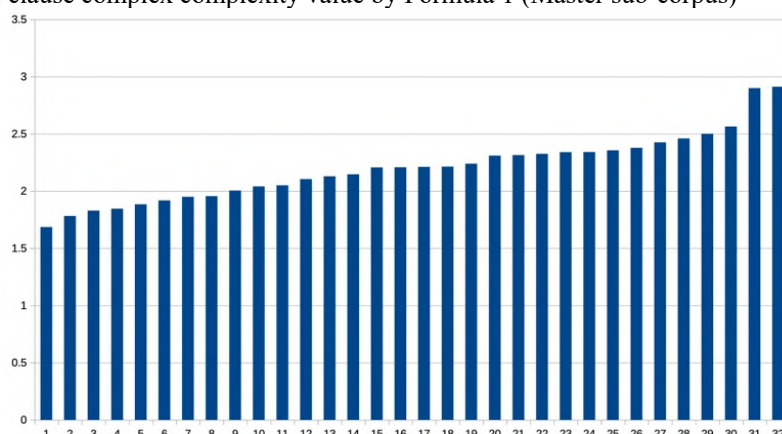


Table 123 The clause complex complexity value by Formula 1 (PhD sub-corpus)

To make it clearer, Table 124 is presented here:

Value Range	1.5-2	2-2.5	2.5-3	3-3.5	3.5-4
M	20	37	19	3	3
D	8	20	4	-	-

Table 124 The number of texts within a complexity value scope in Master and PhD sub-corpora

In both groups, 25% of texts have a complexity value ranging from 1.5 to 2 (not included), most texts have a value ranging from 2-2.5 (not included), with the proportion of the Master group being 46.25% and of the doctoral group being 62.5%, still, some have a value within the scope of 2.5 to 3 (not included) (23.75% in group M and 12.5% in group D). The complexity value in some Master theses can even reach up to 3.5 or higher, with the highest being 3.56, while the highest value in the doctoral group is only 2.91.

As Formula 1 overemphasizes the importance of recursion level, when more clause nexuses on level 1 are chosen in clause complexes, the complexity value would surely decrease as the proportion of clause nexuses on other levels decreases too. The reason for the complexity value ranging mainly from 2 to 2.5 is the high probability of using one and two clause nexuses in a clause complex by students, as suggested by the former section. The higher average complexity value of the Master's group, which exceeds our expectation, is relevant to the sample size as more texts bring a greater variety of writing capacities with

higher probabilities of connecting more clause nexuses on deeper levels, resulting in higher complexity values as shown in Table 122. Then comes a question: does deeper clause recursion indicate the more complex language, or on the clause complex aspect at least?

7.5.6 Significance & Limitation

The biggest problem of the three formulas above lies in the nonconcern of embedded clauses and facts that the relation within a clause and a noun or verb or adjective have not been considered, so the measurement of clause complex complexity reflects more on relations within clauses rather than those in words (phrases) and clauses. These logico-semantic relations are more from a functional instead of structural perspective that the function clauses play in clause complexes present writers' choice in certain registers such as thesis writing. If embedded clauses must be reconsidered, more procedures or practices would be required in annotation as unautomated annotation will be more difficult, more time and energy-consuming, which in fact can also affect research results in a sense. Another measuring method for grammatical categories, taxis, and levels can be matrices that more detailed categories such as additional, temporal, and causal relations are of necessity for investigation, but this method asks for a higher standard for markdown. The next step can be an update for annotation software.

As stated above, the research of "clause complex complexity" is not a supplement or a substitution of "syntactic complexity," instead, the concept is a parallel one for diverse purposes, researching objects from different perspectives by another method and then results in totally different conclusions, with various functioning scopes. The study of clause complexes in a text can thus be calculated and compared, which is a breakthrough and supports writers for meaning-making in focusing on both relation types within clauses and recursion levels that clauses function on. Following studies can combine clausal complexity with word/phrasal-clausal relations complexity; compare complexity values of texts from diverse registers and genres, of chunks that play diverse functions in a text, and of texts written by higher educational levels, etc.; investigate the influence of writing purposes on clause complex choices, resulting in meaningful complexity value diversities; discuss relations of language meaning and form from the sphere of clause complex complexity; explore possibilities for promoting the teaching of academic writing; and endeavor to establish the standard for writing complexity assessment.

7.6 Conclusion for Discussions

The conclusion of this chapter begins with a proverb by Churchill³⁵: "Now this is not the end. It is not even the beginning of the end. But it is, perhaps, the end of the beginning." Though some works have been finished and demonstrated in this chapter, the research in some directions just starts as new concepts have been raised and new angles for solving existing problems have been put forward and commenced.

To be specific, the first section presented general conditions of data outcome in Chapter Five and Six, which covered 1. the grammatical categories distribution of the clause complex system, 2. range, mean and Bayes Factor values of 73 measuring indexes of clause complex grammatical comparison within MA and PhD groups, and 3. lexical choices that mark clause complex relations by comparing category distributions within Master's and doctoral theses, and within the whole corpus and Wu's outcome that more diversities than similarities have been discovered and some interesting phenomena have been dug out; the range of proportions of each grammatical category in the system in the whole corpus have indicated some necessary categories in thesis writing genre with some optional ones; the preference for hypotactic versus paratactic structures or enhancements versus elaborations has been proved to be obvious by judging the Bayes Factor values of all indexes within two groups; the inconsistency of range and mean of these indexes have suggested a multivariate distributed data in both groups, and the comparison of mean values and BF values in each index resulted to a controversial outcome that needs further exploration; and annotation markers in five categories, i.e., elaboration, extension, enhancement, locution and idea have been particularly compared with their numbers in sub-categories, to answer questions raised in Section 1.3 that how the lexical choices influence the selection of relating grammatical relations to realize functions that clause complexes perform in texts, how different are lexicogrammatical choices within Master's and doctoral students, and how to explain these diversities, etc., to provide students with possibilities for improving their grammatical complexity if more hypotactic or complex structures indeed mean development, and to confirm the viewpoint of regarding lexis and grammar as two ends of a cline instead of hyponymy at last.

³⁵ By Winston S. Churchill. From the speech at the Lord Mayor's Day Luncheon, 10th November, 1942.

Section 7.2 then demonstrated four kinds of additions (implicit, unequal, unrelated, and wrongly additions) concluded within academic writing and explained with definitions and examples. In the description of the location of these four additional kinds with their general proportion, except for around 90% of additions being explicit additions, implicit addition, and unequal addition can still be called additions, with wrong additions needing to be excluded. To extend four kinds of additions to any grammatical category in the clause complex system, the location of four layers was revealed as an investigation of peculiarities analysis. Another part briefly stated some kinds of clause complexes with wrong relations or markers in writing and methods to identify and recompose them, i.e., adding, missing, substituting, inconsistent, not-splitting, punctuation mistakes, and wrong usages.

Section 7.3 was concerned more with the differentiation of the implicit clause complex from the explicit one. The theoretical foundation for putting forward the concept and coming up with the division was initially clarified, most of which originated in previous viewpoints on relations within clauses, tacit knowledge, and choice or the probability theory. Before getting right to the point, two groups of close terms were defined, and irrespective of their inappropriateness, i.e., “overt” - “covert” and “communicative” - “tacit,” the counterpart “implicit” versus “explicit” was then confirmed. The average proportion of explicit versus implicit relations in each CC category in texts written by the two groups indicated that implicit enhancement is preferred by doctoral candidates, whereas implicit elaboration is favored by Master’s students. The proportion of explicit versus implicit categories in the whole system suggested that all implicit relations are no more than 5%, in which implicit enhancement accounted only for around 3.5%, followed by implicit extension, implicit projection, and implicit elaboration. Characteristics of the implicit clause complex were concluded, i.e., recessivity, integrality, stability, occasionality, situationality, culturality, relativity, individuality, and exclusivity, followed by the classification of implicit CC from the perspective of annotation: decidable, undecidable, and ambiguous or equivocal ones. Some related structures that are hard to identify were also pointed out with details and examples: unique finite clause complexes, especially clauses connected by words like *and*, *but*, *so*, etc., non-finite ones, particularly when the implicit process is played in dependent clauses and linked by *in* and *without*, one finite clause plus one non-finite clause, implicit projections, etc. The section ended with steps for transforming implicit clause complexes into explicit ones if permitted.

If the discussion of implicit and explicit clause complexes can be treated as a theoretical expansion to SFG and enhancement for practically raising the students’ consciousness in organizing clauses in academic writing, then the generalization and summarization of common patterns for the recursion of clause complexes is another innovation that, on each recursion level, similarities can be more significant than diversities for Master’s and PhD students’ choices, all of which have been clearly demonstrated in Section 7.4. Besides, the clause order of any grammatical category in a clause nexus with corresponding logico-semantic markers was presented to thoroughly deconstruct the clause complex.

In Section 7.5, the concept “the complexity of clause complex” or “the clause complex complexity” has been proposed as an alternative perspective, i.e., a functional or dynamic angel of treating the connection of clauses instead of a structural or static angel that was traditionally called “syntactic complexity.” After origins and the theoretical background were introduced, and the connotation of “complexity” was made clear, the term “the clause complex complexity” was provided with a complete definition, and three formulas synthesizing categories with proportions were modeled based on the definition. In applying the first formula in testing the complexity of texts from MA and doctoral groups, a reversed result has been received, further investigation and improvement of the formula are hence deserved. Though the two other formulas are more delicate on account of the proportion of every sub-category on each level, problems are still obvious when these proportions are calculated according to manual recording. Anyway, the attempt to depict how complex each text is in utilizing clause complexes is worthy of exploration.

Hence, the “Big System” section, “common patterns” section, and “four kinds of additions” are traits of data that generality occupies an important position in forming these sections, though the quantitative method is overwhelming in the first one, the qualitative method dominates in the last one and in describing clause orders in the second, and both qualitative and quantitative methods are adopted in researching for common patterns of clause complex recursions. The “implicit versus explicit” part and “error analysis” part in Section 7.2 pay attention to peculiarities from the corpus that by applying the grounded theory method in the former and case study in the latter, “implicit clause complex” has been depicted in various dimensions and ways of revising wrong clause complexes have been mentioned. The last section, i.e., “the complexity of clause complexes,” is an endeavor for measuring clause complex organization and sophistication in texts, though starting from academic discourse in this research, will not be confined to only this genre in the future.

Chapter 8 Conclusion

8.1 Conclusion

This study has undertaken a comprehensive analysis of the choice of clause complexes in academic thesis writing texts, with a particular focus on the following aspects:

Chapter One presented an introduction to some basic issues about the current study. It outlined the research background in Section 1.1, including the rationale for conducting this research, and discussed the following points: 1. the difficulties that English learners face in writing their theses, especially when they transfer from a Master's student to a PhD candidate and encounter clause combinations of the English language; 2. the importance of focusing on some grammatical categories, such as taxis instead of coordination or subordination, expansion rather than conjunction; 3. the structural diversities of projection from the traditionally called "citation"; 4. the meaning realization of implicit or explicit knowledge of language and the mention of the implicit language, and 5. the complexity of language grammar and lexical choices, and how should we develop a methodology for measuring them. Section 1.2 addressed the purpose of the present investigation, including clause organization and meaning transferring, language development, the clarification of concepts, the adaptability of SFG, and the deepening of theoretical applications. Section 1.3 presented detailed research questions of this study, i.e., explored from the theoretical, methodological, practical, analytical, and applicable aspects. The theoretical significance of the research provided insights into the necessity of excavating research, confirming the direction of research, constructing theoretical analyzing frameworks, adopting a mixed research method, building corpora, annotating texts, verifying the feasibility of quantitative and qualitative analysis, developing theory denotations, and discovering new research directions. The practical significance of the research has been demonstrated that the outcome of Master's and PhD students' theses comparison can be inspiring and instructive for thesis writing, reading texts, teaching writing, Chinese-English translation, machine translation, etc., error analysis can be illuminating for the prevention of further mistakes, the grammatical-lexical analyzing practices can be referential for conducting discourse analysis, the complexity of clause complex can facilitate further texts calculation and comparison, and the clause combining modes from a functional perspective can inform computational linguistics studies.

Chapter Two turned to a review of the relevant literature, encompassing both academic writing and the clause complex. The former has drawn attention to several fundamental concepts in academic writing, including annotations and denotations (such as those related to EAP, ESP, and EAW), research theories of discourse analysis, academic literacies, the cognitive perspective, etc. It has also considered research methods in corpus linguistics from both quantitative and qualitative perspectives, as well as potential research directions in discourse analysis, academic literacies, cognition, identity construction, and so on. The literature review also encompassed writing-related knowledge, including writing purposes and motives, writing modes, writing analysis frameworks, and the writing process. It further included specific research on English thesis writing of Master's and PhD students, conjunction and citation, which are analogous to expansion and projection concepts in the clause complex system, as well as studies on academic writing from a systemic functional linguistic or grammatical perspective. The literature review of clause complexes has demonstrated over one hundred published works in journals, books, online dissertations, and thesis platforms from China and other countries. Given that half of the studies are conducted in China and in Chinese, and that the theme of this project is on Chinese students' conditions and background, the themes of clause complex, projection, and expansion have been intersected with theoretical studies, discourse analysis, English-Chinese language comparison, stylistics, translation, and related studies for literature review. The objective has been to demonstrate the mutual views and conflicts in previous works, identify gaps in past research from methodological, theoretical, and directional aspects, situate this study in related areas, highlight the significance of this study, and indicate the probable creativity that has been presented in subsequent works.

Chapter Three commenced with an exposition of the theoretical dimensions of this project, and an examination of how three directions intersected with and mutually influenced the design, performance, and research analysis. The first area of focus has been systemic functional grammar (SFG), which served as the foundational theory underpinning the entire research project. It has included an in-depth exploration of key concepts such as the definition of the term clause and clause complex, and an examination of the nuances of expansion and projection. The second area of interest has been academic writing (AW), which encompasses a rough depiction of the writing process, including constructing a thesis and identifying effective teaching methodologies. Finally, the third area of focus has been corpus linguistics (CL) with probabilistic theory in SFG. In detail, the SFG section elucidated the rationale

behind its selection as a theoretical foundation, delineated its distinctions from other functional or analogous grammars, illuminated the socio-function of AW, outlined the contents that diversified eight groups of keywords, and identified the focal point of this research: the clause complex. The section has provided definitions and indicated the position of the clause and the clause complex in and around the SFG theoretical frame, and underscored three meta-functions of a clause. The framework of the clause complex has been fully delineated, with Taxis having been subdivided into Hypotaxis and Parataxis, Expansion having been composed of Elaboration (exposition, exemplification, and clarification), Extension (addition, variation, and alternation), and Enhancement (from temporal, spatial, manner, causal, and conditional), and projection having been usually distinguished as quoting or reporting, with examples of their intersections being drawn from both Master's and PhD students' theses. The AW section has considered AW as a register and addressed the following questions: how can SFG be applied to AW, and subsequently influence AW theories and methods; furthermore, how AW practical problems can benefit SFG linguists, and therefore extend the SFG theoretical development. To conduct a qualitative analysis of projection, a framework has been constructed that takes into account several factors, including the Appraisal Theory, the interpersonal function of citation verbs, the writer's intention concerning the author's original meaning, and readers' expectations, the tense/mood/aspect/voice/modality presented in the projection clauses, and other relevant elements. Subsequently, a description of CL has been provided, with particular attention to its narrow sense as influenced by SFG. This encompassed its overall generality, historical development, interrelationships, classification, and research methodologies. Additionally, key terms, particularly those related to probability and instantiation, have been examined. The role of CL as a perspective for SFG studies in this project has also been discussed. In conclusion, a theoretical framework has been outlined and a conclusion has been drawn from a schematic diagram with three circles. This diagram highlighted the emphasis placed on qualitative analysis of register analysis in SFG from SFG & AW research overlap, the execution of quantitative analysis of grammatical ones from SFG & CL crossing, the conduct of lexical analysis from quantitative methods within CL & AW research overlap, and the proposal of a whole clause complex system for this project by scoping the SFG&AW&CL intersection. Therefore, the theoretical foundation has been established for analyzing the practical problems encountered by MA and PhD students in thesis writing in the last section.

Chapter Four has described the corpus selected for analysis in this study and the corresponding analytical methods in detail. The Chinese Academic English (ChAcE) Corpus was constructed, and its data collection conditions, data conversion, and cleaning processes were presented. Additionally, related information was provided. Subsequently, data comprising 80 MA theses (1.28 million words, 28,350 clause complexes) and 32 PhD dissertations (1.44 million words, 28,257 clause complexes) were selected from the ChAcE corpus for clause complexes analysis. Before addressing the chosen corpus, the advantages and disadvantages of qualitative versus quantitative methods were enumerated. A mixed design was then established for the research, and fixed methods were adopted according to the requisite criteria. Descriptive and qualitative methods (case study and grounded theory) were primarily employed for annotation and example analysis; quantitative and comparative methods were chosen for data calculation; experimental summary and modeling analysis were taken for system construction; and an exploratory method was adopted for outcome discussion. In terms of the tools and software, *SysFan* was utilized for the manual annotation of data, the recording of clause complexes and their connecting words with markedness, and the counting of grammatical categories across the entire system. *AntConc* has been proven effective for the calculation of keywords within each clause complex, as well as the exclusion of unnecessary ones. *JASP* proved invaluable for statistical analysis, employing both Independent Samples T-test and Bayesian Independent Samples T-test with figures and diagrams. Following the division of the selected corpus into two groups, i.e., CMAC and CPhD, an overview of its general information was presented, along with an outline of the procedures employed for its importation and annotation. Subsequently, a series of indexes were discussed, including the number of clauses present in each text, the average number of words per clause, the average number of clause complexes within sentences, the average number of clause nexuses per clause complex, the prevalence of expansion and projection structures in each text, the number and proportion of clause nexuses on each recursion level (from 1 to 7), the average proportion of parataxis or hypotaxis in clause nexuses in each text, and so forth. These indexes were counted, calculated, and compared at the Master's and PhD levels. In Section 4.4, the theoretical framework presented in Chapter Three was synthesized with the relevant analytical methods mentioned above, providing an outline for the data analysis.

The data comparison and explanations presented in Chapters Five and Six represent the most crucial elements of this dissertation. Chapter Five commenced with an overview of the general conditions of expansion, subsequently focusing on the composition of expansion, including elaboration, extension, and enhancement. This involved the systematic comparison of quantitative data (including descriptive

statistics, independent samples t-tests, and the Bayesian independent samples t-test for each index) and qualitative information from MA theses and PhD dissertations. In greater detail, the expansion section (5.1) encompassed the average proportion of expansion in all clause nexuses. The indexes were compared in terms of paratactic and hypotactic expansion. The former refers to the average proportion of paratactic expansion in all clause nexuses, paratactic structures, and expansions. The latter, in turn, concerns the average proportion of hypotactic expansion in all clause nexuses, hypotactic structures, and expansions. Additionally, the average proportion of each composition in expansion structures was calculated for two groups. The quantitative analysis of elaboration entailed the calculation of the mean proportion of elaboration in all clause nexuses and conditions of paratactic (i.e., the mean proportion of paratactic elaboration in all clause nexuses, all paratactic structures, paratactic expansions, and all elaborations) or hypotactic elaboration (i.e., the mean proportion of hypotactic elaboration in all clause nexuses, all hypotactic structures, hypotactic expansions, and all elaborations) in two groups.

Qualitative analysis of elaboration in Section 5.3 referred to how elaboration structures, from the classification of exposition, exemplification, clarification, and hypotactic elaboration, were fulfilled through the realization of corresponding markers or lexis by both Master's and doctoral students in their writing with fine distinctions. Though the number of these markers chosen by students was listed and compared, the status of the overuse of which lexis causes a higher proportion of related grammatical structure was not transparent until all data of lexicogrammatical comparisons were comprehensively considered. The quantitative analysis of extension was analogous to that of elaboration, although the numerator of comparing indexes has been changed into paratactic or hypotactic extension, and the denominator of the last group was extension but not elaboration. Then, a qualitative discussion of extension covered additional (positive, negative, and adversative), variational (replative and subtractive), and alternative clause relations, in which the positive addition relation marked by *and* has been proved to be the most important and difficult one for students to master. A qualitative analysis of elaboration in Section 5.3 indicated that the structures of elaboration, including classification of exposition, exemplification, clarification, and hypotactic elaboration, were fulfilled through the realization of corresponding markers or lexis by both Master's and doctoral students in their writing, with notable distinctions. Although the number of markers selected by students was listed and compared, the impact of overusing specific lexis on the prevalence of related grammatical structures remained unclear until all data from lexicogrammatical comparisons were thoroughly examined. The quantitative analysis of extension was analogous to that of elaboration, with the numerator of the comparison indexes modified to include paratactic or hypotactic extension, and the denominator of the last group changed to reflect extension rather than elaboration. Subsequently, a qualitative discussion of extension was undertaken, encompassing additional (positive, negative, and adversative), variational (replative and subtractive), and alternative clause relations. It has been demonstrated that the positive addition relation marked by *and* is the most significant and challenging for students to master. As anticipated, the comparative analyzing results between paratactic and hypotactic enhancement choices by Master's and doctoral students demonstrated that doctoral candidates exhibited a greater preference for the latter. This preference was observed regardless of the parallel measuring index, the average proportion of hypotactic enhancement in all hypotactic expansions, or all enhancement structures (Section 5.6). The qualitative analysis of enhancement was divided into five main functions, as follows: temporal (same time, earlier, or later), spatial (same place), manner (means or comparison), causal (from reason, result, and purpose), and conditional (positive, negative, or concessive). The latter two functions were more significant when students employed causal relations to explain principles, indicate the influence of one matter on another, and present researchers or their intentions to readers. They also used conditional relations to identify sufficient or necessary conditions to fulfill something, highlight something that cannot be realized despite conditions being met, or state the endeavor for something confirmed as unachievable. The preceding sections have demonstrated the similarities and differences between the two groups in their selection of lexicogrammatical structures to convey their meanings and perspectives.

The same data calculation and comparison methods were employed in Chapter Six just as those in Chapter Five, only with categories being changed from hypotactic and paratactic projection to Quoting Locution and Reporting Locution, to Quoting Idea and Reporting Idea; from Locution and Idea to propositions and proposals; and from grammatical choices to lexical (projection keywords) instances. This was done, to horizontally and vertically dig out how different Master's and Doctoral students were referencing and citing other scholars' words directly and indirectly. To be specific, Section 6.1 constituted indexes for comparison of referencing and citing practices within master's theses and doctoral dissertations: the average proportion of projection in all clause nexuses; paratactic projections in clause nexuses, paratactic structures, and all projections; hypotactic projection in all clause nexuses, hypotactic structures and all projections; and the composition of projection, namely, the number and average proportion of locution and idea in all projecting structures. For the quantitative analysis of the

location part, besides hypothesis tests within two groups on the index the average proportion of location in all clause nexuses, paratactic locations were compared from denominators being clause nexuses, paratactic structures, paratactic projections, and locutions to calculate the average proportions in each group; and hypotactic locutions were compared when denominators have been changed to clause nexuses, hypotactic structures, hypotactic projections and again locutions. The qualitative analysis of location was conducted to “projecting verbs”, categorizing them into verbs that lead to proposals or propositions according to the speech functions they perform in a clause complex. By presenting each verb chosen in theses with their frequencies, the condition of which verbs were usually seen in each group, which markers were both common in two groups with high frequencies, which were common but have varieties of their preferences in two groups, and which verbs were unique or peculiar in each group have been plain that the variety came from the number of texts on the one hand and the direction difference within literary studies or linguistics research on the other hand. As not all projection structures appeared in the literature review part of the thesis, it thus needs further investigation. The interpersonal metafunction of projections in thesis writing has been elucidated by integrating the SFG theoretical framework with Appraisal Theory in illustrative examples. Concerning the quantitative analysis of the idea part, apart from transforming the numerator into the paratactic or hypotactic idea and modifying the denominator from location to idea for hypothesis tests within two groups, the method and principle for index comparison remained unaltered. Although most of the thesis exhibits no paratactic ideas, the Bayesian independent samples t-test has been employed to ensure precise results and to supplement mean comparisons. The qualitative analysis of Idea has identified some verbs that lead to the formation of propositions, including: *show, find, indicate, believe, mean, see, think, know, suggest*, etc. The objective has been to identify how projected ideas were originally, objectively, faithfully, or in reverse, personally, subjectively, and forcefully organized in thesis writing and then transmitted to readers to establish a certain relationship with them. The summary for Chapter Six has highlighted some key research findings and their associated results.

In Chapter Seven, the general conditions of data outcome have been discussed in three parts. The first part concerned the grammatical categories’ distribution of the clause complex system. The second part presented the range, mean, and Bayes Factor values of 73 measuring indexes of clause complex grammatical comparison within MA and PhD groups. The third part presented the results of a further analysis of the data. The initial demonstration of lexical choices that marked clause complex relations has been conducted through a comparison of category distributions within Master’s and doctoral theses, as well as within the entire corpus and Wu’s outcome. This comparison has revealed a greater number of diversities than similarities and has also highlighted some interesting phenomena. The range of proportions of each grammatical category in the system in the whole corpus has indicated which categories were essential for effective writing in the thesis genre while others were optional. By evaluating the Bayes Factor values of all indexes within the two groups, it has become evident that there was a preference for hypotactic versus paratactic structures or enhancements versus elaborations. The inconsistency in the range and mean of these indexes has indicated the presence of multivariate distributed data in both groups, and the comparison of mean values and BF values in each index has yielded a controversial outcome that necessitates further exploration. Annotation markers in five categories, i.e., elaboration, extension, enhancement, location, and idea, have been compared with their numbers in sub-categories to address the questions raised in Section 1.3. These include how the lexical choices influence the selection of relating grammatical relations in realizing functions that clause complexes perform in texts, how these lexicogrammatical choices differ between Master’s and doctoral students, how these diversities can be explained, etc. This allows students to enhance their grammatical complexity, assuming that more hypotactic or complex structures represent development. It also corroborated the perspective that lexis and grammar are two ends of a continuum rather than a hierarchical relationship. Section 7.2 presented four types of additions within academic writing texts and the annotation of erroneous clause complexes. The former denoted unrelated, unequal, wrong, and implicit relations in each grammatical category in the clause complex system. These relations were put forward from the addition relation as an investigation of peculiarities analysis. In addition, the location of the four additions and the location of the four layers in any system category have been depicted. The latter indicated some kinds of clause complexes with erroneous relations or markers in writing. It also delineated the means of identifying and then recomposing them (CCs), which included the following: adding, missing, substituting, inconsistent, non-splitting, punctuation mistakes, and wrong usages. Section 7.3 is more concerned with the differentiation of implicit clause complexes from explicit ones, establishing the theoretical foundation for proposing the concept and originating from clause relations, tacit knowledge, and choice or the probability theory. Before delving into the specifics, two groups of closely related terms were delineated. Despite the inappropriateness of these terms, namely “overt” vs. “covert” and “communicative” vs. “tacit,” the counterpart “implicit” versus “explicit” has been validated.

Subsequently, the mean proportion of implicit clause complexes in all CCs in each text within the MA and PhD groups was determined. Additionally, the characteristics of implicit clause complexes, including recessivity, integrality, stability, occasionality, situationality, culturality, relativity, individuality, and exclusivity, were examined. The classification system was also evaluated. Furthermore, some related structures that were challenging to identify were investigated, including unique finite clause complexes, clauses connected by *and*, non-finite ones, especially when the implicit process was played in dependent clauses, one finite clause plus one non-finite, implicit projection, and others. Finally, the process of transforming implicit clause complexes into explicit ones, where feasible, was outlined. The discussion of implicit and explicit clause complexes could be regarded as a theoretical expansion to SFG, which has facilitated the practical development of students' awareness regarding the organization of clauses in academic writing. Moreover, the generalization and summarization of common patterns for clause complexes recursion represented further innovation. This has been illustrated in Section 7.4, at each recursion level, similarities between the choices of master's and doctoral students were more significant than the diversities. In addition, the clause order of any grammatical category in a clause nexus with corresponding logico-semantic markers was presented to provide a comprehensive deconstruction of the clause complex. Section 7.5 proposed the concept of "the complexity of the clause complex" or "the clause complex complexity" as an alternative perspective. This perspective treated the connection of clauses from a functional or dynamic angle rather than the traditional structural or static angle, which has commonly been referred to as "syntactic complexity." Following an introduction to the concept's origins and theoretical background, the term "clause complex complexity" was defined, and three formulas synthesizing categories with proportions were developed based on this definition. An unexpected outcome was encountered when applying the initial formula to assess the complexity of texts from the MA and doctoral groups. However, further investigation and refinement of the formula may have yielded more accurate results. While the two other formulas were more nuanced regarding the proportion of each subcategory at each level, issues were still evident when these proportions were calculated through manual recording. In any case, the attempt to depict the complexity of each text in terms of its use of clause complexes is a worthwhile avenue for further investigation.

In response to the series of questions raised in Section 1.3,

Theoretically:

RQ1T1. The application of clause complexes in text analysis in specific genres, such as academic thesis writing, necessitates a discussion of the theory's applicability, an appropriate corpus, and a corresponding analysis framework. Specifically, the initial step is to establish a framework appropriate for the annotation of the corpus and to select the most suitable qualitative and quantitative analysis methods. Similarly, qualitative analysis of some content requires the construction of a framework, and the subsequent statistical comparative analysis of the research data obtained needs to be conducted. Nevertheless, the foundation of these procedure hinges upon a thorough grasp of the preceding research outcomes and the shortcomings of the clause complex and academic thesis writing.

RQ1T2. Research scope difference of the clause complex compared with concepts such as "conjunction," "coherence," "citation," etc., in traditional grammar (sense), can be referred to Sections 1.1 & 2.1.

Methodologically:

RQ2M1. To ensure the reliability of the text annotation, it is necessary to select at least 30 articles from each group, with a minimum of 100 clause complexes per article. This is because the length of each article varies considerably. A too-large corpus will undoubtedly require a significant investment in manual annotation effort. Conversely, a corpus that is too small may yield significantly biased results.

RQ2M2. The categories of clause complexes can be quantified, as evidenced in Chapters 4, 5, 6, and 7. Nevertheless, the following prerequisites must be met: a complete analysis framework must be established; appropriate annotation recording software, such as SysFan, must be utilized; a certain number of analysis texts must be selected for quantitative analysis; a reasonable number of annotations and comparison plans must be designed; and the purpose and method of quantification must be clarified.

RQ2M3. The research methods that promote clause complex analysis have been outlined in Section 4.2. These include exploratory research (heuristic analysis), comparative analysis, experience summary, grounded theory, and case study in qualitative analysis; descriptive research, corpus linguistics, modeling analysis, etc., in quantitative analysis; and software suitable for clause complex analysis, which includes *SysFan*, *AntConc*, *UAM Corpus Tool*, *JASP*, *SPSS*, *Matlab*, *R*, etc. However, some of these have not been used in this research analysis.

Practically:

RQ3P1. The choice of lexis is realizing the choice of grammar, which in turn is inextricably linked to the specific lexical choices when a particular function is to be realized. Consequently, the realization of logico-semantic relationships within the context of the lexical-grammar continuum is contingent upon

the choice.

RQ3P2. The selected Master's and doctoral theses exhibit certain discrepancies in fundamental aspects, which are delineated in Section 4.3:

The number of words in each sentence (clause complexes) in each text in the Master's group is around 18 - 31, with a mean number of 23.10, and in the doctoral group is about 18 - 33, with a mean number of 25.68. This can be seen in Section 4.3.3.2.

The proportion of clause complexes/sentences in each text in the Master's group is around 39% - 80%, with a mean proportion of 57.06%, and in the doctoral group is about 42% - 67%, with a mean proportion of 56.05%. Conversely, the proportion of clause simplexes/sentences in each text in the Master's group is around 20% - 61%, with a mean proportion of 42.94%, and in the doctoral group is about 33% - 58%, with a mean proportion of 43.95%. This has been demonstrated in Section 4.3.3.4.

The proportion of clause nexuses per clause complexes in each text in the Master's group is around 1.25 - 2.02 with a mean number of 1.56, while in the doctoral group is about 1.32 - 1.76 with a mean number of 1.52. This can be referred to in Section 4.3.3.5.

For the recursion proportion of clause nexuses on each level in each text: on L1, the proportion in the Master's group is around 64.88% while in the doctoral group is about 66.17%; on L2, the proportion in the Master's group is around 25.83% while in the doctoral group is about 25.34%; on L3, the proportion in the Master's group is around 7.33% while in the doctoral group is about 6.64%; on L4, the proportion in the Master's group is around 1.55% while in the doctoral group is about 1.52%. Detailed numbers and other proportions can be seen in Section 4.3.3.6.

The proportion of parataxis per clause nexuses in each text in the Master's group is around 24.39% - 55.72% with a mean proportion of 39%, while in the doctoral group is about 18.12% - 46.20% with a mean proportion of 34.89%. This can be found in section 4.3.3.7. Correspondingly, the proportion of hypotaxis per clause nexuses in each text in the Master's group is around 44.28% - 75.61% with a mean proportion of 61%, while in the doctoral group is about 53.8% - 81.88% with a mean proportion of 65.11%. Details can be referred to in Section 4.3.3.8.

RQ3P3. For choices of grammatical categories, see quantitative analysis in Chapters 5 & 6, as well as that in Section 7.1. Whether PhD dissertations are more complex than MA theses in choosing logico-semantic relations depends on which categories are regarded as markers for the complexity of clause complexes.

RQ3P4. Students do exhibit proclivity when choosing specific words and phrases that mark logico-semantic relations within clause complexes in thesis writing. For a more detailed examination of this phenomenon, please refer to the qualitative analysis in Chapters 5 and 6. The analysis demonstrates that there is a plethora of words and phrases that can be selected to fulfill the same function. Moreover, the same words and phrases can be deployed to achieve subtly different functions in disparate contexts. It is important to note that certain words or phrases may serve multiple functions, and thus, the specific function they perform in a given text must be considered in its context. In general, the choice of expansion lexis is primarily concerned with achieving discourse functions, connecting clauses, and establishing connections between preceding and subsequent elements in a clause complex. Conversely, the choice of projection lexis is predominantly focused on fulfilling interpersonal functions, demonstrating the interaction between individuals, which can also be observed in the writer's affirmation or negation of his/her perspectives, the expression of positive or negative attitudes, the indication of support or opposition, and the anticipation of readers' expectations.

RQ3P5. To ascertain the existence of a distinctive pattern in the thesis writing, it is necessary to undertake a comparative analysis of the register with other registers. Nevertheless, there are notable similarities and differences between the Master's and doctoral students. For example, both groups have a high probability of $1+2$, $\alpha x\beta$, $\alpha'\beta$, $x\beta\alpha$, $\alpha=\beta$, $\alpha'\beta(1+2)$, $x\beta\alpha(1+2)$, $\alpha'\beta(\alpha x\beta)$, $1+2(\alpha x\beta)$, $\alpha=\beta(1+2)$, and so forth. Additionally, there are numerous distinct patterns. At the L3 and above recursion levels, the identification of repeating patterns becomes increasingly challenging. If the number of clause complexes examined in each group is not 500 but 5,000, 50,000, or even larger, the recursion patterns that can be summarized will undoubtedly differ. This will significantly impact the results at L3 and above levels. Nevertheless, the impact of a larger corpus on the common patterns found in L1 and L2 may be relatively limited. For further details, please refer to Section 7.4 of the main text.

Analytically:

RQ4N1. The rationale behind the research findings is presented in Chapters 5, 6, and 7. Given the considerable number of results, it is not feasible to provide a detailed account of each one in this context. The limitations of this thesis will be discussed in Section 8.4.

RQ4N2. Most research results align with the pre-existing expectations, although a few exceptions exist. For instance, Master's students choose more paratactic and hypotactic ideas, whereas doctoral students prefer hypotactic and paratactic locutions to their Master's counterparts. If the choice of ideas over

locutions results in a more complex text, then the projection chosen by Master's students is more intricate than that chosen by doctoral students. Firstly, the number of ideas/locutions is notably low compared to other subcategories. Secondly, the number of Master's texts is 2.5 times that of doctoral texts. Given the discrepancies in writing style between writers, it is expected that the number of projections in Master's texts will exceed that in doctoral texts, with the types remaining consistent. The research findings may slightly differ if the same number of clause complexes are selected from the same corresponding number of texts for data calculation and comparison. As to comparing the results of choosing the overall system category with those of Wu (2000) (see Section 7.1), it has revealed significant discrepancies, which can be attributed to factors such as register, the number of texts, the number of clause complexes, and the statistical methods employed.

Applicable:

RQ5A1. A mathematical model may be constructed to analyze the complexity of clause complexes; however, the parameter problem must first be resolved. This entails determining which grammatical categories are essential within the system and how the proportion of each grammatical category should be quantified. This necessitates some fundamental quantitative research. However, in the absence of a solution to the problem of manually annotating clause complexes, constructing a mathematical model becomes challenging.

RQ5A2. For further details on the practical application of these findings, please refer to Section 8.5.

RQ5A3. The qualitative study of clause complexes is predicated upon establishing an analytical framework, which, however, necessitates a high level of theoretical expertise among researchers. Further discussion is required regarding specific subcategories within the framework and the potential for subcategories to be summarized. The primary objective of the quantitative research is to solve the problem of automatic annotation within the analytical framework. Without computer-assisted annotation, resolving many problems will require significant time and effort, with some issues remaining intractable.

8.2 Contributions of the Study

Most researchers have focused their investigations on grammatical sub-categories of expansion or projection with the intersection of taxis and then stopped, not to mention sub-categories of exemplification or locution, for example. As recursion is crucial for expanding a clause complex, common patterns can illustrate stylistic choices and preferences of writers in conveying meaning, which is also significant but often overlooked. In addition to re-examining the frameworks of projection and expansion in English using the modified analytical framework and incorporating the preceding theoretical points with the real-life examples in the corpora, this research also explains how the sub-modes of projection works and a detailed explication of how projection enables the ellipsis of projecting clauses and substitution of projected clauses. It provides a comprehensive classification of conjunctions and a detailed account of how constituent clauses are combined, the operation of each sub-relation of expansion, and the possibility of ellipsis as a result of the logico-semantic relation of expansion.

In detail, in the introductory chapter, the differences between coordination/subordination and parataxis/hypotaxis are elucidated; the distinction between expansive logico-semantic connecting devices and conjunction/preposition is made; the contrast between projection and citation, as well as between projection verbs and citation verbs is drawn; the difference between tacit knowledge, covert meaning, and the implicit language is presented; and the shortcomings of the concept of syntactic complexity are highlighted, thus providing a robust foundation for subsequent research.

Articles, theses, dissertations, and even book chapters on Clause Complex written by students from countries other than China are referenced and discussed in the Literature Review section to gain a broader overview of the research scope of CC in publication until now. A review of domestic and international research on English academic writing and master's and doctoral writing texts from the perspectives of discourse analysis, academic literacy, and cognitive science provides a more detailed account of the reasons for studying expansion connectors rather than conjunctions and projection rather than citation.

In terms of the theoretical framework that underpins this research, a synthesis of theories and research methods has been employed in three distinct directions: Systemic Functional Grammar, Corpus Linguistics, and Academic Writing, which are collectively constructed to provide a theoretical foundation for understanding and conducting research practices. The rationale for selecting systemic functional grammar as the theoretical foundation of this study is illuminated by a comparative analysis with related functional grammar. The illustration of the social function of academic writing demonstrates the applicability of systemic functional linguistics. Eight groups of related concepts in systemic functional grammar provide a robust theoretical foundation for analyzing grammatical categories within

the system framework. The concepts of the clause and the clause complex, including their positions, distinctions, structures, and the specific constituent of the clause complex, i.e., expansion and projection, along with the concepts of their subcategories, are elucidated through illustrative examples, providing one of the frameworks for the theoretical analysis of this study. The description of grammatical categories supported by writing examples drawn from the corpus attests to the applicability and integrity of the theoretical framework. The evaluative perspective or the Appraisal Theory, which views English academic writing as a register rather than a genre, offers an additional framework for the qualitative analysis of the projection. Finally, the advantages and disadvantages of systemic functional linguistics research from a corpus linguistics perspective highlight the necessity for selecting specific research methods.

It is uncommon for software such as *SysFan* and *JASP* to be utilized in research, particularly within systemic functional linguistics. The former is employed in a limited number of articles for annotation, including the annotation of grammatical metaphors, clauses, and words and phrases. The latter is frequently replaced by alternative software, including *SPSS*, *R*, or the *UAM Corpus Tool*, for statistical operations pertinent to the analysis of corpora. A larger corpus allows for a broader range of statistical possibilities and the identification of more subtle nuances, thereby enhancing the rigor and persuasiveness of the research findings. A comparative analysis of clause complexes in the theses of Chinese Master's and PhD students, as opposed to those from countries with disparate educational backgrounds, represents a rigorous investigation into the evolution of language and its intricacies across varying levels.

In Chapter Five, the quantitative analysis employs a statistical comparison of the choices presented by 37 indicators within and between the expansion-related subcategories as well as their intersections with taxis of over 30,000 clause complexes, as observed in Master and doctoral theses. The findings indicate that doctoral students tend to gravitate towards specific subcategories, including hypotactic enhancement, hypotactic extension, paratactic elaboration, and others. In contrast, Master's students prefer to use subcategories, including hypotactic elaboration, paratactic extension, paratactic enhancement, and so forth in text writing. Furthermore, the choice of subcategories within the system influences the choice of subcategories at the upper level, which is ultimately reflected in the distinction in lexical choice. For instance, doctoral students tend to utilize *by* and *as* to indicate manner: means relationship, *since* for the cause: reason relationship, *in* for the temporal: same time relationship, *though* for the condition: concession one, *as* for manner: comparison, *for* for cause: purpose, and so on in hypotactic enhancements, *whereas* to indicate the addition: adversative relationship in hypotactic extensions. In contrast, Master's students frequently employ the conjunction *and* to present the addition: additive relationship, and *but* for an addition: adversative one in paratactic extensions. They also often utilize *and then* to indicate temporal: later relationship in paratactic enhancements, and *that is to say* for exposition in paratactic elaborations, etc. In addition to the excessive use of *and* as a conjunction and *but*, the primary reason for the discrepancy is the limited range of other relations chosen by master's students. The qualitative analysis of the lexicogrammatical realizations of clause complexes can demonstrate the logico-semantic relationship presented by the lexis by examining the discrepancies in the choice of alternative terms that perform the same sub-function in the system.

In Chapter Six, similarly, through the quantitative statistical comparison of the choice from 27 indicators within and between the projection-related subcategories of the clause complex and the intersection of subcategories with taxis in the Master's and doctoral dissertations, it was found that doctoral students were more inclined to choose some subcategories such as hypotactic locution and paratactic locution, while Master's students more preferred to hypotactic idea and paratactic idea in texts writing. Moreover, the choice of subcategories within the system affects the choice of subcategories at the upper level, which is ultimately reflected in the difference in lexical choice. For example, doctoral students often use words such as *argue*, *point out*, *claim*, *report*, *suggest*, *assert*, etc. in hypotactic locutions, and *explain*, *propose* in paratactic locutions; whereas, Master's students often choose words such as *know*, *think*, *believe*, *mean*, *see*, *prove*, *find out*, *consider*, *realize*, *hope*, *understand* in hypotactic ideas. The probability of choosing paratactic locution and paratactic idea, especially the latter, is extremely low, resulting in a small gap between master and doctoral students in the four projection subcategories.

In general, about the choice of subcategories within the clause complex system as a whole, the distinction between the Master's group and the doctoral one is that the latter is more likely to exhibit a preference for hypotaxis, enhancement in expansion, hypotactic locution in hypotaxis, hypotactic extension in hypotaxis, hypotactic enhancement in hypotaxis, hypotactic enhancement in enhancement in texts writing; while the former has a higher probability in indexes such as idea in projection, extension in expansion, paratactic enhancement in parataxis, hypotactic idea in hypotaxis, hypotactic elaboration in hypotaxis, paratactic extension in extension. The corpus of written texts reveals a prevalence of

hypotaxis, with approximately twice the frequency of parataxis. Expansion is approximately four times as prevalent as projection, with approximately 70% of projections accounted for by ideas. Enhancement accounts for about 44% of expansions, extension accounts for approximately 41%, and elaboration accounts for only around 15%. In all paratactic relations, paratactic extension accounts for around 80%, with paratactic enhancement accounting for approximately 10%. In hypotactic relations, hypotactic enhancement accounts for approximately 51%, with hypotactic ideas accounting for approximately 21% and hypotactic elaboration accounting for nearly 15%. It is noteworthy that over 98% of ideas are hypotactic. Furthermore, approximately 90% of enhancement is hypotactic enhancement, 84% of locution is hypotactic locution, and 80% of elaboration is hypotactic elaboration. In contrast, 90% of extensions are the paratactic extension. These probabilities are contingent upon the intrinsic characteristics of the language itself. The range of sub-category choices within the clause complex system is unprecedented in previous papers. For example, projection accounts for 7% to 36%, while extension accounts for 64% to 93%. A high-probability statistical analysis of texts from other registers or genres would indicate that the final proportion of the two indexes will fall within this range. Considering the statistical methodologies employed in this study, most of the findings are derived from the Independent Samples T-test and the Bayesian Independent Samples T-test, which most effectively illustrate the statistical significance of the results from the Master's group and the doctoral group. Furthermore, the Bayesian factor analysis not only demonstrates the difference between the two groups in the corpus but also provides a measure of the degree of difference, which is not achievable by other statistical methods.

The delineation and proposal of four categories of additions and the potential emergence of additional categories during students' clause complex writing, which reflects students' ability levels, appear straightforward for understanding. However, they are useful in annotating clause complexes and evaluating and comparison. Without this step, the implicit and explicit structures within the context of the writing genre and language (clause complex specifically) structure intersections would be impossible to identify.

The annotation of the chosen corpus demonstrates that not all clause complexes are divided and exemplified by specific words or phrases, thus allowing for the differentiation between explicit clause complexes characterized by relationship markers and implicit clause complexes lacking such markers or exhibiting ambiguity in their relationship markers. This makes the process of annotating the grammatical category of a given clause complex more complicated. The rationale for the adoption of implicit clause complexes can be attributed to many factors, including the migration and transformation of the mother tongue, interlanguage or semi-discourse of English as a Foreign Language (EFL) learners; errors or mistakes in the production of language, whether intentionally or subconsciously; ellipsis in the context of description or explanation; and the intrinsic linguistic characteristics of the language itself. The implicit clause complex, representing a distinct form of clause complexity, exerts a multifaceted influence on researchers, learners, and educators alike. To address this issue, it is necessary to employ a range of methods, including logico-semantic agnation, paraphrasing, addition, deletion, substitution, and others, to facilitate comprehension, annotation, and production. A more complex division of the grammatical structure may facilitate the interpretation of the language is also noteworthy.

Concerning the general clause complex patterns, although only 500 clause complexes have been examined, similarities and differences (peculiarities) in the choice and usage of students in their theses can be identified. The preference for some patterns in both groups motivates the investigation of whether other texts in the same genre also exhibit these preferences or whether these patterns are common or universal in the English language, regardless of the context. This exemplifies the characteristics of recursion in language, which has not been the subject of extensive research to date.

The concept of "clause complex complexity," as distinguished from "syntactic complexity," is defined by its research object and statistical methods. A clause complex may be considered a sentence based on its structural appearance, but it is not a sentence in the conventional sense. It regards a clause as a minimum language unit and then combines clauses with connectives or projection markers independently or dependently. The clause complex also focuses on what functions clause combining plays in conveying meaning, illustrates the deployment of lexis and grammar within the same cline of the continuum, and explains language recursion as continuous clause combinations. The measurement of clause complex complexity is contingent upon the specific clause relations and clause layers in which they reside. This differs from the measurement of syntactic complexity, which focuses on clauses and T-unit ratios or indexes.

A more nuanced grasp of clause complex structures encourages educators to employ a more sophisticated approach to explaining and deconstructing in teaching academic English to EFL students. This includes helping students develop the ability to construct complex sentences and structures that effectively convey their thoughts, read and construe meaning from abstract and complex-seeming texts,

and identify linguistic inconsistencies, especially those in literary translations. Researchers in machine translation can dig out why the meanings of clauses cannot be understood and translated by computers equivalently. They can also explore ways to resolve this problem. Systemic functional grammarians can further research clause complexes to expand our understanding of language, society, and the mind.

There are discernible variations in academic English texts written by Chinese students at the master's and doctoral levels. They are not solely attributable to individual writing improvements and habits but also reflect the impact of educational enhancements and the cognitive absorption of reading materials. Given the disparate educational backgrounds of students at both the Master's and PhD levels, it is pertinent to ascertain whether personal improvements or educational resources are the predominant factors in generating statistically significant differences. Additionally, it is crucial to determine whether students in these two levels who major in other disciplines exhibit variations in writing clause complexes in English and whether these variations are reflected in the statistics presented in this research. Furthermore, it is essential to investigate whether students in these two levels from disparate countries demonstrate consistent performance as indicated by this study. Such discrepancies may be attributed to various factors, including language transfer of the mother tongue, educational background or educational incongruence in classes, personal initiative, and other potential causes. It would be advantageous to examine whether deficiencies can be eliminated by providing students with conscious guidance in writing. Additionally, it invites consideration of whether traditional grammar should be partially or entirely replaced by systemic functional approaches to language when educators select or implement training methods for their students.

In light of the potential for genres and registers to evolve, it is pertinent to consider whether this study continues to offer inspiration and utility in the present context. In other words, can the same methods be used to analyze spoken texts (presentations, seminars, workshops, daily life dialogues, etc.) or written texts such as essays comparisons over time, reports materials, reading feedback, application documents, etc., produced by students from these two levels or from Bachelors to Post-docs or even professors? Despite the absence of statistical analysis relating to the choice of clause complex subcategories across the various sections of the thesis text, the methodologies employed and the theoretical framework established in this study can be directly applied to genre research.

8.3 Implications (for Researchers, Translators, and Teachers)

In light of the contributions and limitations of this research, several points for future inquiry are identified here as warranting particular attention:

1. In the four editions of SFG books edited by M.A.K. Halliday and C.M.I.M. Matthiessen, how do the definitions and descriptions of clause complex categories evolve, and what are the underlying reasons for these changes? Does the concept of the "clause complex" retain its fundamental characteristics? Subsequently, it would be of interest to ascertain whether the concept "clause complex" is also employed in other related functional grammar theories and directions and whether the term "clause" is used in a manner that is consistent with the SFG approach. It would be beneficial to ascertain whether there are any other theories, excluding Functional Grammar, which regard structures beyond a clause and below a paragraph as unique and valuable grammar phenomena, akin to clause complexes.
2. Concerning the categories and divisions of the clause complex, it can be observed that different languages exhibit similarities and peculiarities compared to English. In addition to comparing the categories of expansions and projections with conjunctions and citations in Chinese, a similar typological comparison could be conducted within English and other languages in the same or different language families. The rationale for primarily contrasting English categories with Chinese ones stems from the referential significance of SFG as postulated by Halliday in his study of Chinese grammar, which drew upon the insights of scholars such as Li, Wang and Ming-kai, Kao in China. Furthermore, the concept of *Fuju* in Chinese grammar appears to be analogous to "clause complex" in SFG. However, Hu Zhuanglin and other researchers have advocated that there are significant differences between the two, with the Chinese concept encompassing both shared and distinct features.
3. In instances where the relationship between two clauses that are connected by markers such as *and*, *but*, or *while* is unclear, the analysis of the subcategories of metafunctions that a clause displays in a text can provide insight and facilitate annotation. For example, if two clauses have the same interpersonal or textual metafunction connected by *while*, the clause complex would be better annotated as an addition: adversative rather than an enhancement: conditional. The bidirectional analysis of clause metafunctions and clause complex categories is a far more comprehensive approach. It allows for a more delicate and complex theoretical framework for analyzing clause complexes in discourse, considering the interplay between grammar and lexis in determining the ranking (downranking especially) of choices.
4. Once more, concerning the theoretical framework of clause complexes in a discourse, finite and

non-finite clause divisions can be made in the expansion part, and clauses that are facts in the projection part can be given greater attention. This intends to determine whether they should be included in embedded clauses or are best regarded as a separate category encompassing several sub-classes, including cases, chances, proofs, needs, etc. However, as Halliday (2014, p. 547) notes, the question remains:

“What kind of projection is a fact? It is still a meaning, a semantic abstraction, not some third type differing from both meanings and from wordings (indeed, there is no third level to which it could belong). But it is not a meaning created in anybody’s consciousness, nor is it emitted by any signal source; it is simply got up so as to function as a participant in some other process - typically a relational process, but sometimes also a mental or a verbal one. Not, however, in a material process; facts cannot do things, or have things done to them.”

As embedded clauses can also be classified into two categories: expansion (extending, elaborating, and enhancing) and projection (locution and idea, propositions, and proposals) with intersections of finite and non-finite relations, can it be agnate with hypotaxis and parataxis?

5. An embedded clause may be regarded as a nominalization when the “fact” clauses function as the realization of an element (Carrier or Identifier/Token) in the relational process clause (attributive or identifying) or when the “act” clauses function as the Head or Postmodifier to a Head noun in relational clauses especially attributive ones where the attribute is an evaluative term or identifying ones where they are related to a nominalization or perception, either mental (inert perception) or behavioral (active perception) (Halliday, 2014, p. 504, 540, 549).

“Any clause that has the status ‘projected’ but without any projecting process is a fact and is embedded, either as a nominalization serving as Head or as Postmodifier to a ‘fact’ noun serving as Head. This includes some of those functioning in mental clauses, as mentioned above, and all projections functioning in relational clauses (since a relational process cannot project). It also includes ‘impersonal’ projections such as *it is said...*, *it is believed...*, *it seems...*, where the ‘process’ is not really a process at all, but simply a way of turning a fact into a clause.” (ibid, p. 549)

6. Projection and citation are two distinct concepts according to their research objects (clause relations and utterances/meanings quotation/paraphrase), unique annotation methods (meaning-based and structure-based), and the review of corresponding functions that they play in texts, i.e., providing the information, attributing a proposition to another person, indicating the origin of ideas, concepts or products, or simply identifying an agent within the sentence. However, some citations fall in the research scope of projection and also work the other way around. As both citation and projection annotations depend on the judgment of verbs and quotation marks, there is potential for integrating projection with citation to construct other theoretical frameworks for research. This approach allows for divisions such as integral vs. non-integral citations. Moreover, the investigation of projection in the context of English academic writing discourse analysis can be approached through the lens of citations.

7. In addressing the clause complex diversities inherent in academic discourse, it is essential to determine the most appropriate unit of analysis. This may entail examining the nuances of disciplines, gender, educational levels, chronological change, personal background, or behavior, or it may involve considering the intersection of all these factors. What specific differences should be the subject of discussion? What is the primary factor that influences students’ choices?

8. The potential outcomes of the clause “complex choices” in the context of texts authored by an individual for their Bachelor’s, Master’s, and Doctorate degrees warrant investigation. Additionally, the possibility of obtaining meaningful comparisons by collecting data from a larger sample size, such as 30 or more students, should be explored. This could provide a more robust basis for analysis by allowing for a more comprehensive examination of the texts across the three levels. Such a more scientific research design may yield more intriguing or even unanticipated results. What if the primary factor contributing to a student’s lack of change in their academic writing is the consistency of their educational background, namely, the same university or school and the same supervisor guiding their thesis? The collection of these texts may prove challenging, and it is anticipated that the choices made by students in these texts will not differ significantly from those identified in this research.

9. In the context of grammatical structure annotation, the reliability of the process is enhanced when three or more researchers are involved in the annotation of the same text, with the results of their annotations subsequently compared. Where a researcher is uncertain about the appropriate categorization of a particular structure, the involvement of multiple researchers can provide a higher level of reliability. However, the time and energy required for researchers to complete these tasks also result in increased financial costs. Additionally, the need to address and resolve conflicting opinions necessitates more time for discussion. Theoretically, the annotation of a single text by multiple researchers is a viable approach. Does the involvement of additional researchers enhance the reliability of the annotation process? Furthermore, to what extent can the reliability and validity of the annotations be enhanced, particularly in

terms of the measurement methods employed for comparison?

10. Specific markers, particularly those associated with expansion, are “multivalent,” meaning they can signify either elaboration and extension or extension and enhancement. For further details, please refer to Table 7.3.4. The optimal solution is to identify the nearest finite agnate clause, which means to examine the close agnates to ascertain whether they are exhibiting elaborations, extensions, or enhancements. (Halliday, 2014, p. 487) In the event that the relative clause in question is non-defining, the non-finite form is understood to be serving an elaborating function. In the event that the clause in question is coordinated, the non-finite verb indicates an act of extension. In the event that the clause is enhancing, the non-finite form is enhancing and may be introduced by a conjunctive preposition. (cf. *ibid*, p. 489) However, the specific conditions should be examined in greater detail for each marker, and subsequently, clause complexes can be annotated in accordance with the prevailing context.

11. The concept of “clause complex” has become a valuable tool in discourse analysis. However, if the auto- or semiauto-annotation of clause complexes cannot be achieved, large quantities of text would be either impossible or require a significant investment of time and resources. Therefore, it is imperative to address the challenge of how machines comprehend and construct meaning, which can inform the development of annotation software. Before this, it is essential to make clear how clause complexes should be construed in written, practical texts.

12. From the SysFan annotation interface, one can calculate the types and numbers of grammatical categories on each level. Some categories are highly adopted on levels 1 to 4, while others have never appeared on some levels, especially on levels 5 to 7. For example, hypotactic locutions or elaborations are not observed on these levels. Nevertheless, the requisite outcomes must be obtained through manual counting and calculation.

13. Further research is required to establish a comprehensive theoretical foundation and to define the scope of applying the concept of clause complex complexity.

14. What is the optimal clause-complexity calculating model for a text? The determining factor is the clause complex frames and weights of each category within these frames.

15. The frequency indicates the tendency of most students towards a particular choice, but it does not represent the absolute reality (exact number) of any given text. This implies that students invariably possess unique writing habits and may utilize non-standard expressions or exhibit a deviation from the expected frequency of certain linguistic structures. If an individual prefers the phrase *according to* instead of *somebody says*, the frequency of locution in this particular text may be less than that of others. Similarly, if an individual evinces a penchant for the phrase *the other way around* rather than *and*, the text may exhibit a higher probability of adversative rather than additive or addition. These preferences do affect the results of lexical calculation statistics.

16. The recursion of clauses in clause complexes represents an unexplored area concerning types and amounts of grammatical patterns at each level. This field of study could elucidate how writers organize structures to convey their thoughts and address why some patterns are exclusive to specific levels. The question of why certain pattern is preferred at certain levels also warrants further investigation.

17. The concept of “clause complex” originates in colloquial discourse analysis, which has demonstrated that “sentence” is more applicable to written discourse than to spoken texts, where it is not a particularly useful concept. A comparison of clause complexes in spoken and written registers reveals more intriguing findings.

8.4 Limitations of the Study

It is unavoidable that any research will be constrained by several factors, including the time and energy available to researchers and their synchronic and diachronic cognitive limitations. This is particularly the case for research conducted by younger scholars. It should be noted that some of the limitations discussed in this dissertation have been previously addressed in other chapters. To provide a detailed description:

Though this project is primarily concerned with the comparison of written texts from Master’s and PhD students to obtain statistical results, the texts or original corpus in question originate from different students at two distinct levels, i.e., a comparison of Master’s thesis and PhD dissertation from each student to investigate the individual’s improvement of differences in these two levels and ultimately conclude statistics derived from a sample of over 30 students can provide a more compelling representation. Consequently, this project offers a macroscopical angle on applying academic English in thesis writing.

The structures found in PhD dissertations are typically more complex than those observed in Master’s theses. However, exceptions exist, with some categories in some PhD dissertations appearing relatively straightforward. Nevertheless, these categories often employ more complex vocabulary and

nominalizations (as defined by Hyland, 2009a) to convey the same meaning. Nevertheless, the length of each clause and the number of clause complexes in each text, divided by the number of clause nexuses discovered in that text, can reasonably indicate the overall structure.

This dissertation presents three formulas for calculating the complexity of clause complexes in texts. It is, however, questionable whether these formulas truly reflect the comprehensive conditions that should be taken into account in the clause complex system. It should be noted that the formulas do not consider structures such as non-finite clauses, embedded clauses, nominalization, implicit clause complexes, and so forth. Therefore, the construction of the testing model requires further investigation and evaluation, as some structures may not be of significant importance or exert considerable influence on the calculation results due to their low frequency of usage.

In SFG, a clause is frequently examined from three distinct perspectives: interpersonal, ideational, and textual. These three metafunctions facilitate the classification of a clause into its constituent subcategories. A more detailed analysis of a clause complex, such as one that considers elements like *mood + residue* and *theme + rheme* would require a vast theoretical framework, presenting significant challenges for researchers. This does not imply that the three metafunctions of clauses in a clause complex are inconsequential. On the contrary, they can be instrumental in construing meaning in certain categories, such as implicit ones, additions, expansions, etc.

Due to the considerable volume of data, time constraints, and manual annotation, a strategy was devised whereby half of each text was selected for annotation, either interval clause complexes or the former/latter half of all clause complexes in a text. It remains unclear whether this strategy has any impact on the resulting research.

With annotation, some structures may be summarized and annotated in a manner that differs from the opinions or suggestions of others due to the inherent limitations of personal ability. It is possible to control the congruous annotation of the same structures from beginning to end. Indeed, issues may arise when there is semantic divergence, errors in writing, and mistakes on the writer's side, as well as when ambiguous or implicit clause complexes are encountered. Some issues may remain unresolved.

This research primarily employs an up-down (grammar-lexis) analysis of grammar in texts to demonstrate how meanings are realized through the choice of clause complexes, clauses, logico-semantic relationship markers, recursion times, and other linguistic elements. However, a down-up (grammar-discourse genre/register) analysis is also necessary to fully understand the English academic writing discourse. This will enable us to comprehend how clauses and clause complexes are organized into paragraphs, texts, and the whole discourse because of genre or register restrictions. This topic has been explored in depth by K. Hyland in his published works.

For the expansion of the field of study, this research can be an addition to existing research, not only in terms of grammatical structures but also in methodology and the availability of data drawn from a large corpus. However, regarding the projection aspect, it is impossible to compare the progress of the theories with Zenglei's works. Similarly, it is impossible to say that any practical or applied innovations are superior to the research conducted by certain scholars on applying citations in thesis writing. Although some clause complexes of projection are of structures traditionally regarded as citations, the relationships between them, the circumstances under which they overlap, and the potential for broadening the projection framework remain unaddressed in this research. In light of this overlap, it is also worth further discussion whether the framework of projection in academic writing discourse and even general discourse analysis should be changed and adapted for investigation.

The categorization and analysis of finite and non-finite clause complexes is a challenging endeavor, and as a result, they have not been a primary focus of this research. A similar observation can be made regarding the treatment of "facts" in projection and embedded clause complexes, which have only been marked without further discussion. Indeed, the relations of embedded clauses can be considered analogous to those of normal logico-semantic relations and thus can be calculated. Another question is whether the clause with embedded clause complexes should be regarded as a clause complex or a clause. Should clause complexes in embedded structures be counted as we have traditionally counted clause complexes, that is, as a clause complex formally regarded as a sentence? It is, therefore, pertinent to inquire whether they can be defined as clause nexuses. Given that embedded clauses originate from grammatical metaphor and nominalization, whether this particular structural type should be included in future measurements of clause complex complexity is worth considering.

When a marker such as *and*, *but*, and *while* can connect clauses with different relations and then be annotated as diverse categories as a clause complex, how might one condition them for further annotation? For example, the question of when to regard *and* as an addition: additive relation and when to treat it as a temporal: same time relation is an interesting and useful one for both linguists and learners alike.

Further research is to elucidate the distinctions and similarities between "clause complex complexity" and "syntactic complexity." This should entail the construction of a macroscopic and

microscopic theoretical framework, measurement standards for aforementioned complexity, the theoretical advancements and practical applications of clause complex complexity in discourse, and potentially related research fields.

Some significant theoretical inquiries remain unresolved. These include the diachronic evolution of the definition, description, scope, and markers of clause complexes across diverse editions of Systemic Functional Grammar (SFG) books. Another key area of investigation is the comparison of these elements within clauses and clause complexes, considering the varying perspectives of different SFG scholars and the diverse approaches within functional and traditional theoretical linguistics. Additionally, there is a need to examine Cardiff Grammar on clause and clause complex-related structures.

The grammatical classification of clause complexes in English can be compared with the classification of conjunctions and quotations in Chinese, as outlined in the following. Firstly, Halliday's grammar has its roots in Chinese grammar (for further details, see Hu, 2018, p. 328-40). Secondly, Chinese students will inevitably experience language transfer when writing. The types of applications of Chinese conjunctions are relatively few compared to English, which will have a certain impact on writers. Some writers do not pay sufficient attention to minor words or conjunctions, resulting in loose clause structures and the necessity of the reader to deduce the intended meaning in texts. Some writers even lose components in their clauses while writing, resulting in errors in the connection of clause complexes. Thirdly, writers are unfamiliar with the existence and usage of English conjunctions, so they cannot connect clauses into clause complexes. As a result, the analysis of clause complex annotations in texts becomes more challenging, the results of corresponding lexical and grammatical statistical research are affected, and it becomes evident that writers and teachers do not prioritize clause connection in text writing.

Statistical data can't demonstrate and present every detail, including those resulting from personal habits. However, it can at least explain or illustrate the tendency of preference, diversities, and changes. Research methods that deal with a large corpus can, to some extent, provide enlightenment.

In grammatical analysis, how might two clauses be certified at the same level if they are analyzed layer by layer, from levels 1 to 7? Typically, clauses marked by addition are at the same level and are analogous to two hypotactic (dependent) clauses in two-clause nexuses that are also connected as an additional relation. Another avenue of inquiry is the calculation of grammatical structures and varieties at each level, which is beneficial for translation and machine translation.

Although manual annotation has significant limitations, it remains the optimal choice for clause complex annotation, given the current state of software online. These software programs are unable to fulfill this task based on meaning understanding. For new software (while they can be designed for this specific purpose), as long as the fundamental challenge of understanding the meaning of clause complexes by machine or computer remains unresolved, investing resources in developing software with multiple layers is an inefficient use of time and resources.

As Thompson (2005, p. 321) concludes: "... student writers should look not only at the formal features of citation types but should also consider the implications of different choices, both at a local level (the sentence and the paragraph) and also at a higher discursive level (what to give prominence to, how to maintain the dominance of the authorial voice throughout the text, and how to position oneself in relation to the immediate audience and to the wider disciplinary community)." These scenarios are not exclusive to examining projections relevant to the citation; they also denote the analysis of expansions from the clause complex system. It would, therefore, be beneficial to investigate the role of these lexicogrammatical structures at both the local and discourse levels.

8.5 Suggestions (for Theories, Translation, and Teaching)

The applications and implications of this research have been previously discussed in various academic contexts. They are particularly evident in the following areas: writing, text reading, classroom teaching (clause relationships), text translation (basic cultural comparison), machine translation (computational linguistics), and theoretical development.

When faced with the challenge of determining the relationship between clauses in a clause connection/combination in **writing**, whether two clauses are hypotactically or paratactically placed, whether the expansion relationship is conveyed through addition, extension, or enhancement, and whether the projected contents are considered a locution or an idea in the projection clause complexes, or whether these projected contents are presented as propositions or proposals in projection, the qualitative analysis of the framework of the clause complex category in this thesis can provide insight. When expressing the same meaning, choosing conjunctions, prepositions, verbs, phrases, or connecting clauses in other ways (such as without any obvious markers) is a necessary step in the writing process and this study's focus. In citing literature, it is essential to consider how to convey the author's stance on the

materials referenced and whether to anticipate specific expectations from readers with particular tendencies, such as by projecting propositions or proposals. This study partially addressed the qualitative analysis of the interpersonal function of projecting clauses. Regarding the choice of projecting verbs, such as *suggest*, *demonstrate*, and *show*, which all signify a kind of presentation or indication, a considerable number of writers maintain that these three can be interchanged due to the similarities in their semantics, structure, and function within the clause(s). However, when they are used to demonstrate interpersonal interaction between the writer and the author or the reader, the presented state is also closely related to the context in which they are situated; thus, a generalization of them or lumping them together is not so suitable. Despite the apparent delicacy of the taxonomy of projecting clauses and their relationships in systemic functional linguistics, some general perspectives still cannot be applied when analyzing texts for specific lexical semantic purposes. The examination of implicit clause complexes in this thesis facilitates a deeper understanding of the transfer and other influences from a writer's mother tongue when writing in English, particularly for non-native English speakers. For instance, in addition to the unequal command of the polysemy of English conjunctions or prepositions when connecting clause complex structures in academic English writing by Chinese native speakers, the disparity between these words and the related structural connection words in Chinese represents a significant challenge in lexical selection. Another challenge for writers is the need to make equivalent choices between the limited types of Chinese projecting verbs and the numerous English projecting verbs.

This study contributes to text **reading** through its in-depth analysis and discussion of language meaning construing. Given that English is not the native language of Chinese students, it is unavoidable that they will encounter difficulties when judging and understanding the relationship between clauses in reading texts. In the condition of having an evident linking marker (even of polysemy), the relationship between clauses can be distinguished. However, in the absence of such a word, the evaluation of the relationship between clauses becomes challenging. This issue has been repeatedly highlighted in the study of implicit clause complexes. Students will inevitably feel confused regarding the structure of clauses in a lack of relevant knowledge. In alignment with the writing process, the behavior of reading text also necessitates an understanding of the author's stance and perspective, particularly in the context of projecting propositions or proposals. The author's affirmation, denial, or neutral stance towards the projected content and their expectations regarding the reader's response are all shaped by the choice of the projecting verbs/clause structures and the meaning of other components within the clause. The lack of knowledge regarding this information results in a puzzled reading of the text. Furthermore, it is essential to discern and comprehend the implicit meaning conveyed by the writer. Implicit meaning may be expressed through any grammatical form, including clause complexes. In contrast, as to implicit grammar, the absence of explicitness in structure does not necessarily imply the absence of implicitness in meaning, and vice versa, i.e., the implicitness of meaning may not be presented in implicit structures. When reading a text, it is valuable for readers to consider whether the writer is expressing implicit meaning or whether he/she is expressing explicit or implicit meaning through implicit grammar. To illustrate, the clauses connected by the conjunction *and* can be interpreted as addition, enhancement, or extension relationships. They can also be understood to express relationships of temporal, spatial, or any non-time and non-space relationships. In addition, they can be viewed as substitutes for other relationships and may even become redundant. This belongs to the domain of implicit grammar.

In the field of **teaching**, in addition to improving the theoretical awareness of clause complexes and strengthening the knowledge level of application areas, it is more important for teachers to integrate all aspects of clause complexes involved in the teaching of writing and reading into practice, which will have a subtle influence on students, thereby helping them to understand the meaning of language and express it appropriately. In addition, taking a systematic perspective on the choice of lexis and grammar and teaching it to students can make the construction of the framework of the knowledge system a reality instead of the application of scattered and trivial language knowledge.

When **translating** texts, this study has a certain guiding significance for the equivalent translation of clause complex connection relations, despite that the relevant connection words have not been analyzed in detail from the perspective of language comparison, such as English-Chinese, but only mentioned roughly. However, several strategies, such as adding/deleting/substituting words or phrases, have been proposed for the non-equivalent translation of connection/combination relations. On the one hand, due to the characteristics of the language itself, the clause connection relations in one language may not exist in another language, or even if they do exist, they express different meanings or realize diverse functions. On the other hand, because of the peculiarities of the style conveyed by the language of the translated text, such as literary texts, it is worth exploring whether the connection relations of the clauses that are not obvious (whether they are intentional or unintentional expressions of the author) need to be revealed in another language. The identification/clarification and translation of implicit clause complexes are also problems for translators. If translators are unfamiliar with the relationship between

clause complexes, the translated meaning is bound to be easily misinterpreted, which will also cause some confusion to the readers. When translating projecting clause complexes, the choice of equivalent projecting verbs and the expression of attitude meanings are also challenges for translators. Considering the expression of meanings such as 表明(biao ming), whether to choose words such as *demonstrate*, *indicate*, *show*, *suggest*, or uncommon ones like *present*, *manifest*, *testify* in the Chinese-English translation depends not only on the context within and between clause complexes but also on the writer's writing preferences and style between the lines. For this reason, "Davies claims that contextual frames, realized by projecting clauses, are an important feature in helping to establish the interactional nature of a text, and allowing a writer to express viewpoint both directly and indirectly." (cf. Forey, 2009, p. 167) Thus, the seemingly basic and simple expression of meaning contains rich semantic and pragmatic functional choices.

In **computer-assisted translation**, a key challenge is the ability of computers to recognize and comprehend implicit clause complexes, which are a fundamental aspect of natural language. Given that the current research on implicit clause complexes is still in its infancy and that computer recognition depends largely on relevant knowledge and research, this study has not further addressed the issue. Another issue is computational research on clause complexes. This includes whether the theoretical categories of clause complexes can be automatically annotated and, if so, what language and manner of annotation software should be used. This is contingent upon the taxonomy of categories and the optimization of their theoretical analysis framework on the one hand, and the selection of annotation categories and the difficulty of software development on the other. While there is a consensus regarding the acceptability of automatic syntactic analysis based on context-free grammar and probabilistic grammar, the study of regularity and its translation in the context of clause complex connections and recursion has yet not been addressed. It is necessary to consider whether the probability of selecting the clause complex grammatical categories within and between systems can be used to assist computers in translating clause complex connections and combinations. Moreover, further investigation is required to ascertain the probability of selecting relevant categories varying across a diverse range of texts, encompassing different registers or genres. Halliday & Matthiessen (2004, p. 35) also point out that "[c]omputerized parsing and pattern-matching is now reaching the point where quantitative studies can be undertaken of a number of primary systems in the grammar, using samples large enough to permit comparison among different registers... Not enough work has yet been undertaken along these lines for us to build it into the total picture; but it is a high priority field for future research." Concerning text complexity, only a few researchers have delved into utilizing computer technology to calculate the complexity of clause complexes within a text, intending to distinguish or evaluate the text from a functional grammatical standpoint rather than from the perspective of the t-unit.

In general, this study contributes to the research on clause complexes within the framework of systemic functional grammar. It enhances the coherence of theoretical and applied aspects of clause complex theory. Indeed, the mixed methodology employed in this study is also applicable to investigate other grammatical categories within the framework of systemic functional grammar, both within and among systems.

REFERENCES

- Adjei, A. A., Opoku, K. (2017). The expansion relations of clause complexing (CC) in the editorials of the *Daily Graphic*. *Journal of Literature, Languages and Linguistics*, 30, 12-26.
- Agbaglo, E. (2017). The types and the frequencies of reporting verbs in research articles written by lecturers in a Ghanaian university. *Journal of Literature, Languages and Linguistics*, 34, 51-57.
- Agbaglo, E., Bonsu, E. M. (2022). Functions of reporting verbs in the literature review of Master's theses in the discipline of economics. *Journal of English Language Teaching*, 9(2), 347-359.
- Aktas, R., Cortes, V. (2008) Shell nouns as cohesive devices in published and ESL student writing. *Journal of English for Academic Purposes*, 7, 3-14.
- Allison, D., Cooley, L., Lewkowicz, J., & Nunan, D. (1998). Dissertation writing in action: The development of a dissertation writing support program for ESL graduate research students. *English for Specific Purposes*, 17(2), 199-217.
- Altenberg, B. (1986). Contrastive linking in spoken and written English. In Tottie, G. & Baecklund, I. (eds.), *English in speech and writing*. Stockholm: Almqvist & Wiesel International, pp. 13-40.
- Andersen, T. H., Holsting, A. (2018). Clause complexing in systemic functional linguistics: Towards an alternative description. *Functional Linguistics*, 5(10), 1-25.
<https://doi.org/10.1186/s40554-018-0059-7>
- Anderson, C. (2010). Presenting and evaluating qualitative research. *American Journal of Pharmaceutical Education*, 74, 1-7.
- Andriani, N. (2019). *The logico-semantic relation of clause complexes in the VOA News*. MA Thesis. Medan: University of Muhammadiyah Sumatera Utara.
- Anggara, P. D. (2015). *Tactic system and logico-semantic relations of clause complexes in ASEAN Free Trade Area (AFTA) Agreement*. Dian Nuswantoro University.
- Anggraeni, H. S., Farikah, F. & Indriani, L. (2018). An analysis of parataxis and hypotaxis construction in descriptive text written by first graders of smk bandar sungai buaya malaysia in the year 2017. *Journal of Research on Applied Linguistics Language and Language Teaching*, 1(1), 86-95.
- Armstrong, E. M. (1992). Clause complex relations in aphasic discourse: A longitudinal case study. *Neurolinguistics*, 7(4), 261-275.
- Atkinson, D. (1999). *Scientific discourse in sociohistorical context*. Mahwah, NJ: Lawrence Erlbaum.
- Babbie, E. R. (2010). *The practice of social research*. 12th ed. Belmont, CA: Wadsworth Cengage.
- Bakhtin, M. M. (1981). *The dialogic imagination: Four essays* (translated by C. Emerson & M. Holquist). Austin: University of Texas Press.
- Bakhtin, M. M. (1986). *Speech genres and other late essays*. Austin, TX: University of Texas Press.
- Baldry, A., Thibault, P. (2006). *Multimodal transcription and text analysis*. London: Equinox.
- Bardovi-Harlig, K. (1992). A second look at T-unit analysis: Reconsidering the sentence. *TESOL Quarterly*, 26, 390-395.
- Bardovi-Harlig, K., Bofman, T. (1989). Attainment of syntactic and morphological accuracy by advanced language learners. *Studies in Second Language Acquisition*, 11 (1), 17-34.
- Barghamadi, M. (2021). Reporting verbs in the humanities and medical sciences research articles. *Language Teaching Research Quarterly*, 22, 17-32.
- Barton, D. (1994). *Literacy: An introduction to the ecology of written language*. Oxford: Blackwell.
- Barton, D., Hamilton, M. (1998). *Local literacies*. London: Routledge.
- Bartsch, S., Eckart, R., Holtz, M. & Teich, E. (2005) Corpus-based register profiling of texts from mechanical engineering. In *Proceedings of the Corpus Linguistics Conference CL 2005* (ISSN 1747-9398), University of Birmingham, UK 14-17 July 2005,
<https://www.birmingham.ac.uk/documents/college-artslaw/corpus/conference-archives/2005-journal/languageprocessingandcorpustool/corpusbasedregisterprofiling.pdf>
- Bastin, G. L. (1998). Adaptation. In: Baker, M. (ed.). *Routledge encyclopedia of translation studies*. London & New York: Routledge.
- Bazerman, C. (1988). *Shaping written knowledge*. Madison, WI: University of Wisconsin Press.
- Becher, V. (2011). Von der Hypotaxe zur Parataxe: Ein Wandel im Ausdruck von Konzessivität in neueren populärwissenschaftlichen Texten. In, Breindl, E., Ferraresi, G., Volodina, A. (eds.) *Satzverknüpfungen. Zur Interaktion von Form, Bedeutung und Diskursfunktion*, pp. 181-209. Berlin: de Gruyter.
- Bednarek, M. (2008). *Emotion talk across corpora*. Basingstoke/New York: Palgrave Macmillan.
- Bednarek, M. (2009). Language pattern and attitude. *Functions of Language*, 16(2), 165-192.

- Bednarek, M. (2010). Corpus linguistics and systemic functional linguistics: Interpersonal meaning, identity and bonding in popular culture. In Bednarek, M. & Martin, J. R. (eds.), *New discourse on language: Functional perspectives on multimodality, identity, and affiliation*. Continuum. pp. 237-266.
- Beers, S., Nagy, W. (2007). Syntactic complexity as a predictor of adolescent writing quality: Which measures? Which genre? *Reading and Writing: An Interdisciplinary Journal*, 22(2), 185-200.
- Belcher, D. (2004). Trends in teaching English for Specific Purposes. *Annual Review of Applied Linguistics*, 24(1), 165-186.
- Belcher, D. (2007). Seeking acceptance in an English-only research world. *Journal of Second Language Writing*, 16(1): 1-22.
- Bell, D. (2007). Sentence-initial *and* and *but* in Academic Writing. *Pragmatics*, 17(2): 183-201.
- Berkenkotter, C., Huckin, T. N. (1995). *Genre knowledge in disciplinary communication: Cognition/culture/power*. Hillsdale, NJ: Lawrence Erlbaum.
- Berman, R. A., Nir-Sagiv, B. (2004). Linguistic indicators of inter-genre differentiation in later language development. *Journal of Child Language*, 31(2), 339 - 380.
- Berry, M. (1975). *An Introduction to systemic linguistics: structures and systems* (Vol. I). London: Batsford.
- Berry, M. (2013a). Towards a study of the differences between formal and written English and informal spoken English. In Fontaine, L., Bartlett, T., & O'Grady, G. (Eds.), *Systemic functional linguistics* (pp. 365-383). <http://dx.doi.org/10.1017/cbo9781139583077.022>
- Berry, M. (2013b). Contentful and content light subject themes in informal spoken English and formal written English. In O'Grady, G., Fontaine, L., & Bartlett, T. (Eds.), *Choice in language: Applications in text analysis* (pp. 243-268). London: Equinox.
- Berry, M., Butler, C., Fawcett, R. P. & Huang, G. W. (1996). *Meaning and form: Systemic functional interpretations: Studies for M.A.K. Halliday*. Norwood, N. J.: Ablex.
- Bhatia, V. (2002). A generic view of academic discourse. In Flowerdew, J. (ed.), *Academic discourse* (pp. 21-39). Harlow: Longman.
- Bhatia, V. (2008). Towards critical genre analysis. In Bhatia, V. & Jones, F. (eds.), *Advances in discourse studies*. London: Routledge.
- Biber, D. (1993). Representativeness in corpus design. *Literacy and linguistic computing*, 8(4): 243-57.
- Biber, D. (2006). *University language: A corpus-based study of spoken and written registers*. Amsterdam: Benjamin, pp. 133-176.
- Biber, D. et al. (1999) *Longman grammar of spoken and written English*. Harlow, England: Longman.
- Biber, D., Conrad, S. (2009). *Register, genre and style*. Cambridge: Cambridge University Press.
- Biber, D., Conrad, S., Cortes, V. (2004). If you look at. . .: Lexical bundles in university teaching and textbooks. *Applied Linguistics*, 25 (3), 371-405.
- Biber, D., Conrad, S., Repper, R. (1998). *Corpus linguistics: Investigating language structure and use*. UK: Cambridge University Press.
- Biber, D., Finegan, E. (1988). Adverbial stance types in English. *Discourse Processes*, 11(1), 1-34. <http://dx.doi.org/10.1080/01638538809544689>
- Biber, D., Finegan, E. (1989). Styles of stance in English: Lexical and grammatical marking of evidentiality and affect. *Text*, 9(1), 93-124. <http://dx.doi.org/10.1515/text.1.1989.9.1.93>
- Biber, D., Gray, B., & Poonpon, K. (2011). Should we use characteristics of conversation to measure grammatical complexity in L2 writing development? *TESOL Quarterly*, 45(1), 5-35. <https://doi.org/10.5054/tq.2011.244483>
- Biber, D., Gray, B., & Staples, S. (2014). Predicting patterns of grammatical complexity across language exam task types and proficiency levels. *Applied Linguistics*, 37(5), 639-668. <https://doi.org/10.1093/applin/amu059>
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Harlow: Pearson.
- Bisiada, M. (2013). Changing conventions in German causal clause complexes: A diachronic corpus study of translated and non-translated business articles. *Language in Contrast*, 13(1), 1-27.
- Bloch, J. (2010). A concordance-based study of the use of reporting verbs as rhetorical devices in academic papers. *Journal of Writing Research*, 2(2), 219-244.
- Bloch, J., Chi, L. (1995). A comparison of the use of citations in Chinese and English academic discourse. In Belcher, D. & Braine, G. (Eds.), *Academic writing in a second language: Essays on research and pedagogy* (pp. 231-273). Norwood, NJ: Ablex.
- Blommaert, J. (2005). *Discourse*. Cambridge: Cambridge University Press.
- Bloor, T., Bloor, M. (1995). *The functional analysis of English: A Hallidayan approach*. London: Arnold.
- Blue, G. M. (1988). Individualising academic writing tuition. In Robinson, P. C. (Ed.), *Academic writing: Process and product*. (pp. 129-148). London: MET and British Council.

- Blühdorn, H. (2008). Subordination and coordination in syntax, semantics, and discourse: Evidence from the study of connectives. In Fabricius-Hansen, C., & Ramm, W. (Eds.), *'Subordination' versus 'coordination' in sentence and text: A cross-linguistic perspective* (pp. 59-88). Amsterdam: Benjamins.
- Borg, E. (2000). Citation practices in academic writing. In Thompson, P. (Ed.), *Patterns and perspectives: Insights into EAP writing practice* (pp. 26-42). Reading, UK: Centre for Applied Language Studies.
- Bowen, N. (2018). *Expansion, enhancement, and extension in academic writing*.
<https://doi.org/10.13140/RG.2.2.16478.59204>
- Brémont, C. (1970). Morphology of the French folktale. *Semiotica*, 2(3), 247-276.
- Bresnan, J. (2001). *Lexical-functional syntax*. Oxford: Blackwell.
- Brett, P. (1994). A genre analysis of the results section of Sociology articles. *English for Specific Purposes*, 13(1), 47-59.
- Brew, C., Moens, M. (2004). *Data-intensive linguistics*. The University of Edinburgh.
- Brians, C. L. et al. (2011). *Empirical political analysis: Quantitative and qualitative research methods*. 8th ed. Boston, MA: Longman.
- Brieger, N., Comfort, J. (1996). *Developing business contacts*. Beijing: Foreign Language Teaching and Research Press.
- Brooks, T. A. (1986). Evidence of complex citer motivations. *Journal of the American Society for Information Science*, 37(1), 34-36.
- Brown, P., Levinson, S. (1987). *Politeness: some universals in language usage*. Cambridge: Cambridge University Press.
- Bruce, N. J. (1983). Rhetorical constraints on information structure in medical research report writing. Paper presented at *the ESP in the Arab World Conference*, University of Aston, UK, August, 1983.
- Buckingham, J., Neville, M. (1997). A model of citation options. *Australian Review of Applied Linguistics*, 21(2), 51-66.
- Bulté, B., Housen, A. (2012). Defining and operationalising L2 complexity. In Housen, A., Kuiken, F., & Vedder, I. (Eds.), *Dimensions of L2 performance and proficiency: Investigating complexity, accuracy and fluency in SLA* (pp. 21 - 46). Amsterdam: John Benjamins.
- Bunton, D. (2002). Generic moves in Ph.D. thesis introductions. In Flowerdew, J. (ed). *Academic Discourse*. Harlow: Pearson Education Limited.
- Bunton, D. (2005). The structure of Ph.D. conclusion chapters. *Journal of English for Academic Purposes*, 4(3), 207-24.
- Butler, C. S. (2003a) *Structure and function: A guide to three major structural-functional theories (Part 1), approaches to the simplex clause*. John Benjamins.
- Butler, C. S. (2003b) *Structure and function: A guide to three major structural-functional theories (Part 2), from clause to discourse and beyond*. John Benjamins.
- Byrne, D. (1979). *Teaching writing skills*. London: Longman.
- Byrnes, H., Maxim, H. H., & Norris, J. M. (2010). Realizing advanced foreign language writing development in collegiate education: Curricular design, pedagogy, assessment. *Monograph of the Modern Language Journal*, 94 (Suppl. 1).
- Cadman, K. (1997). Thesis writing for international students: A question of identity? *English for Specific Purposes*, 16 (1), 3-14.
- Caffarel, A., Martin, J. R., Matthiessen, C. M. I. M. (2004). Introduction: Systemic functional typology. In Caffarel, A., Martin, J. R., Matthiessen, C. M. I. M. (eds.), *Language typology: a functional perspective*, pp. 1-76. Amsterdam: Benjamins.
- Campbell, C. (1990). Writing with others' words: using background reading text in academic compositions. In Kroll, B. (ed.), *Second language writing: Research insights for the classroom*. Cambridge: Cambridge University Press. pp. 211 - 230.
- Capra, F. (1996). *The web of life: a new synthesis of mind and matter*. London: Harper Collins.
- Carla, V. (2011). Shades of impersonality: Rhetorical positioning in the academic writing of Italian students of English. *Linguistics and Education*, 22(2), 118-132.
- Carrió-Pastor, M. L. (2013). A contrastive study of the variation of sentence connectors in academic research. *Journal of English for Academic Purposes*, 12(3), 192-202.
- Casanave, C. (1994). Language development in students' journals. *Journal of Second Language Writing*, 3(3), 179 - 201.
- Case, D. O., Higgins, G. M. (2000). How can we investigate citation behavior? A study of reasons for citing literature in communication. *Journal of the American Society for Information Science*, 51(7), 635-645.
- Celce-Murcia, M., Larsen-Freeman, D. (1983). *The grammar book: An ESL/EFL teacher's course*. Japan: Newbury House. pp. 323-334.

- Chafe, W. (1986). Evidentiality in English conversation and academic writing. In Chafe, W. & Nichols, J. (eds.). *Evidentiality: The linguistic coding of epistemology*. Norwood, NJ: Ablex.
- Charisma, Y. (2013). *Functional grammar analysis of clause complexes in the journal articles written by native speakers and non-native speakers*. Muria Kudus University.
- Charles, M. (2003) 'This mystery ...': A corpus-based study of the use of nouns to construct stance in theses from two contrasting disciplines. *Journal of English for Academic Purposes*, 2(4), 313-26.
- Charles, M. (2006a). Phraseological patterns in reporting clauses used in citation: a corpus-based study of theses in two disciplines. *English for Specific Purposes*, 25(3), 310-331.
- Charles, M. (2006b). The construction of stance in reporting Clauses: A Cross-disciplinary Study of Theses. *Applied Linguistics*, 27(3), 492-518.
- Charles, M. (2007) Argument or evidence? Disciplinary variation in the use of the *noun that* pattern. *English for Specific Purposes*, 26(2), 203-18.
- Charles, M. (2012). "Proper vocabulary and juicy collocations": EAP students evaluate do-it-yourself corpus-building. *English for Specific Purposes*, 31(2), 93-102.
- Chen, Muyu. (2023). An error analysis of coordinating conjunction misuse in Chinese ESL learners' writings: A corpus-based approach. *English Language and Literature Studies*, 1(1), 33-43.
- Chen, S., Xuan, W. W., & Yu, W. (2022). Beyond reporting verbs: Exploring Chinese EFL learners' deployment of projection in summary writing. *Sage Open*, 12(2), 1-12.
<https://doi.org/10.1177/21582440221093356>
- Cheng, A. (2006). Analyzing and enacting academic criticism: The case of an L2 graduate learner of academic writing. *Journal of Second Language Writing*, 15(4), 279-306.
- Cheng, A. (2008). Individualized engagement with genre in academic literacy tasks. *English for Specific Purposes*, 27 (4), 387-411.
- Chiu, Yuan-Li (Tiffany). (2016). 'Singing your tune': Genre structure and writer identity in personal statements for doctoral applications. *Journal of English for Academic Purposes*, 21, 48-59.
- Chomsky, N. (1957). *Syntactic structures*. The Hague: Mouton.
- Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: The MIT Press.
- Chovanec, J. (2003). The mixing modes as a means of resolving the tension between involvement and detachment in news headlines. *Brno Studies in English*, 29(1), 51-66.
- Christian, G. (2004). *Analytical chemistry*. U.S.: John Wiley & Sons.
- Christie, F., J. R. Martin (eds.). (1997). *Genres and institutions: Social processes in the workplace and school*. London: Cassell.
- Chubin, D. E., Moitra, S. D. (1975). Content analysis of references: adjunct or alternative to citation counting? *Social Studies of Science*, 5(4), 423-441.
- Clachar, A. (2003). Paratactic conjunctions in Creole speakers' and ESL learners' academic writing. *Word Englishes*, 22(3): 271-289.
- Cloran, C. (2010). Rhetorical unit analysis and Bakhtin's chronotype. *Functions of Language*, 17(1), 29-70. <http://dx.doi.org/10.1075/fo1.17.1.02clo>
- Cloran, C., Stuart-Smith, V., & Young, L. (2007). Models of discourse. In Hasan, R., Matthiessen, C. M. I. M. & Webster, J. J. (Eds.), *Continuing discourse on language: A functional perspective* (vol. 2, pp. 646-670). London: Equinox.
- Coates, J. (1983). *The semantics of modal auxiliaries*. London & Canberra: Croom Helm.
- Coffey, B. (1984). ESP-English for specific purposes. *Language Teaching*, 17 (1), 2-16.
- Coffin, C. (2009). Incorporating and evaluating voices in a film studies thesis. *Writing & Pedagogy*, 1(2), 163-193.
- Coffin, C., Donohue, J. (2012). Academic literacies and systemic functional linguistics: how do they relate. *Journal of English for Academic Purposes*, 11(1), 64-75.
- Coffin, C., O'Halloran, K. (2006). The role of appraisal and corpora in detecting covert evaluation. *Functions of Language*, 13(1), 77-110.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. 2nd ed. Lawrence Erlbaum Associates, Publishers.
- Collins English Dictionary*. 13th edition. (2018). Harper Collins Publishers. See <https://www.collinsdictionary.com/us/dictionary/english/tacit>
- Connor, U. (1996). *Contrastive rhetoric: Cross-cultural aspects of second language writing*. New York: Cambridge University Press.
- Connor, U. (2011). *Intercultural rhetoric in the writing classroom*. Michigan: University of Michigan Press.
- Coombs, V. M. (1986). Syntax and communicative strategies in intermediate German composition. *The Modern Language Journal*, 70(2), 114-124. <https://doi.org/10.1111/j.1540-4781.1986.tb05253.x>

- Cooper, T. C. (1976). Measuring written syntactic patterns of second language learners of German. *The Journal of Educational Research*, 69(5), 176-183.
- Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes*, 23(4), 397-423.
- Couture, B. (1986). *Functional approaches to writing: Research Perspective*. Norwood, NJ: Ablex.
- Craig, T. (1882). *A treatise on projections*. University of Michigan Historical Math Collection.
- Creswell, J. W. (1994). *Research design: Quantitative and qualitative approaches*. U.S: Sage.
- Creswell, J. W., Clark, V. L. P. (2011). *Designing and conducting mixed methods research*, 2nd Edition. SAGE Publications.
- Creswell, J. W., Clark, V. L. P. (2017). *Designing and conducting mixed methods research*, 3rd Edition. SAGE Publications.
- Crewe, W., Wright, C. & Leung, M. W. K. (1985). Connectives: on the other hand, who needs them, though? *Working Papers in Linguistics and Language Teaching*, (8), 61-75.
- Cristofaro, S. (2003). *Subordination*. Oxford: Oxford University Press.
- Crombie, W. (1985a). *Discourse and language learning: a relational approach to syllabus design*. Oxford: Oxford University Press.
- Crombie, W. (1985b). *Process and relation in discourse and language learning*. Oxford: Oxford University Press.
- Crowhurst, M. (1980). Syntactic complexity in narration and argument at three grade levels. *Canadian Journal of Education*, 5(1), 6-13.
- Cureton, E. E. (1956). Rank-biserial correlation. *Psychometrika*. 21(3), 287-290.
- Curry M. J., Lillis T. (2010). Academic research networks: Accessing resources for English-medium publishing. *English for Specific Purposes*, 29(3), 281-295.
- Damanik, E. S. D. (2018). Use of expansion and projection in research paper. *Journal of Language, Literature & Education*, 13(13).
- Davidse, K., Vandelanotte, L. (2011). Tense use in direct and indirect speech in English. *Journal of Pragmatics*, 43(1), 236-250.
- Davies, F. (1988). Reading between the lines: thematic choice as a device for presenting writer viewpoint in academic discourse. *The ESPecialist*, 9(2), 173-200.
- Davies, F. (1994). From writer roles to elements of text: interactive, organisational and topical. In Barbara, L., Scott, M. (eds.) *Reflections on language learning*, pp. 170-83. Clevedon, England: Multilingual Matters.
- Davies, F. (1997). Marked Theme as a heuristic for analysing text-type, text and genre. In Pique, J., Viera, D. (eds.), *Applied languages: Theory and practice in ESP*, pp. 45-71. Valencia: Servei de Publicacions Universitat de Valencia.
- de Magalhães, M. B., Cotterall, S., & Mideros, D. (2018). Identity, voice and agency in two EAL doctoral writing contexts. *Journal of Second Language Writing*, 43, 4-14.
- Declerck, R. (1983). *The structure of infinitive perception verb complements in a transformational grammar*. New York: Plenum Press.
- Denzin, N. K., Lincoln, Y. S. (2000). *Handbook of qualitative research*. 2nd edition. Thousand Oaks, CA: Sage.
- Denzin, N. K., Lincoln, Y. S. (2005). Introduction: The discipline and practice of qualitative research. In Denzin, N. K., Lincoln, Y. S. (eds.), *The Sage handbook of qualitative research*. 3rd edition. Thousand Oaks, CA: Sage.
- Dik, S. C. (1968). *Coordination: its implications for the theory of general linguistics*. Amsterdam: North-Holland.
- Dik, S. C. (1978). *Functional grammar*. North-Holland Linguistic Series, 37. Amsterdam: North-Holland Publishing Company.
- Dik, S. C. (1989). *The theory of functional grammar, Part I: The structure of the clause*. Dordrecht & Providence, RI: Foris.
- Dik, S. C. (1997a). *The theory of functional grammar, Part I: The structure of the clause*. Berlin, New York: De Gruyter Mouton. <https://doi.org/10.1515/9783110218367>
- Dik, S. C. (1997b). *The theory of functional grammar, Part 2: complex and derived constructions*. Berlin, New York: De Gruyter Mouton. <https://doi.org/10.1515/9783110218374>
- Djonov, E. (2005). *Analysing the organisation of information in websites: From hypermedia design to systemic functional hypermedia discourse analysis*. Sydney: University of New South Wales. Unpublished doctoral thesis.
- Dong, Y. R. (1996). Learning how to use citation for knowledge transformation: non-native doctoral students' dissertation writing in science. *Research in the Teaching of English*, 30(4), 428-457.

- Dontcheva-Navratilova, O. (2016). Rhetorical functions of citations in linguistics research articles-a contrastive (English-Czech) study. *Discourse and Interaction*, 9(2), 51-74.
- Dubois, B. L. (1988). Citation in biomedical journal articles. *English for Specific Purposes*, 7(3), 181-193.
- Dudley-Evans, T. (1986). Genre analysis: an investigation of the introduction and discussion sections of M.Sc dissertations. In Coulthard, M. (ed.), *Talking about text* (pp. 128-45). Birmingham: English Language Research, University of Birmingham.
- Dudley-Evans, T. (1995). Common core and specific approaches to the teaching of academic writing. In Belcher, D. & Braine, G. (Eds.), *Academic writing in a second language* (pp. 293-312). Norwood, NJ: Ablex.
- Dueñas P. M. (2012). Getting research published internationally in English: An ethnographic account of a team of Finance Spanish scholars' struggles. *Iberica*, 24, 139-156.
- Duszak, A., Lewkowicz, J. (2008). Publishing academic texts in English: A Polish perspective. *Journal of English for Academic Purposes*, 7(2), 108-120.
- Edmondson, W. (1981). *Spoken discourse*. London: Longman.
- Eggins, S. (1994). *An introduction to systemic functional linguistics*. London: Pinter Publisher.
- Eggins, S. (2004). *An introduction to systemic functional linguistics* (2nd ed.). London: Continuum.
- Eid, F. M. S. (2016.) Functional analysis of clause complex in the Language of news websites texts: A comparative study of two articles. *International Journal of Scientific and Research Publications*, 6(6), 323-334.
- Ellis, R. (2003). *Task-based language learning and teaching*. Oxford, UK: Oxford University Press.
- Enkvist, N. E. (1997). Why we need contrastive rhetoric. *Alternation*, 4(1), 188-206.
- Erickson, F. (1973). What makes school ethnography "ethnographic"? *Anthropology and Education Quarterly*, 15(1), 51-66.
- Esfandiari, R., Sales, S. (2024). Structural and functional characterization of citation practices in academic research writing: A concordance-informed analysis. *Journal of Applied Linguistics and Applied Literature*, 12(1), 1 - 23.
- Esser, J. (2009). *Introduction to English text-linguistics*. Frankfurt: Peter Lang.
- Fahnestock, J. (1983). Semantic and lexical coherence. *College Composition and Communication*, 34(4), 400-416.
- Faigley, L. (1979). Maturity, complexity, and growth in college writing. Paper presented at the annual meeting of the *Conference on College Composition and Communication*, 30th, Minneapolis, Minnesota, April 5 - 7, 1979.
- Fairclough, N. (1989). *Language and power*. London: Longman.
- Fairclough, N. (1992). *Discourse and social change*. Cambridge: Polity Press.
- Fairclough, N. (1995). *Critical discourse analysis*. Harlow: Longman.
- Fairclough, N. (2003). *Analyzing discourse*. London: Routledge.
- Farrokhi, F., Ghandkaran-Shotorban, S. (2014). Clause complexity in applied linguistics research article abstracts by native and non-native English writers: taxis, expansion and projection. *Journal of English Language, Teaching and Learning*, 6(13), 57-70.
- Fawcett, R. P. (1995). *Handbook for the analysis of sentences in English text*. Cardiff Computational Linguistic Unit, University of Wales, Cardiff.
- Fawcett, R. P. (1996). A systemic functional approach to complementation in English. In Berry, M., Butler, C., Fawcett, R. P. & Huang, G. W., *Meaning and form: Systemic functional interpretations: Studies for M. A. K. Halliday*. Norwood, N.J.: Ablex.
- Fawcett, R. P. (2010). A survey of available SFL descriptive frameworks for analyzing complexities of meaning in text. Speech delivered at the *22nd European Systemic Functional Linguistics Conference and Workshop*, July 9 - 11. Koper, Slovenia.
- Feng, Z. (2023). *Formal analysis for natural language processing: A handbook*. Springer Nature Singapore Pte Ltd. https://doi.org/10.1007/978-981-16-5172-4_2
- Ferguson, C. (1994). Dialect, register, and genre: working assumptions about conventionalization. In Biber and Finegan (eds.), *Sociolinguistic perspectives on register*. New York: Oxford University Press. pp. 15-30.
- Fiengo, R. W. (1974). *Semantic conditions on surface structure*. Massachusetts: Massachusetts Institute of Technology Press. Doctoral Dissertation.
- Firth, J. R. (1950). Personality and language in society. *The Sociological Review*, a42(1), 37-52. <https://doi.org/10.1111/j.1467-954X.1950.tb02460.x>
- Firth, J. R. (1957). A synopsis of linguistic theory, 1930-55. In *Studies in Linguistic Analysis* (pp. 1-31). Special volume of the Philological Society. Oxford: Blackwell. [Reprinted as Firth (1968)]
- Firth, J. R. (1968). In Palmer, F. R. (ed.), *Selected papers of J. R. Firth 1952-1959*. London: Longman.

- Firth, J. R. (1969 [1935]). *The technique of semantics. Papers in linguistics: 1934-1951*. London: Oxford University Press. pp. 7-33.
- Fki, N. (2018). Hypotactic enhancing clauses in international treaties. In Sellami-Baklouti, A. & Fontaine, L. (eds.), *Perspectives from systemic functional linguistics*. New York & London: Routledge. pp. 402-422.
- Flahive, D., Snow, B. (1980). Measures of syntactic complexity in evaluating ESL compositions. In Oller, J. W. Jr., Perkins, K. (Eds.), *Research in language testing* (pp. 171 - 176). Rowley, MA: Newbury House.
- Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive developmental inquiry. *American Psychologist*, 34(10), 906-911.
- Flowerdew, J. (1999a). Problems in writing for scholarly publication in English: The case of Hong Kong. *Journal of Second Language Writing*, 8(3), 243-263.
- Flowerdew, J. (1999b). Writing for scholarly publication in English: The case of Hong Kong. *Journal of Second Language Writing*, 8(2), 123-145.
- Flowerdew, J. (2008). Scholarly writers who use English as an additional language: What can Goffman's "Stigma" tell us? *Journal of English for Academic Purposes*, 7(2), 77-86.
- Flowerdew, J. (2015). Revisiting metadiscourse: Conceptual and methodological issues concerning signaling nouns. *Ibérica*, 29(29), 15-34.
- Flowerdew, J., Li, Y. (2009). English or Chinese? The trade-off between local and international publication among Chinese academics in the humanities and social sciences. *Journal of Second Language Writing*, 18(1), 1-16.
- Flowerdew, J., Peacock, M. (Eds.) (2001). *Research perspectives on English for academic purposes*. Cambridge: Cambridge University Press.
- Flowerdew, L. (2003). A combined corpus and systemic-functional analysis of the problem-solution pattern in a student and professional corpus of technical writing. *TESOL Quarterly*, 37(3), 489-511.
- Flowerdew, L. (2005). An integration of corpus-based and genre-based approaches to text analysis in EAP/ESP: Countering criticisms against corpus-based methodologies. *English for Specific Purposes*, 24, 321-332.
- Forey, G. (2002). *Aspects of Theme and their role in workplace texts*. University of Glasgow. Doctoral Thesis.
- Forey, G. (2009). Projecting clauses: interpersonal realisation of control and power in workplace texts. In Forey, G., Thompson, G. (eds.), *Text Type and Texture: In honour of Flo Davies*. London, Oakville: Equinox. pp. 151 - 174.
- Forey, G., Thompson, G. (Eds.). (2010). *Text type and texture*. London: Equinox.
- Foster, P., Skehan, P. (1996). The influence of planning and task type on second language performance. *Studies in Second Language Acquisition*, 18(3), 299-324.
- <http://dx.doi.org/10.1017/S0272263100015047>
- Fox, W., Bayat, M. S. (2007). *A guide to managing research*. Juta Publications.
- Francis, G., Hunston, S. & Manning, E. (1996). *Collins COBUILD grammar patterns 1: Verbs*. London: Harper Collins.
- Fraser, B. (1999). What are discourse markers? *Journal of Pragmatics*, 31(7), 931-952.
- Fries, P. H. (2002). Some aspects of coherence in a conversation. In Fries, P. H., Cummings, M., Lockwood, D., & Spruiell, W. (Eds.), *Relations and functions within and around language* (pp. 346-375). London & New York: Continuum.
- Gall, M. D., Gall, J. P., Borg, W. R. (2007). *Educational research: An introduction* (8th ed.). Boston: Pearson.
- Gao, X. (2016). A cross-disciplinary corpus-based study on English and Chinese native speakers' use of linking adverbials in academic writing. *Journal of English for Academic Purposes*, 24, 14-28.
- Gardezi, S. A., Nesi, H. (2009) Variation in the writing of economics students in Britain and Pakistan: the case of conjunctive ties. In Charles, M., Hunston, S. & Pecorari, D. (eds). *Academic writing: At the interface of corpus and discourse* (pp. 236-250). London: Continuum.
- Gardner, D., Davies, M. (2014). A new academic vocabulary list. *Applied Linguistics*, 35(3), 305-327.
- Gardner, S. (2008). Integrating ethnographic, multidimensional, corpus linguistic and systemic functional approaches to genre description: An illustration through university History and Engineering assignments. In Steiner, E. & Neumann, S. (eds.), *Data and interpretation in linguistic analysis*. 19th European SFL Conference July 2007, 1-34. Saarbrücken: Universität des Saarlandes.
- Gardner, S. (2012). Genres and registers of student report writing: An SFL perspective on texts and practice. *Journal of English for Academic Purposes*, 11(1), 52-63.
- Gardner, S., Holmes, J. (2009). Can I use headings in my essay? Section headings, macrostructures and genre families in the BAWE corpus of student writing. In Charles, M., Pecorari, D. & Hunston, S. (eds.), *Academic writing: At the interface of corpus and discourse*, 251-271. London: Continuum.

- Gerot, L., Wignell, P. (1994). *Making sense of functional grammar*. Sydney: Gerd Stabler.
- Ghadessy, M. (Ed.). (1995). *Thematic development in English texts*. London: Frances Pinter.
- Gilbert, N. G. (1977). Referencing as persuasion. *Social Studies of Science*, 7(1), 113–122.
- Givón, T. (2009). *The genesis of syntactic complexity: Diachrony, ontogeny, neuro-cognition, evolution*. Amsterdam/Philadelphia: John Benjamins.
- Givón, T. (Ed.). (2001). *Syntax: An introduction*, Vol. II. Amsterdam: Benjamins.
- Glaser, B. G. (1978). *Theoretical sensitivity: Advances in the methodology of grounded theory*. Sociology Press.
- Glaser, B., Strauss, A. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Mill Valley, CA: Sociology Press.
- Glass, G. V., Hopkins, K. D. (1984). *Statistical methods in education and psychology*, 2nd Edition. Englewood Cliffs, NJ: Prentice-Hall.
- Gleason, H. A. (1965). *Linguistics and English grammar*. London: Holt, Rinehart, and Winston.
- Gledhill, C. (2000). The discourse function of collocation in research article introductions. *English for Specific Purposes*, 19(2), 115-135.
- Gliquin, G. (2002). Automatic retrieval of syntactic structures: The quest for the Holy Grail. *International Journal of Corpus Linguistics*, 7(2), 183-214.
- Goatly, A. (2004). Corpus linguistics, systemic-functional grammar and literary meaning: A critical analysis of Harry Potter and the Philosopher's Stone. *Revista Ilha do Desterro: A Journal of English Language, Literatures in English and Cultural Studies*, 46, 115-154.
- Goutsos, D. (1997). *Modeling discourse topic: Sequential relations and strategies in expository texts*. Norwood, NJ: Ablex.
- Grabe, W., Kaplan, R. (1997). On the writing of science and the science of writing: Hedging in science text and elsewhere. In Markkanen, R., Schroder, H. (eds) *Hedging and discourse: Approaches to the analysis of a pragmatic phenomenon in academic texts*. Berlin: Walter De Gruyter, pp. 151-67.
- Graham, P. (2004). Predication, propagation, and mediation: SFL, CDA, and the inculcation of evaluative-meaning systems. In Young, L., Harrison, C. (eds), *Systemic functional linguistics and critical discourse analysis* (pp. 53 - 67). London: Continuum.
- Graustein, G., Thiele, W. (1981). Principles of text analysis. *Linguistische Arbeitsberichte*, 31, 3-29.
- Graustein, G., Thiele, W. (1987). *Properties of English texts*. Leipzig: VEB Verlag.
- Gray, B., D. Biber. (2012). Current conceptions of stance. In Hyland, K., Guinda, C. S. (eds.), *Stance and voice in written academic genres*. London: Palgrave Macmillan, pp: 15-33.
- Gregory, M. (2002). Phasal analysis within communication linguistics: Two contrastive discourses. In Fries, P. H., Cummings, M., Lockwood, D., & Spruiell, W. (Eds.), *Relations and functions within and around language* (pp. 316-345). London & New York: Continuum.
- Groom, N. (2020). Attribution and averral revisited: Three perspectives on manifest intertextuality in academic writing. In Thompson, P. (ed.), *Patterns and perspectives: Insights for EAP writing practice*. Reading, UK: CALS, University of Reading.
- Halliday, M. A. K. (1957). *Some aspects of systematic description and comparison in grammatical analysis*. Oxford: Blackwell.
- Halliday, M. A. K. (1959). *The language of the Chinese: Secret history of the Mongols*. Oxford: Blackwell.
- Halliday, M. A. K. (1961). Categories of the theory of grammar. *Word*, 17(3), 241-292. (Reprinted in M.A.K. Halliday (2002) *On Grammar*. Volume 1 in the collected works of M.A.K. Halliday, edited by Jonathan J. Webster. London and New York: Continuum.)
- Halliday, M. A. K. (1966). Some notes in 'deep grammar'. *Journal of Linguistics*, 2(1), 57-67. (Reprinted in M.A.K. Halliday (2002) *On Grammar*. Volume 1 in the collected works of M.A.K. Halliday, edited by Jonathan J. Webster. London and New York: Continuum.)
- Halliday, M. A. K. (1970). Language structure and language function. In Lyons, J. (Ed.), *New horizons in linguistics*. Harmondsworth: Penguin, pp. 140-165.
- Halliday, M. A. K. (1971). Linguistic function and literary style: An inquiry into the language of William Golding's *The Inheritors*. In: Chatman, S. (ed). *Literary Style: A Symposium*. London: Oxford University Press, pp. 330-365.
- Halliday, M. A. K. (1978a). *Language as a social semiotic: the social interpretation of language and meaning*. London: Edward Arnold.
- Halliday, M. A. K. (1978b). Meaning and the construction of reality in early childhood. In Webster, J. J. (ed.), *Continuum*, (4): 113-143.
- Halliday, M. A. K. (1979) Modes of meaning and modes of expression: types of grammatical structure and their determination by different semantic functions. In David J. Allerton et al. (eds), *Function and context in linguistic analysis*. Cambridge: Cambridge University Press. pp. 57-79. (Reprinted in M.A.K.

- Halliday (2002) *On Grammar*. Volume 1 in the collected works of M.A.K. Halliday, edited by Jonathan J. Webster. London and New York: Continuum.)
- Halliday, M. A. K. (1981). Types of structure. In Halliday, M. A. K., Martin, J. R. (eds.) *Readings in systemic linguistics*, pp. 29-41. London: Basford.
- Halliday, M. A. K. (1983). On the transition from child tongue to mother tongue. *Australian Journal of Linguistics*, 3(2), 201-216.
- Halliday, M. A. K. (1985a). *An introduction to functional grammar*. London: Edward Arnold.
- Halliday, M. A. K. (1985b). Spoken and written language. Geelong, Victoria: Deakin University Press.
- Halliday, M. A. K. (1985c). Systemic background. In Benson, J. D., Greaves, W. S. (eds.), *Systemic perspectives on discourse*, Vol. 1: Selected theoretical papers from the Ninth International Systemic Workshop. Ablex. Reprinted in Full in Volume 3 in The collected works of M. A. K. Halliday. London: Continuum.
- Halliday, M. A. K. (1988). On the language of physical science. In Ghadessy, M. (ed.), *Registers of written English: Situational factors and linguistic features* (pp. 162-178). London: Pinter.
- Halliday, M. A. K. (1991). Corpus studies and probabilistic grammar. In Aijmer, K. & Altenberg, B. (eds.), *English corpus linguistics: Studies in honour of Jan Svartvik*. London: Longman.
- Halliday, M. A. K. (1992a). Language as system and language as instance: the corpus as a theoretical construct. In Svartvik, J. (ed.), *Directions in corpus linguistics: Proceedings of Nobel Symposium 82*. 1991, pp. 65-77. Reprinted in Webster, J. J. (ed.), *Collected Works*, vol. 6, *Computational and quantitative Studies*. 2005, pp. 76-92.
- Halliday, M. A. K. (1992b). Systemic grammar and the concept of a "science of language". In *Waiguoyu (Journal of Foreign Languages)*, No. 2 (General Series No. 78), pp. 1-9. Reprinted in Full in Volume 3 in The collected works of M. A. K. Halliday. London: Continuum. p. 209.
- Halliday, M. A. K. (1994a). *An introduction to functional grammar*. 2nd Edition. London: Edward Arnold.
- Halliday, M. A. K. (1994b). Systemic theory. In Asher, R. E., Simpson, J. M. Y. (eds.), *The encyclopedia of language and linguistics*. Vol. 8. Oxford: Pergamon Press, pp. 4505-4508.
- Halliday, M. A. K. (1995b). On language in relation to the evolution of human consciousness, in Sture Allen (ed.), *Of thoughts and words (Proceedings of Nobel Symposium 92: the Relation between Language and Mind)*. London: Imperial College Press. (Reprinted in M.A.K. Halliday (2003), *On language and linguistics*. Volume 3 in the collected works of M. A. K. Halliday, edited by Jonathan J. Webster. London and New York: Continuum.)
- Halliday, M. A. K. (1996). On grammar and grammatics. In Cloran, C., Butt, D. & Hasan, R. (Eds.), *Functional descriptions: Theory into practice*. pp. 1-38. Amsterdam: John Benjamins.
- Halliday, M. A. K. Hu, Zhuanglin Introduction (胡壮麟导读). (2000). *An introduction to functional grammar (功能语法导论)*. Beijing: Foreign Language Teaching and Research Press. (北京: 外语教学与研究出版社).
- Halliday, M. A. K. (2001a). *Language as social semiotic: The social interpretation of language and meaning*. London: Arnold.
- Halliday, M. A. K. (2001b). Towards a theory of good translation. In Steiner, E., Yallop, C. (eds.), *Exploring translation and multilingual text production: Beyond content*. Berlin: Mouton de Gruyter, pp. 13-18.
- Halliday, M. A. K. (2002a). *On grammar*. In Webster, J. J. (ed.), Volume 1 of the collected works of M.A.K. Halliday. London and New York: Continuum.
- Halliday, M. A. K. (2002b). The spoken language corpus. In Aijmer, K. & Altenberg, B. (eds.), *Proceedings of ICAME 2002: The Theory and Use of Corpora*, Göteborg, May 22–26, 2002. Amsterdam: Rodopi.
- Halliday, M. A. K. (2003) *On language and linguistics*. In Webster, J. J. (ed.), Volume 3 of the collected works of M.A.K. Halliday. London and New York: Continuum.
- Halliday, M. A. K. (2004). Introduction: How big is a language? On the power of language. In Webster, J. J. (ed.), *The language of science: Volume 5 in the collected works of M. A. K. Halliday*. London and New York: Continuum.
- Halliday, M. A. K. (2005a). Corpus studies and probabilistic grammar. In Webster, J. J. (ed.), collected works of M. A. K. Halliday, Volume 6: *Computational and quantitative studies*. London: Continuum, pp. 63-75 [originally published in 1991].
- Halliday, M. A. K. (1993/2005b). In J. J. Webster (Ed.), collected works of M. A. K. Halliday. Vol. 6: *Computational and quantitative studies*. London, England: Continuum.
- Halliday, M. A. K. (2005c). Quantitative studies and probabilities in grammar. In Webster, J. J. (ed.), collected works of M. A. K. Halliday, Volume 6: *Computational and quantitative studies*. London: Continuum, pp. 130-56.

- Halliday, M. A. K. (2005d). Towards probabilistic interpretations. In Webster, J. J. (ed.), collected works of M. A. K. Halliday, Volume 6: *Computational and quantitative Studies*. London: Continuum, pp. 42-62 [originally published in 1991].
- Halliday, M. A. K. (2006a). Afterwords. In Thompson, G. & Hunston, S. (Eds.) *System and corpus: Exploring connections*, (pp. 293 - 299). London: Equinox.
- Halliday, M. A. K. (2006b). Some theoretical considerations underlying the teaching of English in China. *The Journal of English Studies (Sichuan International Studies University)* 4, 7-20.
- Halliday, M. A. K. (2008a). *Complementarities in language*. Beijing: The Commercial Press.
- Halliday, M. A. K. (2008b). Working with meaning: Towards an applicable linguistics. In Webster, J. J. (ed.), *Meaning in context*. London: Continuum, pp. 7-23.
- Halliday, M. A. K. (2009a). *Language and society*: Volume 10. In Webster, J. J. (ed.), collected works of M. A. K. Halliday. United Kingdom: Bloomsbury Academic.
- Halliday, M. A. K. (2009b). Methods - techniques - problems. In Halliday, M. A. K. & Webster, J. (Eds.), *Continuum companion to systemic functional linguistic*, (pp. 59 - 86). London: Continuum.
- Halliday, M. A. K. (2013). With Paul J. Thibault (1985). In Martin, J. R. (ed.), *Interviews with M. A. K. Halliday: Language turned back on himself*. London: Bloomsbury Publishing.
- Halliday, M. A. K., Fawcett, R. P. (1987a). Introduction. In Halliday, M. A. K. & Fawcett, R. P. (Eds.), *New developments in systemic linguistics, Vol. 1: Theory and description*. London & New York: Frances Pinter. pp. 1-13.
- Halliday, M. A. K., Fawcett, R. P. (1987b). *New developments in systemic linguistics*. London: Frances Printer.
- Halliday, M. A. K., Hasan, R. (1976a). *Cohesion in English*. London & New York: Longman.
- Halliday, M. A. K., Hasan, R. (1976b). Cohesion in English. *RELC Journal*, 23, 1-17.
- Halliday, M. A. K., Hasan, R. (1985). *Language, context, and text: Aspects of language in a social-semiotic perspective*. Geelong: Deakin University Press.
- Halliday, M. A. K., Hasan, R. (1989). *Language, context, and text: Aspects of language in a social-semiotic perspective* (2nd ed.). Oxford: Oxford University.
- Halliday, M. A. K., James, Z. L. (1993/2005). A quantitative study of polarity and primary tense in the English finite clause. In Halliday, M. A. K. & Webster, J. J. (Eds.), collected works of M. A. K. Halliday, Volume 6: *Computational and quantitative studies* (pp. 93-129). London: Continuum.
- Halliday, M. A. K., Martin, J. R. (1993). *Writing science: Literacy and discursive power*. London: The Falmer Press.
- Halliday, M. A. K., Matthiessen, C. M. I. M. (1999/2006) *Construing experience through meaning: A language-based approach to cognition*. London and New York: Continuum.
- Halliday, M. A. K., Matthiessen, C. M. I. M. (2004). *An introduction to functional grammar*. 3rd Edition. London: Hodder Arnold.
- Halliday, M. A. K., Matthiessen, C. M. I. M. (2014). *Halliday's introduction to functional grammar*. 4th ed. London and New York: Routledge.
- Halliday, M. A. K., McIntosh, A., Strevens, P. (1964). *The linguistic sciences and language teaching*. London: Longmans.
- Halliday, M. A. K., Webster, J. J. (2009) *Continuum companion to systemic functional linguistics*. London/New York: Continuum.
- Halliday, M. A. K., Webster, J. J. (2013) *Halliday in the 21st century*. London: Bloomsbury.
- Hansen-Schirra, S., Neumann, S., Steiner, E. (2007) Cohesive explicitness and explicitation in an English-German translation corpus. *Languages in Contrast*, 7(2), 241-65.
- Harmer, J. (2008). *The practice of English language teaching*. Longman ELT.
- Harvey, A. (1995). Interaction in public reports. *English for Specific Purposes*, 14(3), 189-200.
- Harwood, N. (2009). An interview-based study of the functions of citations in academic writing across two disciplines. *Journal of Pragmatics*, 41, 497-518.
- Hasan, R. (1984). Coherence and cohesive harmony. In Flood, J. (Ed.), *Understanding reading comprehension* (pp. 181-219). Newark: International Reading Association.
- Hasan, R. (1987). The grammarian's dream: Lexis as most delicate grammar. In Halliday, M. A. K. & Fawcett, R. P. (eds.), *New developments in systemic linguistics: Theory and description* (volume 1). London: Pinter. pp. 184-211.
- Hasan, R. (2009) The place of context in a systemic functional model. In Halliday, M. A. K., Webster, J. J. (eds.), *Continuum companion to systemic functional linguistics*. pp. 166-189. London/New York: Continuum.
- Hasan, R. (2015). *Context in the system and process of language: The collected works of Ruqaiya Hasan*, 4. Webster, J. J. (Ed.). London: Equinox.

- Hasan, R. (In Press). *Unity in discourse: The collected works of Ruqaiya Hasan*, 6. Webster, J. J. (Ed.). London: Equinox.
- Hasan, R., Cloran, C. (1990) A sociolinguistic interpretation of everyday talk between mothers and children. In Halliday, M. A. K., Gibbons, J. & Nichola, H. (eds.), *Learning, keeping and using language: Selected papers from the Eighth world congress of applied linguistics*, Sydney, August 16-21, 1987. Amsterdam and Philadelphia: John Benjamins. pp. 67-99.
- Haspelmath, M. (Ed.). (2004). *Coordinating constructions*. Amsterdam: Benjamins.
- Hawes, T., Thomas, S. (1997). Tense choices in citations. *Research in the Teaching of English*, 31(3), 393-414.
- He, Q. (2019). *A corpus-based approach to clause combining in English from the SFL perspective*. Springer.
- Heigham, J., Croker, R. A. (2009). *Qualitative research in applied linguistics: A practical introduction*. Palgrave Macmillan.
- Hemais, B. (2001). The discourse of research and practice in marketing journals. *English for Specific Purposes*, 20(1), 39-59.
- Herke-Couchman, M. (2006) *SFL, corpus and the consumer: An exploration of theoretical and technological potential*. Unpublished Ph.D. Thesis. Macquarie University, Sydney.
- Hewings, A., Lillis, T., Vladimirou D. (2010). Who's citing whose writings? A corpus based study of citations as interpersonal resource in English medium national and English medium international journals. *Journal of English for Academic Purposes*, 9(2), 102-115.
- Hewings, M., Hewings, A. (2002) "It is interesting to note that ...": A comparative study of anticipatory "it" in student and published writing. *English for Specific Purposes*, 21(4), 367-383.
- Hill, S. S., Soppelsa, B. F., West, G. K. (1982). Teaching ESL students to read and write experimental research paper. *TESOL Quarterly*, 16(3), 333-347.
- Hinkel, E. (2003a). Adverbial markers and tone in L1 and L2 students' writing. *Journal of Pragmatics*, 35(7), 1049-1068.
- Hinkel, E. (2003b). Simplicity without elegance: Features of sentences in L1 and L2 academic texts. *TESOL Quarterly*, 37(2), 275-301.
- Hirvela, A. (1997). "Disciplinary portfolios" and EAP writing instruction. *English for Specific Purposes*, 16(2): 83-100.
- Hjelmslev, L. T. ([1953] 1961). *Prolegomena to a theory of language*. United Kingdom: University of Wisconsin Press.
- Ho, V., Li, C. (2018). The use of metadiscourse and persuasion: An analysis of first year university students' timed argumentative essays. *Journal of English for Academic Purposes*, 33, 53-68.
- Hoey, M. (1983). *On the surface of discourse*. London: George Allen & Unwin.
- Hoey, M. (1991). *Patterns of lexis in text*. Oxford: Oxford University Press.
- Hoey, M. (2001). *Textual interaction: An introduction to written discourse analysis*. London: Routledge.
- Hoey, M. (2005). *Lexical priming: A new theory of words and language*. London: Routledge. <http://dx.doi.org/10.4324/9780203327630>
- Hoey, M. (2006). Clause relations. In Keith Brown (ed.), *Encyclopedia of language & linguistics* (Second Edition), pp. 474 - 81. Elsevier Science. <https://doi.org/10.1016/B0-08-044854-2/00496-X>
- Hoey, M., Winter, E. O. (1986). Clause relations and the writer's communicative task. In Couture, B. (ed.), *Functional approaches to writing: research perspectives*. London: Frances Pinter. pp. 120-141.
- Hoffmann, S., Evert, S., Smith, N., Lee, D., and Berglund Prytz, Y. (2008). *Corpus linguistics with BNCweb: a practical guide*. Frankfurt am Main: Lang.
- Holler, A. (2008). German dependent clauses from a constraint-based perspective. In Fabricius-Hansen, C. & Ramm, W. (Eds.), *'Subordination' versus 'coordination' in sentence and text: A cross-linguistic perspective* (pp. 187-216). Amsterdam: Benjamins.
- Holmes, R. (1997). Genre analysis, and the social sciences: An investigation of the structure of research article discussion sections in three disciplines. *English for Specific Purposes*, 16(4), 321-337.
- Holting, A. (2008). Projecting clause complexes and the subjunctive mood as means of projection in German. In Nørgaard, N. (ed.), *Systemic functional linguistics in use*. Odense working papers in language and communication. Vol. 29.
- Holtz, M. (2007). Corpus-based analysis of verb/noun collocations in interdisciplinary registers. *Proceedings of the Corpus Linguistics conference CL 2007*. http://ucrel.lancs.ac.uk/publications/CL2007/paper/14_Paper.pdf
- Holz, M. (2011). *Lexicogrammatical properties of abstracts and research articles: A corpus-based study of scientific discourse from multiple disciplines*. Technischen Universitaet Darmstadt. PhD Dissertation.
- Homburg, T. J. (1984). Holistic evaluation of ESL compositions: Can it be validated objectively? *TESOL Quarterly*, 18(1), 87-107.

- Hood, S. (2006). The persuasive power of prosodies: Radiating values in academic writing. *Journal of English for Academic Purposes*, 5(1), 37-49.
- Hopkins, A., Dudley-Evans, T. (1988). A genre-based investigation of the discussion sections in articles and dissertations. *English for Specific Purposes*, 7(2), 113-121.
[https://doi.org/10.1016/0889-4906\(88\)90029-4](https://doi.org/10.1016/0889-4906(88)90029-4)
- Hopper, P. J., Thompson, S. A. (1980). Transitivity in grammar and discourse. *Language*, 56(2), 251-299.
- Hopper, P. J., Thompson, S. A. (1984). The discourse basis for lexical categories in Universal Grammar. *Language*, 60(4), 703-752.
- Hosseini, R. M., Pour, H. H. (2022). Adversative connectors use in EFL and native students' writing: A contrastive analysis. *The Electronic Journal for English as a Second Language*, 26(1), 1-18.
- Housen, A., Kuiken, F. (2009). Complexity, accuracy, and fluency in second language acquisition. *Applied Linguistics*, 30(4), 461-473.
- Hsu, F. (2017). Chinese characteristics of clause complex: the SFL perspective of achievements from former accounts. In Webster, J. J. & Peng, X. (eds.), *Applying systemic functional linguistics*. Bloomsbury. pp. 151-166.
- Hu, G., Wang, G. (2014). Disciplinary and ethnolinguistic influences on citation in research articles. *Journal of English for Academic Purposes*, 14, 14-28.
- Huang, H. (2020). Clause complexing and characterisation: Joyce's "Two Gallants" revisited. *Language and Literature*, 30(1), 3-20.
- Huddleston, R., Hudson, R., Winter, E. & Henrici, A. (1968). Sentence and clause in scientific English. Report of the research project "The Linguistic Properties of Scientific English." Unpublished report, Department of General Linguistics, University College London.
- Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge: Cambridge University Press.
- Hunston, S. (2013). Systemic functional linguistics, corpus linguistics, and the ideology of science. *Text & Talk*, 33(4-5), 617-640.
- Hunston, S., Francis, G. (2000). *Pattern grammar: A corpus-driven approach to the lexical grammar of English*. Amsterdam: Benjamins. <http://dx.doi.org/10.1075/scl.4>
- Hunston, S., Thompson, G. (2000). *Evaluation in text*. Oxford: Oxford University Press.
- Hunt, K. W. (1964). *Differences in grammatical structures written at three grade levels: The structures to be analyzed by transformational methods*. Tallahassee, FL: Florida State University
- Hunt, K. W. (1965). *Grammatical structures written at three grade levels*. Research Report No. 3. Champaign, IL: National Council of Teachers of English.
- Hunt, K. W. (1970a). Do sentences in the second language grow like those in the first? *TESOL Quarterly*, 4(3), 195-202.
- Hunt, K. W. (1970b). Recent measures in syntactic development. In Lester, M. (Ed.), *Readings in applied transformational grammar* (pp. 179 - 92). New York, NY: Holt, Rinehart, and Winston, Inc.
- Hutchins, J. (1977). On the structure of scientific texts. *UEA papers in Linguistics*. Norwich: University of East Anglia, pp. 18-39.
- Hutchinson, T., Waters, A. (1987). *English for specific purposes: A learning-centred approach*. Cambridge: Cambridge University Press.
- Hyland, K. (1996). Writing without conviction? Hedging in science research articles. *Applied Linguistics*, 17(4), 433-54.
- Hyland, K. (1998). *Hedging in scientific research articles*. Amsterdam: John Benjamins.
- Hyland, K. (1999). Academic attribution: Citation and the construction of disciplinary knowledge. *Journal of Applied Linguistics*, 20(3), 341-367.
- Hyland, K. (2000). *Disciplinary discourse: social interactions in academic writings*. Harlow, UK: Longman.
- Hyland, K. (2001). Humble servants of the discipline? Self-mention in research articles. *English for Specific Purposes*, 20(3), 207-226.
- Hyland, K. (2002a). Activity and evaluation: reporting practices in academic writing. In Flowerdew, J. (Ed.), *Academic discourse* (pp. 115-130). London: Longman.
- Hyland, K. (2002b). Authority and invisibility: Authorial identity in academic writing. *Journal of Pragmatics*, 34(8), 1091-1112.
- Hyland, K. (2002c). Directives: Argument and engagement in academic writing. *Applied Linguistics*, 23(2), 215-239.
- Hyland, K. (2002d). What do they mean? Questions in academic writing. *Text*, 22(4), 529-557.
- Hyland, K. (2003). *Second language writing*. Cambridge: Cambridge University Press.
<https://doi.org/10.1017/CBO9780511667251>

- Hyland, K. (2004a). Disciplinary interactions: Metadiscourse in L2 postgraduate writing. *Journal of Second Language Writing*, 13 (2), 133-151. <https://doi.org/10.1016/j.jslw.2004.02.001>
- Hyland, K. (2004b). Graduates' gratitude: The generic structure of dissertation acknowledgements. *English for Specific Purposes*, 23(30), 303-324.
- Hyland, K. (2005a). *Metadiscourse: Exploring interaction in writing*. New York: Continuum.
- Hyland, K. (2005b). Stance and engagement: A model of interaction in academic discourse. *Discourse Studies*, 7(2), 173-191.
- Hyland, K. (2006). *English for academic purposes: An advanced resource book*. London: Routledge.
- Hyland, K. (2008a). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes*, 27(1), 4-21.
- Hyland, K. (2008b). Academic clusters: text patterning in published and postgraduate writing. *International Journal of Applied Linguistics*, 18(1), 41-62.
- Hyland, K. (2009a). *Academic discourse: English in a global context*. Continuum.
- Hyland, K. (2009b). Writing in the disciplines: Research evidence for specificity. *Taiwan International ESP Journal*, 1(1), 5-22.
- Hyland, K. (2010). Metadiscourse: Mapping interactions in academic writing. *Nordic Journal of English Studies*, 9(2), 125-143.
- Hyland, K. (2012). *Disciplinary identities: Individuality and community in academic discourse*. Cambridge: Cambridge University Press.
- Hyland, K., Hamp-Lyons, L. (2002). EAP: issues and directions. *Journal of English for Academic Purposes*, 1(1), 1-12.
- Hyland, K., Jiang, F. (Kevin). (2016). "We must conclude that...": A diachronic study of academic engagement. *Journal of English for Academic Purposes*, 24, 29 - 42.
- Hyland, K., Jiang, F. K. (2017). Is academic writing becoming more informal? *English for Specific Purposes*, 45, 40-51.
- Hyland, K., Shaw, P. (2016). *The Routledge handbook of English for academic purposes*. Routledge.
- Hyland, K., Tse, P. (2004) Metadiscourse in academic writing: a reappraisal. *Applied Linguistics*, 25(2), 156-177.
- Hyland, K., Tse, P. (2005). Hooking the reader: A corpus study of *evaluative that* in abstracts. *English for Specific Purposes*, 24(2), 123-139.
- Iedema, R. A. M. (1995). *The language of administration: write-it-right industry research report stage three*. Sydney: NSW Dept of Education, Disadvantaged Schools Program, Metropolitan East.
- Iedema, R. A. M. (1997). The language of administration: organizing human activity in formal institutions. In Christie, F. & Martin, J. R. (eds.), *Genre and institutions: Social processes in the workplace and school*, pp. 73-100. London: Cassell.
- Iedema, R. A. M. (2000). Bureaucratic planning and resemiotisation. In Ventola, E. (ed.), *Discourse and the community*, pp. 47-69. Tuebingen: Gunter Narr Verlag Tuebingen.
- Ifadloh, N., Nufus, Z. (2017). The analysis of clause complex in the students' causal conversation. *LET: Linguistics, Literature and Language Teaching Journal*, 7(2), 120-133.
- Isaac, O. N. (2017). The elaborative logico-semantic relation of conjunctions in Ekegusii texts. *Africology: The Journal of Pan African Studies*, 11(1), 218-235.
- Ishikawa, S. (1995). Objective measurement of low-proficiency EFL narrative writing. *Journal of Second Language Writing*, 4 (1), 51-69.
- Issitt, S. (2011) How an L2 learner corpus can identify areas of quantifiable improvement in students' written discourse. Proceedings of the *CL2011 conference*, Birmingham 20-22 July, 2011.
- Ivanic, R. (1998). *Writing and identity the discoursal construction of identity in academic writing*. Amsterdam: John Benjamins Publishing Company.
- Jalilifar, A. (2012). Academic attribution: citation analysis in master's theses and research articles in applied linguistics. *International Journal of Applied Linguistics*, 22(1), 23-41.
- Jalilifar, A., Dabbi, R. (2012). Citation in applied linguistics: Analysis of introduction sections of Iranian master theses. *Linguistik online*, 57(7), 91-104.
- Jeffreys, H. (1961). *The theory of probability*. 3rd Edition. New York: Oxford University Press.
- Jiang, F. (Kevin), Ma, X. (2018). "As we can see": Reader engagement in PhD candidature confirmation reports. *Journal of English for Academic Purposes*, 35, 1-15.
- Jogthong, C. J. (2001). *Research article introductions in Thai: Genre analysis of academic writing*. Unpublished Ph.D. Thesis. West Virginia University.
- Johannessen, J. B. (1998). *Coordination*. Oxford: Oxford University Press.
- Johns, A. M., Paltridge, B., Belcher, D. (2011). Introduction. In Belcher, D., Johns, A. M., Paltridge, B. (eds.), *New Directions in English for specific purposes research*. Michigan: University of Michigan Press.

- Jomaa, N. J. (2017). *Citation uses and perceptions in the literature review of PhD theses by EFL Arab postgraduates*. Unpublished Ph.D. thesis. Malaysia, Sintok: Universiti Utara Malaysia.
- Jomaa, N. J., Bidin, S. J. (2016). Roles of the cited author in citations of the literature review by EFL postgraduates. *International Journal of Applied Linguistics & English Literature*, 5(7), 213-225.
- Jomaa, N. J., Bidin, S. J. (2017). Perspectives of EFL doctoral students on challenges of citations in academic writing. *Malaysian Journal of Learning and Instruction*, 14(2), 177-209.
- Jomaa, N. J., Bidin, S. J. (2019a). Exploring process 'verbs' in EFL postgraduates' citations: A systemic functional linguistics approach. *Indonesian Journal of Applied Linguistics*, 9(1), 188-201.
- Jomaa, N. J., Bidin, S. J. (2019b). Reporting and quoting: Functional analyses of logico-semantic relations of clause complex citations. *3L: The southeast asian journal of English language studies*, 25(1), 158-178.
- Jones, C., Turner, J., Street, B. (eds.). (1999). *Students writing in the University*. Amsterdam: Benjamins.
- Jordan, R. R. (1997). *English for academic purposes: A guide and resource book for teachers*. Cambridge University Press.
- Jutharat, J. (2019). *A functional study of clause complexes in TCI and SSCI journal articles of environmental studies*. Doctoral dissertation. China, Xiamen: Xiamen University Institutional Repository.
- Kaltenbacher, M. (2007) Systemic functional linguistics and corpus analysis: the language of exaggeration in web-sites of tourism. In Gruber, H., Kaltenbacher, M. & Muntigl, P. (eds.), *Empirical approaches to discourse - Empirischeorientierte Ansätze in der Diskursanalyse*. Frankfurt am Main: Peter Lang, pp. 89-117.
- Kameen, P. (1979). Syntactic skill and ESL writing quality. In Yorio, C., Perkins, K., Schachter, J. (Eds.), *On TESOL '79: The Learner in Focus*. Washington, D.C.: TESOL, pp. 343-364.
- Kanoksilapatham, B. (2005). Rhetorical structure of biochemistry research articles. *English for Specific Purposes*, 24(3), 269-292.
- Kanoksilapatham, B. (2007). Rhetorical moves in biochemistry research articles. In Biber, D., Connor, U., Upton, T. A. (eds.), *Discourse on the move: Using corpus analysis to describe discourse structure*. Amsterdam: John Benjamins, pp. 73-120.
- Kaplan, E., Goodglass, H., Weintraub, S. (1983). *Boston naming test*. Philadelphia: Lea and Febiger.
- Kaplan, R. B. (2005). Contrastive rhetoric. In Hinkel, E. (ed.). *Handbook of research in second language teaching and learning*. London: Lawrence Erlbaum Associates, pp. 21-38.
- Karahan, P. (2015). A diagnostic analysis of ELT students' use of connectives. *Social and Behavioral Sciences*, 199, 325-333.
- Kass, R. E., Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90(430), 773-795.
- Kawata, K. (1992). Evaluation of free English composition. *CASELE Research Bulletin*, 22, 49-53.
- Kenny, D. Equivalence. (1998). In Baker, M. (ed.), *Routledge encyclopedia of translation studies*. London & New York: Routledge. pp. 77-80.
- Kinneavy, J. L. (1971). *A theory of discourse: the aims of discourse*. Englewood Cliffs, N.J., Prentice-Hall.
- Knight, N. K. (2008) "Still cool . . . and American too!": an SFL analysis of deferred bonds in internet messaging humour. In Nørgaard, N. (ed.), *Systemic functional linguistics in use*. Odense: Odense working papers in language and communication, Vol. 29, 481-502.
- Kollberg P. (1998). *S-notation - A computer based method for studying and representing text composition* (Lic. thesis). Stockholm University, Stockholm, Sweden.
- Kong, K. (2006). A taxonomy of the discourse relations between words and visual. *Information Design Journal*, 14(3), 207-230.
- Kortmann, B. (1996). *Adverbial subordination: A typology and history of adverbial subordinators based on European languages*. Berlin: Mouton de Gruyter.
- Krathwohl, D. R. (1993). *Methods of educational and social science research: An integrated approach*. New York: Longman/Addison Wesley Longman.
- Kress, G. (2003). *Literacy in the new media age*. London: Routledge.
- Kress, G., Van Leeuwen, T. (2002). *Reading images: The grammar of visual design* (2nd edn). London: Routledge.
- Ku, H. M. (辜鸿铭). (1989). *The discourses and sayings of Confucius*. Shanghai: Kelly and Walsh.
- Kuo, C. H. (1999). The use of personal pronouns: Role relationships in scientific journal articles. *English for Specific Purposes*, 18(2), 121-138.
- Kuo, Chih-Wei, S. W. Cheng, Chih-Hua Kuo. (2011). Citations in research article sections: Integrating forms and functions. *Taiwan Journal of TESOL*, 8(1), 67-94.

- Kurnia, S., Fitriati, S. W., Rukmini, D. (2020). The paratactic and hypotactic constructions in research article abstracts at 2018 ISET proceedings. *English Education Journal*, 10(2), 201-207.
- Kurniawan, C., Luardini, M. A., Karani, E. (2021). The analysis of clause complex of analytical exposition texts written by the English teachers of SMKN 2 Palangka Raya. *Intensive Journal*, 4(2), 24-32.
- Kwan, B. S. C. (2006). The schematic structure of literature reviews in doctoral theses of applied linguistics. *English for Specific Purposes*, 25, 30-55.
- Kwan, B. S. C., Chan, H. (2014). An investigation of source use in the results and the closing sections of empirical articles in Information Systems: In search of a functional-semantic citation typology for pedagogical purposes. *Journal of English for Academic Purposes*, 14, 29 - 47.
- Kwon, M. H., Staples, S., Partridge, R. S. (2018). Source work in the first-year L2 writing classroom: Undergraduate L2 writers' use of reporting verbs. *Journal of English for Academic Purposes*, 34, 86-96.
- Labov, W. (1972). *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.
- Labov, W. (1984). Field methods of the project on linguistic change and variation. In Baugh, J. & Sherzer, J. (eds.), *Language in use*. pp. 28-53. Englewood Cliffs: Prentice Hall.
- Lakatos, I. (1978). *Mathematics, science, and epistemology*. Cambridge: Cambridge University Press.
- Lakoff, R. (1984). The pragmatics of subordination. *Proceedings of the Tenth Annual Meeting of Berkeley Linguistic Society*, pp. 481-491. University of California Berkeley.
- Lauranto, Y. M. (2017). The projected directive construction and object case marking in Finnish. Eesti ja soome-ugri keeleteaduse ajakiri. *Journal of Estonian and Finno-Ugric Linguistics*, 8(2), 155-190.
- Lea, M. R., Street, B. V. (1998). Student writing in higher education: an academic literacies approach. *Studies in Higher Education*, 23(2), 157-172.
- Lea, M. R., Street, B. V. (2006). The "academic literacies" model: Theory and applications. *Theory Into Practice*, 45(4), 368-377.
- Lee, I. (2014). Publish or perish: The myth and reality of academic publishing. *Language Teaching*, 47(2): 250-261.
- Lee, J. J., et al. (2018). Citation practices of L2 university students in first-year writing: Form, function, and stance. *Journal of English for Academic Purposes*, 33, 1-11.
- Lee, O., Goswami, D., Herrington, A., (1983). The discourse-based interview: a procedure for exploring the tacit knowledge of writers in nonacademic settings. In Mosenthal, P., Tamor, L., Walmsley, S.A. (Eds.), *Research on writing: Principles and methods*. Longman, New York, pp. 221-236.
- Lee, P. (1996). *The Whorf theory complex: A critical reconstruction*. John Benjamins.
- Lee, Y. J. (2002). A comparison of composing processes and written products in timed-essay tests across paper-and-pencil and computer modes. *Assessing Writing*, 8(2), 135-157.
- Leech, G. (2015). Descriptive grammar. In Biber, D., Reppen, R. (eds.), *The Cambridge handbook of English corpus linguistics*. Cambridge University Press. pp. 146-60.
- Leech, G. N., Short, M. (1981/2007). *Style in fiction: A linguistic introduction to English fictional prose*. London & New York: Longman.
- Leech, G., Svartvik, J. (1975). *A communicative grammar of English*. London: Longman.
- Lei, J., Hu, G. (2019). Doctoral candidates' dual role as student and expert scholarly writer: An activity theory perspective. *English for Specific Purposes*, 54, 62-74.
- Leijten, M., Van Waes, L. (2013). Keystroke logging in writing research: Using Inputlog to analyze and visualize writing processes. *Written Communication*, 30(3), 358-392.
<https://doi.org/10.1177/0741088313491692>
- Lemke, J. L. (1984). *Semiotics and education* (Toronto semiotic circle monographs, working papers and prepublications). Toronto: Victoria University.
- Lemke, J. L. (1991). Text production and dynamic text semantics. In Ventola, E. (Ed.), *Functional and systemic linguistics: and uses* (pp. 23-38). Berlin & New York: Mouton.
<http://dx.doi.org/10.1515/9783110883527.23>
- Leong, A. P. (2023). Clause complexing in research-article abstracts: comparing human and AI-generated texts. *ExELL*, 11(2), 99-132. <https://sciendocom/article/10.2478/exell-2023-0008>
- Leong, P. A. (2021). Writing in the sciences and humanities: A clause-complex perspective. *WORD*, 67(2), 137-158.
- Li, Q, Zhang, X. (2021). An analysis of citations in Chinese English-major Master's theses and doctoral dissertations. *Journal of English for Academic Purposes*, 51, 100982.
- Li, Y. (1991). *Basic types of clause complexes and discourse connectives: A comparative study between Chinese and English with pedagogical implications*. University of Illinois at Urbana-Champaign. PhD Thesis.
- Li, Y. (2006). A doctoral student of physics writing for publication: A sociopolitically-oriented case study. *English for Specific Purposes*, 25(4), 456-478.

- Lia, P. (2008). Comparing composing processes in writing-only and reading-to-write test tasks. *Assessing Writing*, 13(2), 111-129.
- Lieber, R. (1981). *On the organization of the lexicon*. Bloomington IN: Indiana University Linguistics Club.
- Lillis, T., Hewings, A., Vladimirov, D. & Curry, M. J. (2010). The geolinguistics of English as an academic lingua franca: citation practices across English medium national and English medium international journals. *International Journal of Applied Linguistics*, 20(1), 111-135.
- Lim, J. M. H. (2006). Method sections of management research articles: A pedagogically motivated qualitative study. *English for Specific Purposes*, 25(3), 282-309.
- Lim, Jason Miin-Hwa. (2010). Commenting on research results in applied linguistics and education: A comparative genre-based investigation. *Journal of English for Academic Purposes*, 9(4), 280-294.
- Lim, Jason Miin-Hwa. (2012). How do writers establish research niches? A genre-based investigation into management researchers' rhetorical steps and linguistic mechanisms. *English for Specific Purposes*, 11(3), 229-245.
- Linell, P. (2009). *Rethinking language, mind and world dialogically: Interactional and contextual theories of human sense-making*. Charlotte, NC: Information Age Publishing.
- Litosseliti, L. (2010). *Research methods in linguistics*. Continuum.
- Liu, D. (2008). Linking adverbials: An across-register corpus study and its implications. *International Journal of Corpus Linguistics*, 13(4), 491-518.
- Liu, Q., Deng, L. (2019). Chinese MA student writers' identity construction in citation practices. *Chinese Journal of Applied Linguistics*, 42(3), 365-385.
- Liu, Y., Buckingham, L. (2018). The schematic structure of discussion sections in applied linguistics and the distribution of metadiscourse markers. *Journal of English for Academic Purposes*, 34, 97-109.
- Loban, W. (1976). *Language development: Kindergarten through grade twelve*. (Research Report No. 18). Champaign, Illinois: National Council of Teachers of English, 1976.
- Lock, G. (1995). *Functional English grammar: An introduction for second language teachers*. Germany: Cambridge University Press.
- Loghmani, Z., Ghonsooly, B., & Ghazanfari, M. (2020). Engagement in doctoral dissertation discussion sections written by English native speakers. *Journal of English for Academic Purposes*, 45(1), 100851.
- Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics*, 15(4), 474-496.
- Lu, X., Deng, J. (2019). With the rapid development: A contrastive analysis of lexical bundles in dissertation abstracts by Chinese and L1 English doctoral students. *Journal of English for Academic Purposes*, 39, 21-36.
- Luzón, M. J. (2009). The use of we in a learner corpus of reports written by EFL engineering students. *Journal of English for Academic Purposes*, 8(3), 192-206.
- Lyons, J. (1977). *Semantics*. Cambridge: Cambridge University Press.
- Ma, L. Florence. (2017). Academic writing support through individual consultations: EAL doctoral student experiences and evaluation. *Journal of Second Language Writing*, 43, 72-79.
- Magalhães, M. B., Cotterall, S., & Mideros, D. (2019). Identity, voice and agency in two EAL doctoral writing contexts. *Journal of Second Language Writing*, 43, 4-14.
- Mahboob, A., N. K. Knight (eds.). (2010). *Applicable linguistics*. United Kingdom: Bloomsbury Academic.
- Malinowski, B. (1923). The problem of meaning in primitive languages. In Ogden, C. K., Richards, I. A. (eds.), *The Meaning of Meaning* (pp. 296-336). London: Routledge & Kegan Paul.
- Malinowski, B. (1935). *Coral gardens and their magic*, Vol. 2. London: George, Allen & Unwin.
- Manan, N. A., Noor, N. M. (2014). Analysis of reporting verbs in master's theses. *ICLALIS 2013, Procedia - Social and Behavioral Sciences*, 134, 140-145.
- Mann, W. C., Matthiessen, C. M. I. M., & Thompson, S. A. (1992). Rhetorical structure theory and text analysis. In Mann, W. C., & Thompson, S. A. (eds.), *Discourse description: Diverse linguistic analyses of a fund-raising text*, pp. 39-78. Amsterdam: John Benjamins.
- Mann, W. C., Thompson, S. A. (1986). Relational propositions in discourse. *Discourse Processes*, 9(1), 57-90.
- Mann, W. C., Thompson, S. A. (1987). *Rhetorical structure theory: A theory of text organization*. USC Information Sciences Institute, Technical Report ISI/RS-87-190.
- Mann, W. C., Thompson, S. A. (1988). Rhetorical structure theory: Toward a functional theory of text organization. *Text*, 8(3), 243-281. <http://dx.doi.org/10.1515/text.1.1988.8.3.243>
- Mann, W. C., Thompson, S. A. (Eds.). (1992). *Discourse description: Diverse linguistic analyses of a fund-raising text*. Amsterdam: John Benjamins. <http://dx.doi.org/10.1075/pbns.16>

- Manning, C. D. (2003). Probabilistic syntax. In Bod, R., Hay, J., & Jannedy, S. (Eds.), *Probabilistic linguistics*. Cambridge, Mass.: The MIT Press.
- Mansourizadeh, K. & Ahmad, U. K. (2011). Citation practices among non-native expert and novice scientific writers. *Journal of English for Academic Purposes*, 10(3): 152-161.
- Marley, C. (2002). Popping the question: Questions and modality in written dating advertisements. *Discourse Studies*, 4(1), 75-98.
- Maroko, G. M. (2013). Learning about author positioning in written academic discourse. *Argentinian Journal of Applied Linguistics*, 1(2), 47-60.
- Martin, J. R. (1983). Conjunction: the logic of English text. In Petöfi, J. S. & Sözer, E. (eds.), *Micro and macro connexity of texts*, pp. 1-72. Hamburg: Helmut Buske Verlag.
- Martin, J. R. (1985). *Factual writing: exploring and challenging social reality*. Geelong: Deakin University Press.
- Martin, J. R. (1988). Hypotactic recursive systems in English: Towards a functional approach. In Benson, J., Greaves, W. (eds.), *Systemic functional approaches to discourse: Selected papers from the 12th International Systemics Workshop*. Norwood: Ablex. pp. 240-270.
- Martin, J. R. (1991). Intrinsic functionality: Implications for contextual theory. *Social Semiotics*, 1(1), 99-162.
- Martin, J. R. (1992a). *English text: System and structure*. Amsterdam: Benjamins.
<http://dx.doi.org/10.1075/z.59>
- Martin, J. R. (1992b). Macroproposals: Meaning by degree. In Mann, W. A. & Thompson, S. A. (eds.), *Discourse description: Diverse analyses of a fund-raising text*. Amsterdam: John Benjamins. pp 359-395.
- Martin, J. R. (1996). Types of structure: deconstructing notions of constituency in clause and text. In Hovy, E. H. & Scott, D. R. (eds), *Computational and conversational discourse: Burning issues: An interdisciplinary account*. Heidelberg: Springer. pp. 39-66.
- Martin, J. R. (1997). Analysing genre: functional parameters. In Christie, F. & Martin, J. R. (eds.), *Genres and institutions: Social processes in the workplace and school*. London: Cassell (Open Linguistics Series). pp. 3-39.
- Martin, J. R. (2000). Beyond exchange: Appraisal systems in English. In Hunston, S. & Thompson, G. (eds.), *Evaluation in text: Authorial stance and the construction of discourse*. Oxford: Oxford University Press. pp. 142-175.
- Martin, J. R. (2001). Cohesion and texture. In Schifffrin, D., Tannen, D. & Hamilton, H. E. (Eds.), *Handbook of discourse analysis* (pp. 35-53). London: Blackwell.
- Martin, J. R., David, R. (2003). *Working with discourse: Meaning beyond the clause*. London: Continuum.
- Martin, J. R., Matthiessen, C. M. I. M., Painter, C. (1997). *Working with functional grammar*. United Kingdom: Arnold.
- Martin, J. R., Matthiessen, C. M. I. M., Painter, C. (2010). *Deploying functional grammar*. (J. J. Webster, Ed.). Beijing: Commercial Press.
- Martin, J. R., Rose, D. (2008). *Genre relations: Mapping culture*. London: Equinox.
- Martin, J. R., White, P. R. R. (2005). *The Language of evaluation: Appraisal in English*. Hampshire and New York: Palgrave Macmillan. <https://doi.org/10.1057/9780230511910>
- Martin, J. R., Frances, C. & Rothery, J. (1987). Social processes in education: A reply to Sawyer and Watson (and others). *Working Papers in Linguistics*, No. 5. pp. 116-152.
- Martinec, R., Salway, A. (2005). A system for image-text relations in new (and old) media. *Visual Communication*, 4(3), 337-371.
- Martínez, I. A. (2005). Native and non-native writers' use of first person pronouns in the different sections of biology research articles in English. *Journal of Second Language Writing*, 14(3): 174-190.
- Mason, J. (1996). *Qualitative researching*. Thousand Oaks, CA: Sage.
- Matthiessen, C. M. I. M. (1995). *Lexicogrammatical cartography: English systems*. Tokyo: International Language Sciences Publishers.
- Matthiessen, C. M. I. M. (1998). *Lexicogrammar and collocation: a systemic functional exploration*, ms, Macquarie University, Sydney, draft ii/98, submitted to *Issues in English Grammar*. Hyderabad: Central Institute of English and Foreign Languages.
- Matthiessen, C. M. I. M. (1999). The system of TRANSITIVITY: an exploratory study of text-based profiles. *Functions of Language*, 6(1), 1-51.
- Matthiessen, C. M. I. M. (2002). Combining clauses into clause complexes: A multi-faceted view. In Bybee, J. & Noonan, M. (eds.), *Complex sentences in grammar and discourse: Essays in honor of Sandra A. Thompson*. Amsterdam/Philadelphia: John Benjamins. pp. 235-320.

- Matthiessen, C. M. I. M. (2005). Lexicogrammar in systemic functional linguistics: Descriptive and theoretical developments in the “IFG” tradition since the 1970s. In Hasan, R., Matthiessen, C. M. I. M., & Webster, J. J. (eds.), *Continuing discourse on language: A functional perspective*, Volume 2: pp. 765-858. United Kingdom: Equinox.
- Matthiessen, C. M. I. M. (2006). Frequency profiles of some basic grammatical systems: an interim report. In Thompson, G. & Hunston, S. (Eds.), *System and corpus: Exploring connections*, (pp. 103-142). London: Equinox.
- Matthiessen, C. M. I. M. (2007a). The “architecture” of language according to systemic functional theory: developments since the 1970s. In Hasan, R., Matthiessen, C. M. I. M., & Webster, J. J. (eds.), *Continuing discourse on language*. pp. 505-561. London: Equinox.
- Matthiessen, C. M. I. M. (2007b). The multimodal page: Systemic functional exploration. In Royce, T., & Bowcher, W. L. (eds.), *New directions in the analysis of multimodal discourse*. Mahwah, NJ.: Lawrence Erlbaum and Associates, pp. 1-62.
- Matthiessen, C. M. I. M. (2013). Applicable discourse analysis. In Fang, Y., & Webster, J. J. (Eds.), *Developing systemic functional linguistics: Theory and application* (pp. 138-208). London: Equinox.
- Matthiessen, C. M. I. M. (2019). Register in systemic functional linguistics. *Register Studies*, 1(1), 10-41.
- Matthiessen, C. M. I. M., Bateman, J. A. (1992). *Text generation and systemic-functional linguistics: Experiences from English and Japanese*. London: Pinter Publishers.
- Matthiessen, C. M. I. M., Halliday M. A. K. (2009). *Systemic functional grammar: A first step into the theory*. Beijing: Higher Education Press.
- Matthiessen, C. M. I. M., Teruya, K., Lam, M. (2010). *Key terms in systemic functional linguistics*. London: Continuum.
- Matthiessen, C. M. I. M., Thompson, S. A. (1988). The structure of discourse and “subordination”. In Harman, J. & Thompson, S. A. (eds.), *Clause combining in grammar and discourse*, pp. 275-329. Amsterdam: Benjamins.
- Mauranen, I. A. (1993). Contrastive ESP rhetoric: Metatext in Finnish-English economics texts. *English for Specific Purposes*, 12(1), 3-22.
- Maxwell, J. A. (1992). Understanding and validity in qualitative research. *Harvard Educational Review*, 62(3), 279-301.
- Maxwell, J. A. (2010). Using numbers in qualitative research. *Qualitative Inquiry*, 16(6), 474-482.
- McEnery, T., Wilson, A. (2001). *Corpus linguistics: An introduction*. 2nd Edition. Edinburgh: Edinburgh University Press.
- McGregor, W. (1994). The grammar of reported speech and thought in Gooniyandi. *Australian Journal of Linguistics*, 14(1): 63-92.
- McMilla, J. H., Schumacher, S. (2014). *Research in education: Evidence-based inquiry*. 7th Edition. Pearson.
- McNabb, D. E. (2008). *Research methods in public administration and nonprofit management: Quantitative and qualitative approaches*. 2nd ed. Armonk, NY: M.E. Sharpe.
- McNamara, D. S., Crossley, S. A., & McCarthy, P. M. (2010). Linguistic features of writing quality. *Written Communication*, 27(1), 57-86.
- Merriam, S. B. (2009). *Qualitative research: A guide to design and implementation*. San Francisco, CA: Jossey-Bass.
- Meyer, P. G. (1997). *Coming to know: Studies in the lexical, semantics and pragmatics of academic English*. Tübingen, Germany: Narr.
- Miestamo, M., Sinnemäki, K. & Karlsson, F. (eds.) (2008). *Language complexity: typology, contact, change*. Amsterdam: John Benjamins.
- Miller, C. (1984). Genre as social action. *Quarterly Journal of Speech*, 70(2), 151-167.
- Miller, D. (2006). From concordance to text: Appraising “giving” in Alma Mater donation requests. In Thompson, G. & Hunston, S. (eds.), *System and corpus: Exploring connections*, pp. 248-268. London: Equinox.
- Milton, J., Tsang, E. S. C. (1993). A corpus-based study of logical connectors in EFL students’ writing: Directions for further research. In Pemberton, R. & Tsang, E. S. C. (eds.), *Studies in lexis: Working papers from a seminar*. Hong Kong: Hong Kong University of Science & Technology Language Centre. pp. 215-246.
- Mochizuki, N., Starfield, S. (2021). Dialogic interactions and voice negotiations in thesis writing groups: An activity systems analysis of oral feedback exchanges. *Journal of English for Academic Purposes*, 50, 100956.
- Mohamed, N. (2016). Use of conjunctions in argumentative essay by ESL undergraduates. *Academia Journal UiTMT*, 5(1), 1-13.

- Moore, S. (2003). Writers' retreats for academics: exploring and increasing the motivation to write. *Journal of Further and Higher Education*, 7(3), 333-343.
- Moravcsik, M. J., Murugesan, P. (1975). Some results on the function and quality of citations. *Social Studies of Science*, 5(1), 86-92.
- Moreton, E. (2007). A corpus-informed study of the narrative of Mary Prince. *Proceedings of the Corpus Linguistics conference CL 2007*.
<https://www.pala.ac.uk/uploads/2/5/1/0/25105678/moreton2007.pdf>
- Morey, R. D., Rouder, J. N., Pratte, M. S., & Speckman, P. L. (2011). Using MCMC chain outputs to efficiently estimate Bayes factors. *Journal of Mathematical Psychology*, 55(5), 368-378.
- Morley, G. D. (2000). *Syntax in functional grammar: An introduction to lexicogrammar in systemic linguistics*. London: Continuum.
- Morton, J., Storch, N. (2019). Developing an authorial voice in PhD multilingual student writing: The reader's perspective. *Journal of Second Language Writing*, 43, 15-23.
- Msuya, E. A. (2019). Analysis of citation verbs in EFL academic writing: The case study of dissertations and theses at the University of Dares Salaam, Tanzania. *Iranian Journal of Applied Language Studies*, 11(2), 141-164.
- Muftah, H. (2014). Conjunctions in argumentative writing of Libyan tertiary students. *English Language Teaching*, 7(3), 108-120.
- Muijs, D. (2010). *Doing quantitative research in education with SPSS*. 2nd edition. London: SAGE Publications.
- Mulderrig, J. (2011). The grammar of governance. *Critical Discourse Studies*, 8(1), 45-68.
- Munday, J. (2008). *Style and ideology in translation: Latin American writing in English*. London; New York: Routledge.
- Munday, J. (2012). *Evaluation in translation: critical points of translator decision-making*. London; New York: Routledge.
- Murray, J. D. (1997). Connectives and narrative text: The role of continuity. *Memory & Cognitive*, 25(2), 227-236.
- Murray, R., Moore, S. (2006). *The handbook of Academic writing: A fresh approach*. New York: Open University Press.
- Myers, G. (1989). The pragmatics of politeness in scientific articles. *Applied Linguistics*, 10(1), 1-35.
- Nagar, R., Fine, J. (2013). Clause complex manifestation in depression. *Text & Talk*, 33(4-5), 595-615. <https://doi.org/10.1515/text-2013-0027>
- Nesbitt, C. (1994). *Construing linguistic resources: consumer perspectives*. Ph.D. thesis, University of Sydney.
- Nesbitt, C., Plum, G. (1988). Probabilities in a systemic-functional grammar: The clause complex in English. In Fawcett, R. P., Young, D. (eds.), *New developments in systemic linguistics*, Vol. 2: Theory and Application, pp. 6-38, London: Pinter Publishers.
- Nesi, H., Gardner, S. (2012). *Genres across the disciplines: Student writing in higher education*. Cambridge: Cambridge University Press.
- Newmeyer, F. J. (2000). *Language form and language function*. United Kingdom: MIT Press.
- Ngongo, M. (2018). Taxis and logico-semantic relation in undergraduate students' English theses writing text: A systemic functional linguistics approach. *Journal of Arts, Science & Commerce*, 4(2): 146-152.
- Nguyen T. T. L., Pramoolsook, I. (2015). Reporting verbs in literature review chapters of TESOL Master's theses written by Vietnamese postgraduates. *ESP Today*, 3(2): 196-215.
- Nguyen, T. T. L., Pramoolsook, I. (2016). Citation in Vietnamese TESOL: Analysis of master's thesis introduction chapters. *The Asian ESP*, 11(1), 95-120.
- Norris, J. M., Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics*, 30(4), 555-578.
- Noviandari, N. S. (2019). *Logico semantic relation analysis of clause complex based on SFL in CNN news*. Thesis. Ponorogo State Institute of Islamic Studies (IAIN Ponorogo).
- Noviandari, N. S., Yuwono, D. (2020). Logico semantic relation analysis of clause complex in CNN news. *English Language Teaching, Applied Linguistics and Literature*, 1(1), 24-31.
- Nuyts, J. (1983). On the methodology of a functional language theory. In Dik, S. C. (Ed.), *Advances in Functional Grammar*. Dordrecht: Foris. pp. 369-386.
- O'Donnell, M. J. (2008). The UAM Corpus Tool: Software for corpus annotation and exploration. In Callejas, B., Carmen, M. et al. (eds.), *Applied linguistics now: Understanding language and mind / La Lingüística Aplicada Hoy: Comprendiendo el Lenguaje y la Mente*. pp.1433 - 1447. Universidad de Almería.
- O'Dowd, E. (1992). The syntactic metaphor of subordination: A typological study. *Lingua*, 86(1), 46-80.

- O'Dowd, E. (2012). The development of linguistic complexity: a functional continuum. *Language Teaching*, 45(3), 329-346.
- O'Keeffe, A., McCarthy, M. J., Carter, R. A. (2007). *From corpus to classroom: Language use and language teaching*. Cambridge: Cambridge University Press.
- Okamura, A. (2008). Citation forms in scientific texts: Similarities and differences in L1 and L2 professional writing. *Nordic Journal of English Studies*, 7(3), 61-81.
- Onwuegbuzie, A. J., Frels, R. K. (2010). *An examination of the frequency rate of the verb "found" throughout scholarly publications*. Unpublished Manuscript.
- Ortega, L. (2003). Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. *Applied Linguistics*, 24(4), 492-518.
- Ortega, L. (2015). Syntactic complexity in L2 writing: Progress and expansion. *Journal of Second Language Writing*, 29, 82-94.
- Osisanwo, W. (1999). *An introductory analytical grammar of English*. Lagos: Femolus-Fetop Publishers.
- Oster, S. (1981). The use of tenses in "reporting past literature" in EST. In Selinker, L., Tarone, E., & Hanzeli, V. (Eds.), *English for academic and technical purposes* (pp.76-90). Rowley, MA: Newbury House Publishers, Inc.
- Othman, W., Al Qutob, D. (2022). Taxis and logico-semantic relations in English Arabic translation. *TRAlinea*, 24, 1-10. <https://doi.org/10.37745/gjahss.2013/vol10n8119>
- Ozturk, I. (2007). The textual organization of research article introductions in applied linguistics: Variability in a single discipline. *English for Specific Purposes*, 26, 25-38.
- Pagano, A. S., Magalhães, C. M. and Alves, F. (2004) Towards the construction of a multilingual, multifunctional corpus: factors in the design and application of CORDIAL. *Tradterm*, 10: 143-162.
- Palmer, F. R. (1986). *Mood and modality*. Cambridge, UK: Cambridge University Press.
- Palmer, F. R. (1990). *Modality and the English modals*. 2nd Ed. London: Longman.
- Paltridge, B. (1997). Thesis and dissertation writing: Preparing ESL students for research. *English for Specific Purposes*, 16(1), 61-70.
- Paltridge, B., Harbon, L., Hirsch, D., Shen, H., Stevenson, M., Phakiti, A., & Woodow, L. (2009). *Teaching academic writing: An introduction for teachers of second language writers*. Ann Arbor: University of Michigan Press.
- Panggabean, R. Y. B. (2020). *The investigation of parataxis and hypotaxis construction on TEDx talks by Josh Kaufman*. Medan: University of Muhammadiyah Sumatera Utara. MA Thesis.
- Parret, H., Chafe, W. L. (1974). *Discussing language*. Netherlands: Mouton.
- Parsons, G. (1991). Cohesion coherence: Scientific texts. In Ventola, E. (Ed.), *Functional and systemic linguistics: Approaches uses* (pp. 415-429). <http://dx.doi.org/10.1515/9783110883527.415>
- Patpong, P. (2006). *A systemic functional interpretation of Thai grammar: an exploration of Thai narrative discourse*. Unpublished Ph.D. Thesis. Macquarie University, Sydney.
- Patton, M. Q. (1990). *Qualitative evaluation and research method*. US: Sage.
- Paul, S. (2010). Writing in an electronic age: A case study of L2 composing processes. *Journal of English for Academic Purposes*, 9(4), 295-307.
- Pecorari, D. (2006). Visible and occluded citation features in postgraduate second-language writing. *English for Specific Purposes*, 25(1), 4-29.
- Peng, J. (2019). Authorial voice constructed in citation in literature reviews of doctoral theses: Variations across training contexts. *Journal of English for Academic Purposes*, 37, 11-21.
- Peng, X. (2016). English morphemic constituents working for discourse wording: Extending rank scale from "clause (complex)" up to "text (type)". *International Journal of English Linguistics*, 6(3), 38-60.
- Petrić, B. (2007). Rhetorical functions of citations in high- and low-rated master's theses. *Journal of English for Academic Purposes*, 6(3), 238-253.
- Petrić, B., Harwood, N. (2013). Task requirements, task representation, and self-reported citation functions: An exploratory study of a successful L2 student's writing. *Journal of English for Academic Purposes*, 12(2), 110-124.
- Pickard, V. (1995). Citing previous writers: What can we say instead of "say"? *Hong Kong Papers in Linguistics and Language Teaching*, 18, 89-102.
- Plakans, L., Gebriel, A. (2012). A close investigation into source use in integrated second language writing tasks. *Assessing Writing*, 17(1), 18-34.
- Plum, G., Cowling, A. (1987) Social constraints on grammatical variables: tense choice in English. In Steele, R. & Threadgold, T. (eds.), *Language Topics: Essays in Honour of Michael Halliday*, Vol II. Amsterdam/Philadelphia: John Benjamins, pp. 281-305.
- Polanyi, M. (1958). *Personal knowledge: Towards a post-critical philosophy*. In Nye, M. J. (ed.), Chicago: University of Chicago Press.

- Polanyi, M. (1964). *Personal knowledge: Towards a post-critical philosophy*. New York: Harper and Row.
- Polio, C. G. (1997). Measures of linguistic accuracy in second language writing research. *Language Learning*, 47(1), 101-143.
- Polio, C. G. (2001). Research methodology in second language writing research: The case of text-based studies. In Silva, T. & Matsuda, P. K. (eds.), *On second language writing*. Mahwah, NJ: Lawrence Erlbaum. pp. 91-115.
- Posteguillo, S. (1999). The schematic structure of computer science research articles. *English for Specific Purposes*, 18(2), 139-160.
- Pretorius, E. J. (2006). The comprehension of logical relations in expository texts by students who study through the medium of ESL. *System*, 34(3), 432-450.
- Prior, P. (1998). *Writing/Disciplinarity: A sociohistoric account of literate activity in the academy*. Mahwah, NJ: Lawrence Erlbaum.
- Quirk, R., Greenbaum, S. (1973). *A university grammar of English*. London: Longman.
- Quirk, R., Greenbaum, S. (1987). *A university grammar of English: based on A grammar of contemporary English*. Harlow: Longman.
- Quirk, R., Greenbaum, S. (1990). *A student's grammar of the English language*. Pearson Education Ltd.
- Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A comprehensive grammar of the English language*. London/New York: Longman.
- Rabab'ah, G., Al-Marshadi, A. (2013). Integrative vs. non-integrative citations among native and non-native English Writers. *International Education Studies*, 6(7), 78-87.
- Radford, A. (1999). *Linguistics: An introduction*. United Kingdom: Cambridge University Press.
- Ramanathan, V., Kaplan, R. (2000). Genres, authors, discourse communities: Theory and application for (L1 and) L2 writing instructors. *Journal of Second Language Writing*, 9(2), 171-191.
- Raputri, E. (2022). The use of clause complexes in IJAL research article abstracts. The proceeding of *the International Conference of Humanities and Social Science*, 1(1), 141-147.
- Ravid, D. (2005). Emergence of linguistic complexity in written expository texts: Evidence from later language acquisition. In Ravid, D. & Bat-Zeev Shyldkrot, H. (Eds.), *Perspectives on language and language development* (pp. 337 - 355). Dordrecht: Kluwer.
- Riazi, A. M. (2016). *The Routledge encyclopaedia of research methods in applied linguistics*. Routledge.
- Richards, J. C., Platt, J., Platt, H. (1992). *Longman dictionary of language teaching and applied linguistics*. London: Longman.
- Richards, R. T. (1988). Thesis/dissertation writing for EFL students: An ESP course design. *English for Specific Purposes*, 7(3). 171-180.
- Robert, W., Buker, S. (1990). *Writing up research*. Englewood Cliffs, NJ: Prentice Hall.
- Robinson, P. C. (1980). *ESP (English for Specific Purposes)*. Oxford: Pergamon Press.
- Rodrigues-Júnior, A. S. (2018). Clause complexes as the basis for construing the narrative point of view in translation context. *D.E.L.T.A.*: 34(1), 41-54.
- Rodríguez-Vergara, D. (2015). Clause combining in research articles in Spanish and English: A systemic-functional analysis. *US-China Foreign Language*, 13(7), 471-482.
- Roeper, T., Speas, M. (eds.). (2014). *Recursion: Complexity in cognition*. Dordrecht: Springer.
- Rogers, R., I. Schaenen. (2014). Critical discourse analysis in literacy education: A review of the literature. *Reading Research Quarterly*, 49(1): 121-143.
- Roget's 21st century thesaurus. 3rd edition. (2013). the Philip Lief Group.
- Römer, U. (2005) "This seems counterintuitive, though ...": Negative evaluation in linguistic book reviews by male and female authors. In Bonelli, E. T., Camiciotti, G. D. L. (eds.), *Strategies in academic discourse*. Amsterdam/Philadelphia: John Benjamins, pp. 97-116.
- Rosenbaum, P. S. (1967). *The grammar of English predicate constructions*. Cambridge, Massachusetts: Massachusetts Institute of Technology Press.
- Rosenwasser, D., Stephen, J. (2009). *Writing analytically* (5th ed.). Boston: Thomson Wadsworth.
- Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t-tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, 16, 225-237.
- Rowley-Jolivet, E., Carter-Thomas, S. (2014). Citation practices of expert French writers of English: issues of attribution and stance. In Lyda, A., Warchal, K. (Eds.), *Occupying niches: Interculturality, cross-culturality and aculturality in academic research* (pp.17-34). Springer International Publishing.
- Rubinson, J. F., Rubinson, K. A. (1998). *Contemporary chemical analysis*. New Jersey: Prentice Hall.
- Rudestam, K. E., Newton, R. R. (2014). *Surviving your dissertation: A comprehensive guide to content and process*. SAGE Publications.

- Rukmini, D. (2010). The logico-semantic relation of clause complexes in the abstracts of the final project reports produced by the English department students. *Ragam Jurnal Pengembangan Humaniora*, 10(3), 107-115.
- Rukmini, D. (2014). The quality of clause complexes in article abstracts written by Semarang State University graduate students. *The New English Teacher*, 8 (1), 121-131.
- Rydén, M. (1974). On notional relations in the relative clause complex. *English Studies*, 55(6), 542-545.
- Salager, F. (1994). Hedges and textual communicative function in medical English written discourse. *English for Specific Purposes*, 13(2), 149-170.
- Salager-Meyer, F. (1999). From “Mr. Guthrie is profoundly mistaken...” to “Our data do not seem to confirm the results of a previous study on...”: A diachronic study of polemicity in academic writing (1810- 1995). *Iberica*, (1), 5-28.
- Salager-Meyer, F. (2008). Scientific publishing in developing countries: Challenges for the future. *Journal of English for Academic Purposes*, 7 (2), 121-132.
- Salager-Meyer, F., Defives, G. (1998) From the gentleman’s courtesy to the scientist’s caution: A diachronic study of hedges in academic writing (1810-1995). In Fortanet, I., Posteguillo, D., Palmer, J. C., & Coll, J. F. (eds.), *Genre studies in English for academic purposes*. Castello de la Plana: Publicacions de la Universitat Jaume I, pp. 133-172.
- Sampson, G. R., Gil, D. & Trudgill, P. (eds.) (2009). *Language complexity as an evolving variable*. Oxford University Press.
- Samraj, B. (2002). Introductions in research articles: Variations across disciplines. *English for Specific Purposes*, 21(2), 1-17.
- Samraj, B. (2005). An exploration of a genre set: Research article abstracts and introductions in Two disciplines. *English for Specific Purposes*, 24(2), 141-156.
- Samraj, B. (2008). A discourse analysis of master’s theses across disciplines with a focus on introductions. *Journal of English for Academic Purposes*, 7(1), 55-67.
- Samraj, B. (2013). Form and function of citations in discussion sections of master’s theses and research articles. *Journal of English for Academic Purposes*, 12 (4), 299-310.
- Santos, J. A. L. (2018). On political science students’ academic prose: Reporting verbs in the undergraduate thesis literature review. *Veridian E-Journal*, 11(5), 248-259.
- Saragih, A. (2007). *Introduction to systemic grammar*. Medan: State University of Medan.
- Sarip, S. K., Hidayat, Y. (2019). The clause complex construction realised in the short story entitled “The Witch’s Brew”. *Journal of Applied Linguistics*, 1(1), 1-14.
- Sauerland, U., Trotzke, A. (2011). Biolinguistic perspectives on recursion: Introduction to the special issue. *Biolinguistics*, 5(1-2), 1-9.
- Schembri, N. (2009). Citation practices: Insights from interviews with six undergraduate students at the University of Malta. In O’Brien, L. J. & Giannoni, D. S. (Eds.), *University of reading: Language studies working papers* (Vol. 1, pp. 16-24). University of Reading.
- Schiffrin, D. (1986). Functions of “and” in discourse. *Journal of Pragmatics*, 10(1), 41-66.
- Schiffrin, D. (1987). *Discourse markers*. Cambridge: Cambridge University Press.
- Schiffrin, D. (2006). Discourse marker research and theory: Revisiting and. In Fischer, K. (ed.), *Approaches to discourse particles*. Amsterdam: Elsevier, pp. 315-338.
- Schleppergrell, M. (1992). Subordination and linguistic complexity. *Discourse Processes*, 15(1), 117-313.
- Schleppergrell, M. (1996). Conjunction in spoken English and ESL writing. *Applied Linguistics*, 17(3), 271-285.
- Schmied, J. (1993). Qualitative and quantitative research approaches to English relative constructions. In Souter, C. & Atwell, E. (Eds.), *Corpus-based computational linguistics* (pp. 85-96). Amsterdam: Rodopi.
- Scollon, R. (1998). *Mediated discourse as social interaction: a study of news discourse*. London: Longman.
- Scott, M. (1996). *Wordsmith tools 4*. Oxford: Oxford University Press.
- Scott, M., Tribble, C. (2006). *Textual patterns: Key words and corpus analysis in language education*. Amsterdam: John Benjamins.
- Sellami-Baklouti, A. (2011). The impact of genre and disciplinary differences on structural choice: Taxis in research article abstracts. *Text & Talk*, 31(5), 503-523.
- Sellami-Baklouti, A. (2013). A probabilistic approach to choice: The impact of contextual factors on the tactic system in research article abstracts. In O’Grady, G., Bartlett, T., Fontaine, L. (eds.), *Choice in language: Applications in text analysis*, pp. 215-42. London: Equinox.
- Sellke, T., Bayarri, M., and Berger, J. O. (2001). Calibration of *p* values for testing precise null hypotheses. *The American Statistician*, 55(1), 62-71.

- Setia, E., Sutjaja, I. G. M., Saragih, A., and Putrayadnya, I. B. (2009). Clause complex and experiential realization in court texts (Bali Bomb Case I): A systemic functional linguistics study. *E-journal of Linguistics*, 3(1).
- Shannon, C. E., Weaver, W. (1949). *The mathematical theory of communication*. Urbana, IL: The University of Illinois Press, pp. 1-117.
- Sharma, A. (1980). Syntactic maturity: Assessing writing proficiency in a second language. In Silverstein, R. (Ed.), *Occasional papers in linguistics*, No. 6 (pp. 318 - 325). Carbondale, IL: Southern Illinois University.
- Sharoff, S. (2017). Corpus and systemic functional linguistics. In Bartlett, T., O'Grady, G. (eds.), *The Routledge handbook of systemic functional linguistics*. London & New York: Routledge. pp. 533-546.
- Shaw, P. (1992). Reasons for the correlation of voice, tense and sentence function in reporting verbs. *Applied Linguistics*, 13(3), 302-319.
- Shen, L., Carter, S., Zhang, Lawrence J. (2019). EL1 and EL2 doctoral students' experience in writing the discussion section: A needs analysis. *Journal of English for Academic Purposes*, 40, 73-86.
- Shi, L. (2004). Textual borrowing in second-language writing. *Written Communication*, 21(2), 171-200.
- Shi, W. & Gu, Z. (2023). A feature analysis of clause complex in scientific news from *China Daily*: A case study of clean energy news reports. *Scholars International Journal of Linguistics and Literature*, 6(4), 197-201.
- Shooshtari, Z. G., Jalilifar, A. R. (2010). Citation and the construction of subdisciplinary knowledge. *The Journal of Teaching Language Skills*, 2(1), 45-66.
- Simon, H. A. (1962) The architecture of complexity. In *Proceeding of the American Philosophical Society*, 106(6), 467-482.
- Simpson-Vlach, R., Ellis, N. (2010). An academic formulas list: New methods in phraseology research. *Applied Linguistics*, 31(4), 487-512.
- Sinclair, J. M. (1970). *A course of spoken English: Grammar*. Oxford: Oxford University Press.
- Sinclair, J. M. (1985). On the integration of linguistic description. In van Dijk, T. A. (ed.), *Handbook of discourse analysis*, Vol. 2. London: Academic Press.
- Sinclair, J. M. (2004). *Trust the text: Language, corpus and discourse*. London/New York: Routledge.
- Sinclair, J. M. (Ed.) (1991). *Corpus, concordance, collocation*. Oxford, UK: Oxford University Press.
- Singh, K. (2007). *Quantitative social research methods*. Los Angeles, CA: Sage.
- Skoog D. A. et al. (2000). *Analytical chemistry: An introduction*. US: Saunders College Publishing.
- Slembrouck, S. (2001). Explanation, interpretation and critique in the analysis of discourse. *Critique of Anthropology*, 21(1), 33-57.
- Soler-Monreal, C., Gil-Salom., L. (2011). A cross-language study on citation practice in PhD theses. *International Journal of English Studies*, 11(2), 53-75.
- Sorensen, M. N. (1997). *Logical Connectors*.
<https://staff.washington.edu/marlynell/grammar/logicalconnectors.html>.
- Souter, C. (1990). Systemic-functional grammars and corpora. In Aarts, J. & Meijs, W. (eds.), *Theory and practice in corpus linguistics*, pp. 179-211. Amsterdam: Rodopi.
- Sriniwass, S. (2009). Methods and procedures for a systemic functional linguistic analysis: An investigation into clause complexing relations. *Journal of Modern Languages*, 19(1), 131-156.
- Srinon, U., White, P. n.d. Clause complex analysis on Thai university students' English essays in a longitudinal genre-based course. *Proceedings of the International Conference: Doing Research in Applied Linguistics*.
- Stake, R. E. (1995). *The art of case study research*. London: Sage Publications Ltd.
- Stanley, R. M. (1984). The recognition of macro structure: A pilot study. *Reading in a Foreign language*, 2(1), 156-168.
- Stapleton, P. (2010). Writing in an electronic age: A case study of L2 composing processes. *Journal of English for Academic Purposes*, 9(4), 295-307.
- Stapleton, P. (2012). Gauging the effectiveness of anti-plagiarism software: An empirical study of second language graduate writers. *Journal of English for Academic Purposes*, 11(2), 125-133.
- Starfield, S. (2004). "Why does this feel empowering?" Thesis writing, concordancing, and the corporatizing university. In Norton, B., & Toohey, K. (eds.), *Critical pedagogies and language learning*. Cambridge: Cambridge University Press, pp. 138-156.
- Stewart, M. F., Grobe, C. H. (1979). Syntactic maturity, mechanics of writing, and teachers' quality ratings. *Research in the Teaching of English*, 13(3) 207-215.
- Stoddard, S. (1991). *Text and texture: Patterns of cohesion*. Norwood, NJ: Ablex.
- Stotesbury, H. (2003). Evaluation in research article abstracts in the narrative and hard sciences. *Journal of English for Academic Purposes*, 2(4), 327-341.
- Strauss, A. L. (1987). *Qualitative analysis for social scientists*. Cambridge: Cambridge University Press.

- Stuart, B. (2007). *Alice's adventures in wonderland*. Beijing: Foreign Language Teaching and Research Press.
- Stubbs, M. (1996). Towards a modal grammar of English: a matter of prolonged fieldwork. In Stubbs, M., *Text and corpus analysis: Computer-assisted studies of language and culture*. pp. 196-229. Oxford: Blackwell.
- Sükan, S., Mohammadzadeh, B. (2022). Challenges of writing theses and dissertations in an EFL context: Genre and move analysis of abstracts written by Turkish M.A. and Ph.D. students. *Frontiers in Psychology*, 13. <https://doi.org/10.3389/fpsyg.2022.925420>
- Sulistyaningrum, S. D., Rasyid, Y. (2015). The logico-semantic relation of students' presentation in acceleration program of SMA labschool Jakarta. *International Journal of Language Education and Culture Review*, 1(1), 41-54.
- Sullivan, P., Porter, J. (1997). *Opening spaces: Writing technologies and critical research practices*. Greenwood Publishing Group.
- Sunardi, R. Y. (2017a). *Clause complex analysis in the 1945 Constitution of the State of the Republic of Indonesia*. Semarang: Dian Nuswantoro University.
- Sunardi, R. Y. (2017b). Logical meaning realisations in the agreement text on common effective preferential tariff scheme for the asean free trade area. *The 1st Education and Language International Conference Proceedings Center for International Language Development of Unissula*, 1(1), 842-850.
- Swales, J. (1981). *Aspects of article introduction*. Birmingham: The University of Aston, Language Studies Unit.
- Swales, J. (1986). *Citation analysis and discourse analysis*. *Applied Linguistics*, 7(1), 39-56.
- Swales, J. (2004). *Research genres: Explorations and applications*. Cambridge: Cambridge University Press.
- Swales, J. M., Ahmad, U.K., Chang, Y.-Y., Chavez, D., Dressen, D. F. and Seymour, R. (1998). Consider this: The role of imperatives in scholarly writing. *Applied Linguistics*, 19(1), 97-121.
- Swales, J. M. (1990). *Genre analysis: English in academic and research settings*. Cambridge [England]; New York: Cambridge University Press.
- Swales, J. M. (2014). Variation in citational practice in a corpus of student biology papers: From parenthetical plonking to intertextual storytelling. *Written Communication*, 31(1), 118-141.
- Tâm, N. T. M. (2013a). *Logico-semantic relationship in English and Vietnamese clause complexes*. Hanoi: Vietnam National University. PhD Thesis.
- Tâm, N. T. M. (2013b). The notion of clause complex in systemic functional linguistics. *VNU Journal of Foreign Studies*, 29(4), 25-36.
- Tang, E., Ng, C. (1995). A study on the use of connectives in ESL students' writing. *Perspectives* (City University of Hong Kong), 7(2), 105-122.
- Tang, R., John, S. (1999). The "I" in identity: Exploring writer identity in student academic writing through the first person pronoun. *English for Specific Purposes*, 18 (Supplement 1), S23-S39.
- Teasdale, J. D. (1993). Emotion and two kinds of meaning: Cognitive therapy and applied cognitive science. *Behavioural Research and Therapy*, 31(4), 339-354.
- Teddle, C., Tashakkori, A. (2009). *Foundations of mixed methods research*. Los Angeles, CA: Sage.
- Teich, E. (2009). Linguistic computing. In Halliday, M. A. K. & Webster, J. (eds), *Continuum companion to systemic functional linguistics*. London: Equinox, pp. 113-127.
- Teich, E., Fankhauser, P. (2005). Exploring lexical patterns in text: lexical cohesion analysis with WordNet. In Dipper, S., Götze, M. & Stede, M. (eds), *Heterogeneity in focus: Creating and using linguistic databases*. Potsdam: Universität Potsdam, pp. 129-145.
- Terkel, S. (1974). *Working*. London & New York: The New Press.
- Teruya, K. (2006). *A systemic functional grammar of Japanese*. London: Continuum.
- Teubert, W. (2003). Writing, hermeneutics, and corpus linguistics. *Logos and Language*, 4, 1-17.
- Theo van Leeuwen. (1996). The representation of social actors. In Caldas-Coulthard, C. R., & Coulthard, M.(eds.), *Texts and practices: Readings in critical discourse analysis*, pp. 32-70. London: Routledge.
- Thibault, P. (1987). An interview with Michael Halliday. In Steele, R., & Threadgold, T. (eds.), *Language topics: essays in honour of Michael Halliday*, Vol. 2, pp. 599-627. Amsterdam: Benjamins.
- Thomas, C. D., Yin, R. K. (2018.) *Case study research and applications: Design and methods*. SAGE Publications.
- Thomas, S., Hawes, T. P. (1994). Reporting verbs in medical journal articles. *English for Specific Purposes*, 13(2), 129-148.
- Thompson, G. (1994a). Propositions, projections and things. Paper presented at the 21st ISFC, Gent 1-5 August 1994.
- Thompson, G. (1994b). *Reporting Collins Cobuild English guides 5*. London: Harper Collins Publishers.
- Thompson, G. (1996a). *Introducing functional grammar*. London: Edward Arnold.

- Thompson, G. (1996b). Voices in the text: discourse perspectives on language reports. *Applied Linguistics*, 17(4), 501-530.
- Thompson, G. (2001). Interaction in academic writing: Learning to argue with the reader. *Applied Linguistics*, 22(1), 58-78.
- Thompson, G. (2004). *Introducing functional grammar (2nd Edition)*. London: Arnold.
- Thompson, G. (2013). Labelling discourse acts: Interpersonal projection across registers. In Fang, Y., & Webster, J. J. (Eds.), *Developing systemic functional linguistics: Theory and application* (pp. 209-226). London: Equinox.
- Thompson, G. (2014). *Introducing functional grammar (3rd ed.)*. London: Routledge.
<https://doi.org/10.4324/9780203785270>
- Thompson, G., Hunston, S. (2000). Evaluation: An introduction. In Hunston, S. & Thompson, G. (eds.), *Evaluation in text: Authorial stance and the construction of discourse*. Oxford: Oxford University Press. pp. 1-27. <https://doi.org/10.1093/oso/9780198238546.003.0001>
- Thompson, G., Hunston, S. (2006). (eds.) *System and corpus. Exploring connections*. London: Equinox.
- Thompson, G., Ye, Y. (1991). Evaluation in the reporting verbs used in academic papers. *Applied Linguistics*, 12(4), 365-382.
- Thompson, P. (2000). Citation practices in PhD theses. In Burnard, L. & McEnery, T. (Eds.), *Rethinking language pedagogy from a corpus perspective* (pp. 91-101). Frankfurt: Peter Lang.
- Thompson, P. (2001). *A pedagogically-motivated corpus-based examination of PhD theses: macrostructure, citation practices and uses of modal verbs*. The University of Reading. Doctoral Dissertation.
- Thompson, P. (2002). Manifesting intertextuality in the PhD theses. *Revista Canaria De Estudios Ingleses*, 44, 97-114.
- Thompson, P. (2005). Points of focus and position: intertextual reference in PhD theses. *Journal of English for Academic Purposes*, 4, 307-323.
- Thompson, P., Tribble, C. (2001). Looking at citations: Using corpora in English for academic purposes. *Language Learning and Technology*, 5(3), 91-105. <https://dx.doi.org/10.125/44568>
- Thurston, J., Candlin, C. (1997). *Exploring academic English: A workbook for student essay writing*. Sydney: National Centre for English Language Teaching and Research.
- Tomori, O. S. (1977). *Morphology and syntax of present-day English: An introduction*. Ibadan: Heinemann Edition Ltd.
- Tribble, C. (2009). Writing academic English: a survey review of current published resources. *ELT Journal*, 63(4), 400-417.
- Triki, N. (2018). Revisiting the metadiscursive aspect of definitions in academic writing. *Journal of English for Academic Purposes*, 37, 104-116.
- Trung, N. T., Hoa, N. T. Q. (2020). A systemic-functional analysis of parataxis and hypotaxis in academic IELTS samples essays. *Journal of Science and Technology*, 18(12), 43-36.
- Tucker, G. H. (1998). *The lexicogrammar of adjectives: A systemic functional approach to lexis*. London: Cassell.
- Tucker, G. H. (2006). Systemic incorporation: On the relationship between corpus and systemic functional grammar. In Thompson, G. & Hunston, S. (eds.), *System and corpus: Exploring connections*, pp. 81-102. London: Equinox.
- Tucker, P. (2003). Evaluation in the art-historical research article. *Journal of English for Academic Purposes*, 2(4), 291-312.
- Uba, S. Y. (2020). Semantic categories of reporting verbs across four disciplines in research articles. *English Language Teaching*, 13(1), 89-98.
- Unsworth, S. (2007). L1 and L2 acquisition between sentence and discourse: Comparing production and comprehension. *Lingua*, 117(11), 1930-1958. <https://doi.org/10.1016/j.lingua.2006.11.009>
- Un-udom, S., Un-udom, N. (2020). A corpus-based study on the use of reporting verbs in applied linguistics articles. *English Language Teaching*, 13(4), 162-169.
- Vallauri, E. L. (1995). A simple test for theme and rheme in the clause complex. *Language Sciences*, 17(4), 357-378.
- van der Hulst, Harry (ed.). (2010). *Recursion and human language*. Berlin/New York: Mouton de Gruyter.
- Van Dijk, T. A. (1972). *Some aspects of text grammars*. Berlin: De Gruyter Mouton.
<https://doi.org/10.1515/9783110804263>
- Van Dijk, T.A. (1977). Semantic macro-structures and knowledge frames in discourse comprehension. In Just, M. A., & Carpenter, P. A. (eds.), *Cognitive processes in comprehension (1st ed.)*. Psychology Press.
<https://doi.org/10.4324/9781315798851>

- van Doorn, J., Ly, A., Marsman, M., & Wagenmakers, E.-J. (2020). Bayesian rank-based hypothesis testing for the rank sum test, the signed rank test, and Spearman's rho. *Journal of Applied Statistics*, 47(16), 2984-3006.
- van Leeuwen, T. (2005). *Introducing social semiotics: An introductory textbook*. London & New York: Routledge.
- Van Valin, R. D., LaPolla, R. J. (1997). *Syntax: Structure, meaning, and function*. Cambridge: Cambridge University Press.
- Van Valin, R. D. (2001). *An introduction to syntax*. United Kingdom: Cambridge University Press.
- Verstraete, J. C. (2007). *Rethinking the coordinate-subordinate dichotomy: Interpersonal grammar and the analysis of adverbial clauses in English*. Berlin: de Gruyter Mouton.
- Vilha, M. (1999). *Medical writing: Modality in focus*. Amsterdam: Rodopi.
- Vinkler, P. (1987). A quasi-quantitative citation model. *Scientometrics*, 12(1-2), 47-72.
- Voloshinov, V. N. (1995). *Marxism and the philosophy of language*, Bakhtinian thought: An introductory reader. Dentith, S., Matejka, L. & Titunik, I. R. (trans.), London: Routledge.
- Vorvilas, G. (2014). Logico meanings in multimedia learning materials: A multimodal discourse analysis. *Acta Didactica Napocensia*, 7(2), 25-40.
- Vorvilas, G., Vergidis, D., & Ravanis K. (2011). Multimodal microgenres for designing learning objects. *Themes in Science & Technology Education*, 4(2), 89-102.
- Vovk, V. G. (1993). A logic of probability, with application to the foundations of statistics. *Journal of the Royal Statistical Society Series B*, 55(2), 317-351.
- Vyatkina, N., Hirschmann, H., & Golcher, F. (2015). Syntactic modification at early stages of L2 German writing development: A longitudinal learner corpus study. *Journal of Second Language Writing*, 29, 28-50.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.
- Wallis, S., Nelson G. (2001). Knowledge discovery in grammatically analysed corpora. *Data Mining and Knowledge Discovery*, 5, 307-340.
- Wang, F. M. (2020). An analysis of the use of clause complex in scientific journals. *Journal of Education & Entrepreneurship*, 7(1), 56-64.
- Wang, G., Hu, G. (2022). Citations and the nature of cited sources: A cross-disciplinary and cross-linguistic study. *SAGE Open*, 12(2), 1-14.
- Watson-Gegeo K. A. (1988). Ethnography in ESL: Defining the essentials. *TESOL Quarterly*, 22(4), 575-592.
- Weatherall, R. (2019). Writing the doctoral thesis differently. *Management Learning*, 50(1), 100-113. <https://doi.org/10.1177/1350507618799867>
- Webster, J. J. (2002). *Text linguistics*. Chinese version translated by Jiujiu XU. Beijing: China Press of Social Sciences.
- Webster, J. J. (2014). Text and texture. In Halliday, M. A. K. & Webster, J. J. (Eds.), *Text linguistics: The how and why of meaning* (pp. 181-425). London: Equinox.
- Webster's new world college dictionary*. 4th edition. (2010). Houghton Mifflin Harcourt.
- Wegener, P. (1885). *Untersuchungen über die Grundfragen des Sprachlebens*. Reprint 1991. New York, NY: Benjamins.
- Weissberg, R., Buker, S. (1990). *Writing up research: Experimental research report writing for students of English*. Englewoods Cliff, NJ: Prentice Hall Regents.
- White, E. B. (1952). *Charlotte's web*. New York: Harper & Row Publishers Inc.
- White, L. (2003). *Second language acquisition and universal grammar*. Cambridge: Cambridge University Press.
- White, P. R. R. (2015). *An introductory tour through appraisal theory: The appraisal website*. http://www.grammatics.com/appraisal/appraisaloutline/framed/appraisaloutline.htm#P4_50
- Whorf, B. L. (1937[1945]). Grammatical categories. *Language*, 21(1), 1-11. In Lee, P. (1996). *The Whorf theory complex: A critical reconstruction*. John Benjamins.
- Whorf, B. L. (1956). In Carroll, J. B. (ed.), *Language thought and reality: selected writing of Benjamin Lee Whorf*. Cambridge, MA: The MIT Press.
- Widdowson, H. (2004). *Text, context, pretext*. Oxford: Blackwell.
- Wignell, P., Martin, J. R., Eggins, S. (1993). The discourse of geography: Ordering and explaining the experiential world. In Halliday, M. A. K. & Martin, J. R. (Eds.), *Writing science: Literacy and discursive power* (pp. 136 - 165). Pittsburgh, PA: University of Pittsburgh Press.
- Wilcox, K. C., Jeffery, J. (2014). Adolescents' writing in the content areas: National study results. *Educational Theory and Practice Faculty Scholarship*, 17. https://scholarsarchive.library.albany.edu/etap_fac_scholar/17

- Wilkins, S., Hazzam, J., Lean, J. (2021). Doctoral publishing as professional development for an academic career in higher education. *The International Journal of Management Education*, 19(1), 100459.
- Willis, T., et al. (2010). The geolinguistics of English as an academic lingua franca - citation practices across English medium national and English medium international journals. *International Journal of Applied Linguistics*, 20(1), 111-135.
- Winter, E. (1971). Connection in science material: a proposition about the semantics of clause relations. In *Science and technology in a second language: papers from a seminar held at the University of Birmingham from 27th to 29th March 1971*. London: Centre for Information on Language Teaching. pp. 41-52.
- Wodak, R. (ed.). (1989). *Language, power and ideology: Studies in political discourse*. Amsterdam: John Benjamins.
- Wolfe-Quintero, K., Inagaki, S. & Kim, H. Y. (1998). *Second language development in writing: Measures of fluency, accuracy, and complexity*. Honolulu, HI: University of Hawaii Press.
- Wong, A. T. Y. (2005). Writers' mental representations of the intended audience and of the rhetorical purpose for writing and the strategies that they employed when they composed. *System*, 33(1), 29-47.
- Wood, L., Jones, C. L., Williams, A. (2002a). *Pass Cambridge BEC preliminary student book*. Beijing: Economic Science Press.
- Wood, L., Sanderson, P., Williams, A. (2002b). *Pass Cambridge BEC vantage student book*. Beijing: Economic Science Press.
- Wu, C. (2000). *Modelling linguistic resources: a systemic functional approach*. PhD thesis. Macquarie University.
- Wu, C. (2009). Corpus-based research. In Halliday, M. A. K., & Webster, J. J. (eds.), *Continuum companion to systemic functional linguistics*. Continuum. pp: 128-42.
- Wu, C., Fang, J. (2007). The semiotics of university introductions in Australia and China. In Barbara, L. & Sardinha, T. (eds.), *Proceedings of the 33rd International Systemic Functional Congress (PUCSP, São Paulo, Brazil)*. ISBN 85-283-0342-X. pp. 568-591.
https://www.pucsp.br/isfc/proceedings/Artigos%20pdf/27acd_fang_568a591.pdf
- Wulansari, A. (2017). The analysis of clause complex in *The Snow of Kilimanjaro*: A systemic functional discourse analysis. The conference paper on *Conference on Language and Language Teaching*.
- Xu, F. (2016). The developmental features of citation competence in L2 academic writing: Cross-sectional and longitudinal evidence. *Journal of Foreign Languages*, 39(3), 73-82.
- Xuan, W. Winfred, Chen, S. (2020). Taking stock of accumulated knowledge in projection studies from systemic functional linguistics: A research synthesis. *Functional Linguistics*, 7, 1-19.
- Yallop, C. (2001). The construction of equivalence. In Steiner, E.; Yallop, C. (eds.), *Exploring translation and multilingual text production: Beyond content*. Berlin; New York: Mouton de Gruyter. pp. 229-246.
- Yang S. (2014). Expressions of modality in the clause complex of Chinese: "Marked" markers of the logico-semantic relation of expansion. In *Annual Review of Functional Linguistics*. Beijing: Higher Education Press, (12), 39-53.
- Yang, L. (2013). Evaluative functions of reporting evidentials in English research articles of Applied Linguistics. *Open Journal of Modern Linguistics*, 3(2), 119-126.
- Yang, L., Shi, L. (2003). Exploring six MBA students' summary writing by introspection. *Journal of English for Academic Purposes*, 2(3), 165-192.
- Yang, R., Allison, D. (2003). Research articles in applied linguistics: Moving from results to conclusions. *English for Specific Purposes*, 22(4), 365-385.
- Yanuar, E. C. (2013). *Analysis of clause complex in English textbooks written by native and non-Native speakers*. University of Muria Kudus. MA Thesis.
- Yeh, C. (2009). Student researchers' citation behavior. *Taiwan Journal of TESOL*, 6(1), 1-21.
- Yeh, C. (2010). Citation practices in TESL research articles: A comparative study. *Asian ESP Journal*, 6(2), 40-64.
- Yin, R. K. (1981). The case study crisis: Some answers. *Administrative science quarterly*, 26(1), 58-65.
- Yin, R. K. (2009). *Case study research, design and method*. 4th edition. London: Sage Publications Ltd.
- Young, L., Harrison, C. (2004). Introduction. In Young, L., & Harrison, C. (eds), *Systemic functional linguistics and critical discourse analysis* (pp. 1-11). London: Continuum.
- Yuniar, R. (2018). Logico-semantic relation of clauses used in analytical exposition written by the third semester students of English department. *Journal of Research on Applied Linguistics Language and Language Teaching*, 2, 1(11), 121-128.

- Zappavigna, M., Dwyer, P. & Martin, J. R. (2008). Syndromes of meaning: exploring patterned coupling in a NSW Youth Justice Conference. In Mahboob, A. & Knight, N. (eds), *Questioning linguistics*. Newcastle: Cambridge Scholars Publishing, pp. 164-185.
- Zhou, Y., Abidin, M. J. B. Z. (2023). Identifying taxis and logico-semantic relations in Chinese EFL writing sample. *English Language Teaching*, 16(9), 1-12.
- Zitzewitz, P. W., Elliott, T. G., Haase, D. G., et al. (2005). *Physics: Principles and problems*. USA: The McGraw-Hill Companies, Inc. Chapter Four: Forces in One Dimension.
- Bai, Liru. (白丽茹). (2021). On the potential relationship between the explicit logical connection awareness of the English texts and the English writing proficiency (英语篇章显性逻辑连接意识与英语写作水平潜在关系研究). Beijing: *Foreign Language Teaching and Research* (北京: 外语教学与研究), 53(5), 733-744.
- Bao, Gui. (鲍贵). (2009). Syntactic complexity in EFL learners essays: A multidimensional perspective (英语学习者作文句法复杂性变化研究). Beijing: *Foreign Language Teaching and Research* (北京: 外语教学与研究), 41(4), 291-297.
- Cao, Cong. (曹聪). (2011). *A functional study of projection in narratives of marine incident and accident reports* (海事事故报告陈述中投射的功能研究). Dalian: Dalian Maritime University (大连: 大连海事大学). Master's thesis.
- Chang, Lumin. (常路敏). (2014). *A study of projection in Charlotte's Web* (儿童文学作品《夏洛的网》中投射现象的研究). Linfen: Shanxi Normal University (临汾: 山西师范大学). Master's thesis.
- Chen, Anling. (陈安玲). (1997). The stylistic function of clause complexes (小句复合体的语体功能). Shanghai: *Journal of Foreign Languages* (上海: 外国语), (4), 62-65.
- Chen, Anling. (陈安玲). (2000). The textual function of clause complexes (小句复合体的语篇功能). Dalian: *Foreign Languages and Their teaching* (大连: 外语与外语教学), (5), 15-17.
- Chen, Huiyuan., Wu, Xudong. (陈慧媛, 吴旭东). (1998). The effect of task complexity and task condition on EFL writing performance (任务难度与任务条件对EFL写作的影响). Guangzhou: *Modern Foreign Languages* (广州: 现代外语), (2), 27-39.
- Chen, Mingfang. (陈明芳). (2005). The projection meaning of language: from the aspect of English verbs (从英语动词的体义看语言的投射意义). Dalian: *Foreign Languages and Their teaching* (大连: 外语与外语教学), (5), 56-60.
- Cheng, Hua. (程华). (2014). On the relationship of projection and style under the perspective of SFL (功能语言学视角下投射和语体关系研究). Suihua: *Journal of Suihua University* (绥化: 绥化学院学报), 34(12), 62-64.
- Cheng, Xiaotang. (程晓堂). (2005). On the clause relation of clause complexes (论小句复合体中的小句关系). Harbin: *Foreign Language Research* (哈尔滨: 外语学刊), (4), 60-65.
- Ding, Jianxin. (丁建新). (2000). A systemic study of projection in English clause complexes (英语小句复合体投射系统之研究). Guangzhou: *Modern Foreign Languages* (广州: 现代外语), 23(1), 45-57.
- Ding, Suping. (丁素萍). (2007). The semantic compositional scale and its realizations in ideational metafunction (概念功能中的语义级差与体现形式研究). Dalian: *Foreign Languages and Their teaching* (大连: 外语与外语教学), (9), 4-8.
- Ding, Suping., Shu, Wei. (丁素萍, 舒伟). (2013). On the projecting system: a linguistic analysis of *Alice's Adventures in Wonderland* (投射系统研究 ---- 以《爱丽丝奇境漫游记》中的言语分析为例). Dalian: *Foreign Languages and Their teaching* (大连: 外语与外语教学), (1), 22-26.
- Ding, Zhanping. (丁展平). (2002). Hedging in English academic writing (英语学术论文中的通言研究). Hangzhou: *Journal of Zhejiang University* (Humanities and Social Sciences) (杭州: 浙江大学学报) (人文科学版) 32(6), 108-114.
- Dong, Mei., Li, Zhongxia. (董梅, 李忠霞). (2011). Two queries on the projecting clause complexes (关于投射型小句复合体的两个疑问). Shijiazhuang: *Youthful Years* (石家庄: 青春岁月), (3), 69.
- Feng, Zongxin. (封宗信). (2012). Linguistic indeterminacy and fuzziness in systemic-functional grammar (语言的不确定性与系统功能语法中的模糊性). Harbin: *Foreign Language Research* (哈尔滨: 外语学刊), (5), 1 - 5.

- Fu, Xiaoli. (符小丽). (2011). The textual analysis of the clause complex system framework (小句复合体理论框架内的语篇分析). Guilin: *Journal of Guilin College of Aerospace Technology* (桂林: 桂林航天工业高等专科学校学报), (1), 94-98.
- Gaun, Limei. (关丽妹). (2020). Knowledge construction of English and Chinese physics textbooks from the perspective of clause complex (小句复合体视角下英汉物理教材的知识建构). Liuzhou: *Journal of Liuzhou Vocational and Technical College* (柳州: 柳州职业技术学院学报), 20(2), 131-136.
- Ge, Feiyun. (葛飞云). (2020). *Clause translation from the perspective of clause relations: A case study based on E-C translation of short story Katu* (从小句关系理论看小句的翻译 ---- 以短篇小说《开图》汉译为例). Nanjing: Nanjing University (南京: 南京大学). Master's thesis.
- Ge, Tianshuang., Wang, Ziyi. (葛天爽, 王姊祎). (2023). Authorial voice construction in English academic writing from the perspective of literature citation (基于文献引用视角的英语学术写作中写作者声音的建构). *Modern Linguistics* (现代语言学), 11(4), 1279-1289.
- Guo, Chunjie. (郭纯洁). (2015). *Research methods of modern linguistics* (现代语言学研究方法). Beijing: Science Press (北京: 科学出版社).
- Guo, Xiaohong. (郭晓红). (2016). *A study of projection in Chinese and American English news discourse from the perspective of systemic functional linguistics* (系统功能语言学视角下的中美英语新闻语篇投射研究). Jian: Shandong Normal University (济南: 山东师范大学). Master's thesis.
- Hao, xinggang., Li, Huaijuan. (郝兴刚, 李怀娟). (2015). Research on projecting system of clause complexes in English hard news (英语硬新闻中小句复合体投射系统研究). Jiangsu: *Jiangsu Foreign Language Teaching and Research* (苏州: 江苏外语教学研究), (1), 60-64.
- Haung, Guowen., Xiao, Junhong. (黄国文, 肖俊洪). (1996). *Aspects of English complex sentences: From sentence to text* (英语复合句 ---- 从句子到语篇). Xiamen: Xiamen University Press (厦门: 厦门大学出版社).
- He, Ting. (何婷). (2022). A contrast of logico-semantic expressions of English and Chinese clause complexes in prose discourse: A case study of rush and Its English version (英汉小句复合体在散文语篇的逻辑语义表达对比 ---- 以《匆匆》语篇及其译文为例). *Modern Linguistics* (现代语言学), 10(9), 1970-1977.
- He, Wei. (何伟). (2002). On recursiveness, embedding and functional syntactic reanalysis (递归、嵌入与功能句法再分析). Harbin: *Foreign Language Research* (哈尔滨: 外语学刊), (3): 64-69.
- He, Wei., Cheng, Ming. (何伟, 程铭). (2021). Logical system from an ecolinguistics perspective (生态语言学视角下的逻辑关系系统). Nanjing: *Journal of PLA University of foreign languages* (南京: 解放军外国语学院学报), 44(3), 51-59.
- He, Xiaowen., Luo, Zhiyong., Hu, Zijuan., Wang, Ruiqi. (何晓文, 罗智勇, 胡紫娟, and 王瑞琦). (2021). Automatic recognition of sentence boundary based on clause complex (基于小句复合体的句子边界自动识别研究). Beijing: *Journal of Chinese Information Processing* (北京: 中文信息学报), 35(5), 1-8.
- Hu, Honghui. (胡红辉). (2024). Research on the projecting language of Xu Yuanchong's English translated version of *Lunyu* under the perspective of SFL (系统功能视域下许渊冲《论语》英译投射语言研究). Urumchi: *Language and Translation* (乌鲁木齐: 语言与翻译), 1, 43-49.
- Hu, Zhuanglin. (胡壮麟). (2002). Halliday's model of grammatical metaphor (评语法隐喻的韩礼德模式). In Huang, Guowen., Yang, Binjun. (eds.), *Text, language function and language teaching* (黄国文, 杨炳钧(编). 语篇、语言功能、语言教学). Guangzhou: Sun Yat-sen University Press. (广州: 中山大学出版社), pp. 88-105.
- Hu, Zhuanglin. (胡壮麟). (2018). *The Chinese origin of Halliday's academic thoughts and its return* (韩礼德学术思想的中国渊源和回归). Beijing: Foreign Language Teaching and Research Press (北京: 外语教学与研究出版社).
- Hu, Zhuanglin., Zhu, Yongsheng., Zhang, Delu., Li, Zhanzi. (eds.) (胡壮麟, 朱永生, 张德禄, 李战子). (合编). (2005). *Introduction to Systemic Functional Linguistics* (系统功能语言学概论). Beijing: Peking University Press (北京: 北京大学出版社).

- Huang, Guowen. (黄国文). (1996). A syntactic-semantic analysis of the enhanced theme construction in English (英语强势主位结构的句法----语义分析). Beijing: *Foreign Language Teaching and Research* (北京: 外语教学与研究), (3), 44-48.
- Huang, Guowen. (黄国文). (1997). Thematic analysis of English existential processes (英语存在过程的主位分析). In Hu, Zhuanglin., Fang, Yan (eds.), *Progress of functional linguistics in China* (胡壮麟, 方琰主编. 功能语言学在中国的进展). Beijing: Tsinghua University Press. (北京: 清华大学出版社).
- Huang, Guowen. (黄国文). (1999). *Study of English language problems* (英语语言问题研究). Guangzhou: Sun Yat-sen University (广州: 中山大学出版社).
- Huang, He. (黄荷). (2017). The logico-semantic relation of the “longest sentence in English literature”: from the perspective of functional stylistics (“英语文学最长句”中的逻辑语义关系: 功能文体学视角). Xi'an: *Foreign Language Education* (西安: 外语教学), 38(6), 37-41.
- Huang, Ziyun. (黄紫筠). (2015). *A study on the realization of logico-semantic relations in English as a second language academic writing* (二语学术写作中的逻辑语义关系实现研究). Nanchang: Nanchang University (南昌: 南昌大学). Master's thesis.
- Ji, Xiaoling. (纪小凌). (2009). A study of the development of Chinese EFL learners' written English (英语学习者书面语发展研究). Guangzhou: *Modern Foreign Languages* (广州: 现代外语), (32)2, 178-185.
- Jia, Peipei., Zeng, Lei. (贾培培, 曾蕾). (2021). A study of projecting sources in news discourse (新闻语篇中投射源的研究). Beijing: *Foreign Language in China* (北京: 中国外语), 18(4), 40-46.
- Jia, Peipei., Zhuang, Jingyuan. (贾培培, 张敬源). (2015). On the recursiveness of projection (投射的递归性研究). Beijing: *Foreign Language in China* (北京: 中国外语), 12(5), 41-48.
- Jiang, Ting., Yang, Xia. (蒋婷, 杨霞). (2018). The contrastive analysis of writers' identity construction in English and Chinese legal research articles: from the perspective of engagement system (英汉法律类学术论文中作者身份构建的对比研究 ---- 以介入系统为视角). Xi'an: *Journal of Xi'an International Studies University* (西安: 西安外国语大学学报), 26(4), 10-15.
- Ju, Yumei. (鞠玉梅). (2016). The subject features of reporting clauses and the construction of authority identity in academic writings (学术写作中引述句的主语特征与身份构建研究). Beijing: *Foreign Language Teaching and Research* (北京: 外语教学与研究), 48(6), 926-936.
- Li, Changzhong. (李长忠). (2006). The textual features of the Chinese and English classical poetry: on the inter-clausal relation from Halliday's functional grammar (从功能语言学的句际关系看英汉语古典诗歌语篇特征). Dalian: *Foreign Languages and Their teaching* (大连: 外语与外语教学), (12), 5-8.
- Li, Juyuan. (李桔元). (2007). Projection and experience-construing (投射与经验识解). Beijing: *Foreign Language in China* (北京: 中国外语), 4(1), 36-41.
- Li, Li. (李莉). (2010). *Projection in news discourse from the perspective of systemic-functional linguistics-compared with English fiction* (从功能语言学角度分析新闻中的投射). Dalian: Dalian Maritime University (大连: 大连海事大学). Master's thesis.
- Li, Ningning. (李宁宁). (2021). *A comparative study on the explicitation in the translation of cognitive projecting verbs of Hong Lou Meng in the Yang's and Hawkes' versions* (《红楼梦》心理投射动词英译的显化比较研究 ---- 以杨译本和霍译本为例). Nanning: Guangxi University for Nationalities (南宁: 广西民族大学). Master's thesis.
- Li, Taizhun. (李太珠). (2020). The analysis of the logico-semantic relations and their teaching implications in business English listening and speaking discourse (商务英语听说语篇的逻辑语义关系分析及其对教学的启示). Wuhan: *English Square* (武汉: 英语广场), (10), 32-34.
- Li, Xiaohong. (李晓红). (2017). The contrastive analysis of discourse strategies in abstracts of doctoral dissertations in linguistics (语言学博士学位论文摘要的语篇策略对比分析). Nanjing: *Journal of PLA University of foreign languages* (南京: 解放军外国语学院学报), 40(3), 19-27.
- Liang, Lujin. (梁鲁晋). (2004). Phenomena in projection systems (投射系统中的现象). Harbin: *Foreign Language Research* (哈尔滨: 外语学刊), (4), 55-58.
- Liao, Yiqing. (廖益清). (2006). The functional linguistic analysis of English projecting clause complexes (英语投射型小句复合体的功能语言学分析). Harbin: *Foreign Language Research* (哈尔滨: 外语学刊), (1), 59-67.

- Lin, Xiaoping. (林晓萍). (2018). *Error analysis of Chinese-English machine translation under the framework of clause complex theory* (小句复合体理论框架下的汉英机器翻译错误分析). Guangzhou: Guangdong University of Foreign Studies (广州: 广东外语外贸大学). Master's thesis.
- Liu, Dan. (刘丹). (2017). *Discourse measurement method based on quantitative analysis of clause relations* (基于小句关系定量分析的语篇测量方法). Harbin: Northeast Agricultural University (哈尔滨: 东北农业大学). Master's thesis.
- Liu, Tana. (刘塔娜). (2017). *A contrastive study on clause relations of English and Chinese news discourse* (英汉新闻语篇小句关系的对比研究). Shenyang: Northeast Normal University (沈阳: 东北师范大学). Master's thesis.
- Liu, Xiangdong. (刘向东). (2020). On the recontextualization and functional variation of projection (投射系统的再语境化模式及功能变异). Jinan: *Shandong Foreign Language Teaching* (济南: 山东外语教学), 41(2), 40-52.
- Liu, Yanxia. (刘艳霞). (2003). The textual analysis of clause complexes in a dialogue from *Wuthering Heights* under SFG (用系统功能语法中的小句复合体理论对《呼啸山庄》中的一段对话进行语篇分析). Chongqing: *The Journal of English Studies* (重庆: 英语研究), 2(3/4), 23-34.
- Lou, Baocui. (娄宝翠). (2011). Reporting verbs in learners' English Master's theses (学习者英语硕士论文中的转述动词). Nanjing: *Journal of PLA University of foreign languages* (南京: 解放军外国语学院学报), 34(5), 64-8.
- Lou, Baocui. (娄宝翠). (2013). A contrastive analysis of reporting verbs and stance expressions in academic papers of Chinese and English undergraduate students (中英大学生学术论文中转述动词及立场表达对比分析). Jinan: *Shandong Foreign Language Teaching* (济南: 山东外语教学), 2, 50-55.
- Lou, Baocui. (娄宝翠). (2017). A study of adverbials of result in Chinese learners' academic writing (学习者英语学术论文结果副词使用特征研究). Nanjing: *Journal of PLA University of foreign languages* (南京: 解放军外国语学院学报), 40(3), 70-8.
- Lv, Shuxiang. (ed.) (吕叔湘编). (2002). *A comparative study on English translations of old Gems* (中诗英译比录). Beijing: Chung Hwa Book Company (北京: 中华书局).
- Ma, Rong., Qin, Xiaoqing. (马蓉, 秦晓晴). (2014). *Text borrowing strategies of citations in Second Language academic writing: A case study of English Master's thesis* (二语学术写作引用话语的文本借用策略特征 ---- 以英专硕士论文为例). Xi'an: Foreign Language Education (西安: 外语教学), 35(6), 50-54.
- Ma, Rong., Qin, Xiaoqing. (马蓉, 秦晓晴). (2016). Citation features in academic English writing based on classroom read-to-write tasks (基于课堂读写任务的学术英语写作引用特征研究). Wuhan: *Foreign Language and Literature Research* (武汉: 外国语文研究), 2(2), 84-92.
- Ming, Jianping. (明建平). (2020). The comparison of paratactic and hypotactic clause complexes in English and Chinese (英汉小句复合体形合与意合取向对比). Shanghai: *Educational Linguistic Studies* (上海: 教育语言学研究), 54 - 66.
- Mu, Furong. (穆芙蓉). (2009). An analysis of the clause complexes in English advertisements (英语广告词中的小句复合体特征分析). Beijing: *China Science and Technology Information* (北京: 中国科技信息), (20), 228-229.
- Peng, Xuanwei. (彭宣维). (2000). *A comprehensive comparison between English and Chinese texts* (英汉语篇综合对比). Shanghai: Shanghai Foreign Language Education Press (上海: 上海外语教育出版社).
- Qi, Wenjing. (齐文静). (2016). *A comparative and contrastive study on logico-semantic relations between English and Chinese clauses in text* (英汉小句在语篇中的逻辑语义关系比较与对比研究). Tianjin: Tianjin University of Technology (天津: 天津理工大学). Master's thesis.
- Qin, Xiaoqing., Wen, Qiufang. (秦晓晴, 文秋芳). (2007). *EFL writing of college English majors in China: A developmental perspective* (中国大学生英语写作能力发展规律与特点研究). Beijing: China Social Sciences Press (北京: 中国社会科学出版社).
- Qiu, Yanping. (裘燕萍). (2007). Scare quotes and their evaluative functions in journalistic genre (部分投射及其在新闻语类中的评价功能). Shanghai: *Journal of Foreign Languages* (上海: 外国语), (3), 32-38.

- Ren, Minzhen. (任明珍). (2020). *Study on projection in Chinese and American English news discourse: A case of reports on the south China sea in THE NAVYTIMES and THE CHINA* (中美英语新闻语篇中的投射研究) ---- 以美国海军时报网和中国日报网涉南海新闻报道为研究对象). Zhenjiang: Jiangsu University (镇江: 江苏大学). Master's thesis.
- Song, Yanmei., Wang, Huoyan. (宋瑞梅, 汪火焰). (2020). A comparative study of citations in the literature review part of graduates' dissertations (硕博连读研究生学位论文文献综述中的引用对比研究). *Nanjing: Foreign Language Research* (南京: 外语研究), 37(2), 49-55.
- Su, Jianhong. (苏建红). (2006a). A contrastive study of Chinese and English clause complexes (汉英小句复合体对比研究). Shanghai: Shanghai International Studies University (上海: 上海外国语大学). Master's thesis.
- Su, Jianhong. (苏建红). (2006b). A contrastive study of cohesion with English and Chinese clause complexes (英汉小句复合体内部衔接对比研究). *Suzhou: Journal of Jiangsu University* (苏州: 江苏大学学报), 8(6), 64-68.
- Sun, Jianqiu. (孙建秋). (1998). *Study guide for the use of developing business contacts* (中级商务英语学习指导). Beijing: Foreign Language Teaching and Research Press (北京: 外语教学与研究出版社).
- Sun, Yansu. (孙厌舒). (2016). A study on citing practice in English academic writings by EFL learners (学习者学术英语写作中的引用行为研究). *Harbin: Foreign Language Research* (哈尔滨: 外语学刊), 1, 132-136.
- Sun, Yinghui. (孙迎晖). (2010). Generic structure of Chinese MA thesis introductions (中国学生英语专业硕士论文“前言”部分的语类结构模式分析). *Beijing: Foreign Language in China* (北京: 中国外语), 7(6), 54-60.
- Tan, Xiaochen., Dong, Rongyue. (谭晓晨, 董荣月). (2007). The influence of task type and language proficiency on oral accuracy and complexity (任务类型和语言水平对英语专业大学生口语准确性和复杂性的影响). *Nanjing: Journal of PLA University of foreign languages* (南京: 解放军外国语学院学报), 5, 54-58. DOI:10.3969/j.issn.1002-722X.2007.05.009.
- Tang, Jincan., Fan, Weiwei. (唐金灿, 樊葳葳). (2008). The influence of native thinking on English academic paper writing (母语思维对英语学术论文写作的影响). *Wuhan: Research in Higher Education of Engineering* (Supplement) (武汉: 高等工程教育研究) (增刊), 55-58.
- Tang, Lina. (唐丽娜). (2016). *A functional study of projections in American diplomatic speeches* (美国外交发言稿中投射的功能研究). Lanzhou: Northwest Normal University (兰州: 西北师范大学). Master's thesis.
- Wang, Genlian. (王根莲). (2014a). An interpretation of the syntactics and semantics of “fact” projection (“事实”投射的句构与功能语义阐释). *Beijing: Journal of University of Science and Technology Beijing* (北京: 北京科技大学学报), 30(3), 10-16.
- Wang, Genlian. (王根莲). (2014b). The rankshift phenomenon of projecting system (投射系统的级转移现象). *Nanjing: Foreign Languages Research* (南京: 外语研究), (5), 7-11.
- Wang, Jing. (王净). (2008). The logico-semantic function of English “projecting” clause complexes (英语“投射”小句复合体的逻辑-语义功能初探). *Haikou: Journal of Hainan Radio & TV University* (海口: 海南广播电视大学学报), (3), 17-20.
- Wang, Jingjing., Jiang, Feng. (王晶晶, 姜峰). (2019). The study of stance construction in academic writing by Chinese Ph.D. students of science and engineering and journal article writers (中国理工科博士生学术论文写作立场建构研究). *Shanghai: Foreign Language World* (上海: 外语界), 3, 23-31.
- Wang, Jingjing., Lv, Zhongshe. (王晶晶, 吕中舌). (2017). The study of self-mentions in academic writing by Chinese Ph.D. students of science and engineering and journal article writers (理工科博士生学术英语写作中的作者自我指称语研究). *Shanghai: Foreign Language World* (上海: 外语界), (2), 89-96.
- Wang, Li. (王丽). (2013). A brief study of projecting clause complexes in different discourse genres (不同语篇体裁中投射型小句复合体浅析). *Jingzhou: Journal of Yangtze University* (荆州: 长江大学学报), 26(7), 103-104, 163.

- Wang, Quanzhi. (王全智). (2008). A comparative study of clause complexes and *Fuju* (小句复合体与复句的比对研究). Dalian: *Foreign Languages and Their teaching* (大连: 外语与外语教学), (11), 9-12.
- Wang, Quanzhi. (王全智). (2015). Differences of clause complexes and *Fuju* under the perspective of finiteness (限定性视角下的小句复合体与复句之研究差异). Xi'an: *Foreign Language Education* (西安: 外语教学), 36(4), 37-40.
- Wang, Ruiqi. Luo, Zhiyong., Liu, Xiang., Han, Ruifang., Li, Shuxin. (王瑞琦, 罗智勇, 刘祥, 韩瑞昉, 李舒馨). (2021). Machine reading comprehension based on clause complex (基于小句复合体的中文机器阅读理解研究). The proceeding of the *twentieth China National Conference on Computational Linguistics* (第二十届中国计算语言学大会论文集), pp. 723-735.
- Wang, Xiaoyan. (汪晓燕). (2011). *Projection theory of functional grammar and English reading teaching* (功能语法的投射理论与英语阅读教学). Qingdao: Ocean University of China (青岛: 中国海洋大学). Master's thesis.
- Wang, Yannan. (王亚南). (2014). *A comparative study on the use of logical connectives in academic writing* (基于语料库的学术语篇逻辑连接词使用对比研究). Xinxiang: Henan Normal University (新乡: 河南师范大学). Master's thesis.
- Wang, Yueli., Xu, Hongliang. (王月丽, 徐宏亮). (2019). A study on the developmental features of the use of first-person pronouns in academic writing and writer identity construction among Chinese English learners (中国英语学习者学术写作中第一人称使用发展特征与身份构建研究). Beijing: *Foreign Language Education in China* (北京: 外语教育研究前沿), 2(3), 58-64.
- Wu, Xiaochun. (吴晓春). (2012). *A corpus-based study of the logical connectors in Chinese EFL argumentative writing* (基于语料库的中国外语学习者议论文写作中逻辑连接词的使用研究). Changchun: Northeast Normal University (长春: 东北师范大学). Doctoral thesis.
- Wu, Yafei. (吴雅菲). (2019). A complexity study of logical relationship of clause complex: A comparison of the difficulty between the original and simplified version of *Alice's Adventures* (小句复合体关系体系之复杂性研究 ---- 《爱丽丝漫游奇遇记》原版与简写版难易度对比). Tianjin: *Journal of Tianjin Foreign Studies University* (天津: 天津外国语大学学报), 26(5), 79-86.
- Xiao, Junhong. (肖俊洪). (2000). On "hypotaxis" from the systemic functional grammar (从功能句法角度看“主从关系”). Dalian: *Foreign Languages and Their teaching* (大连: 外语与外语教学), (4), 17-20.
- Xie, Lei. (谢雷). (2011). A research on metafunction of projection clause in news discourse (投射小句在新闻语篇中的元功能研究). Harbin: *Northeast Agricultural University* (哈尔滨: 东北农业大学). Master's thesis.
- Xing, Fukun., Xu, Jianing. (邢富坤, 徐佳宁). (2023). Investigation of the clause complexes transfer and generation capability from Chinese to English for machine translation (面向机器翻译的汉英小句复合体转换生成能力调查). The proceeding of the *22nd China National Conference on Computational Linguistics* (第二十二届中国计算语言学大会论文集), pp. 102-112.
- Xu, Fang. (徐昉). (2011). A corpus study of identity chunks in Chinese students academic English writing (中国学生学术英语写作中的身份语块的语料库研究). Nanjing: *Foreign Language Research* (南京: 外语研究), 3, 57-63.
- Xu, Fang. (徐昉). (2012). Citation features in English empirical research discourse (实证类英语学术研究话语中的文献引用特征). Shanghai: *Journal of Foreign Languages* (上海: 外国语), 35(6), 60-68.
- Xu, Fang. (徐昉). (2016). The developmental features of citation competence in L2 academic writing: Cross-sectional and longitudinal evidence 二语学术写作的引证能力及其发展特征: 截面与历时证据. Shanghai: *Journal of Foreign Languages* (上海: 外国语), 39 (3), 73-82.
- Xu, Fang. (徐昉). (2017). International publishing and professional identity construction by Chinese researchers on foreign language teaching (国际发表与中国外语教学研究者的职业身份建构). Dalian: *Foreign Languages and Their teaching* (大连: 外语与外语教学), 1, 26-32.
- Xue, Junjie. (薛俊洁). (2011). *The study of expansion relation of clauses in MARPOL73/78 convention annexes* (MARPOL73/78 公约附则中小句扩展关系研究). Dalian: Dalian Maritime University (大连: 大连海事大学). Master's thesis.

- Yang, Hong. (杨红). (2009). A summary of the study on clause complex in China (国内小句复合体研究综述). *Guangzhou: Educational Research on Foreign Languages and Arts* (广州: 外语艺术教育研究), (12), 36-39.
- Yang, Xiaomei. (杨晓妹). (2008). *A study on cohesive devices within the English clause complex* (英语小句复合体内部衔接手段研究). Baoding: Hebei University (保定: 河北大学). Master's thesis.
- Yang, Xinran. (杨欣然). (2015). The authorial first person pronouns and writers' identity construction in second language academic writing (二语学术写作中的自我指称与作者身份建构). Dalian: *Foreign Languages and Their teaching* (大连: 外语与外语教学), 4, 50-56.
- Ye, Dandan. (叶丹丹). (2018). *Error analysis of English-Chinese machine translation output under the framework of clause complex theory* (小句复合体理论框架下的英汉及其翻译译文错误分析). Guangzhou: Guangdong University of Foreign Studies (广州: 广东外语外贸大学). Master's thesis.
- Zeng, Lei. (曾蕾). (2000a). Projection and discourse analysis from the perspective of functional linguistics (从功能语言学角度看“投射”与语篇分析). Dalian: *Foreign Languages and Their teaching* (大连: 外语与外语教学), (11), 15-17.
- Zeng, Lei. (曾蕾). (2000b). A functional approach to projection clause nexus in English and Chinese (英汉“投射”小句复合体的功能与语义分析). Guangzhou: *Modern Foreign Languages* (广州: 现代外语), (2), 163-173.
- Zeng, Lei. (曾蕾). (2002). From the logical to the experiential: An alternative ideational way into the analysis of “projection” in SFG (从逻辑功能到经验功能 ---- 扩展“投射”现象的概念功能模式). Guangzhou: *Modern Foreign Languages* (广州: 现代外语), 25(3), 269-275.
- Zeng, Lei. (曾蕾). (2003). A study of the syntactic and semantic features of “projection” from the perspective of ideational metaphor in SFG (论系统功能语法中“投射”概念隐喻句构及其语义特征). Guangzhou: *Modern Foreign Languages* (广州: 现代外语), 26(4), 351-357.
- Zeng, Lei. (曾蕾). (2007). “Projection” in academic discourse: from the perspective of grammatical metaphor (从语法隐喻视角看学术语篇中的“投射”). Harbin: *Foreign Language Research* (哈尔滨: 外语学刊), (3), 46-49.
- Zeng, Lei. (曾蕾). (2008). An approach to the metaphorical meanings of the tense of projecting verbs in English academic discourse (英语学术语篇中投射动词时态的语法隐喻). Beijing: *Journal of University of Science and Technology Beijing* (北京: 北京科技大学学报), 24(2), 104-108.
- Zeng, Lei. (曾蕾). (2016). From projection clause nexus to projection text: A case study of *Lunyu* and its English version (从投射小句复合体到投射语段 ---- 以《论语》原文与译文的对等分析为例). Guangzhou: *Modern Foreign Languages* (广州: 现代外语), 39(1), 42-51.
- Zeng, Lei., Hu, Jin. (曾蕾, 胡瑾). (2007). An approach to the multimodality of “projection in academic discourse (学术话语中的多模式“投射”). Loudi: *Journal of Hunan Institute of Humanities Science and Technology* (娄底: 湖南人文科技学院学报), (5), 141-145.
- Zeng, Lei., Liang, Hongyan. (曾蕾, 梁红艳). (2007). Multimodal metaphorical modes of “projection” in academic discourse (学术话语中多模式投射“隐喻式”). Xi'an: *Journal of Xi'an International Studies University* (西安: 西安外国语大学学报), 15(4), 14-18.
- Zeng, Lei., Liang, Hongyan. (曾蕾, 梁红艳). (2012). A study of facts in systemic functional linguistics (“事实”定位及其投射系统). Guangzhou: *Modern Foreign Languages* (广州: 现代外语), 35(1), 23-29.
- Zeng, Lei., Yang, Muwen. (曾蕾, 杨慕文). (2016). Discussion on the construction of projection analyzing framework in graphic discourse (图文语篇中的投射分析框架构建探讨). Xi'an: *Journal of Xi'an International Studies University* (西安: 西安外国语大学学报), 24(4), 35-38.
- Zeng, Lei., Yang, Muwen. (曾蕾, 杨慕文). (2019). A study of projection system in academic comics (学术漫画的投射系统研究). Guangzhou: *Modern Foreign Languages* (广州: 现代外语), 42(5), 610-622.
- Zeng, Lei., Yu, Hui. (曾蕾, 于晖). (2005). Constructing the interpersonal meaning and its scale modes of “projection signal” (“投射符号”的人际意义及其等级模式之构建). Xi'an: *Foreign Language Education* (西安: 外语教学), 26(6), 25-29.
- Zhang, Chunyan., Ran, Yongping. (张春燕, 冉永平). (2024). A corpus-based analysis of the dynamic logico-semantic relations of And-clause complexes (基于语料库的and小句复合体的逻辑语义动态性研究). Shanghai: *Technology Enhanced Foreign Language Education* (上海: 外语电化教学), 1, 85-92.

- Zhang, Liyin., Chen, Xinren. (张立茵, 陈新仁). (2020). A study on citation practices by expert writers as evaluators (专家作为评价者的引用行为研究). Dalian: *Foreign Languages and Their teaching* (大连: 外语与外语教学), 6, 121-130.
- Zhang, Shuhui. (张书慧). (2009). The classification and their functional semantic features of English "projection" clauses (谈英语‘投射’小句复合体的分类及其功能语义特征). Hohhot: *Journal of Language and Literature Studies* (呼和浩特: 语文学刊), (12), 11-2, 39.
- Zhang, Xuezhen. (张学贞). (2022). *A contrastive study of Chinese-English-Japanese NT clause complexes based on a self-built trilingual comparable corpus* (基于汉英日三语可比语料库的NT小句复合体对比研究). Qingdao: Qingdao University (青岛: 青岛大学). Master's thesis.
- Zhang, Yaoting. (张耀庭). (2011). A study of clause complexes in systemic functional grammar (系统功能语法中的小句复合体研究). Suihua: *Journal of Suihua University* (绥化: 绥化学院学报), 31(3), 177-190.
- Zhang, Yubo. (张玉波). (2016). *Re-describing the finite system in the English clause from the systemic functional perspective* (英语小句限定系统的功能语言学再描写). Chongqing: Southwest University (重庆: 西南大学). Doctoral Thesis.
- Zhao, Lu. (赵璐). (2008). *A functional study on the interpersonal meanings of projection in academic articles* (学术论文中投射的人际意义的功能研究). Shijiazhuang: Hebei Normal University (石家庄: 河北师范大学). Master's thesis.
- Zhao, Mengjuan., Zhong, Lanfeng. (赵梦娟, 钟兰凤). (2016). Citation research in English academic papers written by students and researchers (学生研究者英语学术论文中的引用研究). Fuzhou: *Foreign Language and Literature Studies* (福州: 外国语言文学), 33(1), 32-41.
- Zhao, Wanchang. (赵万长). (2006). *Projection and its application in language teaching* (投射理论及其在语言教学中的应用). Shanghai: Shanghai International Studies University (上海: 上海外国语大学). Master's thesis.
- Zhao, Yiyang., Zhang, Le. (赵亦阳, 张乐). (2022). A corss-disciplinary contrastive study on evaluator identity construction in journal article citations (期刊论文引用行为中评价者身份构建的学科对比研究). Shanghai: *Journal of University of Shanghai for Science and Technology* (上海理工大学学报), 44(2), 141-149.
- Zheng, Shutang. (郑树棠). (2001). *New horizon college English: Reading and writing 1* (新视野大学英语读写教程I). Beijing: Foreign Language Teaching and Research Press (北京: 外语教学与研究出版社).
- Zhong, Kexin. (仲可欣). (2023a). The analysis of clause complex features in English paper abstracts (英文论文摘要的小句复合体特征分析). Hefei: *Overseas English* (合肥: 海外英语), 1, 15-17.
- Zhong, Kexin. (仲可欣). (2023b). *A corpus-based study of English clause complex features in research article abstracts* (基于语料库的英文学术论文摘要小句复合体特征研究). Qingdao: Qingdao University (青岛: 青岛大学). Master's thesis.
- Zhong, Weiwei. (仲微微). (2020). How images and texts work in the translation of narrative picture books (叙事绘本翻译中图文逻辑语义关系的处理). Jining: *Journal of Jining University* (济宁: 济宁学院学报), 41(5), 25-32.
- Zhongwei. (仲伟). (2020). A contrastive study of the logical relations and their realizations between German and Chinese clauses: A systemic functional linguistics approach (系统功能语言学视角下的德汉小句逻辑关系及其体现形式对比研究). Beijing: *Beijing Foreign Studies University* (北京: 北京外国语大学). Doctoral thesis.
- Zhu, Yue. (朱跃). (2005). Primary clause, secondary clause and message conveyance: A comment on Halliday's clause complex theory (首要子句、次要子句与信息传递 --- 兼评Halliday的子句复合体思想.) Sichuan: *Journal of Sichuan International Studies University* (四川: 四川外语学院学报), 21(2), 74-77.
- Zhumu, Langma. (朱穆郎玛). (2018). *Identification of dependency relations between clauses within clause complexes based on the MOOD system* (语气系统下小句复合体小句依存关系识别研究). Dalian: Dalian University of Technology (大连: 大连理工大学). Master's thesis.

APPENDIX

Appendix 4.3.2.1

Year	M	D
2005	3	2
2006	3	1
2007	3	3
2008	3	3
2009	3	1
2010	7	4
2011	7	2
2012	7	1
2013	7	2
2014	7	1
2015	5	1
2016	5	1
2017	5	4
2018	5	2
2019	5	2
2020	5	2
	80	32

Direction	Number
Culture	6
Literature	5
Listening	4
Speaking	3
Reading	4
Writing	3
Phonology	4
Morphology	5
Syntax	4
Semantics	4
Pragmatics	3
Methodology	5
	50

Table A4.3.2.1-1 The number of theses chosen in each year in M and D groups and in each direction in M group (2005 to 2014)

CMAC				CPhD			
Name	Words	CC	½ CC	Name	Words	CC	½ CC
CMAC05ME_21	10342	458	229	CPhD05_03	59503	2159	1080
CMAC05PR_24	18504	809	405	CPhD05_08	30333	1304	652
CMAC05WR_20	13668	550	275	CPhD06_04	64331	2502	1251
CMAC06LIT_31	19168	818	409	CPhD07_04	27061	1141	571
CMAC06MO_18	13343	538	269	CPhD07_08	49745	1529	765
CMAC06SY_17	16186	643	322	CPhD07_11	33873	1341	671
CMAC07CU_35	16549	569	285	CPhD08_01	36280	1582	791
CMAC07LI_20	16908	719	360	CPhD08_05	33763	1320	660
CMAC07RE_19	15559	769	385	CPhD08_12	47983	1712	856
CMAC08CU_33	15541	863	432	CPhD09_02	41028	1564	782
CMAC08SE_25	12444	609	305	CPhD10_01	22769	998	499
CMAC08SP_18	12422	535	268	CPhD10_09	37235	1688	844
CMAC09ME_24	15563	730	365	CPhD10_15	58466	2380	1190
CMAC09PH_21	14226	518	259	CPhD10_18	35096	1307	654
CMAC09PR_28	15540	581	291	CPhD11_06	44692	1644	822
CMAC10LIT_15	14475	715	358	CPhD11_07	43471	1654	827
CMAC10LIT_21	13496	739	370	CPhD12_09	46742	1758	879
CMAC10LIT_7	17803	759	380	CPhD13_03	52119	1845	923
CMAC10PR_19	15466	714	357	CPhD13_10	55091	1883	942
CMAC10SE_18	15851	638	319	CPhD14_02	39172	1850	925
CMAC10SP_11	12941	658	329	CPhD15_01	45987	1732	866
CMAC10SY_9	14323	684	342	CPhD16_01	45319	2104	1052
CMAC11CU_15	16849	886	443	CPhD17_01	49098	1927	964

CMAC11LI_9	14908	557	279	CPhD17_02	51799	2732	1366
CMAC11ME_16	12657	569	285	CPhD17_03	68483	2757	1379
CMAC11PH_9	13583	530	265	CPhD17_04	59421	2234	1117
CMAC11PR_10	17985	627	314	CPhD18_0001	42301	2064	1032
CMAC11SE_19	24725	1188	594	CPhD18_0002	45364	1602	801
CMAC11WR_9	15562	684	342	CPhD19_0001	27315	993	497
CMAC12LI_6	16629	623	312	CPhD19_0002	61683	1915	958
CMAC12ME_5	14310	694	347	CPhD20_0001	37415	1243	622
CMAC12MO_3	11819	519	260	CPhD20_0002	47373	2038	1019
CMAC12PH_1	9979	479	240	32	1440311	56502	28257
CMAC12SE_5	13517	543	272				
CMAC12SP_5	17699	574	287				
CMAC12SY_13	17731	862	431				
CMAC13CU_21	14575	684	342				
CMAC13CU_27	15753	738	369				
CMAC13LI_12	14933	692	346				
CMAC13ME_11	15206	620	310				
CMAC13RE_10	15393	716	358				
CMAC13RE_2	10280	469	235				
CMAC13SY_7	12586	524	262				
CMAC14CU_8	11238	576	288				
CMAC14LIT_25	15288	695	348				
CMAC14MO_11	14026	675	338				
CMAC14PH_17	13542	652	326				
CMAC14PR_4	24486	862	431				
CMAC14RE_15	14339	672	336				
CMAC14WR_4	14001	658	329				
CMAC15_01	21546	994	497				
CMAC15_02	15103	790	395				
CMAC15_03	15076	516	258				
CMAC15_04	15983	527	264				
CMAC15_05	21647	976	488				
CMAC16_01	13100	552	276				
CMAC16_02	20333	970	485				
CMAC16_03	19448	920	460				
CMAC16_04	17232	755	378				
CMAC16_05	19296	834	417				
CMAC17_01	11361	522	261				
CMAC17_02	19843	736	368				
CMAC17_03	16891	583	292				
CMAC17_04	12129	563	282				
CMAC17_05	18148	652	326				
CMAC18_01	12375	400	200				
CMAC18_02	11804	509	255				
CMAC18_03	18797	930	465				
CMAC18_04	24775	1341	671				
CMAC18_05	16891	809	405				
CMAC19_01	16662	831	416				

CMAC19_02	15038	584	292				
CMAC19_03	19105	903	452				
CMAC19_04	15066	594	297				
CMAC19_05	17315	742	371				
CMAC20_01	30614	1614	807				
CMAC20_02	25484	1295	648				
CMAC20_03	14105	538	269				
CMAC20_04	15667	594	297				
CMAC20_05	18130	909	455				
80	1282881	56667	28350				

Table A4.3.2.1-2 The number of words, clauses & annotated clause complexes (sentences) in the CMAC and CPhD corpus

Appendix 4.3.3.6.1

ID	CC Nexuses (1-7)	1	2	3	4	5	6	7
CMAC05ME_21	193	123	50	13	5	2		
CMAC05PR_24	314	213	77	20	3	1		
CMAC05WR_20	220	143	55	18	2	1	1	
CMAC06LIT_31	339	203	106	24	3	2	1	
CMAC06MO_18	246	159	68	14	5			
CMAC06SY_17	306	188	83	31	4			
CMAC07CU_35	302	187	83	20	11	1		
CMAC07LI_20	328	193	91	31	11	1	1	
CMAC07RE_19	257	189	58	10				
CMAC08CU_33	229	181	40	6	2			
CMAC08SE_25	201	145	45	8	2	1		
CMAC08SP_18	213	144	55	13	1			
CMAC09ME_24	317	206	77	30	3	1		
CMAC09PH_21	149	105	37	6	1			
CMAC09PR_28	309	208	77	20	3	1		
CMAC10LIT_15	230	164	52	12	2			
CMAC10LIT_21	276	195	63	16	2			
CMAC10LIT_7	489	242	155	67	18	7		
CMAC10PR_19	341	207	96	27	9	1	1	
CMAC10SE_18	431	216	136	57	14	6	1	1
CMAC10SP_11	251	180	55	16				
CMAC10SY_9	335	207	97	27	4			
CMAC11CU_15	291	203	70	13	4	1		
CMAC11LI_9	377	211	127	34	4	1		
CMAC11ME_16	290	177	84	23	5	1		
CMAC11PH_9	205	141	49	9	4	1	1	
CMAC11PR_10	314	200	84	24	6			
CMAC11SE_19	525	321	147	48	7	2		
CMAC11WR_9	244	184	53	7				
CMAC12LI_6	409	230	125	43	8	2	1	
CMAC12ME_5	312	188	92	28	4			
CMAC12MO_3	231	152	63	14	2			
CMAC12PH_1	159	117	38	4				

CMAC12SE_5	223	147	60	14	2			
CMAC12SP_5	385	204	121	44	14	2		
CMAC12SY_13	326	232	78	13	3			
CMAC13CU_21	168	134	30	4				
CMAC13CU_27	341	215	90	26	9	1		
CMAC13LI_12	259	167	63	20	9			
CMAC13ME_11	311	200	86	21	4			
CMAC13RE_10	400	241	105	46	8			
CMAC13RE_2	221	138	57	17	5	2	2	
CMAC13SY_7	273	173	73	18	7	2		
CMAC14CU_8	168	132	29	7				
CMAC14LIT_25	408	214	124	46	19	5		
CMAC14MO_11	209	156	43	10				
CMAC14PH_17	213	149	48	14	2			
CMAC14PR_4	671	342	194	82	32	14	6	1
CMAC14RE_15	264	179	67	18				
CMAC14WR_4	261	178	56	20	5	2		
CMAC15_01	529	333	147	36	11	2		
CMAC15_02	254	193	55	6				
CMAC15_03	220	131	57	24	6	2		
CMAC15_04	275	171	69	27	5	3		
CMAC15_05	619	340	190	72	15	2		
CMAC16_01	262	171	64	21	4	2		
CMAC16_02	573	337	172	49	11	4		
CMAC16_03	304	211	68	23	2			
CMAC16_04	368	236	96	31	5			
CMAC16_05	350	247	82	20	1			
CMAC17_01	215	146	56	11	2			
CMAC17_02	434	251	118	45	14	4	1	1
CMAC17_03	341	195	104	35	5	2		
CMAC17_04	222	153	54	12	3			
CMAC17_05	379	199	111	47	13	9		
CMAC18_01	203	127	52	20	3	1		
CMAC18_02	209	135	60	11	2	1		
CMAC18_03	376	251	98	26	1			
CMAC18_04	545	367	140	33	5			
CMAC18_05	322	222	86	14				
CMAC19_01	340	227	79	24	5	4	1	
CMAC19_02	366	204	113	38	11			
CMAC19_03	309	217	67	18	6	1		
CMAC19_04	319	189	87	31	10	2		
CMAC19_05	329	198	86	34	10	1		
CMAC20_01	533	389	116	25	3			
CMAC20_02	590	376	160	44	9	1		
CMAC20_03	222	151	58	12	1			
CMAC20_04	348	205	101	36	5	1		
CMAC20_05	236	189	39	7	1			

Table A4.3.3.6.1-1 The number of clause nexus on each recursion level in each text in CMAC group

ID	CC Nexuses (1-7)	1	2	3	4	5	6	7
CPhD05_03	1139	719	316	87	14	3		
CPhD05_08	459	332	107	19	1			
CPhD06_04	1078	699	264	81	27	6	1	
CPhD07_04	415	279	106	25	4	1		
CPhD07_08	834	474	232	89	27	11	1	
CPhD07_11	533	352	133	35	11	2		
CPhD08_01	739	485	187	57	10			
CPhD08_05	550	370	129	41	8	1	1	
CPhD08_12	883	540	256	74	11	2		
CPhD09_02	684	447	179	49	8	1		
CPhD10_01	315	223	72	14	4	1	1	
CPhD10_09	503	355	113	29	6			
CPhD10_15	855	585	212	45	10	3		
CPhD10_18	753	430	198	86	28	10	1	
CPhD11_06	717	470	190	42	12	3		
CPhD11_07	752	451	219	62	19	1		
CPhD12_09	706	452	183	58	11	2		
CPhD13_03	836	545	215	60	15	1		
CPhD13_10	914	586	235	65	22	4	2	
CPhD14_02	672	450	170	42	8	2		
CPhD15_01	805	509	215	66	13	2		
CPhD16_01	838	578	195	55	7	1	1	1
CPhD17_01	887	562	231	67	24	3		
CPhD17_02	1220	760	334	93	30	2	1	
CPhD17_03	1004	718	222	45	15	4		
CPhD17_04	909	656	209	40	4			
CPhD18_0001	723	546	158	18	1			
CPhD18_0002	623	447	139	31	6			
CPhD19_0001	506	320	141	39	5	1		
CPhD19_0002	1014	620	285	81	23	5		
CPhD20_0001	469	344	111	10	3	1		
CPhD20_0002	787	539	205	41	2			

Table A4.3.3.6.1-2 The number of clause nexus on each recursion level in each text in CPhD group

Appendix 4.3

Descriptive Statistics	Table 4.3.3.1-1		Table 4.3.3.2-1		Table 4.3.3.3-1	
	Clauses No.		Words/Clauses		Clause Complex No.	
	D	M	D	M	D	M
Valid	32	80	32	80	32	80
Missing	0	0	0	0	0	0
Mean	1765.688	708.337	25.678	23.102	495.094	201.05
Std. Error of Mean	80.385	22.981	0.56	0.354	23.811	6.789
Median	1722	673.5	25.973	22.212	479.5	194
Mode ^a	993	684	18.96	18.008	447	189
Std. Deviation	454.727	205.545	3.166	3.168	134.695	60.72
Variance	206776.93	42248.809	10.026	10.038	18142.797	3686.959
Skewness	0.392	1.866	0.074	0.744	0.091	1.353
Std. Error of Skewness	0.414	0.269	0.414	0.269	0.414	0.269
Kurtosis	-0.133	5.064	0.028	-0.209	-0.496	1.9
Std. Error of Kurtosis	0.809	0.532	0.809	0.532	0.809	0.532
Shapiro-Wilk	0.973	0.852	0.985	0.938	0.98	0.873
P-value of Shapiro-Wilk	0.571	< .001	0.931	< .001	0.799	< .001
Range	1764	1214	13.574	12.93	537	284
Minimum	993	400	18.96	18.008	223	105
Maximum	2757	1614	32.534	30.938	760	389
Sum	56502	56667			15843	16084

^a More than one mode exists, only the first is reported

(continue)

Descriptive Statistics	Table 4.3.3.4-1		Table 4.3.3.5-1		Table 4.3.3.7-1		Table 4.3.3.8-1	
	Clause Complex Perc.		C Nexuses per CC		Parataxis CC Per.		Hypotaxis No.	
	D	M	D	M	D	M	D	M
Valid	32	80	32	80	32	80	32	80
Missing	0	0	0	0	0	0	0	0
Mean	56.045	57.064	1.518	1.559	34.888	39.005	488.219	193.125
Std. Error of Mean	1.058	0.971	0.019	0.019	1.123	0.844	24.956	8.022
Median	55.84	54.763	1.525	1.548	34.944	37.691	485	181
Mode ^a	42.062	39.181	1.324	1.249	18.124	24.39	199	147
Std. Deviation	5.984	8.688	0.107	0.172	6.354	7.547	141.171	71.752
Variance	35.806	75.479	0.011	0.03	40.372	56.964	19929.144	5148.313
Skewness	-0.221	0.258	0.352	0.604	-0.282	0.315	0.343	1.098
Std. Error of Skewness	0.414	0.269	0.414	0.269	0.414	0.269	0.414	0.269
Kurtosis	-0.19	-0.389	-0.088	0.248	0.203	-0.469	0.397	1.266
Std. Error of Kurtosis	0.809	0.532	0.809	0.532	0.809	0.532	0.809	0.532
Shapiro-Wilk	0.982	0.982	0.974	0.969	0.979	0.976	0.985	0.924
P-value of Shapiro-Wilk	0.849	0.302	0.63	0.05	0.757	0.141	0.934	< .001
Range	24.512	40.169	0.435	0.772	28.075	31.331	654	340
Minimum	42.062	39.181	1.324	1.249	18.124	24.39	199	78
Maximum	66.574	79.35	1.759	2.021	46.199	55.721	853	418
Sum							15623	15450

^a More than one mode exists, only the first is reported

Appendix 4.3-1 The descriptive statistics of the number of clauses, the number of words in each clause, the number of clause complexes, the clause complex percentage, the clause nexuses number in each clause complex, the parataxis percentage in all clause complexes, and the hypotaxis percentage in all clause complexes in CMAC and CPhD groups

Descriptive Statistics	Table 4.3.3.6.1-1		Table 4.3.3.6.1-2					
	C Nexuses (1-7) No.		CNs L1 No.		CNs L2 No.		CNs L3 No.	
	D	M	D	M	D	M	D	M
Valid	32	80	32	80	32	80	32	80
Missing	0	0	0	0	0	0	0	0
Mean	753.813	316.575	495.094	201.05	192.531	83.713	51.438	24.813
Std. Error of Mean	38.14	12.472	23.811	6.789	10.935	4.092	4.105	1.767
Median	752.5	307.5	479.5	194	196.5	77	47	20.5
Mode ^a	315	341	447	189	215	55	41	20
Std. Deviation	215.75	111.549	134.695	60.72	61.859	36.601	23.22	15.808
Variance	46547.899	12443.235	18142.797	3686.959	3826.515	1339.65	539.157	249.901
Skewness	0.053	1.166	0.091	1.353	0.221	1.044	0.11	1.302
Std. Error of Skewness	0.414	0.269	0.414	0.269	0.414	0.269	0.414	0.269
Kurtosis	-0.288	1.203	-0.496	1.9	-0.081	0.808	-0.85	2.008
Std. Error of Kurtosis	0.809	0.532	0.809	0.532	0.809	0.532	0.809	0.532
Shapiro-Wilk	0.986	0.906	0.98	0.873	0.978	0.922	0.968	0.9
P-value of Shapiro-Wilk	0.946	< .001	0.799	< .001	0.744	< .001	0.44	< .001
Range	905	522	537	284	262	165	83	78
Minimum	315	149	223	105	72	29	10	4
Maximum	1220	671	760	389	334	194	93	82
Sum	24122	25326	15843	16084	6161	6697	1646	1985
^a More than one mode exists, only the first is reported								
^b All values are identical								

(continue)

Descriptive Statistics	Table 4.3.3.6.1-3							
	CNs L4 No.		CNs L5 No.		CNs L6 No.		CNs L7 No.	
	D	M	D	M	D	M	D	M
Valid	32	70	26	42	8	11	1	3
Missing	0	10	6	38	24	69	31	77
Mean	12.156	6.243	2.808	2.452	1.125	1.545	1	1
Std. Error of Mean	1.518	0.636	0.523	0.393	0.125	0.455	NaN	0
Median	10.5	5	2	2	1	1	1	1
Mode ^a	4	2	1	1	1	1	1	1
Std. Deviation	8.588	5.317	2.669	2.549	0.354	1.508	NaN	0
Variance	73.749	28.273	7.122	6.498	0.125	2.273	NaN	0
Skewness	0.708	2.165	2.029	2.98	2.828	3.104	NaN ^b	NaN ^b
Std. Error of Skewness	0.414	0.287	0.456	0.365	0.752	0.661	0	1.225
Kurtosis	-0.608	7.089	4.098	10.377	8	9.838	NaN ^b	NaN ^b
Std. Error of Kurtosis	0.809	0.566	0.887	0.717	1.481	1.279	0	∞
Shapiro-Wilk	0.912	0.8	0.746	0.603	0.418	0.432	NaN ^b	NaN ^b
P-value of Shapiro-Wilk	0.013	< .001	< .001	< .001	< .001	< .001	NaN ^b	NaN ^b
Range	29	31	11	13	1	5	0	0
Minimum	1	1	0	1	1	1	1	1

Maximum	30	32	11	14	2	6	1	1
Sum	389	437	73	103	9	17	1	3
^a More than one mode exists, only the first is reported								
^b All values are identical								

Appendix 4.3-2 The descriptive statistics of the number of clause nexuses in total and that on each recursion level (from 1 to 7) in CMAC and CPhD groups

Descriptive Statistics	Table 4.3.3.6.1-4							
	CNs L1 Per.		CNs L2 Per.		CNs L3 Per.		CNs L4 Per.	
	D	M	D	M	D	M	D	M
Valid	32	80	32	80	32	80	32	80
Missing	0	0	0	0	0	0	0	0
Mean	66.174	64.879	25.338	25.827	6.636	7.326	1.523	1.552
Std. Error of Mean	0.813	0.775	0.365	0.391	0.356	0.303	0.152	0.129
Median	65.59	64.614	25.627	26.055	7.138	6.987	1.312	1.342
Mode ^a	56.835	49.489	21.853	24.324	2.132	5.405	1.17	0
Std. Deviation	4.597	6.933	2.062	3.494	2.013	2.713	0.862	1.158
Variance	21.131	48.072	4.253	12.205	4.054	7.361	0.743	1.342
Skewness	-0.015	-0.021	-0.162	-0.377	-0.082	0.212	0.611	0.718
Std. Error of Skewness	0.414	0.269	0.414	0.269	0.414	0.269	0.414	0.269
Kurtosis	-0.383	-0.178	-0.908	0.27	0.639	-0.539	0.177	-0.023
Std. Error of Kurtosis	0.809	0.532	0.809	0.532	0.809	0.532	0.809	0.532
Shapiro-Wilk	0.982	0.99	0.963	0.981	0.961	0.983	0.957	0.938
P-value of Shapiro-Wilk	0.855	0.818	0.324	0.267	0.284	0.347	0.23	< .001
Range	18.684	30.596	7.269	17.162	9.289	11.339	3.58	4.769
Minimum	56.835	49.489	21.853	16.525	2.132	2.362	0.138	0
Maximum	75.519	80.085	29.122	33.687	11.421	13.701	3.718	4.769
^a More than one mode exists, only the first is reported								

Appendix 4.3-3 The descriptive statistics of clause nexus percentage on the recursion level 1-4 in CMAC and CPhD groups

Appendix 5.1

Descriptive Statistics	Table 5.1.1-1		Table 5.1.1.1-1		Table 5.1.2-1		Table 5.1.2.1-1		Table 5.1.2.2-1		Table 5.1.2.3-1	
	Expansion		Expansion-CC Perc.		Expansion Para		Expansion Para/CC Perc.		Expansion Para/Para Perc.		Expansion Para/Exp. Perc.	
	D	M	D	M	D	M	D	M	D	M	D	M
Valid	32	80	32	80	32	80	32	80	32	80	32	80
Missing	0	0	0	0	0	0	0	0	0	0	0	0
Mean	611.594	257.913	81.023	81.384	257.188	120.412	33.957	38.008	97.358	97.465	41.818	46.691
Std. Error of Mean	unknown	unknown	0.621	0.601	15.604	5.674	1.133	0.834	0.811	0.433	1.27	0.944
Median	602.5	246	80.94	82.409	271	107	33.801	36.943	99.576	99.174	41.562	45.528
Mode ^a	729	257	75.155	81.818	184	70	18.124	23.182	100	100	23.288	28.333
Std. Deviation	179.248	93.734	3.51	5.38	88.268	50.748	6.411	7.456	4.59	3.875	7.185	8.442
Variance	32129.67	8786.056	12.323	28.943	7791.254	2575.385	41.104	55.585	21.072	15.013	51.624	71.27
Skewness	0.029	1.193	0.176	-0.583	-0.139	1.579	-0.116	0.286	-2.924	-1.864	-0.205	0.085
Std. Error of Skewness	0.414	0.269	0.414	0.269	0.414	0.269	0.414	0.269	0.414	0.269	0.414	0.269
Kurtosis	-0.573	1.331	-0.869	0.26	-1.179	3.271	-0.135	-0.493	10.338	2.951	0.329	-0.259
Std. Error of Kurtosis	0.809	0.532	0.809	0.532	0.809	0.532	0.809	0.532	0.809	0.532	0.809	0.532
Shapiro-Wilk	0.976	0.905	0.966	0.975	0.949	0.875	0.981	0.977	0.627	0.706	0.986	0.987
P-value of Shapiro-Wilk	0.685	< .001	0.386	0.125	0.136	< .001	0.834	0.149	< .001	< .001	0.938	0.589
Range	679	451	12.255	27.67	307	264	27.294	31.068	22.193	15.942	33.712	37.871
Minimum	249	120	75.155	64.368	85	50	18.124	23.182	77.807	84.058	23.288	28.333
Maximum	928	571	87.41	92.038	392	314	45.418	54.25	100	100	57	66.204
Sum	19571	20633			8230	9633						

^a More than one mode exists, only the first is reported

(continue)

Descriptive Statistics	Table 5.1.3.1-1			Table 5.1.3.1-1			Table 5.1.3.2-1			Table 5.1.3.3-1		
	Expansion Hypo			Expansion Hypo/CC Perc.			Expansion Hypo/Hypo Perc.			Expansion Hypo/Exp. Perc.		
	D	M		D	M		D	M		D	M	
Valid	32	80		32	80		32	80		32	80	
Missing	0	0		0	0		0	0		0	0	
Mean	354.406	137.5		47.066	43.376		72.299	71.152		58.182	53.309	
Std. Error of Mean	19.473	6.037		0.976	0.823		0.894	0.943		1.27	0.944	
Median	352.5	128.5		46.812	43.645		72.075	73.05		58.438	54.471	
Mode ^a	133	85		34.263	46.041		62.774	69.231		43	33.796	
Std. Deviation	110.158	54.001		5.52	7.362		5.057	8.436		7.185	8.442	
Variance	12134.83	2916.101		30.47	54.193		25.575	71.162		51.625	71.27	
Skewness	0.377	1.08		0.096	-0.017		0.214	-0.806		0.205	-0.085	
Std. Error of Skewness	0.414	0.269		0.414	0.269		0.414	0.269		0.414	0.269	
Kurtosis	-0.168	1.291		0.273	0.355		-0.498	0.699		0.329	-0.259	
Std. Error of Kurtosis	0.809	0.532		0.809	0.532		0.809	0.532		0.809	0.532	
Shapiro-Wilk	0.97	0.925		0.987	0.991		0.982	0.946		0.986	0.987	
P-value of Shapiro-Wilk	0.499	< .001		0.963	0.864		0.86	0.002		0.938	0.589	
Range	465	259		25.438	38.836		20.877	42.972		33.712	37.87	
Minimum	133	54		34.263	25.172		62.774	47.753		43	33.796	
Maximum	598	313		59.701	64.008		83.651	90.725		76.712	71.667	
Sum	11341	11000										

^a More than one mode exists, only the first is reported

(continue)

Descriptive Statistics	Table 5.1.4.1-1			Table 5.1.4.1-2			Table 5.1.4.2-1			Table 5.1.4.2-2			Table 5.1.4.3-1			Table 5.1.4.3-2		
	Elaboration No.			Elaboration/Exp. Perc.			Extension No.			Extension/Exp. Perc.			Enhancement No.			Enhancement/Exp. Perc.		
	D	M		D	M		D	M		D	M		D	M		D	M	
Valid	32	80		32	80		32	80		32	80		32	80		32	80	
Missing	0	0		0	0		0	0		0	0		0	0		0	0	
Mean	87.063	39.55		14.409	15.293		242.75	107.425		39.386	42.032		281.781	110.938		46.205	42.674	
Std. Error of Mean	6.23	2.185		0.827	0.615		15.522	4.628		1.36	0.833		15.935	5.351		1.09	0.756	
Median	77.5	39		13.895	14.841		247	98		37.844	41.667		282.5	101		45.954	42.606	
Mode ^a	77	41		7.399	4.734		120	64		19.452	41.667		193	141		33	25	
Std. Deviation	35.241	19.544		4.678	5.499		87.804	41.396		7.691	7.447		90.142	47.865		6.167	6.766	
Variance	1241.931	381.972		21.883	30.244		7709.484	1713.665		59.158	55.457		8125.66	2291.047		38.031	45.774	
Skewness	0.768	0.867		0.781	0.264		-0.006	1.344		-0.005	0.312		0.46	1.454		0.038	0.359	
Std. Error of Skewness	0.414	0.269		0.414	0.269		0.414	0.269		0.414	0.269		0.414	0.269		0.414	0.269	
Kurtosis	0.669	1.025		0.44	-0.34		-1.107	2.403		0.484	0.231		0.233	2.331		-0.387	1.139	
Std. Error of Kurtosis	0.809	0.532		0.809	0.532		0.809	0.532		0.809	0.532		0.809	0.532		0.809	0.532	
Shapiro-Wilk	0.949	0.939		0.954	0.984		0.955	0.905		0.98	0.99		0.973	0.877		0.994	0.982	
P-value of Shapiro-Wilk	0.135	< .001		0.189	0.403		0.197	< .001		0.814	0.767		0.589	< .001		1	0.319	
Range	153	91		19.998	23.266		329	213		36.715	39.694		417	230		26.293	38.927	
Minimum	24	7		7.399	4.734		71	50		19.452	24.658		105	50		33	25	
Maximum	177	98		27.397	28		400	263		56.167	64.352		522	280		59.293	63.927	
Sum	2786	3164					7768	8594					9017	8875				

^a More than one mode exists, only the first is reported

Appendix 5.1 The descriptive statistics of comparing the number of expansions, the proportion of expansion in all clause complexes, the number of paratactic expansion, the proportion of paratactic expansion in all clause complexes, paratactic expansion in parataxis, paratactic expansion in all expansions, the number of hypotactic expansions, hypotactic expansion in all clause complexes, hypotactic expansion in hypotaxis, hypotactic expansion in all expansions, the number of elaborations, the proportion of

elaboration in all expansions, the number of extension, the proportion of extension in all expansions, the number of enhancements, the proportion of enhancement in all expansions in CMAc and CPhD groups

Appendix 5.2

Descriptive Statistics	Table 5.2.1-1		Table 5.2.2-1		Table 5.2.2.1-1		Table 5.2.2.2-1		Table 5.2.2.3-1		Table 5.2.2.4-1	
	Elaboration-CC Perc.		Para Elaboration		Elaboration Para Perc.		Para Elab.-Para Perc.		Ela. Para/Exp. Para Perc.		Ela. Para-Ela. Perc.	
	D	M	D	M	D	M	D	M	D	M	D	M
Valid	32	80	32	80	32	80	32	80	32	80	32	80
Missing	0	0	0	0	0	0	0	0	0	0	0	0
Mean	11.687	12.429	18.281	8.262	2.313	2.57	6.615	6.633	6.853	6.815	2	12.5
Std. Error of Mean	0.68	0.509	2.539	0.848	0.264	0.184	0.693	0.46	0.732	0.47	19	18.95
Median	11.037	12.405	13.5	6	1.925	2.231	5.907	5.513	5.95	6.105	19.655	22.659
Mode ^a	6.163	14.545	2	5	0.358	1.087	1.049	3.125	5.435	3.125	1.738	1.795
Std. Deviation	3.848	4.556	14.36	7.589	1.492	1.645	3.922	4.112	4.141	4.208	9.833	16.053
Variance	14.808	20.758	206.209	57.588	2.226	2.707	15.378	16.907	17.144	17.705	96.682	257.701
Skewness	0.694	0.393	1.062	3.439	0.699	1.546	0.395	1.244	0.419	1.242	0.095	1.763
Std. Error of Skewness	0.414	0.269	0.414	0.269	0.414	0.269	0.414	0.269	0.414	0.269	0.414	0.269
Kurtosis	-0.002	-0.171	-0.027	15.715	-0.655	4.218	-0.946	1.945	-0.995	2.076	-0.609	3.377
Std. Error of Kurtosis	0.809	0.532	0.809	0.532	0.809	0.532	0.809	0.532	0.809	0.532	0.809	0.532
Shapiro-Wilk	0.955	0.98	0.855	0.668	0.901	0.896	0.947	0.913	0.936	0.917	0.98	0.828
P-value of Shapiro-Wilk	0.193	0.245	<.001	<.001	0.007	<.001	0.12	<.001	0.058	<.001	0.809	<.001
Range	15.159	20.481	50	52	4.77	9.905	13.276	22.128	13.354	23.009	37.45	75
Minimum	6.163	4.225	2	0	0.358	0	1.049	0	1.079	0	2	0
Maximum	21.322	24.706	52	52	5.128	9.905	14.325	22.128	14.433	23.009	39.45	75
Sum			585	661								

^a More than one mode exists, only the first is reported

(continue)

Descriptive Statistics	Table 5.2.3-1			Table 5.2.3.1-1			Table 5.2.3.2-1 T			Table 5.2.3.3-1		
	Hypo Elaboration			Elab. Hypo-CC Perc.			Hypo Elab.-Hypo Perc.			Ela. Hypo/Exp. Hypo Perc.		
	D	M		D	M		D	M		D	M	
Valid	32	80		32	80		32	80		32	80	
Missing	0	0		0	0		0	0		0	0	
Mean	68.781	31.288		9.374	9.858		14.336	15.908		19.795	22.408	
Std. Error of Mean	4.484	1.867		0.608	0.485		0.804	0.711		1.046	0.951	
Median	64.5	30.5		8.78	9.308		14.014	15.613		20.042	22.067	
Mode^a	66	29		4.175	1.061		6.667	2.116		8.936	17.857	
Std. Deviation	25.365	16.702		3.438	4.339		4.549	6.355		5.915	8.506	
Variance	643.402	278.942		11.821	18.83		20.693	40.392		34.983	72.346	
Skewness	0.527	0.72		1.549	0.275		0.758	0.168		0.59	-0.01	
Std. Error of Skewness	0.414	0.269		0.414	0.269		0.414	0.269		0.414	0.269	
Kurtosis	0.52	1.288		3.522	-0.162		0.673	0.241		0.587	0.174	
Std. Error of Kurtosis	0.809	0.532		0.809	0.532		0.809	0.532		0.809	0.532	
Shapiro-Wilk	0.953	0.962		0.884	0.98		0.946	0.987		0.965	0.99	
P-value of Shapiro-Wilk	0.171	0.018		0.002	0.227		0.111	0.594		0.38	0.775	
Range	110	89		16.721	20.41		18.854	32.98		26.064	40.915	
Minimum	20	2		4.175	1.061		6.667	2.116		8.936	2.614	
Maximum	130	91		20.896	21.471		25.521	35.096		35	43.529	
Sum	2201	2503										

^a More than one mode exists, only the first is reported

Appendix 5.2 The descriptive statistics of comparing the proportion of elaboration in all clause complexes, the number of paratactic elaborations, the proportion of paratactic elaboration in all clause nexuses, paratactic elaboration in parataxis, paratactic elaboration in all paratactic expansions, paratactic elaboration in all elaborations, the number of hypotactic elaborations, hypotactic elaboration in all clause nexuses, hypotactic elaboration in hypotaxis, and hypotactic elaboration in all hypotactic expansions in CPhD groups

Appendix 5.4

Descriptive Statistics	Table 5.4.1-1		Table 5.4.2-1		Table 5.4.2.1-1		Table 5.4.2.2-1		Table 5.4.2.3-1		Table 5.4.2.4-1	
	Extension/CC Perc.		Para Extension		Extension Para/CC Perc.		Extension Para/Parataxis Perc.		Ext. Para/Exp. Para Perc.		Extension Para/Extension Perc.	
	D	M	D	M	D	M	D	M	D	M	D	M
Valid	32	80	32	80	32	80	32	80	32	80	32	80
Missing	0	0	0	0	0	0	0	0	0	0	0	0
Mean	31.97	34.187	213.406	98.313	28.106	31.203	80.097	80.014	82.25	82.085	87.545	90.98
Std. Error of Mean	1.177	0.717	14.128	4.491	1.161	0.748	1.366	0.827	1.205	0.751	0.808	0.674
Median	31.615	33.446	215	87	26.094	30.415	80.298	81.394	81.315	83.108	88.386	91.779
Mode^a	15.139	30.909	260	64	13.006	16.585	75	57.021	71.765	78.571	80.189	100
Std. Deviation	6.659	6.41	79.921	40.167	6.569	6.692	7.729	7.396	6.814	6.713	4.573	6.032
Variance	44.346	41.086	6387.346	1613.357	43.156	44.779	59.735	54.702	46.428	45.069	20.912	36.391
Skewness	0.023	0.378	0.004	1.422	0.127	0.368	-0.614	-0.613	0.126	-0.845	-0.078	-1.306
Std. Error of Skewness	0.414	0.269	0.414	0.269	0.414	0.269	0.414	0.269	0.414	0.269	0.414	0.269
Kurtosis	0.172	-0.372	-1.195	2.907	-0.26	-0.379	0.917	0.525	-1.331	0.917	-0.97	3.248
Std. Error of Kurtosis	0.809	0.532	0.809	0.532	0.809	0.532	0.809	0.532	0.809	0.532	0.809	0.532
Shapiro-Wilk	0.974	0.973	0.948	0.9	0.975	0.976	0.945	0.974	0.934	0.957	0.962	0.921
P-value of Shapiro-Wilk	0.62	0.093	0.126	<.001	0.644	0.142	0.106	0.107	0.052	0.009	0.311	<.001
Range	29.941	26.789	290	223	27.764	29.267	34.16	38.101	21.76	35.83	16.274	34.615
Minimum	15.139	22.086	61	34	13.006	16.585	57.219	57.021	71.765	59.292	80.189	65.385
Maximum	45.08	48.875	351	257	40.77	45.852	91.379	95.122	93.525	95.122	96.463	100
Sum			6829	7865								
^a More than one mode exists, only the first is reported												

(continue)

Descriptive Statistics	Table 5.4.3-1		Table 5.4.3.1-1		Table 5.4.3.2-1		Table 5.4.3.3-1		Table 5.4.3.4-1	
	Hypo Extension		Extension Hypo/CC		Extension Hypo/Hypotaxis		Ext. Hypo/Exp. Hypo Perc.		Extension Hypo/Extension	
	D	M	D	M	D	M	D	M	D	M
Valid	32	80	32	80	32	80	32	80	32	80
Missing	0	0	0	0	0	0	0	0	0	0
Mean	29.344	9.113	3.864	2.984	5.999	4.845	8.258	6.881	12.455	9.02
Std. Error of Mean	2.6	0.668	0.243	0.212	0.402	0.328	0.52	0.475	0.808	0.674
Median	26.5	8	3.992	2.678	6.166	4.332	8.587	6.097	11.614	8.221
Mode^a	11	8	1.21	0	2	0	3.03	0	3.537	0
Std. Deviation	14.707	5.977	1.374	1.894	2.271	2.937	2.941	4.251	4.573	6.032
Variance	216.297	35.721	1.888	3.586	5.159	8.628	8.648	18.075	20.912	36.391
Skewness	0.919	1.096	0.356	1.011	0.658	0.777	0.385	1.1	0.078	1.306
Std. Error of Skewness	0.414	0.269	0.414	0.269	0.414	0.269	0.414	0.269	0.414	0.269
Kurtosis	0.248	1.812	0.401	1.733	1.433	0.752	0.462	2.673	-0.97	3.248
Std. Error of Kurtosis	0.809	0.532	0.809	0.532	0.809	0.532	0.809	0.532	0.809	0.532
Shapiro-Wilk	0.911	0.932	0.971	0.943	0.954	0.96	0.967	0.94	0.962	0.921
P-value of Shapiro-Wilk	0.012	< .001	0.54	0.001	0.189	0.013	0.42	< .001	0.311	< .001
Range	53	31	6.344	9.859	10.936	14.789	13.462	24.706	16.274	34.615
Minimum	10	0	1.21	0	2	0	3.03	0	3.537	0
Maximum	63	31	7.554	9.859	12.936	14.789	16.492	24.706	19.811	34.615
Sum	939	729								

^a More than one mode exists, only the first is reported

Appendix 5.4 The descriptive statistics of comparing the proportion of extension in all clause complexes, the number of paratactic extensions, the proportion of paratactic extension in all clause nexuses, paratactic extension in parataxis, paratactic extension in all paratactic expansions, paratactic extension in all extensions, the number of hypotactic extensions, hypotactic extension in all clause nexuses, hypotactic extension in hypotaxis, hypotactic extension in all hypotactic expansions, and hypotactic extension in all extensions in CMAC and CPhD groups

Appendix 5.6

Descriptive Statistics	Table 5.6.1-1		Table 5.6.2-1		Table 5.6.2.1-1		Table 5.6.2.2-1		Table 5.6.2.3-1		Table 5.6.2.4-1	
	D	M	D	M	D	M	D	M	D	M	D	M
Valid	32	80	32	80	32	80	32	80	32	80	32	80
Missing	0	0	0	0	0	0	0	0	0	0	0	0
Mean	37.365	34.768	25.5	13.838	3.537	4.234	10.645	10.819	10.898	11.1	9.527	12.368
Std. Error of Mean	0.833	0.693	1.702	1.114	0.235	0.244	0.92	0.547	0.914	0.567	0.608	0.691
Median	37.071	35.209	25	11	3.538	3.76	11.297	10.039	11.676	10.329	10.168	11.422
Mode^a	33.333	18.621	25	11	1.393	0.379	4.127	7.018	4.276	10	3.257	10
Std. Deviation	4.712	6.202	9.625	9.963	1.329	2.182	5.207	4.895	5.17	5.075	3.441	6.177
Variance	22.207	38.464	unknown	unknown	unknown	unknown	27.11	23.964	26.732	25.752	11.839	38.152
Skewness	0.167	0.457	0.515	1.916	0.146	1.067	1.132	0.565	1.047	0.699	-0.137	0.875
Std. Error of Skewness	0.414	0.269	0.414	0.269	0.414	0.269	0.414	0.269	0.414	0.269	0.414	0.269
Kurtosis	0.376	1.583	0.816	5.221	-0.131	1.898	1.661	0.458	1.504	0.936	-0.311	1.515
Std. Error of Kurtosis	0.809	0.532	0.809	0.532	0.809	0.532	0.809	0.532	0.809	0.532	0.809	0.532
Shapiro-Wilk	0.991	0.968	0.969	0.83	0.958	0.941	0.897	0.975	0.894	0.967	0.958	0.954
P-value of Shapiro-Wilk	0.993	0.04	0.485	< .001	0.244	0.001	0.005	0.111	0.004	0.038	0.24	0.006
Range	22.752	38.639	44	59	5.361	12.059	21.755	23.542	21.606	25.883	14.044	32.209
Minimum	26.295	18.621	6	1	1.393	0.379	4.127	1.22	4.276	1.22	3.257	1.124
Maximum	49.047	57.26	50	60	6.754	12.438	25.882	24.762	25.882	27.103	17.301	33.333
Sum			816	1107								
^a More than one mode exists, only the first is reported												

(continue)

Descriptive Statistics	Table 5.6.3-1			Table 5.6.3.1-1			Table 5.6.3.2-1			Table 5.6.3.3-1			Table 5.6.3.4-1		
	Hypo Enhancement		Enh. Hypo/CC		Enh. Hypo/Hypo		Enh. Hypo/Expa. Hypo Perc.		Enh. Hypo/Enh.						
	D	M	D	M	D	M	D	M	D	M	D	M	D	M	
Valid	32	80	32	80	32	80	32	80	32	80	32	80	32	80	
Missing	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
Mean	256.281	97.1	33.828	30.534	51.965	50.399	71.947	70.711	90.473	87.632					
Std. Error of Mean	15.54	4.711	0.818	0.696	0.915	1.123	1.073	1.128	0.608	0.691					
Median	256	90	33.196	30.998	52.172	50.13	70.541	69.952	89.832	88.578					
Mode^a	260	50	23.24	16.207	42.579	40	61.429	47.059	82.699	80					
Std. Deviation	87.906	42.14	4.626	6.224	5.176	10.044	6.072	10.089	3.441	6.177					
Variance	7727.499	1775.787	21.397	38.742	26.793	100.885	36.865	101.786	11.839	38.152					
Skewness	0.64	1.553	0.202	0.565	0.298	0.516	0.276	0.204	0.137	-0.875					
Std. Error of Skewness	0.414	0.269	0.414	0.269	0.414	0.269	0.414	0.269	0.414	0.269					
Kurtosis	0.554	3.34	0.535	1.746	-0.176	0.407	-0.466	0.033	-0.311	1.515					
Std. Error of Kurtosis	0.809	0.532	0.809	0.532	0.809	0.532	0.809	0.532	0.809	0.532					
Shapiro-Wilk	0.965	0.882	0.986	0.971	0.976	0.977	0.969	0.991	0.958	0.954					
P-value of Shapiro-Wilk	0.373	<.001	0.938	0.069	0.678	0.166	0.471	0.832	0.24	0.006					
Range	406	224	22.63	37.781	21.089	49.11	23.019	49.673	14.044	32.209					
Minimum	99	40	23.24	16.207	42.579	29.197	61.429	47.059	82.699	66.667					
Maximum	505	264	45.87	53.988	63.668	78.307	84.448	96.732	96.743	98.876					
Sum	8201	7768													

^a More than one mode exists, only the first is reported

Appendix 5.6 The descriptive statistics of comparing the proportion of enhancement in all clause complexes, the number of paratactic enhancements, the proportion of paratactic enhancement in all clause nexuses, paratactic enhancement in parataxis, paratactic enhancement in all paratactic expansions, paratactic enhancement in all enhancements, the number of hypotactic enhancement, hypotactic enhancement in all clause nexuses, hypotactic enhancement in hypotaxis, hypotactic enhancement in all hypotactic expansions, and hypotactic enhancement in all enhancements in CMAc and CPhD groups

Appendix 6.1

Descriptive Statistics	Table 6.1.1-1		Table 6.1.1.1-1		Table 6.1.2-1		Table 6.1.2.1-1		Table 6.1.2.2-1		Table 6.1.2.3-1		Table 6.1.3-1	
	Projection No.		Proj./Clause Nexuses Perc.		Para Projection No.		Proj. Para/Clause Nexuses Perc.		Proj. Para/Parataxis Perc.		Proj. Para/Projection Perc.		Hypo Projection No.	
	D	M	D	M	D	M	D	M	D	M	D	M	D	M
Valid	32	80	32	80	32	80	32	80	32	80	32	80	32	80
Missing	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Mean	142.21 9	58.625	18.977	18.606	8.406	3.025	0.931	0.994	2.642	2.526	4.876	5.754	133.81 3	55.725
Std. Error of Mean	8.521	2.959	0.621	0.602	2.877	0.501	0.288	0.166	0.811	0.431	1.533	0.977	7.58	2.932
Median	138.5	52	19.06	17.591	1	1	0.163	0.298	0.424	0.826	0.844	1.439	130	51.5
Mode ^a	126	71	12.59	18.182	0	0	0	0	0	0	0	0	146	34
Std. Deviation	48.2	26.468	3.51	5.389	16.274	4.483	1.629	1.483	4.59	3.851	8.673	8.736	42.877	26.226
Variance	2323.2 1	700.54 1	12.323	29.039	264.83	20.101	2.654	2.198	21.072	14.832	75.229	76.318	1838.4 8	687.82 2
Skewness	0.828	1.298	-0.176	0.571	3.477	1.924	3.222	1.846	2.924	1.855	3.387	2.127	0.666	1.37
Std. Error of Skewness	0.414	0.269	0.414	0.269	0.414	0.269	0.414	0.269	0.414	0.269	0.414	0.269	0.414	0.269
Kurtosis	1.642	2.342	-0.869	0.274	14.442	3.572	13.097	3.123	10.338	2.928	14.261	5.001	0.649	2.996
Std. Error of Kurtosis	0.809	0.532	0.809	0.532	0.809	0.532	0.809	0.532	0.809	0.532	0.809	0.532	0.809	0.532
Shapiro-Wilk	0.954	0.906	0.966	0.977	0.565	0.717	0.607	0.722	0.627	0.708	0.595	0.703	0.965	0.906
P-value of Shapiro-Wilk	0.186	<.001	0.386	0.152	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	0.371	<.001
Range	226	136	12.255	27.989	83	19	8.267	6.548	22.193	15.942	44.624	40.909	189	141
Minimum	66	18	12.59	7.643	0	0	0	0	0	0	0	0	66	13
Maximum	292	154	24.845	35.632	83	19	8.267	6.548	22.193	15.942	44.624	40.909	255	154
Sum	4551	4690			269	242							4282	4458

More than one mode exists, only the first is reported

^a More than one mode exists, only the first is reported

(continue)

Descriptive Statistics	Table 6.1.3.1-1			Table 6.1.3.2-1			Table 6.1.3.3-1			Table 6.1.4.1-1			Table 6.1.4.2-1			Table 6.1.4.2-2		
	Proj. Hypo/Clause Nexuses Perc.			Proj. Hypo/Hypotaxis Perc.			Proj. Hypo/Projection Perc.			Locution No.			Locution/Projec tion Perc.			Idea/Projection Perc.		
	D	M		D	M		D	M		D	M		D	M		D	M	
Valid	32	80		32	80		32	80		32	80		32	80		32	80	
Missing	0	0		0	0		0	0		0	0		0	0		0	0	
Mean	18.046	17.668		27.701	28.91		95.124	94.433		44.219	15.525		29.487	27.594		43.1	70.513	
Std. Error of Mean	0.661	0.623		0.894	0.939		1.533	0.972		5.81	1.015		2.387	1.432		5.487	2.387	
Median	18.465	16.957		27.925	26.976		99.156	98.582		39.5	14		26.477	26.241		95	38	
Mode^a	10.259	6.468		16.349	22.807		100	100		14	12		8.654	16.216		95	31	
Std. Deviation	3.738	5.57		5.057	8.4		8.673	8.696		32.866	9.076		13.501	12.807		31.04	22.212	
Variance	13.969	31.023		25.575	70.555		75.229	75.62		1080.1	82.379		182.28	164.02		963.48	493.35	
Skewness	-0.197	0.57		-0.214	0.798		-3.387	-2.204		2.097	1.057		0.666	0.532		0.159	-0.666	
Std. Error of Skewness	0.414	0.269		0.414	0.269		0.414	0.269		0.414	0.269		0.414	0.269		0.414	0.269	
Kurtosis	-0.787	0.584		-0.498	0.741		14.261	5.333		5.639	1.061		0.382	0.514		-0.708	3.685	
Std. Error of Kurtosis	0.809	0.532		0.809	0.532		0.809	0.532		0.809	0.532		0.809	0.532		0.809	0.532	
Shapiro-Wilk	0.975	0.978		0.982	0.947		0.595	0.689		0.802	0.927		0.954	0.979		0.976	0.896	
P-value of Shapiro-Wilk	0.634	0.179		0.86	0.002		<.001	<.001		<.001	<.001		0.191	0.204		0.676	<.001	
Range	14.586	29.164		20.877	42.972		44.624	40.909		157	44		56.4	66.667		118	126	
Minimum	10.259	6.468		16.349	9.275		55.376	59.091		9	0		8.654	0		37	6	
Maximum	24.845	35.632		37.226	52.247		100	100		166	44		65.054	66.667		155	132	
Sum										1415	1242					3136	3448	
^a More than one mode exists, only the first is reported																		

Appendix 6.1 The descriptive statistics of comparing the number of projections, the projection in all clause nexuses, the number of paratactic projections, the paratactic projection in all clause nexuses, the paratactic projection in parataxis, the paratactic projection in all projections, the number of hypotactic projections, hypotactic projection in all clause nexuses, hypotactic projection in hypotaxis, hypotactic projection in all projections, the number of locutions, the proportion of locution in all projections, the number of ideas, the proportion of idea in all projections in CMAC and CPhD groups

Appendix 6.2

Descriptive Statistics	Table 6.2.1-1		Table 6.2.2-1		Table 6.2.2.1-1		Table 6.2.2.2-1		Table 6.2.2.3-1		Table 6.2.2.4-1	
	Locution/C Nexuses Perc.		Para locution		Locution Para/Clause Nexuses Perc.		Locution Para/Parataxis Perc.		Locution Para/Proj. Para Perc.		Locution Para/Locution Perc.	
	D	M	D	M	D	M	D	M	D	M	D	M
Valid	32	80	32	80	32	80	32	80	32	80	32	80
Missing	0	0	0	0	0	0	0	0	0	0	0	0
Mean	5.548	4.962	7.125	2.4	0.783	0.803	2.218	2.07	91.311	88.478	10.512	12.53
Std. Error of Mean	0.506	0.279	2.502	0.435	0.248	0.149	0.694	0.392	2.705	2.822	2.405	2.038
Median	4.657	4.803	1	0.5	0.119	0.085	0.307	0.232	100	100	4.13	1.667
Mode^a	3.989	0	0	0	0	0	0	0	100	100	0	0
Std. Deviation	2.863	2.494	14.151	3.893	1.403	1.336	3.924	3.505	15.301	25.24	13.604	18.226
Variance	8.199	6.22	200.242	15.154	1.968	1.784	15.395	12.288	234.114	637.053	185.082	332.192
Skewness	1.184	0.783	3.519	2.245	3.325	2.336	3.031	2.266	-1.707	-2.434	1.713	1.894
Std. Error of Skewness	0.414	0.269	0.414	0.269	0.414	0.269	0.414	0.269	0.414	0.269	0.414	0.269
Kurtosis	1.156	0.732	14.572	5.256	13.814	5.929	11.214	5.188	1.784	5.143	3.967	4.127
Std. Error of Kurtosis	0.809	0.532	0.809	0.532	0.809	0.532	0.809	0.532	0.809	0.532	0.809	0.532
Shapiro-Wilk	0.901	0.961	0.551	0.673	0.597	0.664	0.618	0.653	0.635	0.526	0.776	0.732
P-value of Shapiro-Wilk	0.006	0.017	< .001	< .001	< .001	< .001	< .001	< .001	< .001	< .001	< .001	< .001
Range	11.748	13.057	72	18	7.171	6.548	19.251	15.942	50	100	59.504	88.889
Minimum	1.859	0	0	0	0	0	0	0	50	0	0	0
Maximum	13.607	13.057	72	18	7.171	6.548	19.251	15.942	100	100	59.504	88.889
Sum			228	192								
^a More than one mode exists, only the first is reported												

(continue)

Descriptive Statistics	Table 6.2.3-1		Table 6.2.3.1-1		Table 6.2.3.2-1		Table 6.2.3.3-1		Table 6.2.3.4-1	
	Hypo location		Location Hypo/Clause Nexuses Perc.		Location Hypo/Hypotaxis Perc.		Location Hypo/Proj. Hypo Perc.		Location Hypo/Location Perc.	
	D	M	D	M	D	M	D	M	D	M
Valid	32	80	32	80	32	80	32	80	32	80
Missing	0	0	0	0	0	0	0	0	0	0
Mean	37.094	13.125	4.765	4.159	7.339	6.853	27.177	24.416	89.488	86.22
Std. Error of Mean	4.191	0.83	0.377	0.213	0.558	0.345	2.026	1.227	2.405	2.306
Median	34.5	12	4.141	3.954	6.906	6.873	25.81	22.997	95.87	95.834
Mode^a	14	12	1.786	0	2.344	5.556	8.654	25	100	100
Std. Deviation	23.708	7.428	2.134	1.906	3.156	3.081	11.462	10.977	13.604	20.627
Variance	562.088	55.174	4.553	3.633	9.959	9.495	131.367	120.488	185.082	425.47
Skewness	2.156	0.801	1.03	0.237	0.798	0.137	0.335	0.687	-1.713	-2.113
Std. Error of Skewness	0.414	0.269	0.414	0.269	0.414	0.269	0.414	0.269	0.414	0.269
Kurtosis	7.394	0.389	0.95	-0.288	0.408	-0.484	-0.356	1.374	3.967	5.151
Std. Error of Kurtosis	0.809	0.532	0.809	0.532	0.809	0.532	0.809	0.532	0.809	0.532
Shapiro-Wilk	0.822	0.952	0.914	0.987	0.938	0.989	0.965	0.972	0.776	0.713
P-value of Shapiro-Wilk	< .001	0.005	0.014	0.616	0.068	0.74	0.367	0.078	< .001	< .001
Range	123	35	9.034	8.504	13.131	13.75	43.111	64.706	59.504	100
Minimum	9	0	1.786	0	2.344	0	8.654	0	40.496	0
Maximum	132	35	10.82	8.504	15.475	13.75	51.765	64.706	100	100
Sum	1187	1050								

^a More than one mode exists, only the first is reported

Appendix 6.2 The descriptive statistics of comparing the proportion of locution in all clause nexuses, the number of paratactic locution, paratactic locution in all clause nexuses, paratactic locution in parataxis, paratactic locution in all paratactic projections, paratactic locution in all locutions, the number of hypotactic locution, hypotactic locution in all clause nexuses, hypotactic locution in hypotaxis, hypotactic locution in all hypotactic projections, and hypotactic locution in all locutions in CMAC and CPhD groups

Appendix 6.3

Locution	M	D
Proposal	136	109
admit	-	1
advocate	10	3
affirm	-	1
agree	3	4
announce	1	-
argue	11	19
ask	4	-
claim	-	5
conclude	-	1
contend	-	1
demand	1	-
insist	1	-
joke	1	-
maintain	1	4
point out	2	3
propose	11	10
recommend	-	4
remind	1	-
reproach	1	-
require	24	2
say	3	-
state	4	-
stress	-	2
suggest	43	44
tell	13	4
warn	1	1

Table 6.3.1-1 The number of proposals led by each verbal lexis in all locutions in two groups

Locution	M	D
Proposition	938	1094
acknowledge	1	3
add	1	1
admit	20	16
affirm	3	-
announce	4	-
answer	3	2
argue	85	268
ask	30	14
assert	9	20
blame	4	-
claim	69	96
clarify	4	-
complain	1	5
conclude	56	39
contend	-	9
convey	6	3
convince	6	7
declare	6	8
deny	2	4

disagree	2	-
discuss	9	7
emphasize	11	14
explain	45	60
express	6	3
insist	7	7
interpret	2	3
mention	1	-
maintain	11	11
point	8	-
point out	126	145
propose	25	43
put	4	7
put forward	12	2
quote	-	5
recognize	-	4
remark	1	-
report	9	44
reply	8	2
say	188	118
shout	2	-
state	62	53
stress	15	5
suggest	22	38
summarize	10	4
talk about	-	1
tell	42	23

Table 6.3.1-2 The number of propositions led by each verbal lexis in all locutions in two groups

Appendix 6.4

Descriptive Statistics	Table 6.4.1-1		Table 6.4.2		Table 6.4.2.1		Table 6.4.2.2		Table 6.4.2.3		Table 6.4.2.4	
	Idea/C Nexuses Perc.		Para idea		Idea Para/C Nexuses Perc.		Idea Para/Parataxis Perc.		Idea Para/Proj. Para Perc.		Idea Para/Idea Perc.	
	D	M	D	M	D	M	D	M	D	M	D	M
Valid	32	80	32	80	32	80	32	80	32	80	32	80
Missing	0	0	0	0	0	0	0	0	0	0	0	0
Mean	13.429	13.643	1.281	0.625	0.148	0.191	0.425	0.456	8.689	11.522	1.44	1.617
Std. Error of Mean	0.675	0.576	0.458	0.159	0.052	0.048	0.155	0.118	2.705	2.822	0.609	0.388
Median	12.591	13.063	0	0	0	0	0	0	0	0	0	0
Mode^a	6.474	3.774	0	0	0	0	0	0	0	0	0	0
Std. Deviation	3.818	5.154	2.593	1.426	0.293	0.432	0.88	1.057	15.301	25.24	3.443	3.471
Variance	14.577	26.567	6.725	2.035	0.086	0.187	0.774	1.117	234.114	637.053	11.854	12.045
Skewness	0.14	0.794	2.546	3.324	2.285	3.185	2.581	3.446	1.707	2.434	3.557	2.666
Std. Error of Skewness	0.414	0.269	0.414	0.269	0.414	0.269	0.414	0.269	0.414	0.269	0.414	0.269
Kurtosis	-0.928	0.735	6.57	12.944	4.599	12.411	6.569	15.162	1.784	5.143	13.913	7.676
Std. Error of Kurtosis	0.809	0.532	0.809	0.532	0.809	0.532	0.809	0.532	0.809	0.532	0.809	0.532
Shapiro-Wilk	0.962	0.955	0.576	0.504	0.583	0.516	0.563	0.5	0.635	0.526	0.483	0.546
P-value of Shapiro-Wilk	0.316	0.007	< .001	< .001	< .001	< .001	< .001	< .001	< .001	< .001	< .001	< .001
Range	13.782	26.877	11	8	1.096	2.589	3.636	6.667	50	100	16.923	18.182
Minimum	6.474	3.774	0	0	0	0	0	0	0	0	0	0
Maximum	20.256	30.651	11	8	1.096	2.589	3.636	6.667	50	100	16.923	18.182
Sum			41	50								

^a More than one mode exists, only the first is reported

(continue)

Descriptive Statistics	Table 6.4.3.1			Table 6.4.3.1-1			Table 6.4.3.2-1			Table 6.4.3.3-1			Table 6.4.3.4-1		
	Hypo idea		Idea Hypo/C Nexuses Perc.	Idea Hypo/Hypotaxis Perc.		Idea Hypo/Proj. Hypo Perc.	Idea Hypo/Idea Perc.		M	D		M	D		M
	D	M		D	M		D	M		D	M		D	M	
Valid	32	80	32	80	32	80	32	80	32	80	32	80	32	80	80
Missing	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Mean	96.719	42.475	13.281	13.452	20.362	21.981	72.823	75.3	98.56	98.383					
Std. Error of Mean	5.51	2.477	0.69	0.581	0.99	0.899	2.026	1.268	0.609	0.388					
Median	93.5	37	12.153	12.781	20.659	20.436	74.19	77.003	100	100					
Mode^a	74	30	5.378	3.774	8.571	7.5	48.235	75	100	100					
Std. Deviation	31.168	22.151	3.901	5.197	5.598	8.041	11.462	11.34	3.443	3.471					
Variance	971.434	490.683	15.215	27.013	31.335	64.651	131.367	128.589	11.854	12.045					
Skewness	0.145	1.558	0.078	0.824	0.186	1.03	-0.335	-0.695	-3.557	-2.666					
Std. Error of Skewness	0.414	0.269	0.414	0.269	0.414	0.269	0.414	0.269	0.414	0.269					
Kurtosis	-0.777	3.892	-0.831	0.758	-0.153	1.276	-0.356	1.103	13.913	7.676					
Std. Error of Kurtosis	0.809	0.532	0.809	0.532	0.809	0.532	0.809	0.532	0.809	0.532					
Shapiro-Wilk	0.973	0.892	0.968	0.953	0.988	0.929	0.965	0.971	0.483	0.546					
P-value of Shapiro-Wilk	0.598	< .001	0.458	0.005	0.966	< .001	0.367	0.069	< .001	< .001					
Range	117	126	14.878	26.877	25.249	40.799	43.111	64.706	16.923	18.182					
Minimum	37	6	5.378	3.774	8.571	7.5	48.235	35.294	83.077	81.818					
Maximum	154	132	20.256	30.651	33.82	48.299	91.346	100	100	100					
Sum	3095	3398													
^a More than one mode exists, only the first is reported															

Appendix 6.4 The descriptive statistics of comparing the proportion of ideas in all clause nexuses, the number of paratactic ideas, paratactic ideas in all clause nexuses, paratactic ideas in parataxis, paratactic ideas in all paratactic projections, paratactic ideas in all ideas, the number of hypotactic ideas, hypotactic idea in all clause nexuses, hypotactic idea in hypotaxis, hypotactic idea in all hypotactic projections, and hypotactic idea in all ideas in CMAc and CPhD groups

Appendix 6.5

Idea	M	D
Proposal	62	35
assume	1	-
believe	1	3
concern	-	1
decide	1	-
discover	-	1
expect	8	5
hint	-	1
hope	37	14
imply	-	1
mean	1	1
propose	2	2
regard	-	1
sense	1	-
show	-	4
suggest	4	-
think	3	-
turn out	1	-
understand	1	-
wish	1	1

Table 6.5.1-1 The number of proposals led by each verbal lexis in all ideas in two groups

Idea	M	D
Proposition	2323	1989
afraid	8	2
assume	32	58
believe	211	116
consider	56	25
decide	4	5
deem	1	-
demonstrate	48	42
determine	16	39
discover	10	22
expect	-	3
feel	19	11
find	242	265
find out	43	11
guess	2	-
identify	4	7
imply	31	44
indicate	225	253
intend	2	-
investigate	20	28
judge	7	2

know	172	72
mean	196	138
predict	8	20
prove	72	31
realize	52	23
reflect	20	13
remember	1	1
see	168	114
show	331	346
suggest	70	164
suppose	3	6
think	198	97
understand	50	31
wonder	1	-

Table 6.5.1-2 The number of propositions led by each verbal lexis in all ideas in two groups

Appendix 7.4

M2		D2	
$\alpha'\beta(1+2)$	11	$\alpha'\beta(1+2)$	11
$x\beta\alpha(\alpha x\beta)$	8	$x\beta\alpha(1+2)$	10
$1+2(\alpha x\beta)$	7	$\alpha'\beta(\alpha x\beta)$	9
$x\beta\alpha(1+2)$	5	$\alpha x\beta(1+2)$	8
$x\beta\alpha(\alpha=\beta)$	5	$1+2(\alpha x\beta)$	6
$\alpha=\beta(1+2)$	4	$1=2(1+2)$	4
$\alpha=\beta(\alpha x\beta)$	4	$\alpha=\beta(1+2)$	4
$\alpha'\beta(\alpha x\beta)$	4	$\alpha=\beta(\alpha'\beta)$	4
$x\beta(1+2)\alpha$	4	$1+2(1+2)$	3
$x\beta\alpha(\alpha'\beta)$	4	$1+2(\alpha'\beta)$	3
$1+2(1+2)$	3	$\alpha'\beta(\alpha x\beta)$	3
$1+2(\alpha'\beta)$	3	$\alpha'\beta(\alpha+\beta)$	3
$1=2(1+2)$	3	$\alpha'\beta(\alpha=\beta)$	3
$\alpha'\beta(\alpha+\beta)$	3	$\alpha x\beta(\alpha x\beta)$	3
$\alpha x\beta(1+2)$	3	$x\beta(1+2)\alpha$	3
$\alpha x\beta(\alpha x\beta)$	3	$x\beta(\alpha'\beta)\alpha$	3
$1+2(1=2)$	2	$x\beta(\alpha x\beta)\alpha$	3
$1+2(x\beta\alpha)$	2	$x\beta\alpha(\alpha=\beta)$	3
$1x2(\alpha x\beta)$	2	$x\beta\alpha(\alpha x\beta)$	3
$\alpha(1+2)x\beta$	2	$1+2(1x2)$	2
$\alpha=\beta(1x2)$	2	$1+2+3$	2
$\alpha=\beta(\alpha'\beta)$	2	$1=2(1+2+3)$	2
$\alpha'\beta(1x2)$	2	$1x2(\alpha x\beta)$	2
$\alpha'\beta(x\beta\alpha)$	2	$\alpha'\beta(1+2)$	2
$x\beta(\alpha x\beta)\alpha$	2	$\alpha=\beta(\alpha x\beta)$	2
$x\beta\alpha(1x2)$	2	$\alpha x\beta(\alpha'\beta)$	2
$1''2(1+2)$	1	$x\beta\alpha(1''2)$	2
$1''2(\alpha x\beta)$	1	$1''2(1+2)$	1
$1(1+2)x2$	1	$1''2(\alpha x\beta)$	1
$1(1+2+3+4)x2$	1	$1(1+2)+2+3$	1
$1(\alpha'\beta)+2$	1	$1(1+2)=2$	1
$1(\alpha=\beta)+2(\alpha=\beta)$	1	$1(\alpha'\beta)+2(\alpha x\beta)$	1
$1(\alpha=\beta)+2(x\beta\alpha)$	1	$1(\alpha x\beta)+2$	1
$1(\alpha'\beta)+2(\alpha'\beta)$	1	$1(\alpha x\beta)+2(\alpha x\beta)$	1
$1(\alpha x\beta)+2(1+2)$	1	$1+2(1=2)$	1
$1(\alpha x\beta)x2$	1	$1+2(\alpha+\beta)$	1
$1(\alpha x\beta)x2(1+2)$	1	$1+2(\alpha'\beta)$	1
$1(x\beta\alpha)+2$	1	$1'2(1+2)$	1
$1+2(\alpha'\beta)$	1	$1x2(\alpha'\beta)$	1
$1+2(\alpha+\beta)$	1	$1x2x3x4(1+2)$	1
$1+2+3(\alpha x\beta)$	1	$\alpha'\beta(1+2+3)$	1
$1=2(\alpha+\beta)$	1	$\alpha'\beta(\alpha+\beta)$	1

$1=2(\alpha'\beta)$	1	$\alpha''\beta(\alpha=\beta)$	1
$1x2(1+2)$	1	$\alpha''\beta(x\beta\alpha)$	1
$1x2(\alpha=\beta)$	1	$\alpha(1+2)+\beta(\alpha'\beta)$	1
$\alpha''\beta(1+2)$	1	$\alpha(1+2)=\beta$	1
$\alpha''\beta(\alpha=\beta)$	1	$\alpha(1+2)=\beta(\alpha'\beta)$	1
$\alpha(1+2)+\beta$	1	$\alpha(1+2)x\beta$	1
$\alpha(1+2)=\beta(1x2)$	1	$\alpha(1x2)+\beta(1x2)$	1
$\alpha(\alpha+\beta)=\beta$	1	$\alpha(\alpha'\beta)=\beta(\alpha'\beta)$	1
$\alpha(\alpha'\beta)=\beta(1+2)$	1	$\alpha(\alpha+\beta)=\beta$	1
$\alpha(\alpha'\beta)x\beta(\alpha x\beta)$	1	$\alpha(\alpha=\beta)+\beta$	1
$\alpha(\alpha'\beta)x\beta(x\beta\alpha)$	1	$\alpha(\alpha x\beta)=\beta$	1
$\alpha(\alpha x\beta)+\beta$	1	$\alpha(\alpha x\beta)x\beta(1+2)$	1
$\alpha(\alpha x\beta)=\beta$	1	$\alpha+\beta(\alpha x\beta)$	1
$\alpha(\alpha x\beta)x\beta$	1	$\alpha'\beta(1+2+3)$	1
$\alpha(\alpha x\beta x c)x\beta(1'2)$	1	$\alpha'\beta(1=2)$	1
$\alpha+\beta(1+2)$	1	$\alpha'\beta(1x2)$	1
$\alpha+\beta(1=2)$	1	$\alpha'\beta(\alpha'\beta)$	1
$\alpha+\beta(\alpha'\beta)$	1	$\alpha'\beta(\alpha x\beta x c)$	1
$\alpha=\beta(\alpha=\beta)$	1	$x\beta(1+2)\alpha(\alpha'\beta)$	1
$\alpha'\beta(1+2+3)$	1	$x\beta(\alpha'\beta)\alpha(\alpha x\beta)$	1
$\alpha'\beta(\alpha=\beta)$	1	$x\beta\alpha(1=2)$	1
$\alpha'\beta(\alpha'\beta)$	1	$x\beta\alpha(1x2)$	1
$\alpha x\beta(\alpha+\beta)$	1	$x\beta\alpha(x\beta\alpha)$	1
$\alpha x\beta(\alpha=\beta)$	1		
$\alpha x\beta(\alpha'\beta)$	1		
$\alpha x\beta x c(\alpha x\beta)$	1		
$x\beta(1+2)\alpha(1+2)$	1		
$x\beta(1+2+3+4)\alpha$	1		
$x\beta(1=2)\alpha$	1		
$x\beta(\alpha=\beta)\alpha(\alpha=\beta)$	1		
$x\beta(\alpha'\beta)\alpha(\alpha'\beta)$	1		
$x\beta\alpha(\alpha'\beta)$	1		
$x\beta\alpha(x\beta\alpha)$	1		

Table A7.4.1 Categories of the clause complex pattern on Level 2 in two groups with their numbers

M3		D3	
$1+2(\alpha x\beta(\alpha'\beta))$	2	$\alpha'\beta(\alpha x\beta(\alpha x\beta))$	2
$\alpha=\beta(\alpha x\beta(\alpha'\beta))$	2	$x\beta\alpha(\alpha'\beta(1+2))$	2
$x\beta\alpha(1+2(\alpha x\beta))$	2	$1(\alpha x\beta)+2(\alpha x\beta(1+2))$	1
$x\beta\alpha(\alpha'\beta(1+2))$	2	$1+2(1+2(\alpha x\beta))$	1
$1''2(\alpha x\beta(\alpha=\beta))$	1	$1+2(1x2(1x2))$	1
$1(\alpha(1+2)x\beta)+2$	1	$1+2(\alpha=\beta(\alpha x\beta))$	1
$1(\alpha x\beta(1+2))+2$	1	$1+2(\alpha'\beta(1+2))$	1
$1+2(\alpha=\beta(1+2))$	1	$1+2(\alpha x\beta(1+2))$	1
$1+2+3+4+5(\alpha'\beta(1+2))$	1	$1=2(1=2(1+2))$	1
$1=2(1(\alpha x\beta)+2(\alpha=\beta))$	1	$1=2(\alpha x\beta(1+2))$	1
$1=2(\alpha'\beta(1+2))$	1	$1'2(x\beta\alpha(1x2))$	1
$1x2(\alpha'\beta(1+2))$	1	$1x2(\alpha'\beta(\alpha x\beta))$	1
$1x2(x\beta\alpha(1+2))$	1	$\alpha''\beta(\alpha x\beta(\alpha x\beta))$	1
$\alpha''\beta(1x2(\alpha x\beta))$	1	$\alpha=\beta(1+2+3+4(\alpha'\beta))$	1
$\alpha''\beta(\alpha(\alpha x\beta)=\beta)$	1	$\alpha=\beta(\alpha+\beta(\alpha x\beta))$	1
$\alpha''\beta(\alpha=\beta(1+2))$	1	$\alpha=\beta(\alpha'\beta(\alpha x\beta))$	1
$\alpha''\beta(\alpha x\beta(1+2))$	1	$\alpha=\beta(\alpha x\beta(1+2))$	1
$\alpha=\beta(1+2(1x2))$	1	$\alpha=\beta(\alpha x\beta(\beta\alpha))$	1
$\alpha'\beta(1(\alpha x\beta)+2)$	1	$\alpha'\beta(1(\alpha=\beta)+2(\alpha=\beta))$	1
$\alpha'\beta(1+2(\alpha'\beta))$	1	$\alpha'\beta(1+2(\alpha x\beta)+3)$	1
$\alpha''\beta(1=2(\alpha+\beta))$	1	$\alpha'\beta(1=2(1+2+3+4+5))$	1
$\alpha'\beta(\alpha(1+2)x\beta)$	1	$\alpha'\beta(\alpha(1+2+3)x\beta(1+2))$	1

(continue)

$\alpha'\beta(\alpha x\beta(1+2))$	1	$\alpha'\beta(\alpha x\beta(1+2))$	1
$\alpha x\beta(1+2(\alpha x\beta))$	1	$\alpha'\beta(\alpha x\beta(1+2+3))$	1
$\alpha x\beta(\alpha=\beta(1+2))$	1	$\alpha'\beta(\alpha x\beta(1=2))$	1
$\alpha x\beta(\alpha'\beta(1+2))$	1	$\alpha'\beta(\alpha x\beta(\alpha'\beta))$	1
$\alpha x\beta(\alpha'\beta(1x2))$	1	$\alpha x\beta(\alpha'\beta(\alpha x\beta))$	1
$\alpha x\beta(\alpha'\beta(\alpha+\beta))$	1	$x\beta(1+2)\alpha(\alpha x\beta(1=2))$	1
$\alpha x\beta(\alpha'\beta(\alpha x\beta))$	1	$x\beta(\alpha'\beta(1+2))\alpha$	1
$x\beta(1(\alpha'\beta)+2)\alpha(1''2)$	1	$x\beta(\alpha'\beta(1+2))\alpha(\alpha x\beta)$	1
$x\beta(1+2(\alpha=\beta))$	1	$x\beta(\alpha x\beta)\alpha(\alpha+2(1x2))$	1
$x\beta(\alpha+\beta)\alpha(\alpha(1+2)=\beta)$	1	$x\beta\alpha(1(\alpha'\beta)+2)$	1
$x\beta\alpha(1(\alpha'\beta)+2)$	1	$x\beta\alpha(1+2(\alpha x\beta))$	1
$x\beta\alpha(1(\alpha+\beta)x2(1+2))$	1	$x\beta\alpha(1+2(x\beta\alpha))$	1
$x\beta\alpha(1+2(1+2))$	1	$x\beta\alpha(1+2+3(\alpha=\beta))$	1
$x\beta\alpha(\alpha'\beta(\alpha=\beta))$	1	$x\beta\alpha(\alpha=\beta(1+2))$	1
$x\beta\alpha(\alpha(1+2)+\beta(1+2))$	1	$x\beta\alpha(\alpha=\beta(1x2))$	1
$x\beta\alpha(\alpha(1+2)=\beta)$	1	$x\beta\alpha(\alpha'\beta(1+2))$	1
$x\beta\alpha(\alpha=\beta(1x2))$	1	$x\beta\alpha(\alpha'\beta(\alpha x\beta))$	1
$x\beta\alpha(\alpha'\beta(\alpha+\beta))$	1	$x\beta\alpha(\alpha x\beta(1x2))$	1
$x\beta\alpha(\alpha x\beta(1=2))$	1		
$x\beta\alpha(\alpha x\beta(\alpha x\beta))$	1		

Table A7.4.2 Categories of the clause complex pattern on Level 3 in two groups with their numbers