

Published as: Fawcett, Robin P., 2006b. 'Establishing the grammar of "typicity" in English: an exercise in scientific inquiry', Part 2'. *Educational Research on Foreign Languages & Art*. No. 3. Guangzhou: Guangdong Teachers College of Foreign Language and Arts, 3-34.

Establishing the grammar of 'typicity' in English: an exercise in scientific inquiry

Robin P. Fawcett

Centre for Language and Communication Research
Cardiff University

Abstract

This paper has two aims: (i) to develop a more adequate model than is currently available for the syntax and semantics of a central area of English grammar for which there are still, surprisingly, no adequate accounts in the literature - i.e. the meanings and forms of 'typicity' - and (ii) to examine critically the adequacy of the methods used in such descriptive studies in general and in this study in particular. The discussion assumes the desirability of taking a functional approach to language and, while the model used here is Systemic Functional Grammar, the discussion does not depend on a prior understanding of that model, and it should be relevant to any functionally-oriented model of language. Part 1 introduced the framework within which the key discussion that follows here will take place, but in doing this it provided an overview of 'selection' in the English nominal group that constitutes a major publication in its own right. In Part 2 the focus is almost entirely on (i) the forms and meanings of 'typicity' in English, and (ii) the methods available to the twenty-first century linguist in seeking to determine which of various alternatives provides the best way to model the facts.

Part 2

'Typicity': exploring the methodology for deciding between alternative models

Part 1 included the following: (i) an outline of the overall methodology that I assume to be needed in order to make progress in science in general and linguistics in particular; (ii) a summary of what a Systemic Functional Grammar (SFG) is and how it works; (iii) an introduction to the overall structure of the English nominal group, as it is seen in the 'Cardiff Grammar' version of SFG; (iv) a review of the surprisingly sparse literature on the aspects of the nominal group to be examined here; and finally, at greater length, (v) a description of the concept of 'selection' in the English nominal group, in which we recognized no fewer than ten different types of determiner. The question with which we ended was this: 'Is typicity also a type of selection?'

5 Is there a ‘typic determiner’ - and what alternative analyses should we consider?

5.1 The data to be explained

The structure in English for which I originally proposed the concept of the **typic determiner** (Fawcett 1980:220) answers the apparently simple question: ‘What type or types of thing?’. Typical examples of ‘typic nouns’ that expound the head of a nominal group that fills such a determiner are shown in the underlined portions of (1a) to (3a) - which are repeated here from Section 1.1. The question is:

‘How are such “typic nouns” related, syntactically and semantically, to the words that precede and follow them?’

- (1a) (The system needs) a few different sorts of documents.
 (2a) (You need to determine) the appropriate type of insulin the person should use.
 (3a) (This is) one of the first of the new varieties of GM wheat.

Given the analyses in Section 4, the obvious way to model such phenomena is to say that a nominal group whose head is *type* (or a semantically similar word) fills the typic determiner, this being followed by the selector *of*. However, there are several reasons to consider a different and more radical proposal.

What are the main ‘typic nouns’ of English? The first three in the following list, which all express a ‘general’ meaning, are by far the most frequent. Those on the second line are less frequent ‘general’ ones, and those on the third are similar - but they need to be understood in their ‘product’ rather than their ‘process’ sense. The items in the fourth line tend to occur with certain classes of object, e.g. *this brand of washing powder / icecream, a new breed of cattle / chicken, a different make of car / television, another strain / variety of rose / wheat*, etc. And those in the fifth line are some of the principal technical terms for ‘type’ used in biology.¹

type, sort, kind;
class, category; sub-class, sub-category, sub-type;
classification, categorization, sub-classification, sub-categorization;
brand, breed, form, make, strain, variety, version, genre etc.,
*phylum, class, order, family, genus, species, variety, biotype etc.*²

However, when we think systemically and semantically (rather than merely structurally at the level of form) we become aware that there are other forms which express ‘typicity’, as well as those illustrated in (1a) to (3a). Most importantly, we need a description that provides for certain **covert** realizations of ‘typicity’ which go unnoticed much of the time. An unproblematical example is *They’ve brought out a new stamp*, where the meaning is clearly ‘a new type of stamp’. The high frequency of such examples suggests that ‘typicity’ occurs

¹But see the discussion in Section 5.7.2 of the possibility of expanding this list to include words such as *size, shape, colour* etc.

²Notice that the word *example* and its near-synonyms are not listed here - as they are in the equivalent list in Matthiessen 1985:657. ‘Example’ and ‘type’ are in fact two different - if frequently confused - concepts. *Example* typically occurs as the head of the nominal group in one sub-type of **representational determiner** (see Section 4.3).

much more frequently than is implied by the peripheral place it occupies in most grammars.³

In the rest of Section 5, I shall (i) summarize my previous approach to this area of the grammar; (ii) present the evidence that led me to consider a radical alternative syntax for it; (iii) state four hypotheses which, if supported by the evidence of corpus data, could be interpreted as strengthening the claims of the new model; (iv) describe the corpus data used to test the hypotheses; (v) report on (a) how far these data supported the hypotheses and (b) the further evidence that appeared; and finally (vi) decide between four alternative models of the syntax of this part of the nominal group.

5.2 My previous approach to ‘typicity’

The description of the English nominal group on which this paper draws were originally prepared as part of the ‘guidelines’ for use by members of my research team on a large text analysis project (Fawcett & Perkins 1980a-d, 1981). Parts of this description have appeared in Fawcett 1974-6/81, 1980, 2000 and forthcoming b. In these works I treated examples such as (1a) to (3a) as just another determiner that uses the pattern of ‘selection’ that we met in Section 4. It is an approach that has served my research team, myself and many generations of students well in various large- and small-scale text analysis projects. Indeed, it is my proposal of this structure for handling examples with *type of* that we find reflected in Matthiessen’s account of this part of the grammar (1995:657).

However, as I reviewed the examples in the above-mentioned notes in preparation for writing this paper, I was struck by certain patterns which led me to consider a radically different way of modelling their syntax.

At this point I decided to consult the three recent comprehensive grammars of English (Quirk et al 1985, Biber et al 1999, and Huddleston & Pullum 2002) to see what they had to say about the semantics and syntax of ‘typicity’. Disappointingly, there was little of value on either the semantics or the syntax of *type*, etc, most assuming without discussion that the structure should be what I have termed the ‘prepositional group as qualifier’ construction. However, Biber et al (1999:258) do at least comment that ‘it is not clear how these structures should be analyzed’, and they go on to suggest, interestingly, that ‘there are indications that species nouns [i.e. *type*, etc] may be felt to be subordinate in much the same way as a determiner’ - a viewpoint which seems to reflect the proposals in Fawcett 1974-/81 and 1980, but which they presumably reached independently.⁴

5.3 Thinking the unthinkable: ‘selection’ between heads

³There are yet more semantically related variants - including, alongside *this type of document*, (i) *a document of this type*, (ii) *this document type*, (iii) *such a document* and (iv) *the like(s) of this document*. While the system network and realization rules given in Sections 5.5 and 5.12 can be set within a wider network that handles these, they are not covered in this paper. (However, we refer to Type (ii) in Section 5.7.2.)

⁴The topic on which all three ‘big grammars’ focus is one which is realized in morphology - and which has long been a preoccupation of traditional and prescriptive grammars. It is the question of so-called ‘agreement’ in examples such as *these types of oil are v. these type of oil are* Quirk et al 1985 (pp. 248-52) are less helpful than usual on the meaning of such items. They suggest that nominal groups with *type* etc, are a sub-type of the construction that is here called the **partitive determiner**. Thus they describe *a new sort of computer* as ‘partition in respect of quality’ while *a piece of cake* is said to be ‘partition by quantity’. Yet it is clear that, when typic nouns are followed by *of*, they should not be grouped with the ‘partitive’ nouns that we met in Section 4 - if only for the structural reasons given there. It is perhaps significant that Biber et al (1999:256-70), in their approach to this area of grammar, break with their usual practice of adopting the categories of Quirk et al 1985. Instead, they treat this phenomenon - which they term a ‘species noun’ - as one to be considered in its own right. Their main contribution to the present grammar was to provide useful information on the probability of occurrences of *sort(s)*, *kind(s)* and *type(s)* and *variety/ies* in the four major registers around which their work is structured, and I have incorporated these probabilities in the generative version of the grammar presented in Sections 5.5 and 5.12.

To explain my route to the new conceptualisation of the semantics and syntax of typicality to be considered here, it will be helpful to introduce a new term: the **typic head**. We saw in Section 4 that the major function of the noun at the head of a nominal group is to spell out the place of the **widest referent** in the language's taxonomy of the **cultural classification** of 'things'. For the present discussion, then - and without prejudging the nature of their structural relationship to each other - we shall say that, in (1a), the word *sorts* expounds the **typic head**, and the word *documents* expounds the **cultural classification head** (or **cc head** for short). Note, then, that I am not at this stage suggesting an answer to the question of how the 'typic head' relates to the structures with which it occurs. That is precisely the question that we are investigating in this section - together with other questions that will arise in the course of this exploration.

The first intimations that it might be valuable to consider an alternative structure came when I noticed the **place** in the nominal group that the typic head typically occupied in the small corpus with which I was working at the time. These suggested that, unlike all the determiners introduced in Section 4, the typic head typically occurs IMMEDIATELY BEFORE the cc head. In other words, it seemed that there were, typically, NO MODIFIERS BETWEEN THE TYPIC HEAD AND THE CC HEAD. (We shall return later to consider how strongly that generalization should be stated.)

If this was so, what was the reason? In search of an answer, I re-examined (i) the semantics of the typic and cc heads and (ii) their relationship to each other - and I was struck by how closely their meanings are related. The main function of the cc head is to identify the referent in terms of the language's **cultural classification** of 'things' - and so in terms of the 'class' of 'thing' that it is. And the word *class* is a member of the list of items that function as the typic head. So, when a noun fills the head of a nominal group, as in *this ant*, it means something like 'The referent is a member of the class "ant"'. And, in a nominal group with a **typic head** such as *this type of ant*, THE SUBSTANTIVE REFERENT IS A SUB-TYPE (OR SUB-CLASS) OF THE TYPE (OR CLASS) SPECIFIED IN THE HEAD.⁵ And, taking this to its logical conclusion, we might adopt the position that the 'selection' is made in terms of the type of **meaning** realized in the cc head.

Sections 5.4 to 5.10 will explore this question: 'What are the implications of this way of viewing the data for the semantics and syntax of the nominal group?'

5.4 Four hypotheses

As I began to think about the data in terms of this general hypothesis, I noticed other patterns in examples such as (1a) to (3a) that seemed to fit neatly with it. After collecting 100 further examples (using the google search mechanism; see Section 5.6), I formulated the following four hypotheses, each of which I then tested them against the evidence of two very much larger corpora.

Hypothesis 1 In nominal groups that contain a **typic head**, the modifiers will PRECEDE not only the cc head (as ed) but also the typic head - with the result that there will be no modifiers between the typic head and the cc head.⁶ Thus, if we do find modifier-like items between these two elements, the prediction is that they will be part of a **compound noun** - and so PART OF the **cc head**. In other words, the possible modifier and the cc head will constitute EITHER (i) an established compound

⁵Strictly speaking, then, we should call the typic head the 'sub-typic head', but we shall stick with the shorter form of 'typic head'.

⁶More accurately, between the *of* that follows the typic head and the cc head.

noun - OR (ii) a **nonce-formation** of one (i.e. an item created by the Performer ‘for the nonce’ (Quirk et al 1985:1522). **Example:** in (3a), i.e. *one of the first of the new varieties of GM wheat*, the initials *GM* (which stand for ‘genetically modified’) are part of the recently-formed compound noun *GM wheat*.⁷

Hypothesis 2 There will similarly be no determiners between the typic head and the cc head. **Examples:** (1a) to (3a).

Hypothesis 3 The usual ‘describing’ relationship of the qualifier(s) to the cc head will be unaffected, since qualifiers always occur after it. However, the substantive referent is often expressed in the **typic head**,⁸ so that we should expect that if it requires a qualifier to classify it, this will occur after the cc head (since there is no other place for it to go). **Example:** in *the appropriate type of insulin (that) the person should use* (2a), the qualifier *(that) the person should use* classifies the referent of *type (of insulin)* - and not *insulin*.

Hypothesis 4 (This was originally motivated solely by personal interest, but we shall find that it too is relevant to the general hypothesis.) There is a dialectal variation between forms such as *different sorts of documents* - as in (2a) - and *different sorts of document* (with no final ‘s’). **Example:** The google corpus example in (1a) illustrates the ‘plural form’ usage in *documents*, while at least some British middle-class speakers of English in their sixties in 2006 (including myself) prefer the ‘singular form’ *document*.

5.5 The variables to be investigated: the TYPICITY network as the source

The next task is to identify the variables that we need to consider. In principle, we might investigate variations at the levels of either **form** or **meaning** - or both. In SFL, where the heart of the grammar is the system network of choices between meanings, we prefer, whenever we can do so with confidence, to state the variables in terms of the options in meaning of the system network - while at the same time paying close attention to their realizations at the level of form. The features that specify the main variables used here are shown in the right half of the system network in Figure 8, and we shall shortly meet realizations of all four types. However, as the entry condition to the TYPICITY network shows, each applies to both [count_cc] and [mass_cc] things, so that there are eight combinations to consider. And, when [overt_type] is selected, these are multiplied by the number of the choices between *type*, *sort*, *species* etc.

The network is given in the form in which it would be used in the generative version of the grammar, so that we can use it later on for this purpose. The three associated realization rules, (62.2), (80) and (80.01), are given in Section 5.12.⁹

⁷Notice that there was no hypothesis at that stage of the investigation about modifiers that precede the typic head. However, they were to emerge later as a significant factor (see Section 5.7).

⁸But it is also expressed in (i) the part of the nominal group that precedes it and (ii) the cc head, since a ‘type’ must be a ‘type of something’.

⁹The probabilities on the features in the systems are derived from various sources; those for [singular_cc] v [plural_cc] and for [singular_overt_type] v [plural_overt_type] are derived from Biber et al 1999 (pp. 334 and 256 respectively), and those for [overt_type] v [covert_type] and [singular_covert_type] v [plural_covert_type] are estimates - the latter being based on the probabilities for [singular_cc] v [plural_cc].

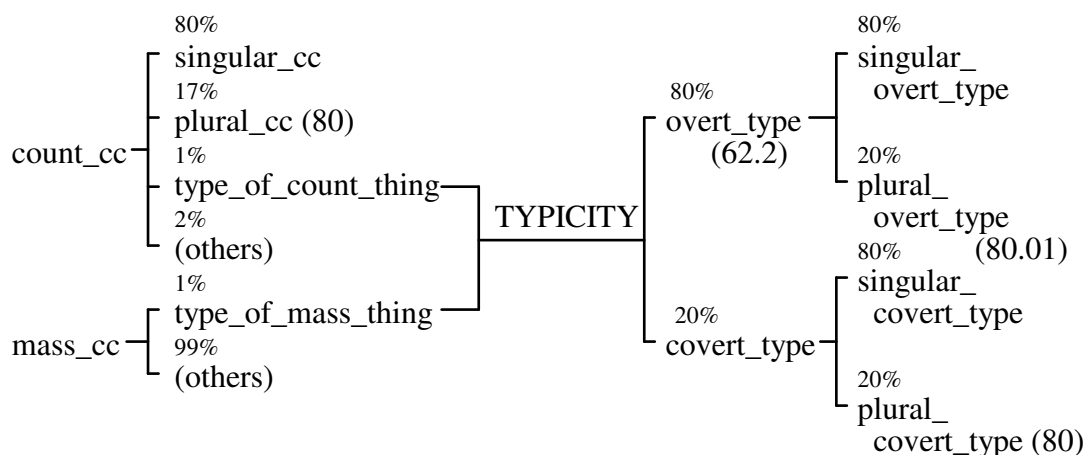


Figure 8 : The system network for TYPICITY

We can relate the network in Figure 8 to Examples (1a) to (3a) as follows (working through the network from left to right):

- 1 A [thing] in the system network for the CULTURAL CLASSIFICATION (cc) of ‘things’ is either [count_cc], as in (1a), or [mass_cc] as in (2a) and (3a).¹⁰
- 2 If either [type_of_count_thing] or [type_of_mass_thing] is chosen, the TYPICITY sub-network is entered. So in (1a) [type_of_count_thing] has been chosen, and in (2a) and (3a) [type_of_mass_thing] has.
- 3 Notice that the system entered from [count_cc] also includes [singular_cc] and [plural_cc]. This means that when [type_of_count_thing] is chosen the usual choice of [singular_cc] or [plural_cc] is not available. In other words, ‘typic’ nominal groups do NOT choose between making the cc head ‘singular’ or ‘plural’. (Thus, if there is variation in the cc head - as Hypothesis 4 suggests there is - this model predicts that it doesn’t realize a choice between meanings, but is due to some extraneous factor, e.g. the divided usage of different dialects and/or idiolects.)
- 4 We now come to the first system in the TYPICITY network - which, despite its primacy in the network, we have so far barely mentioned. The choice it offers is between [overt_type] and [covert_type] - and it is the choice of [overt_type] that is illustrated in (1a) to (3a).¹¹ To see the difference that the choice of [covert_type] makes, compare (1a) to (3a) with (1b) to (3b):¹²

(1b) (The system needs) a few different documents.

(2b) (You need to determine) the appropriate insulin the person should use.

(3b) (This is) one of the first of the new GM wheats.

- 5 Each of the two features of [overt_type] and [covert_type] then enters a system that

¹⁰This is a slight over-simplification, because some ‘things’ are [pair_only] (such as ‘scissors’ and ‘trousers’) and some are [plural_only] (such as ‘police’). The system to which these lead also includes the feature [type_of_count_thing], so that they too enter the TYPICITY system.

¹¹The realization rule 62.2 on [overt_type] ensures that one of the items *type*, *sort*, *kind*, *variety*, *species* and so on is generated on re-entry to the network (see Section 5.12).

¹²These examples, which are vital to the development of a full picture of what the model must be able to handle, were generated from (1a) to (3a) by ‘thought experiments’. This is a clear example of the need to use this source of evidence as a complement to evidence from corpora.

resembles the usual NUMBER system for ‘count’ noun senses. But this is a different system; it is the ‘type’ that is either ‘singular’ or plural’, so that the choice gets realized in the typic head - not the cc head. In (2a), then, [singular_overt_type] has been chosen, while in (1a) and (3a) [plural_overt_type] has.

6 The network from which the items *type*, *sort*, *species*, etc are generated is implemented in the full version of the grammar, but there is no space to show it here. However, it is partly inferable from the examples in Section 5.1, and Section 5.12 gives typical realization rules.

It will now be clear that modelling TYPICITY is not only a matter of identifying an appropriate way to represent the structure round a typic head. As (1b) to (3b) show, each of (1a) to (3a) has an equivalent form in which the meaning of ‘typicity’, while still present, is expressed **covertly**. We must therefore also identify an appropriate structure for cases where the ‘typicity’ is covert.

Interestingly, (1b) is ambiguous, in that these items can also be generated without entering the TYPICITY system. Does (1b) mean ‘many different types of document’, or simply ‘many different documents’? The context (a technical description of a computer system) suggests the former interpretation, but the latter is in principle possible too. This explains my earlier suggestion that the meaning of ‘typicity’ is much more frequent than many linguists - and especially corpus linguists, with their understandable preoccupation with evidence at the level of form - have so far assumed. And a functional grammar must attend to meaning as well as form.

5.6 Obtaining and supplementing relevant corpus data

To test the four hypotheses, I used three types of corpus evidence. The first was the ‘second level’ source of corpus data found in Biber et al 1999 - the first major grammar to be built around the concept of corpus frequencies. On pp. 256-7 they provide an interesting set of data on the probabilities for ‘typicity’ - but only at a fairly high level of generalization and - unsurprisingly - only in relation to items that are overtly realized.¹³ The main value of their data for the present study was that it established that, if I searched in a corpus for *sort(s)*, *kind(s)* and *type(s)*, I would probably cover around 95% of examples of nominal groups containing a typic noun.

The second type of corpus evidence was the vast (but unstructured) google corpus that is available to every user of the internet. A pilot study of the first 100 examples was used to check the coverage of the semantic categories derived from the system network in Figure 8, assuming that [overt_type] had been selected, and to formulate the four hypotheses. Next, I checked a further 200 examples to look explicitly for counter-examples to the hypotheses. (I shall comment on the latter in the next section.)

Thirdly, I checked the interim findings based on these pilot studies by a far larger sample from the COBUILD corpus. There were 500 examples of each of *type(s) of*, *sort(s) of* and *kind(s) of*. While all 500 examples with *type(s)* were ‘typic’, only 300 (60%) of those with *kind(s)* were, and only 200 (40%) of those with *sort(s)* were. I thus had a further corpus of 1000 examples of what we may call ‘typic nominal groups’.¹⁴ In addition, as a further test of Hypothesis 2, I examined all cases of *sort(s)*, *kind(s)* or *type(s)* followed by *of the* in the 56 million word version of the COBUILD corpus. But almost all these were what Biber et al

¹³It is extremely challenging - though not always impossible - to devise ways of interrogating a corpus for unrealized items, as in cases of ‘covert typicity’ such as (1b) to (3b), and for the purposes of the present study it did not seem necessary to attempt this major task.

¹⁴The test for whether *sort* or *kind* express ‘typicity’ is to replace the item by *type* - since *type of* cannot be used as a vagueness marker.

term ‘vagueness markers or hedges’ (1999:257), e.g. *He’d like sort of the goals widened.* (We shall explore the important implications of the few exceptions in Section 5.7.2.)

Here, contrary to established practice in corpus-based studies, I shall not illustrate my main points with corpus examples - valuable though these are for many purposes, especially as a source of counter-examples. Instead, I shall present ‘homogenized’ forms, in order to bring out the similarities and differences clearly. The corpora show that all four of the types being considered here exist in large numbers, and the VARIATIONS IN FREQUENCY ARE REFLECTED IN THE PERCENTAGES ON THE FEATURES IN THE NETWORK. A set of four such examples is set out below as (4a) to (7a), under the heading of **with [overt-type]**.

My next step was to derive from these a directly equivalent set with the feature [covert_type] - these being set out below as (4b) to (7b). Let us be explicit about the evidence for these. They are invented examples. But this doesn’t mean that there is no evidence for them. Each was produced by a careful ‘thought experiment’, in which I first constructed an example that would be equivalent to the original, but with covert typicity, and then asked myself whether it illustrated a frequent meaning-form unit in English. My judgement in each case was that it did. If I had had doubts, I would have conducted ‘informant-testing’ experiments.¹⁵ So, while we have no direct corpus evidence for (4b) to (7b) as we have for (4a) to (7a), I am confident that very substantial numbers of these types occur in natural texts.

The imaginary context for all of these examples is: *Scientists have recently discovered*

with [overt_type]

google COBUILD

(4a) a new type of ant	[singular_overt_type, count_cc]		
(5a) a new type of oil	[singular_overt_type, mass_cc]		
(6ai) two new types of ants	[plural_overt_type, count_cc]	99%	70%
(6aii) two new types of ant	[plural_overt_type, count_cc]	1%	30%
(7a) two new types of oil	[plural_overt_type, mass_cc]		

with [covert_type]

(4b) a new ant	[singular_covert_type, count_cc]
(5b) a new oil	[singular_covert_type, mass_cc]
(6b) two new ants	[plural_covert_type, count_cc]
(7b) two new oils	[plural_covert_type, mass_cc]

Figures are given for (6ai) to (6aii), because Hypothesis 4 requires this information (see Section 5.7.4).

¹⁵I would first use corpus evidence to check if I thought would help, but in the case of ‘covert typicity’ it would only work for ‘mass’ things with a ‘plural covert type’.

5.7 Interpreting the findings of the corpus searches

Let us now see how each of the four hypotheses fared.¹⁶

5.7.1 Hypothesis 1: findings

This hypothesis, you will recall, predicted that any apparent modifiers between the typic head and the cc head would be interpretable as elements of a compound noun. The corpus of 300 google examples included eight possible counter-examples, as shown in the underlined portions of (1) to (8) below.

- (1) a particular type of inflow band
- (2) various types of star clusters
- (3) one sort of reactive agent
- (4) four major categories of concept maps
- (5) types of variable stars
- (6) this type of joint resolution
- (7) different kinds of free email
- (8) the types of plant and animal communities living in and on the soil

Examples (1) to (5) appear to be technical terms in various fields, and so to be reasonably strong candidates for the status of ‘compound noun’. The problem with such data is that it is notoriously hard for someone from outside a specialist field to know what is a technical term in it and what is not. Expressions such as *inflow band*, *star clusters* and *concept maps* are typical compound nouns, being made up of two adjacent nouns. But compound nouns can also consist of an adjective followed by a noun, as in (3), so that we should not rule out (5) to (7). In Chapter 15 of Fawcett forthcoming a, I suggest a battery of tests to help identify compound nouns, one of which is the **pronunciation test** (aka the ‘stress test’). It supports the present analysis in several cases, i.e. *inflow band* is pronounced ‘*inflow band* (not ‘*inflow band*’), and the same holds for ‘*star clusters*, ‘*concept maps* and probably ‘*plant and animal communities*. And those of (1) to (8) that don’t pass the pronunciation test probably pass the **semantic test** - though only a member of the community using this particular technical vocabulary could decide with confidence. It is therefore seems possible that the underlined portions of (5) to (8) are technical terms in their respective fields, and so also compound nouns.

Because of the number of these borderline cases, I then made a detailed study of similar examples in the 1000 examples of the COBUILD corpus. Surprisingly, over 10% of these were either (i) a clear case of a compound noun functioning as cc head (*cervical cap*, *Christmas card*, *climate change* etc) or (ii) a ‘modifier-head pair’ with such a strong collocational link (if only for a specialist group) as to suggest that they function as compound nouns (*automatic weapons*, *biosafety suit*, *business loans* etc). These two sets of examples constituted about 3% of all typic nominal groups.

However, there was a third group that was twice as large, in which there was a strong

¹⁶The first three hypotheses are also supported by the fact that all the corpus examples cited in Biber et al (1999:257) and in the relevant entries in the *COBUILD English Dictionary* conform to these predictions. There is one apparent counter-example in Biber et al, i.e. *His eyes had a kind of icy brilliance about them*. Here the modifier *icy* occurs immediately before the cc head, yet seems unlikely to be ‘classifying’. By coincidence, however, the following section concerns vagueness markers. Their examples are all from the register of conversation, with none from fiction - the register from which the above example comes. The ‘typicity’ test of replacing *kind* by *type* produces a very odd-sounding result, i.e. *His eyes had a type of icy brilliance about them*, and this suggests that we might in this case interpret *type of* as a literary ‘vagueness marker’.

case for recognizing a modifier between *of* and the cc head. These included:

American novelists, arabesque marquetry, clear criteria, cognitive errors, confidential material, good advice, gimmicky Irish-American showband, ingratiating nod, largish coil, spectacular match, titanic battle.

Some of these have no collocational ties (e.g. *American novelists*) and, while others do (e.g. *good advice*), they don't pass the tests that identify compound nouns. It is therefore clear that Hypothesis 1 must be either revised or rejected.

I propose this revision: that the grammar permits modifiers which serve the **classifying** function (see Section 2.4) to occur here. In other words, for the purpose of defining the 'class' of 'thing' from which the 'sub-type' (or 'sub-class') is selected, ANY MODIFIER THAT OCCURS BETWEEN *of* AND THE CC HEAD IS DEEMED TO BE PART OF THE CLASSIFICATION OF THE WIDEST REFERENT. This in turn suggests that they should be modelled in a similar way to compound nouns, i.e. as a nominal group with a 'modifier + head' structure that fills the head. And this analysis, notice, requires only a minimal modification to the basic claim of Hypothesis 1 (i.e. that modifiers do not occur between the typic and cc heads).

Your response might well be to point to adjectives in the above examples that are not typically used to **classify** (e.g. *good, gimmicky, ingratiating, spectacular, titanic*). Indeed, two are purely **affective** modifiers, (*good* and *spectacular*). But these are the only two in 1000 examples, and we must give weight to this fact too. The Cardiff Grammar does this, recognizing that ANY modifier can serve the **classifying** function - even an **affective** modifier, in certain cases. And a close examination of the above cases, in their context, suggests that these modifiers do indeed classify the widest referent - and so are, in effect, part of it.

In all these cases, then, a modifier may occur between the typic and the cc head. And with this proviso Hypothesis 1 can stand.

However, while studying the 1000 COBUILD examples, I noticed a second pattern that is interpretable as evidence for the possible new approach. It concerns the modifiers that occur BEFORE the **typic head**, but we shall consider this phenomenon now because their meanings relate, like those that occur directly BEFORE the cc head, to the **cc head**.

Well over a fifth of the typic heads in the 1000 examples were preceded by modifiers - which is within the expected range.¹⁷ Predictably, almost 90% of them clearly describe the referent of the **typic head** that immediately followed them - typical modifiers being *different* (40%), *certain* (17%), *other* or *another* (10%), *various* (8%) and *particular* (7%).¹⁸ Interestingly, however, a significant minority (over 10%, so almost 3% of the 1000 COBUILD examples) contain modifiers whose semantic values appear to describe the referent of the **cc head**. Their position is surprising, because THE PERFORMER COULD INSTEAD HAVE PLACED THEM IMMEDIATELY BEFORE THE CC HEAD. Typical examples (with the relevant modifier underlined) are as follows.¹⁹

(9) similar psychosocial types of illness (cp similar types of psychosocial illness)

¹⁷This assumes that modifiers occur with typic heads in the same proportion as with other types of 'common noun' head. But this varies according to register: in academic and news registers it is around 40%, but it is only 8% in conversation (figures based on Biber et al 1999:578).

¹⁸The rest were spread among a variety of different items, some being semantically related to those listed above, e.g. *same, varying*. Others reflect frequent parameters for identifying subtypes (*new, traditional, older; rare, common, obscure*), and others importance (*main, predominant*). A further set expressed affective evaluation (*special, exciting, popular*). There are two other types of modifier that precede the typic head, and these will be mentioned briefly in Section 5.7.2.

(10) this callous type of killing (cp this type of callous killing)

(11) an inefficient and ineffective type of person (cp a type of inefficient and ineffective person)

(12) the longer term strategic type of research (cp the type of longer term strategic research)

Why, we may ask, do speakers of English sometimes place such modifiers BEFORE THE TYPIC HEAD? Is this an area of variation in dialect and idiolect - possibly leading to language change? Might the same person use both structures? Only further research can tell us. One possible explanation is playfulness - i.e. it is fun to test the 'rules' of our syntax, as poets do. But another could be that we sense the close semantic relationship between the meaning of 'sub-class' expressed in the typic head and the meaning of 'cultural classification' (i.e. 'class') expressed in the cc head, so that we try to avoid separating the two types of head (other than by *of*) by placing the modifier before the typic head, so raising questions about which is the 'main' head.²⁰

To summarize: in the surprisingly high proportion of 10% of examples with modifiers, a modifier that we might have expected to occur immediately before the cc head position is in fact placed BEFORE THE TYPIC HEAD. We should therefore consider what a structure of the nominal group designed to handle such phenomena might be like, and we shall do this in Section 5.10.

Both of the findings noted here suggest that, even though we have had to modify the original hypothesis slightly to allow classifying modifiers to occur, Hypothesis 1 does indeed express a clear tendency in modern English.

5.7.2 Hypothesis 2: findings

The second hypothesis predicted that there would be no determiners between the typic head and the cc head. In the small google corpus there was just one case which challenged it:

(13) What type of a virus do you consider the most dangerous?

But in the COBUILD corpus there were ten further cases with *a*, including:

(14) If you want to use that kind of a metaphor

(15) I do believe that it is some sort of a right.

And in the full google corpus there are almost three million examples containing *type(s) of a* Interestingly, in the hundred or so cases I inspected the item *a(n)* can always be omitted without a change of meaning. This suggests that this may be one of those areas of English in which there are - at least for the moment - alternative forms that express the same meaning. So the concept of variation in dialect and idiolect - possibly leading to language change - may be relevant here, as it is in the findings on Hypotheses 1 and 4 (see Section 5.7.4).²¹

If the position was that the quantifying determiner was always expounded by *a(n)* in such cases, the grammar could accommodate it by treating *a metaphor* in (15) and *a right* in (16)

²⁰

²¹One possible way in which examples such as (14) and (15) may have come into use is as follows. Since the word *a* is pronounced as a schwa, such examples may be a representation of a spoken form that adds a schwa between *of* and the initial consonant of the following word - and subsequently this minimal phonetic addition has been transferred to the written form as *a* (rather as a primary school child may write *He must of seen me*). A second (and complementary) possibility is that forms such as *a metaphor* and *a right* may be being used here to express the concept of the 'whole class' of the referent (aka 'generic reference').

as a nominal group that is embedded in the head - rather as with the **m h** structures in Section 5.7.1. However, a further search in google revealed small but still substantial numbers of equally valid examples such as (16) and (17):

- (16) Insects are only one type of several potential biological control agents.
 (17) The operator compares the type of two expressions.

The conclusion to be drawn from these examples is that the grammar must provide for the presence of a quantifying determiner between the typic and the cc heads, and we shall consider this requirement in Section 5.10.

But what about the deictic determiner *the* in this position? There were no such examples in the small google corpus, but on the model of (13) I created What type of the virus do you consider the most dangerous? This I judged to be acceptable, so I then made a special search in 500 COBUILD examples of *type / sort / kind of the xxx*. I found only one clear case, i.e. (18).²²

- (18) The psychosocial type of the condition (will determine the outcome).

However, I then consulted the full google corpus, which registered almost 8.6 million cases of *type(s) of the* - showing clearly that here too there is a phenomenon that requires explanation.

But how far is this a case simply an equivalent of (13) to (17) with a deictic determiner? Consider first the semantics of *the* in (18). We can infer that the referent of *the condition* has already been referred to in the text (or is recoverable from the context of situation), and that it is this that enables it to be referred to here by *the condition*. This, then, is the only explanation needed for the occurrence of *the* before the cc head. But does this affect the relationship between *the psychosocial type* and *of the condition*?

At this point we need to pause, in order to note that there are two types of modifier that precede a typic head, in addition to those examined in Section 5.7.1. Both are typically filled by a nominal group (which usually has just a head). In the first sort, the modifier states the **variable** (or 'attribute') that is relevant to specifying the 'type of xxxxx', as in (19):

- (19) What is the content type of the file?
 = 'the type of the file, in terms of the type of its content'

But it may also be an adjective - as is *psychosocial* in (18). So *psychosocial* in (18) serves essentially the same function as the noun *content* in (19), i.e. both identify the 'dimension of variation' within which the 'type' is to be specified.

But in the second sort of 'thing' that fills the modifier that precedes the typic head is very different. Surprisingly, it states the cultural classification of the referent - i.e. it is THE NOUN THAT TYPICALLY OCCURS AS THE CC HEAD. Compare (20a) with the more familiar structure of (20b):

- (20a) Map the element type to a root table.
 (20b) Map the type of element to a root table.

In the case of examples such as (20a), then, the grammar cannot avoid the fact that THE MEANING OF 'TYPE' IS TREATED AS THE HEAD OF THE WHOLE NOMINAL GROUP. This implies that it is functioning as a 'cultural classification' IN ITS OWN RIGHT - i.e. as a particularly

²²All the cases of *sort of* and *kind of* were vagueness markers.

abstract class of ‘thing’.²³ Probably the most frequent example of this structure is *the blood type* (= ‘the type of blood’) - though for many users this now functions as a compound noun.²⁴

The fact that typic nouns sometimes occur as the cc head of a nominal group, as in (20a), raises the question of whether the typic head in (18) and (19), though semantically different, should also be analyzed as a cc head.

The picture is further complicated by the fact that other variables in the specification of ‘things’ - other ‘dimensions of variation’ - can be co-ordinated with *the type*, such as *the size* in (21a). This strongly suggests that *type* and *size* are the heads of two co-ordinated nominal groups. If so, the full form of the underlined portion of (21a) that showed this ellipsis would be as in (21b).

(21a) The treatment varies with the size and type of the tumor.

(21b) the size (of the tumor) and (the) type of the tumor

Interestingly, this suggests is that ‘size’ can be construed as a type of ‘type’ - just as ‘psychosocial type’ in (18) and ‘content type’ in (19) are types of ‘type’. And this in turn opens up a whole new area of investigation, i.e. the possibility that concepts such as ‘size’ in *What size of shoes do you wear?* are in fact realized as typic heads. Indeed, we can answer the question *What type of thing is it?* by using virtually any of the many ways in which a ‘thing’ can be specified - so by a modifier, a qualifier (so using a full clause) and even some types of determiner. As a textbook example (literally) of this, consider this text, which is headed *Types of volcano*:

Volcanologists have classified volcanoes into groups based on the shape of the volcano, the materials they are built of, and the way the volcano erupts.

Clearly, more research needs to be done on these sub- and super-types of ‘typicity’ before a comprehensive and principled analysis emerges. Regretfully, we cannot reach a clear position on these matters here.²⁵ But we do need to note that it is at least arguable that, in nominal groups that contain the *type of the xxxx* pattern, the ‘typic head’ functions as the ‘cc head’ of the whole nominal group (with what follows analyzed as a qualifier). This would mean that such cases needn’t be treated as counter-examples to the claim made in Hypothesis 2.

In summary, then, we can say that English prefers not to have a determiner between the typic and cc heads. But the grammar must nonetheless provide for certain cases in which quantifying and deictic determiners appear to occur in this position - perhaps by providing for alternative structures within the cc head in some cases. But in others - e.g (16) - it seems that a **qd** occurs between the two types of head, and it would be useful if the grammar provided for this (e.g. in the way suggested in Section 5.10).

²³Notice too that, if the referent of *the element* had been more readily recoverable, it would have been referred to by *it*, as in *Map its type to the root table* - so giving us a second realization of ‘typic’ meaning with *type* as the cc head.

²⁴We shall not consider cases on the boundary between such cases and cases in which the typic selector and the following head are ellipsed, as in *There are three types of volcano. The first type is ...*

²⁵These areas of ‘typicity’ require further investigation, perhaps as a PhD topic, and the current grammar needs to be expanded to accommodate them.

5.7.3 Hypothesis 3: findings

The third hypothesis predicted that if a qualifier follows the cc head it would normally describe the referent of the typic head (plus what precedes and follows it, including the cc head). This proved to be overwhelmingly the case, typical examples being:

- (22) the type of fish (that) they like catching
 (23) the sort of detail (that) I could never get from photos

In (22), for example, (*that*) *they like catching* completes the meaning of *the type (of fish)* - not *fish*.

Occasionally, however, the corpus provided examples where the qualifier's function was to complete the meaning begun in the cc head, as in:

- (24) a very different type of agreement with former Cold War rivals
 (25) different types of behaviour from men

In these and many other similar examples the cc head is an 'event noun' - so that it is predictable that a qualifier will be introduced to express a Participant Role associated with the event, as here. Thus such cases are not counter-examples to the claim, but examples of a functionally different structure.

The significance of this last finding is considerable, because it clearly suggests, once again, that we should treat the typic head as an element of the matrix nominal group - rather than locating it in an embedded nominal group. I shall explain why this is so when I introduce the possible new structure in Section 5.10.

5.7.4 Hypothesis 4: findings

The fourth hypothesis was concerned with a possible minor dialectal variation - i.e. the presence or absence of 's' in examples such as the following (repeated below from Section 5.6):

google COBUILD

- | | | |
|-----------------------------|-----|-----|
| (6ai) two new types of ants | 99% | 70% |
| (6aii) two new types of ant | 1% | 30% |

I added this fourth hypothesis because of my puzzlement at the difference between (i) my own usage (e.g. *types of determiner*, in this paper) and (ii) the evidence from (a) Quirk et al (1985:764-5), (b) Huddleston & Pullum (2002:352-3) and (c) the original small google corpus. Even though the two great 'comprehensive' grammars devote most of their discussion of this construction to the morphology of 'agreement' (which (6aii) challenges) the surprising fact is that neither provides for (6aii). And google yielded only four examples. So was my usage an illegitimate variant of Standard English, I wondered?

Biber et al (1999:255-6), however, do allow for it (e.g. *certain types of car*) - but only as a marked form that fails to follow what they present as an 'agreement principle'.²⁶ However,

²⁶The rule implementing this 'agreement principle' states: 'If the "type" is "plural", make the cc head "plural" too - but only if the cultural classification is "count", as in *two sorts of ants*, and not "mass". Thus, despite the existence of *these oils* (in the sense of 'these types of oil') we shall not find cases of **these sorts of oils*. Thus, while Biber et al present their 'agreement principle' as applying generally, it does not in fact apply when the cc head is 'mass'.

in a study of 500 cases of *types of* in the COBUILD corpus, I was relieved to find a 30% usage of the pattern I use, as in (6aii) above. So my usage is perhaps not so aberrant as its omission in two of the three great grammars implies!²⁷

In contrast, all three big grammars point out that we shall attest forms with a different type of lack of ‘agreement’, as in *These sort of parties are dangerous*. While I found just one such example in the small google corpus (*these type of files*) and only two in the COBUILD corpus of 1000, there were almost 0.8 million examples in the full google corpus.

All such variants, however, are matters of morphological variation according to dialect or idiolect - and they may be instances of language change. They should therefore be handled in the realization rules (i.e. NOT as choices between meanings in the system network). So how, you might ask, are they relevant to a study of the semantics and syntax of this construction? See Section 5.10 for the answer.

5.8 The implications of the findings: five statements

We can now replace the FOUR hypotheses by the following FIVE statements - each of which needs to be taken into account when deciding how we may most appropriately represent the structure of such examples, as we shall in Section 5.10. (Statement 1a arises from a modifier pattern noted while investigating Hypothesis 1.) The ‘description’ avoids commitment to any one structure, so is simply in terms of the string of items and unit classes (as in the products of COBUILD, such as Francis et al 1996 & 1998).

Statement 1 In a nominal group that contains a **typic head**, there will usually (with a 95% probability of all typic nominal groups) be no **modifier** between the *of* that follows it and the **cc head** (which may be expounded by a simple or a compound noun). When there is one, it is functioning as a **classifying** modifier that describes the referent denoted by the cc head.²⁸

Statement 1a Occasionally (with a 3% probability) a modifier which appears to describe the referent of the cc head - and which we would therefore expect to occur immediately before it - occurs BEFORE THE TYPIC HEAD.

Statement 2 There will almost always (over 99%) be **no determiner** between the typic and cc heads. Nonetheless there are several structures in which quantifying and deictic determiners do occur in this position and, while no final decision can be made at this point on the analysis of all of them, it may well be useful if the grammar provides for the possibility of modelling THE REALIZATION OF TYPICITY AS OCCURRING BEFORE THE QUANTIFYING DETERMINER.

Statement 3 When a **qualifier** follows the head, it will probably (over 90%) describe the substantive referent, i.e. the referent of the **typic head** in the examples considered in this paper (and so also of the rest of the nominal group that precedes and follows it, including the cc head, since a ‘type’ must be a ‘type of something’). Exceptions occur when the function of the qualifier is to fill out the meaning of the cc head - e.g. in nominalizations, where it is expounded by a lexical verb or an ‘event’ noun.

Statement 4 There is dialectal or idiolectal variation (so possibly language change) if (i) the typic head has a ‘plural’ meaning and (ii) the cc head denotes a ‘count’ referent. The variants are:

²⁷Could it be that Huddleston and Pullum, who are both British in upbringing but who have both worked elsewhere for most of their professional lives (Huddleston in Australia, Pullum in the US) have in this respect lost touch with their British roots? But that wouldn’t explain the omission in Quirk et al 1985.

²⁸Since this is not the focus of this paper, let me say at this point that the resulting **m h** structure is to be treated as a **ngp** embedded as the **head** of the matrix **ngp**.

- 1 *these sorts of comments* (the most frequent),
- 2 *these sorts of comment* (half as frequent, possibly most frequent in British English) and
- 3 *these sort of comments* (relatively rare in standard English because this form is stigmatized, but increasingly used).

However, this usage does not extend to mass nouns. (See Section 10 for the relevance of these findings to the main issue with which this paper is concerned.)

Thus the three examples in (1a) to (3a) with which we started Section 5 are indeed typical ‘typic nominal groups’ - except that I have found no corpus examples that remotely match the complexity before the typic head in the invented example of (3a). Nonetheless, I have found, over the years, many cases with two and occasionally more such determiners and their selectors that precede a typic head, so I am satisfied that the overall model needs to provide for examples such as (3a).

We turn now to the question of how these proposals should be formalized in the grammar itself. We shall first discuss the alternative ways of representing the **syntax** of ‘typic’ constructions - both overt and covert - and then I shall set out the realization rules that translate the **semantic features** from the network in Figure 8 into the **formal structures** that we are about to meet.

In the functional approach to language used here, we assume that the representations at the levels of form and meaning are mutually dependent - *tout se tient*, as Meillet (1937) expressed it. And this gives a central role to the realization rules that relate the two. Any discussion of alternative structures, therefore, necessarily involves asking questions about (i) how directly such structures correspond to the meanings they express, and (ii) whether they would cause problems for the realization rules.

Since it is simpler, we shall start with the structure required when the typicity is **covert**. Then in the following section we shall address the trickier problem of deciding on the structure to represent **overt** typicity.

5.9 The structure of examples with ‘covert typicity’

How should ‘covert typicity’ - as in (4b to (7b) in Section 5.6 - be realized at the level of form?

The type of example in which it is most obvious that ‘covert typicity’ is present is the type when the cc head is expounded by an inherently **mass** noun, as in *a new oil* (5b) and *two new oils* (7b). In such examples, the paradoxical effect is that the head that realizes the meaning ‘mass’ (i.e. *oil*) also signals the meaning of ‘singular’ or ‘plural’ - and so apparently also the meaning ‘count’. In other words, it is the **cc head** that expresses the meaning of ‘singular covert type’ or ‘plural covert type’ - meanings which, when ‘typicity’ is realized OVERTLY, are realized in the **typic head** (as *type* or *types* etc). And, because nominal groups with a ‘mass’ referent do not express the meanings of ‘singular’ and ‘plural’ (except in the present case and *the two teas* case) we recognize immediately that these meanings apply here to the covert meaning of ‘type’.

However, ‘covert typicity’ also occurs with **count** nouns, as in *a new ant* and *two new ants* in (4b) and (6b) in Section 5.6. But the ‘typicity’ is usually less obvious. In our initial example of *a few new documents*, for example - which we met in (1b) in Section 5.5 - it isn’t immediately clear whether the referent is a ‘count’ thing with a ‘plural type’ that is ‘covert’, or simply a ‘plural’ thing - though the context suggests the first. But in an example such as *Six new stamps are on sale at the post office* the meaning of ‘covert typicity’ is quite obvious, because we know from our beliefs about what post offices sell that the Performer is referring to ‘types of stamp’.

Moreover, the far greater frequency of ‘count’ nouns over ‘mass’ nouns suggests strongly

that there are likely to be many more cases of ‘covert typicity’ than most grammars appear to assume, so that ‘typicity’ may well be far more central to understanding the English nominal group than is commonly assumed.²⁹

The key question for this section is: ‘Is the syntax of *a new ant* in (1) below - which is like (4b) in Section 5.6 - different from that in (2) below?’

(1) Scientists have recently discovered a new ant.

(2) A new ant is coming to help (spoken while watching a nature film of ants transporting a leaf).

Is there any alternative to treating *ant* as the head in both? It might at first seem desirable to treat *a new ant* in (1) as derived by ellipsis from *a new type of ant*, with *type of* ellipsed. But now consider (3) - which is like (4b) in Section 5.6.

(3) Scientists have recently discovered two new oils.

Clearly, this cannot be an ellipsed version of **two new types of oils*, because we do not use this form (as noted in Section 5.7). This demonstrates that ‘covert typicity’ really is a choice in the **semantics** (as modelled in Figure 8) - and that it is not to be explained as ellipsis at the level of **form**. We therefore use the same structure (**qd m h**) for both (1) and (2).

Thus all such examples share the same syntax but are, in principle, ambiguous between the two semantic interpretations.

5.10 The structure of examples with ‘overt typicity’

We come now to the most challenging question of this paper: ‘What structure should be used to represent nominal groups with ‘overt typicity’, such as *two new types of ant(s)* in (1)?

(1) (Scientists have recently discovered) two new types of ant(s).

We shall consider in turn each of four possible structures, bringing to bear on the decision two major types of evidence: (i) findings from the corpus studies described in Sections 5.4 to 5.8, and (ii) factors involved in integrating the possible structure into the full, generative version of the grammar. However, the first type of evidence inevitably involves the second, as we found in Section 4.2.³⁰

5.10.1 Four candidate structures

I suggest that there are four main candidate structures (though more could be invented). We shall begin by considering two that we have already rejected in Section 4.2 for a formally similar structure - that of *five of those books*.

Option A (as in Section 4.2) is the ‘prepositional group as qualifier’ construction (as

²⁹It would be an interesting project to try to establish what proportion of apparently simple nominal groups in fact carry the meaning of ‘covert typicity’.

³⁰Thus, while traditional ‘formal tests’ can help in deciding between alternative structures in a ‘text-descriptive’ grammar that doesn’t have a ‘theoretical-generative’ counterpart (see Fawcett 2000:78-81 for these terms), they are regarded here supplementary guides to determining the structures required as outputs from the realization rules. They are replaced here by (i) the criterion of elegance, i.e., roughly, simplicity in the operation of the grammar and (ii) the principle of maximizing the transparency of the representation at the level of form of semantic relationships between elements.

illustrated in Figure 4). Section 4.2 cited five reasons why Option A would be an extremely poor choice for modelling *five of those books*, most being derived from the basic evidence of how a SFG works. Essentially THE SAME REASONS APPLY IN THE PRESENT CASE. The main difference comes in the second major reason cited in Section 4.2. Here it is not choices in ‘quantity’ that are conditional on the choice of ‘singular’, ‘plural’ or ‘mass’ (as realized in the head), but the probabilities in the network of choices in ‘type of type’ (e.g. *type* v *species* v *brand* v *genre* v *strain* etc), since these vary with the ‘cultural classification’ of the head. And in this case too it is important to make the choice of a meaning realized in the head ON THE FIRST TRAVERSAL OF THE NETWORK - because the choices in the ‘type’ network can only be safely entered when the grammar knows the ‘class of thing’ of the widest referent. (We shall return to this point in Section 5.10.3.)

But in a typic nominal group there is a further reason. This is that in *a new ant* - whether it is ‘simple’ or realizes ‘covert typicity’ - the word *ant* has to be the head of the nominal group (as we saw in Section 5.9). So, to enable the relevant realization rule to capture the generalization that *ant* realizes the ‘cultural classification’ of the ‘thing’ in all these cases, we must treat *ants* in (1) as the head of the matrix nominal group.

For these six compelling reasons, then, we reject the use of the ‘prepositional group as qualifier’ construction for modelling such examples.

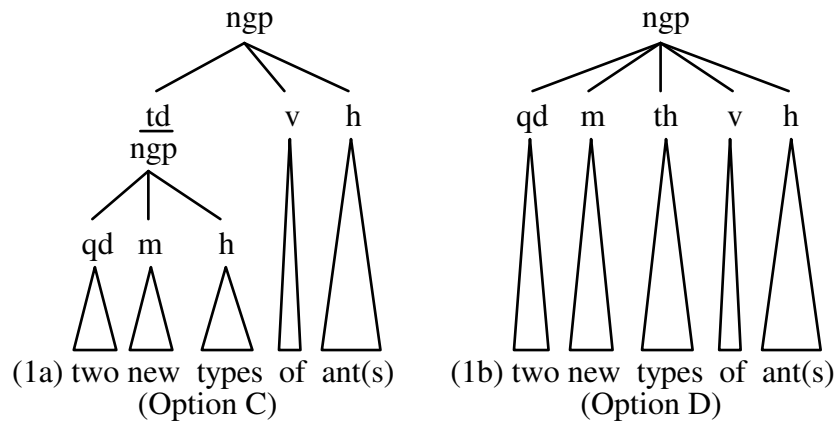
Option B is the mirror image of Option A. In such a structure, *of* would function as a ‘postposition’, and in this case it is the typic head that would be buried one or two layers down. Again, we considered this structure in Section 4.2 for *five of those books* and we rejected it. Here we reject it again - and for similar reasons. The first is that we often need to generate examples such as the short version of (2), in which the cultural classification is recoverable - i.e. with neither the cc head nor the preceding selector *of*. This is most simply achieved if both are elements of the same unit, as in Options C and D. The second reason to reject this option is that, like Option A, it has the disadvantage that it introduces at least one additional layer of structure that is avoided in Options C and D.

(2) What type (of ant) is that?

We come now to the two final candidate structures. **Option C** is to treat overt ‘typicity’ as a type of ‘selection’ - so on a par with the structures that realize meanings such as ‘part’ or ‘representation’ (which we met in Section 4.3). Its structure is therefore like that of Option C in Figure 9. This is the structure that I have assumed, for the last thirty years or so, to be the best way to represent the semantic relations involved - but which I have recently begun to query. Essentially, it represents the view that, just as a **representational determiner**, as in *a recent photo of her house* or *his concept of democracy*, is one type of ‘selection by abstraction’, the **typic determiner** in *a new type of house* and *this variety of democracy* is another.

However, we have seen that ‘typicity’ involves certain factors that make it significantly different from most of the determiners described in Section 4.3. One is the existence of both **covert** and **overt** versions - though this occurs, arguably, with ‘representation’, as we saw in Section 4.3. Another is the concept introduced in Section 5.3, i.e. the notion that the relationship BETWEEN THE TWO ITEMS THAT EXPOUND THE TWO HEADS (the **typic head** and the **cultural classification (cc) head**) may be best modelled as one of ‘selection’ between heads. It was to investigate this notion that I undertook the corpus studies described in Sections 5.4 to 5.8.

However, the structure that most directly reflects the new interpretation of the semantic relations involved is **Option D**, i.e. the attractively simple one shown in (1b) in Figure 9 (where **th** stands for **typic head** and **h** stands for **cc head**).



Key (possible new categories only):
 td = typic determiner (preferred structure)
 th = typic head (rejected structure)
 h = head (also described here as ‘cultural classification head’)

Figure 9: Two ways of representing the structure around the typic head

The choice therefore lies between Options C and D. As before, there are two types of evidence to take into account: (i) the findings of the corpus studies and (ii) the requirements of the full, generative version of the grammar.

5.10.2 Evaluating Options C and D: (i) evidence from the corpus studies

Let us begin by asking how well each structure matches the requirements of the data that have emerged from the corpus studies (as summarized in Section 5.8).

Statement 1 summarizes a view of the structure in which no modifiers may occur between the typic and cc heads. Those which at first appear to do so are interpreted as part of the ‘classification’ of the referent, and so as occurring WITHIN the **cc head** (either as a compound noun or as an embedded nominal group with a **m h** structure). Option D can accommodate this fact by allowing a further embedded nominal group at the cc head (if there is no existing compound noun). Precisely the same analysis, however, could be used with Option C. And Option C would also be able to model any modifiers that describe the referent of the cc head as occurring between the **td** and the **h**, if this were after all to prove desirable. So on this criterion the evidence doesn’t point clearly to either option.

However, **Statement 1a** points out a problem for Option C - namely, that we sometimes find a **modifier** which describes the **cc head** - but which actually occurs BEFORE THE TYPIC HEAD. The question is therefore ‘How can the grammar ensure that in such cases the adjectives generated for such a modifier are semantically appropriate to the cc head?’ Recall (from Note 3 in Section 2.4) that in this version of SFG adjectives do not expound modifiers directly; they expound the apex of a quality group that fills the modifier (because they can be ‘tempered’). The most straightforward solution to the problem is therefore to HAVE THE CC HEAD IN THE SAME UNIT AS THE MODIFIER, as in Option D. Then the realization rule for the feature which inserts the **modifier** into the structure can also do the following. Referring to the relevant features in the CULTURAL CLASSIFICATION network, it can state for each (using preference rules) what features are more and less likely to be selected on re-entry to the network to generate the adjective at the **apex** of the **quality group** that fills the modifier. However, the grammar required for Option C (in which the quality group is one layer lower in the structure) can handle such cases with only a little additional complexity, so Option D

has only a small advantage over Option C.³¹

Statement 2 points out that, while there are usually no **determiners** between the typic and cc heads, there are several structures with quantifying and deictic determiners in this position whose semantics and syntax require further study - with a strong case already for some. For this reason it would be desirable to provide for the possibility of modelling the realization of ‘typicity’ as occurring BEFORE THE QUANTIFYING DETERMINER - other things being equal.. This is therefore an argument - though one based at this point only on the likely need for future extensions of the grammar - for Option C.

Statement 3 makes essentially the same requirement on the grammar as Statement 1a. When there is a **qualifier** (which follows the cc head), the grammar typically needs to ensure that its internal semantics is appropriate to the **typic head** rather than the ccc head - even though it isn’t adjacent to it. The grammar operates most straightforwardly if the qualifier is a sister element to the typic head. But in this respect Options C and D are equivalent, because in Option C any qualifier of the typic head would be shown as an element of the nominal group that fills the **td** (as we saw in Section 5.12). Option C, however, involved a type of complexity in the realization rule that is avoided in Option D, i.e. the discontinuity that arises from the fact that the qualifier occurs after the cc head. Thus the qualifier of *types* occurs after *ant(s)*. This is quite a strong argument for Option D.³²

Statement 4 described three co-existing patterns in the morphology of the typic and cc heads, where the differences reflect register and dialectal differences rather than differences of meaning. (Two of them are shown in Figure 9.) This is further support for the picture suggested in Statements 1a and 2, i.e. that there is uncertainty as to where the ‘headship’ of a typic nominal group lies. And this in turn could perhaps be interpreted as support for the radical new analysis in Option D (though we should consider whether other concepts, such as the combination of ‘substantive referent’ and ‘widest referent’, may provide an equally appropriate framework for expressing this uncertainty). Moreover, if the **cc head** is to ‘agree’ with the **typic head** (as in the most frequent pattern of *two new types of ants*), this ‘agreement’ is a little simpler for the grammar to provide for IF THE TWO HEADS ARE ELEMENTS OF THE SAME UNIT. But realization rule for Option C can be adapted to handle it with the addition of one line, so it is not a major criterion.

On balance, then, the findings of the corpus studies favour Option D. In other words, at this stage of the investigation of ‘typicity’, there seemed to be several advantages in treating all four elements of modifier, typic head, head and qualifier (together with any preceding determiners) as sister elements of the same unit.

5.10.3 Evaluating Options C and D: (ii) their use in generating typic nominal groups

When the corpus evidence that seemed to favour Option D began to grow, I decided that I must test its viability in the generative version of the grammar. I therefore wrote new realization rules for the basic ‘typicity’ network in Figure 8, and the result was a satisfyingly elegant grammar. The problem was that the rules that generated Option C were equally elegant - so this exercise failed to resolve the issue.

However, when at the end of the corpus studies there was still no clear conclusion, I went back to the two versions of the generative grammar, in order to try to refine them to the point

³¹The grammar already provides for cases when the selection of a given feature on one traversal of the network influences the selection of a feature on a later traversal of the network - even when it is realized two layers lower in the structure.

³²However, I should point out that this doesn’t make a new requirement on the grammar, since it already has to provide for a small number of other types of discontinuous unit (e.g. the underlined quality group in *The Prime Minister is a more important person than the Chancellor*).

where they could also provide for the variation in the probabilities on the features in the sub-network from which the items *type*, *sort*, *breed*, *strain* etc. are generated. It was then that I discovered that Option D would require a significantly more complex grammar than Option C.

The specific problem is that of how to model the fact that, if the cc head is *sheep*, the typic head is much more likely to be *breed* than, say, *brand*, *make* or *strain*. And one of the goals of the present grammar is to model probabilistic ‘rules’ such as these as well as ‘all-or-nothing’ rules. What makes this difficult for Option D is that the features from which both *sheep* and *breed* are generated are located in the CULTURAL CLASSIFICATION network. In Option D the typic and cc heads are elements of the same unit - so that, if the network is to generate both *sheep* to expound the cc head and *breed* to expound the typic head, it would need to be entered twice ON THE SAME TRAVERSAL OF THE NETWORK.

It is certainly POSSIBLE to construct a SFG that enters the same network more than once when generating a unit (and we have experimented with that approach in the COMMUNAL Project). And it is probably also possible to provide for the probabilities to be changed in conjunction with this (though so far as I know no SFG has yet implemented this). But our experience in COMMUNAL has convinced us that the overall grammar is significantly simpler - and so more elegant and easier to ‘service’ - if it operates on the principle that IT ONLY ENTERS THE NETWORK ONCE IN GENERATING EACH UNIT. Other things being equal, therefore - and we think they are - it is DESIRABLE to generate the exponent of the **cc head** on one traversal of the sub-network for the CULTURAL CLASSIFICATION of ‘things’, and the exponent of the **typic head** on a SUBSEQUENT traversal of the network. So, while it is not impossible to generate them both as elements of the same unit, the adoption of Option D would make the grammar considerably more complex - and so add to its fragility.

At this point in the investigation, then, the decision between Options C and D was still fairly evenly balanced, with the weight of the evidence now somewhat in favour of Option C - but with the fascination of the new concept that underlies the new syntax still exerting a significant pull.

5.10.4 Evaluating Options C and D: (iii) the fresh evidence and the decision

It was at this point - with the corpus studies completed but with no convincing result - that I encountered, entirely by accident, an example of a typic nominal group with a new structure. It provided such persuasive evidence of the need for Option C that the various tentative arguments for Option D were immediately swept away.³³

The relevant example is the underlined portion of (4):

(4) The first [i.e. *this type of callous killing*] could mean ‘this particular sub-type of the type of killing I am calling callous’.³⁴

Let us call this a ‘type of type’ structure. As soon as I noticed (4), I did a search of the google corpus for similar examples - and I quickly found a small but significant body of them. This confirmed that, while the ‘type of type’ structure is infrequent, it is fully legitimate - and so one that it is essential for the grammar to provide for. Examples from google that are like (4) include (with the two typic heads underlined):

(5) each type of several types of health facilities

³³Ironically, this vital example arrived as part of a comment in an email from Chris Butler, in his role as one of the editors of this volume - and its unintended effect was to delay somewhat the completion of this paper.

³⁴This example occurred in the context of a discussion of Example (10) in Section 5.7.1.

- (6) (a class can be) a type of many classes (of entities)
 (7) (it was) another type of the same species (of xxxx)

It is satisfying to be able to say that the system network in Figure 3 can be used to generate such structures without any alteration. This is because it already allows re-entry to the overall network to generate the embedded nominal group that fills the **td**, as in Option C in Figure 9. It is because of this that the embedded nominal group can itself be given a typic determiner which is in turn filled by a nominal group. Structurally, this type of embedding is similar to that noted in Section 4.4. In contrast, Option D is simply incapable of handling 'type of type' structures, because it doesn't have the embedded nominal group that is needed to model the recursion.³⁵

The analysis of each of (4) to (7) is broadly similar, i.e. as in Figure 10. Interestingly, both (5) and (6) happen to exemplify the need to locate the typic determiner before the quantifying determiner - a point to which we shall return in the next section.

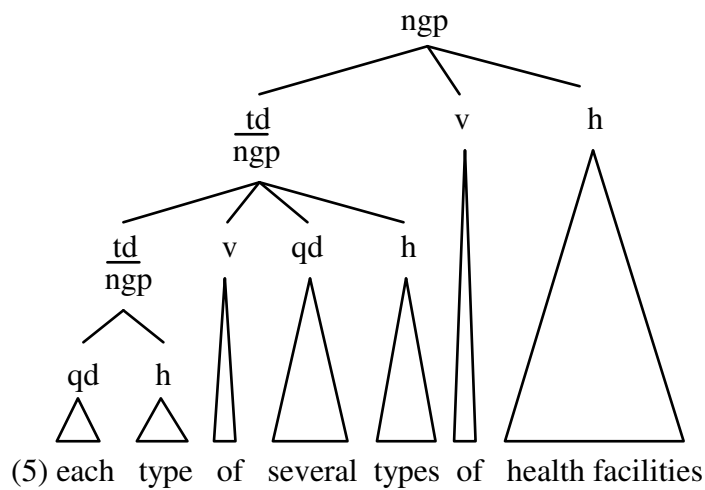


Figure 10: Embedding in the typic determiner

This section, then, has identified the need to use Option C rather than Option D to represent this aspect of the nominal group. We have seen in the discussions of the findings of the corpus studies that the problems for the realization rules that this decision brings are relatively small. We can therefore now abandon Option D with some confidence that the grammar that uses Option C has been - or can be - extended to handle such patterns.

5.11 The place in the nominal group of the typic determiner

What should be the position of the typic determiner in the sequence of the determiners? In my earlier descriptions of the English nominal group I had assumed that it occurred immediately before the deictic determiner - probably influenced by the proximity of the typic head to the cc head in practically all examples, as noted in Section 5.3. But the evidence from my corpus studies suggests that it should be located much earlier, at least before the quantifying determiner.

We have also noted that, like the **representational** and **partitive determiners**, it is

³⁵You might think that it could add typic heads recursively to the same nominal group, but this would run up against one of the founding principles of the grammar - namely, that each element in a unit serves a different function.

always and only filled by a nominal group. Such determiners are typically - but not exclusively - the first determiner in the overall nominal group in which they occur.

Here we treat the typical determiner as occurring at the start of the nominal group, so as immediately preceding the representational determiner.³⁶ This means that any determiner that may APPEAR to occur before it should in fact be treated as embedded in it, i.e. as a determiner in the nominal group whose head is *type*, etc.

5.12 The realization rules that generate these structures

In the last two sections we have established the principles that have led to the decision to use Option C, and here I shall state the basic realization rules that generate such structures from the system network in Figure 8. The reason for doing this is to enable you to test the general claims made in the last few sections about the elegance with which this grammar operates.

There are two sets of rules: (i) those that apply when building the **matrix nominal group**, and (ii) those that apply on re-entry (as specified in Rule 62.2) to generate the **embedded nominal group** (which fills the **td** of the matrix nominal group). I shall first give the rules themselves, then ‘translate’ them into English, with brief comments. Here is the first set:

```
62.2 : overt_type :
      td @ 5,
      for td prefer [thing, .... abstract_thing, .... thing_as_type],
      if singular_overt_type then for td prefer [singular_cc],
      if plural_overt_type then for td prefer [plural_cc],
      for td re_enter_at_entity,
      tv @ 6, tv < "of".37
```

```
80 : plural_cc or plural_covert_type :
      if not ['irregular noun senses'] then h <+ "+s".38
```

```
80.01 : plural_overt_type :
      if count_cc
      then (if not ['irregular noun senses'] then h <+ "+s").
```

Here is the second set of rules - i.e. those that state the realizations in the head of the nominal group that is generated to fill the **td** in the matrix nominal group. Note that Rule 80 applies again - this time to generate the ‘plural’ form of *type* (or a near-synonym) that expounds the head of the embedded nominal group.

```
78.001 : informal_type_c:           h < "sort".
78.002 : general_type_c :           h < "kind".
78.003 : general_technical_type_c : h < "type".
76.012 : unmarked_variety_c or biol_variety_c : h < "variety".
```

³⁶Regrettably, there is no space to discuss the reasons here.

³⁷In the generative version of the grammar, the selector - which is represented in the text-descriptive version simply as ‘v’ - is distinguished from any other selectors that may be generated for the same nominal group by labelling it as **tv**, i.e. as the ‘typic selector’.

³⁸The plural forms of ‘irregular nouns’ such as *man*, *woman* and *child* are generated by a different rule, which consults a table of the relevant forms. This also provides for the irregular plural of *species*, when it occurs in the embedded nominal group that fills the **td**.

76.004 : class_c or bioI_class_c :	h < "class".
76.017 :type_of_typically_tame_animal_c :	h < "breed".
76.025 : species_c :	h < "species".

....etc, for over a score other near-synonyms of *type*.

Rule 80 (as above).

These few rules generate the eight basic variants provided for by the system network in Figure 8 (four for each of 'count_cc' and 'mass_cc'), as well as a sample of the variations on *type*.

Each **realization rule** is composed of one or more **statements**. Their purpose is to specify that, under the conditions specified in the realization statement (which may be simply the choice of the feature to which the rule is attached) one or more **operations** get executed. They are therefore 'statements' about what is to be done, so they are, in effect, commands.

Here is the 'translation' of Rule 62.2, which contains seven 'statements':

1 If the feature 'overt_type' is selected, the element **typic determiner (td)** is to be located at **Place 5** in the structure of the nominal group currently being generated. (The role of the numbered Places is to get the elements in the right order, and when this is done the unused ones are stripped away.)

The grammar then instructs itself to do three things when it re-enters the system network - which it must do, to choose the 'selection expression'

2 The grammar instructs itself to select the bundle of features - [thing] and seven other features, then [type_of_thing] - that will lead to the generation of a head such as *type* or *species*. This **preference re-setting rule** applies unconditionally.

3 The grammar then instructs itself that, if [singular_overt_type] is chosen on the current traversal, the feature [singular_cc] must be chosen on re-entry.

4 Similarly, if [plural_overt_type] is chosen, [plural_cc] must be selected.

Thus each of the second and third instructions has a condition.

5 The grammar instructs itself to re-enter the overall system network at its initial entry condition. i.e. at the feature [entity].

Without this, (2) to (4) would not happen.

6 Finally, the grammar locates the **typic selector (tv)** - i.e. the selector that gets introduced when there is a typic determiner - at Place 6.... and

7 it specifies that it is to be expounded by the item *of*.³⁹

Rule 80 is much simpler. It states:

'If either of the features [plural_c] or [plural_overt_type] is selected - and so long as the 'noun sense' that is being selected on the same traversal of the network forms its

³⁹You may be wondering why the grammar needs to specify the 'number' of the embedded 'typic' nominal group when generating the matrix nominal group, since the choice between [singular_cc] and [plural_cc] needs to be made for the lower unit in any case. 'Isn't it rather clumsy,' you might ask, 'to first make the choice for the matrix nominal group and then to transfer it, via the two 'preference' statements, to the embedded nominal group?' The answer is that the choice must in fact be made for the matrix nominal group in its own right, in order to ensure that the grammar generates the plural form of the head of the matrix nominal group when the 'number' of the 'types' is 'plural'. Thus the present grammar generates *two types of ants* rather than *two types of ant* - as is required by the statistically dominant dialect of English, according to the google and COBUILD corpora. (The grammar of my dialect is even simpler, however: the choice is only made for the embedded nominal group, because the head of the matrix nominal group always uses the base form of the noun, as in *two types of ant*.)

‘plural’ variant in a **regular** manner - the **head** (which will already have been expounded by an item such as *ant* or *oil*) is to be further expounded by the addition of the suffix ‘s’.

For the many noun senses whose plural form is **irregular**, as in as *men*, *women* and *children*, the plural form is generated by a special rule containing similar conditions, which we omit here.⁴⁰

This one rule (and its ‘irregular’ equivalent) elegantly provide the ‘plural’ forms of the head in three different situations. The rule realizes: (i) the simple ‘plural’ form in the matrix nominal group (e.g. *two ants*); (ii) the ‘plural’ form in the matrix nominal group (when the meaning of ‘typicity’ is ‘covert’ and ‘plural’, e.g. *two oils* and, in this sense, *two ants*); and (iii) - since the rule applies to the embedded nominal group as well as the matrix nominal group - the ‘plural’ forms for *type* and its near-synonyms in the embedded nominal group (e.g. *two types of ant(s)* - even including the ‘irregular’ plural form of *species*, as in *two new species of ant(s)*).

Rule 80.01 is equally elegant and economical (and also has an ‘irregular plural form’ equivalent). It states:

‘If, when the feature [plural_overt_type] is chosen, the feature [count_cc] has also been chosen - i.e. if the referent is ‘plural’- then (as in Rule 80), if the ‘noun sense’ is not realized by an irregular plural, the head is given the suffix ‘s’.’

Notice that the rule neatly avoids generating *two types of oils*, because it only applies to [plural_cc] things, and not to [mass_cc] things.

Interestingly, this rule is only needed in those dialects of English that use the form *two new types of ants*, in contrast with *two new types of ant* (the majority, according to both google and COBUILD corpora). No such ‘agreement’ rule is needed in dialects such as my own that use the second form, so that their grammar is simpler than the ‘agreement’ version by one rule.

Any other elements and items that may be required for the matrix nominal group are generated by the normal rules for generating nominal groups. These include the head of the matrix nominal group (e.g. *ant* or *oil*), the full range of types of determiner and ‘selection’ described in Section 4, and modifiers and qualifiers.

Now we turn to the rules that apply to the nominal group embedded at **td**. We shall take Rule 62.31 as an example of the set of forty or so rules that provide the ‘base’ of the item that expounds the **head**. It simply states:

‘If the feature ‘informal_type’ is chosen, the head will be expounded by *sort*.’

As with the matrix nominal group, the other elements and items that may be required for the embedded nominal group that fills the **td** are generated by the normal rules for generating nominal groups. Again, these include the full range of types of determiner and ‘selection’ described in Section 4 (though with their probabilities greatly reduced) - while the exponent of the head is, of course, preselected (by Rule 62.2). Other nominal group elements such as modifiers and qualifiers may be generated - but only within the restrictions identified in the relevant statements in Section 5.8. Thus any qualifiers in a nominal group that fills a **td** will be discontinuous, occurring after the head of the matrix nominal group.

In summary, we can say that the attractive simplicity of the structures described in

⁴⁰The plural of *variety* - i.e. *varieties* - is generated by a low-level ‘spelling adjustment’ rule that converts all cases of plural nouns ending in *ys* to *ies*.

Sections 5.9 and 5.10 is matched by the simplicity of the realization rules that generate them.

6 Conclusions

6.1 Overview of the grammar of ‘typicity’

The first aim of this paper was to develop a more adequate model than those currently available for the syntax and semantics of ‘typicity’. I shall now offer a very brief summary of my proposals for this, and then I shall turn to the second aim: to make a critical evaluation of the methods used in investigations of language in general, and in this paper in particular.

The proposals set out here combine with the existing generative grammar of the English nominal group, as described in the Cardiff Grammar, to form an even more comprehensive generative grammar of the nominal group. And from this, of course, we can derive a text-descriptive version of the grammar, i.e. a framework for describing nominal groups at the levels of both form and meaning.⁴¹

The most difficult decisions were made in Sections 4.2 and 5.10, where we considered the arguments for and against different structures for modelling (i) ‘selection’ in general (as in *five books* and *a large number of those books*) and (ii) ‘typic’ nominal groups such as *two new types of ant(s)*, etc. In the first case we rejected the traditional ‘prepositional group as qualifier’ structure (Figure 4) in favour of the ‘selection’ model (Figure 6), and for ‘typicity’ we rejected both the ‘prepositional group as qualifier’ analysis and the possible ‘two head’ structure - again in favour of the ‘selection’ model (Option C, as illustrated in (1b) in Figure 9 and in Figure 10).

However, there are no ‘purely formal’ criteria for choosing between alternative structures, and our criteria have in practice depended on two characteristics that are desirable in any functional grammar: a **syntax** that reflects the semantic relationships involved as transparently as possible, and a set of **realization rules** for ‘translating’ meanings into forms that are as economical as possible. And in SFL these two work in concert with an equally carefully structured **system network**.

How can we summarize the **theoretical-generative** version of the grammar proposed here? The best way is simply to refer you to the relevant sections and diagrams. The overall model is summarized in Figure 1. Figure 8 in Section 5.5 gives the system network for TYPICITY, Section 5.12 states the associated realization rule, and Option C in Figure 9 and Figure 19 show typical resulting structures. Thus, when evaluating a generative SFG, it is necessary to take account of all three of these: system network, realization rules and the structural outputs at the level of form.

And how can we summarize the **text-descriptive** version of grammar - i.e. the version that is adapted for use in describing texts? We can summarize the proposals for modelling the structures of the **overt** forms of the determiners as follows (the elements being listed in the sequence in which they occur in the matrix nominal group):

⁴¹Sometimes, in the development of a new area of grammar, the text-descriptive grammar is produced first. But when this happens it should be regarded as the first stage in the development of a full, generative grammar. Any such initial text-descriptive grammar is a temporary model, and it is only when a text-descriptive grammar is derived from a full theoretical-generative grammar that it should be regarded as stable.

element	item	unit (s)
typic determiner		ngp (h < <i>type</i> , etc)
representational determiner		ngp (h < <i>photo</i> . etc)
partitive determiner		ngp (h < <i>back</i> , etc)
fractionative determiner	<i>half</i> or	ngp (h < <i>fifth</i> etc)
quantifying determiner	item or	ngp or quantity group
ordinative determiner		quality group (apex < <i>fifth</i> etc)
superlative determiner		quality group (apex < <i>finest</i> etc)
totalizing determiner	<i>all</i>	
quality-introducing determiner	<i>those</i>	
deictic determiner	item or	genitive cluster
head (aka ‘cc head’ in Section 5)	noun	ngp (occasionally)

As a final example, consider *one of the first of the new varieties of GM wheat* (i.e. (3a) from Section 1.1). Using linear representations of its first and second layers of structure, its analysis is:

(3ai) ngp: one of the first of the new varieties [td] of [v] GM wheat [h]

(3aii) ngp at **td**: one [qd] of [v] the first [od] of [v] the dd] new [m]varieties [h]

In a tree diagram format, then, the overall structure resembles that in Figure 10 - but the complex structure within the **td** is more like that in Figure 7.

A grammar that provides only structures (as in the first two editions of *IFG*) can be used for text description, so long as its criteria for identifying categories are sufficiently clear (i.e. it is a **text-descriptive grammar**). But it should be regarded as a ‘place-holder’ description until its proposers develop both its system networks and its realization rules - i.e. until it is presented as a **theoretical-generative grammar** - so that we can see that it actually ‘works’.⁴²

To summarize: the structure used here, as in Option C in Figure 9, is significantly simpler than that suggested by other scholars - who, when they express a view at all, almost always press into service the standard ‘prepositional group as qualifier’ analysis (using an analysis similar to that in Figure 4). The attractive simplicity of the structure proposed here is the result of the pleasing economy of the realization rules - and so in turn with the carefully crafted system network. The three, as I have emphasized throughout, are interdependent - so that, while it is tempting to characterize a grammar by the elegance of one or other component, the three in fact need to be evaluated as one composite package.

We have seen that the realization rules for the ‘prepositional group as qualifier’ structure would be far more complex than those given here, largely because many of the choices in other elements of the matrix nominal group depend on the choices in ‘cultural classification’ and ‘number’ that are expressed in the head - and the head would in that approach be buried two layers down, in an embedded nominal group.⁴³

The structure used here, then, has these twin advantages: (i) it is both simple and as transparent a reflection of the semantic relationships involved as one could hope to find - which makes it ideal for use in text analysis - and (ii) it is generated by simple realization

⁴²Halliday & Matthiessen 2004 adds some networks, but there is no attempt to provide systematic coverage of ALL the networks, nor to provide realization rules. This is understandable, since that work’s goal is to be a text-descriptive grammar, not a theoretical-generative one.

⁴³It would not be impossible to do this, and there are occasional places in the full grammar where it is necessary. But it would involve introducing a degree of additional complexity that any linguist would prefer to avoid, unless it is absolutely necessary. And, as the explicit grammar set out here shows, it isn’t.

rules - so that it contributes importantly to the elegance of the generative version of the grammar.

6.2 Methodology in linguistics in general and SFL in particular

Finally, we shall consider briefly the methods used in this investigation - and their lessons for future work. Let us first establish the overall framework in which, I suggest, the science of linguistics should be undertaken. This will be a more formal and more general description than the informal characterization of how we should make progress in linguistics given in Section 1.3. Essentially, the methods are those likely to be used by a scientist working in any discipline. They are (expressed here in a way that attempts to transcend different viewpoints within the philosophy of science):

- 1 the initial observation of patterns of behaviour in data (which presupposes, minimally, a set of tentative 'pre-theoretical' categories);
- 2 adapting or building a theory-based model (or part of one) that could account for those patterns (so, in linguistics, this must include a 'grammar' of meaning and form that is capable of being used to generate and analyze text-sentences, and ultimately to be a component in an overall model that explains the generation of whole texts, including interactive dialogue);
- 3 testing the model against the evidence of a much larger quantity of data (obtained from either systematic observation or experimentation);
- 4 changing, refining, and/or occasionally rejecting and replacing part (or even all) of the model
- 5 ... and so on, repeating 3 and 4 in a potentially unending attempt to develop a fully comprehensive and maximally insightful model.

In principle, a linguist - or any other type of scientist - can undertake any such programme of research on their own. But they are likely to make better progress when they operate with a small team of researchers, and when they test their findings through comparisons with others working in the same field through publication.

I shall now say a little about the various types of evidence that I have drawn on in this study, in relation to the phases of research outlined above. Some of the conclusions that I draw will, I should warn you, go against current fashion.

For me, Phase 1 occurred in the 1970s. As with most linguists who wished to use real data at the time, the 'patterns of behaviour in data' that I observed were noted down from (i) the everyday language in use around me, and (ii) the real life texts that my students and I were trying to analyze. But there was also (iii) the evidence of patterns noted by other linguists and found in their books and papers - most notably in (a) Quirk et al 1972 (the first of the big modern grammars) and (b) the various Scale and Category grammars being written in the 1960s and early 1970s.⁴⁴ And I must also have used a fourth type of evidence - invented examples - about which I shall have more to say in a moment.

Phases 3 and 4 in the development of a model recur cyclically, and the most recent recurrence in relation to 'typicity' in English occurred for me in the years 2004-5. By then the types of evidence available to the scientist of language were very different. The value of corpus linguistics was very well established, and Sections 5.6 to 5.8 describe how I was able to draw quickly and easily on the data available in large corpora. This, then, was my major

⁴⁴ All were, of course, derived from Halliday 1961 (see Fawcett 2000:26f for this early stage of the history of SFL). Thus my 'tentative categories' in Phase 1 came from these works. They were therefore only 'pre-theoretical' in the sense that they were based on an earlier theory than the one from which my present categories are derived (for which see Fawcett 2000).

source of the ‘much larger quantity of data’ required for Phase 3 of the research, as I refined the model. And this was the ‘systematic observation’ of the data referred to in the summary above of what happens in Phase 3 of the scientific approach.

However, we should note that these days the use of corpus data doesn’t necessarily depend on spending long hours on primary research. Evidence from corpus linguistics is increasingly available through what Neale (2002:187) has called ‘the “second level” use of corpora’ - e.g. as found in the *COBUILD Dictionary* (Sinclair 1987/95), the detailed descriptions in Francis, Hunston & Manning 1996 & 1998, and Biber et al 1999. Neale 2006 describes how we discovered the existence of a whole new class of Process types in English through the ‘second level’ use of corpora. So the boundary between what I have earlier called the ‘big grammars’ and the evidence from large corpora is disappearing. However, we have seen in Section 2 that the large grammars have been very little help in the present study, so that I found it necessary to consult the primary data of COBUILD and google - and this situation seems certain to continue for detailed studies for some years to come. Nonetheless, we must count ‘consulting the literature’ as a second potential source of insights in scientific inquiry.

Next I want to point out the central role played by a third type of evidence - probably in all subject areas. It is one that has a long history in linguistics but that is currently out of fashion.⁴⁵ It is what I have referred to above as ‘invented examples’.

The formal linguists of the 1960s and onwards dignified invented examples by describing them as the results of ‘native speaker intuition’. But the most accurate and most useful name for this source of data is, I suggest, **thought experiments**. The value in using this label is to remind us to check that, when we use this source of data, we exercise proper scientific care.

What, precisely, is involved in a ‘thought experiment’? Let us assume that you are considering an example that could be analyzed in more than one way - as in the case of the ‘typic nominal groups’ considered here. The first step is to try to create in your mind a number of possible alternative texts - texts that might have occurred at that point but didn’t. You will probably do this first at the level of **form** - but, as a systemicist, you should also be thinking about the contrasts in **meaning** that are expressed in the forms.⁴⁶ The second step is to evaluate the status of the possible alternatives. Do native users of English say or write such things? If so, with what dialect and register restrictions? And with what probabilities? and so on. If you have any doubts, test your results against the evidence of corpora and/or informant testing. The basic principle, then, is to distrust your natural assumption that you ‘know’ what people (including yourself) actually say (as I was reminded when I investigated Hypothesis 4 in Section 5.4). I suspect that all linguists in fact carry out thought experiments - and if we do it is best to acknowledge it, and to check regularly that we are doing it in a scientific manner.

As you may have noticed, thought experiments have been an essential tool in the present study. Without them I could not, in Section 5.6, have derived the crucial examples of **covert typicity** in (4b) to (7b) from (4a) to (7a). As a second example of the value of using thought experiments, I would point out that I could have waited for ever to establish the canonical sequence of the determiners, if I had restricted myself solely to the evidence of recorded

⁴⁵This is probably the result of (i) an understandable reaction to the over-dependence on invented examples by formal linguists from the 1960s onwards, and (ii) the current fashionable emphasis - which is fully justified - on the value of corpus-based research.

⁴⁶However, one of the lacks in SFL publications is that so far none provide a full set of clear, user-friendly system networks, with guidance in how to use them. The Cardiff Grammar has comprehensive system networks designed for the computer grammar, and I intend to make these available in a user-friendly form in Fawcett in preparation. It would be good to have equivalent documentation from the Sydney Grammar in due course, to facilitate comparison. (Sadly, the networks in Matthiessen 1995 and Halliday & Matthiessen 2004 do not meet these criteria.)

texts. To resolve this question I needed BOTH to look at a great many texts AND to conduct a lot of thought experiments. It was only by combining these two methods that I was able to discover the patterns that lie hidden behind the first impression that is conveyed by the high frequency items.⁴⁷ And I have used thought experiments at various other points in the study, e.g. in creating *what type of the virus* from *what type of a virus* in Section 5.7.2.

However, in the current widespread appreciation of the value of corpora, it is easy to dismiss thought experiments as an outmoded tool. In fact, I suggest, they are still one of the best and most natural tools available to the linguist. This is because a scientist building a model of any phenomenon should always be on the lookout for counter-examples, and thought experiments are the quickest way to find them.

Thought experiments have a particularly natural place in SFL. A SYSTEMICIST SHOULD ALWAYS BE THINKING SYSTEMICALLY - i.e. you should consider not only the data before you but also the bits of text that might have been used at that point but were not - AND IT IS BY THOUGHT EXPERIMENTS THAT WE IDENTIFY THESE.

So far we have noted the following three sources of evidence for deciding between alternative structures: (i) large corpora, (ii) the literature on English grammar (which increasingly includes 'second level' corpus evidence) and (iii) thought experiments. But there are two further sources of evidence that I have used to help me to decide between the alternative structures for representing 'typicity'. And one of these, we have found in this paper, inevitably invades discussions of the weight to be attached to the findings of the first three.

This fourth type of evidence is one that has been important in the present study, i.e. the degree of complexity that the structure introduces to the model when it operates as a **generative grammar** - and specifically the complexity of the **realization rules** through which **meanings** come to be expressed as **forms** (as described earlier in this section). As in all scientific modelling, we look for economy and clarity in the way the grammar operates, but in linguistics we also look for a representation in syntax that reflects, as transparently as possible, the semantic relationship of 'selection' between the referents, and so of elements of the nominal group, as described in this paper.⁴⁸

However, this paper is, as I said at the start, the story of the use FIVE types of evidence. In particular, it is the story of how the fifth type of evidence made the crucial difference when deciding between the structures of Options C and D. This fifth type is in fact the type with which my study of 'typicity' started, back in the 1970s. Moreover, it is another type which, together with the use of thought experiments (aka 'invented examples' and 'native speaker intuition'), is often stigmatized by linguists as 'unscientific' in today's corpus-oriented culture. It is the linguistic data of our **direct experience** - the spoken and written acts of everyday communication and thought that surround us all the time. Indeed, if in mid-2005 I hadn't noticed in an email from a friend the counter-example to the model I was then considering, I would probably not have become convinced - as I now am - that the structure presented here is the best that can be proposed in the current state of our knowledge.

What are the lessons of this study for the methodology of linguistics? There are at least the following six.

First, the exploratory use of large corpora provides many examples that we should not expect to be able to think up without the help of that technology. Indeed, some of these

⁴⁷As another example, I might point out that it was because the 18th century grammarians didn't look beyond the high frequency of the items *the* and *a(n)* that they developed the model of a system for the 'article' with its two features 'definite' and 'indefinite' - and this, as we saw in Section 4, is misleading.

⁴⁸It is these factors, in my view, that are reflected indirectly in tests that are supposedly 'purely formal'. If form and meaning are mutually defining, as Saussure's concept of the 'sign' suggests, then there is no point in trying to set up tests that are 'purely formal' or 'purely semantic'. As I have remarked earlier, citing Meillet (1937), 'tout se tient'.

examples were counter-examples to my initial hunches, and some suggested new patterns of 'typicity' that we need to provide for.

Second, we should not assume that even the most comprehensive of modern grammars (such as Huddleston & Pullum 2002) grammars will have adequate statements at the levels of meaning and form for every topic.

The third lesson, which is perhaps less obvious, is that it is inadequate merely to evaluate the merits of alternative structures as insightful or not on the basis of how we feel about them. We must also pay attention to how they would operate in a generative version of the grammar that turns meanings into forms.

The fourth lesson is that we should not be afraid to create new examples - i.e. 'invented examples' - by thinking systemically, and then testing them through either (i) thought experiments and/or (ii) by checking in large corpora - if only via google - such as to see if they actually occur.

And the fifth lesson is, of course, that the grammarian should continue to cultivate an ear and an eye for the real-life example that will expand or undermine one's current model of a language. The fact is that in the present case my corpora - large though they were - happened not to include the vital counter-example that would have made me rethink my position. So I am confirmed by the experience of this investigation of 'typicity' in an old habit, i.e. that I need to maintain an expectant alertness to the potential evidence of the data of my everyday encounters, whether spoken or written. But since this is all part of the pleasure of linguistics, adding this fifth source of evidence to the other four is no great hardship.

And what is the overall moral of this tale - and so the sixth and final lesson? It is that one should not assume that every single new idea that one has, however exciting it may at first appear, will turn out to be valid. Every new idea needs to be tested, using all the means at our disposal - as we have here tested the (partly) new idea of the 'two-headed nominal group'. And only if the new idea passes these tests should it be accepted. into the description - and, if it also has implications for the theory, into the theory. This is a principle that must be applied to every new idea, no matter how eminent its progenitor. It is as important to be able to reject new ideas - but only on sound grounds - as it is to be able to accept new ideas.

References

- Biber, D., Johansson, S., Leech, G., Conrad, S., and Finegan, E., 1999. *Longman Grammar of Spoken and Written English*. Harlow: Pearson.
- Butler, C. S., 2003a. *Structure and Function: An introduction to three major structural-functional theories. Part 1: Approaches to the simplex clause*. Amsterdam: John Benjamins.
- Butler, C. S., 2003b. *Structure and Function: An introduction to three major structural-functional theories. Part 2: From clause to discourse and beyond*. Amsterdam: John Benjamins.
- Downing, A. 1991. 'An alternative approach to theme: a systemic-functional perspective'. In *WORD* 42.2, 119-143.
- Downing, A. 1996. 'The semantics of *get*-passives'. In Hasan, R., Cloran, C., and Butt, D. (eds.) *Functional descriptions: Theory in Practice*. Amsterdam: John Benjamins, 179-206.
- Downing, A., and Locke, P., 1992. *A University Course in English Grammar*. New York: Prentice Hall.
- Downing, A., and Locke, P., 2006. *English Grammar: A University Course*. 2nd. edn. Abingdon and New York: Routledge.
- Fawcett, R.P., 1974-6/81. *Some Proposals for Systemic Syntax* (revised version of 1974, 1975b and 1976a). Cardiff: Polytechnic of Wales (now University of Glamorgan).
- Fawcett, R.P., 1980. *Cognitive Linguistics and Social Interaction: Towards an Integrated Model of a Systemic Functional Grammar and the Other Components of an Interacting Mind*. Heidelberg: Julius Groos and Exeter University.
- Fawcett, R.P., 2000. *A Theory of Syntax for Systemic Functional Linguistics*. Current Issues in Linguistic Theory 206. Amsterdam: John Benjamins.
- Fawcett, R.P., forthcoming a. *Functional Syntax Handbook: Analyzing English at the Level of Form*. London: Equinox.
- Fawcett, R., forthcoming b. 'Modelling 'selection' between referents in the English nominal group: an essay in scientific inquiry in linguistics'. In Butler, C.S., Hidalgo Downing, R., and Lavid, J., *Functional Perspectives on Grammar and Discourse: Papers in Honour of Prof. Angela Downing*. Amsterdam: John Benjamins.
- Fawcett, R.P., forthcoming c. *Functional Semantics Handbook: Analyzing English at the Level of Meaning*. London: Equinox.
- Fawcett, R.P., and Perkins, M.R., 1980a-d. *Child Language Transcripts 6-12, Volume 1: Six Year Olds, Volume 2: Eight Year Olds, Volume 3: Ten Year Olds, Volume 4: Twelve Year Olds*. Pontypridd, Wales: Polytechnic of Wales (now University of Glamorgan).
- Fawcett, R.P., and Perkins, M.R., 1981. 'Project report: Language development in 6- to 12-year-old children. In *First Language* 2, 75-9.
- Francis, G., Hunston, S., & Manning, E., 1996. *Collins COBUILD Grammar Patterns 1: Verbs*. London: HarperCollins.
- Francis, G., Hunston, S., and Manning, E., 1998. *Collins COBUILD Grammar Patterns 2: Nouns and Adjectives*. London: HarperCollins.
- Fawcett, R.P., Tucker, G.H., and Lin, Y.Q., 1993. 'How a systemic functional grammar works: the role of realization in realization'. In Horacek, H., and Zock, M., (eds.), 1993, *New Concepts in Natural Language Generation*, London: Pinter, 114-86.
- Halliday, M.A.K., 1961/76. 'Categories of the theory of grammar'. *Word* 17, 241-92. Reprinted as Bobbs-Merrill Reprint Series No. Language-36, and in part in Halliday, M.A.K., 1976, *System and Function in Language: Selected Papers by M.A.K. Halliday* (ed. G.R. Kress), London: Oxford University Press, 84-7.

- Halliday, M.A.K., 1985. *An Introduction to Functional Grammar*. London: Arnold.
- Halliday, M.A.K., 1994. *An Introduction to Functional Grammar* (Second Edition). London: Arnold.
- Halliday, M.A.K., 1996. 'On grammar and grammatics'. In Hasan, R., Cloran, C., and Butt, D. (eds.) *Functional descriptions: Theory in Practice*. Amsterdam: John Benjamins, 1-38.
- Halliday, M.A.K., and Matthiessen, C.M.I.M., 1999 *Construing Experience through Meaning: a Language-Based Approach to Cognition*. London: Cassell Academic.
- Halliday, M.A.K., and Matthiessen, C.M.I.M., 2004, *An Introduction to Functional Grammar* (Third Edition). London: Arnold.
- Huddleston, Rodney D, and Pullum, Geoffrey K. (principle authors) 2002. *The Cambridge Grammar of English*. Cambridge: Cambridge University Press.
- Matthiessen, C.M.I.M., 1995. *Lexicogrammatical Cartography: English Systems*. Tokyo: International Language Sciences Publishers.
- Meillet, A., 1937. *Introduction a l'Etude Comparative des Langues Indo-Européennes*. 8me ed. Paris: Libriere Hachette.
- Neale, A., 2002. *More Delicate TRANSITIVITY: Extending the PROCESS TYPE system networks for English to include full semantic classifications*. PhD Thesis. Cardiff: School of English, Communication and Philosophy, Cardiff University. Available on request via neale@qmul.ac.uk
- Neale, A., 2006. 'Matching corpus data and system networks'. In Thompson, G., & Hunston, S., (eds.), *System and Corpus: Exploring Connections*. London: Equinox.
- Prakasam, V., 1996. "'NGp of NGp" constructions: a functional-structural study. In Berry, M., Butler, C.S., Fawcett, R.P., and Huang, G.W. (eds.) *Meaning and Form: Systemic Functional Interpretations*. Meaning and Choice in Language: Studies for Michael Halliday. Norwood, NJ: Ablex, 567-83.
- Quirk, R., Greenbaum, S., Leech, G., and Svartvik, J., 1972. *A Grammar of Contemporary English*. London: Longman.
- Quirk, R., Greenbaum, S., Leech, G., and Svartvik, J., 1985. *A Comprehensive Grammar of the English Language*. London: Longman.
- Sinclair, John, (editor-in-chief), 1987/95. *Collins COBUILD English Language Dictionary*. London: Collins.
- Sinclair, John, (editor-in-chief), 1990. *Collins COBUILD English Grammar*. London: HarperCollins.
- Sinclair, J., 1991. *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- Tucker, G.H., 1996. 'Cultural classification and system networks: a systemic functional approach to lexis'. In Berry, M., Butler, C.S., Fawcett, R.P., and Huang, G.W. (eds.) *Meaning and Form: Systemic Functional Interpretations*. Meaning and Choice in Language: Studies for Michael Halliday. Norwood, NJ: Ablex, 533-536.
- Tucker, G.H., 1998. *The Lexicogrammar of Adjectives: a Systemic Functional Approach to Lexis*. London: Cassell Academic.