

A corpus-based stylistic study of newspaper English

by

Lesley Evelyn Jeffries

Submitted in accordance with the requirements for the  
degree of PhD.

The University of Leeds

Department of Linguistics and Phonetics

September 1989

## Abstract

This study is based on a corpus of 2400 clauses taken from British national newspapers in 1986 and stored in a computer database with each clause coded for a number of grammatical (and some semantic) features. These features relate to the verb phrase (e.g. finiteness), the clause (e.g. subordination) and the subject (e.g. form).

In the first stage of the investigation the database is described in terms of the features coded therein. The scope of the description is on three levels. First, the data are described in total and are considered to constitute a representative sample of newspaper English. Secondly, the database is split into three pre-determined sub-databases according to their text-type. These are: news articles, editorials and readers' letters. A pattern is discovered of 'letters-as-norm' with the other text-types on different sides of the average. Thirdly, the database is split on a different dimension according to the eight different newspapers included in the sampling. A pattern of three groups of newspapers; 'quality', 'central' and 'popular', is found for some features.

The second section exploits the database primarily as an example of written English, rather than emphasising its newspaper origins. Here some problems of description, which have implications for the debate about the division between syntax and semantics, are explored.

The first such 'problem' arises out of a study of the environment of copula 'BE' and concerns the borderline between the grammatical functions of subject and subject complement. Some well-known differences are confirmed and some new ones discovered. A small area of overlap, however, remains.

The second problem is the familiar difficulty of deciding when an -en form is an adjective and when it remains a participle. It is argued that the contexts of -en forms are often influential in their interpretation as adjectival or verbal forms.

The third problem concerns the sequential verbs (sometimes called 'catenative' verbs) which govern a following non-finite verb phrase. These verbs, which defy attempts to classify them syntactically, are shown to be amenable to semantic classification. The question of restrictions on sequences of more than two verb phrases (i.e. two sequential verbs + one 'normal' verb) is explored and some tentative conclusions are reached.

### Acknowledgements

There are many people who have given me moral support throughout the research for, and writing of this thesis. They are to be found among my family and friends and among the staff of the University of Leeds; in the Department of Linguistics and Phonetics, the School of English and the Computing Service. I am grateful for this support.

There are some who have played a particularly important part and deserve my special thanks. They include David Barber, who supervised my work from the beginning but unfortunately left the University just 3 months before its completion. My thanks for a shorter, but more intensive period of supervision go to Prof. G. Sampson who stepped into the gap left by David and had some extremely useful contributions to make on both academic and practical matters.

I owe a great debt also to Loreto Todd who was under no official obligation to help me, but was always ready to give me her time and help. Finally, I would like to thank Dave Webb for his help on statistical matters and in the word-processing and printing, particularly in the latter stages when the tasks were tedious and my stamina was failing.

# A corpus-based stylistic study of Newspaper English

## Contents

<b>Chapter 1</b>	<b><u>Introduction</u></b>	<b>1</b>
<b>1.1</b>	<b>The Thesis</b>	<b>1</b>
1.1.1	'Newspaper English' and 'style'.....	1
1.1.2	A corpus-based study.....	7
1.1.3	A descriptive study.....	11
1.1.4	A selective study.....	13
1.1.5	Structure of the thesis.....	14
<b>1.2</b>	<b>Background Studies</b>	<b>15</b>
1.2.1	Passive verb phrases in Mechanical Engineering texts.....	15
1.2.2	Passive verb phrases in letters to the press and other text-types.....	17
1.2.3	Influence of background studies on the thesis.....	19
<b>1.3</b>	<b>The Pilot Study</b>	<b>21</b>
1.3.1	The Data.....	21
1.3.2	Computer entry, storage and retrieval .	22
1.3.3	Choice of features to be coded.....	23
1.3.4	Practical problems of coding.....	29
1.3.5	Problems of formal categories.....	31
1.3.6	Enquiries from the data - numerical....	35
1.3.7	Enquiries from the data - exploratory..	44
1.3.8	The environment of copula BE.....	45
1.3.9	The borderline between adjectives and participles.....	50
1.3.10	Verb phrases in sequence.....	61
<b>1.4</b>	<b>Literature</b>	<b>69</b>
1.4.1	Corpus linguistics.....	69
1.4.2	Syntax and semantics.....	72
1.4.3	Categorical Indeterminacy.....	77
1.4.4	Borderlines: subject/complement.....	80
1.4.5	Borderlines: adjective/verb.....	82
1.4.6	Borderlines: full/auxiliary verbs.....	85
<b>1.5</b>	<b>The Main Database</b>	<b>89</b>
1.5.1	Data collected and coded.....	89

1.5.2	Features coded.....	91
1.5.3	Extracting information from the database.....	97
<b>Chapter 2</b>	<b><u>Aspects of styles in the corpus</u></b>	<b>99</b>
<b>2.1</b>	<b>Description of the whole database</b>	<b>99</b>
2.1.1	Main verbs and the copula 'BE'.....	101
2.1.2	Finiteness.....	102
2.1.3	Verb phrase types.....	104
2.1.4	Passive verb phrases.....	105
2.1.5	Verb phrase sequences.....	108
2.1.6	Clause types and functions.....	108
2.1.7	Subject type.....	110
2.1.8	Agent type.....	113
2.1.9	Subjects and agents - form.....	115
<b>2.2</b>	<b>Comparative description by text-type</b>	<b>118</b>
2.2.1	Features showing no difference between text-type.....	118
2.2.2	Features which differ between text-type.....	122
<b>2.3</b>	<b>Comparative description by newspaper</b>	<b>142</b>
2.3.1	The copula 'BE'.....	142
2.3.2	Modal auxiliaries.....	145
2.3.3	Passive verb phrases.....	147
2.3.4	Verb phrase sequences.....	148
2.3.5	Subordinate clauses.....	150
2.3.6	Subject type.....	152
2.3.7	Subject form.....	155
<b>Chapter 3</b>	<b><u>The copula 'BE' and its environment</u></b>	<b>159</b>
<b>3.1</b>	<b>Types of following context</b>	<b>161</b>
<b>3.2</b>	<b>Noun phrase complements</b>	<b>165</b>
3.2.1	The problem.....	166
3.2.2	The investigation.....	168
3.2.3	End-weight.....	178
3.2.4	Negation and number.....	183
3.2.5	Anaphora.....	188
3.2.6	End-focus.....	189
<b>3.3</b>	<b>Conclusions about NP subjects and NP complements</b>	<b>190</b>

<u>Chapter 4</u>	<u>The borderline between adjectives and participles</u>	194
4.1	Decisions arising from the pilot study	197
4.2	Method of investigation	200
4.3	Results and analysis	202
4.4	Syntactic and semantic features of the borderline	213
4.5	The adjective/ participle borderline - summary	223
<u>Chapter 5</u>	<u>Verb phrases in sequence</u>	226
5.1	VPs in sequence - a problem of description	226
5.2	Sequential verb phrases and other features	228
5.3	The grammatical status of sequential verbs	232
5.4	Restricted class of non-final VPs in sequences	238
5.5	Non-final sequential verbs - a partial semantic analysis	242
5.6	The ordering of longer sequences	253
<u>Chapter 6</u>	<u>Conclusion</u>	259
6.1	Aims of the thesis	259
6.2	Results of the thesis	260
6.2.1	The style of newspaper English.....	260
6.2.2	Styles of different text-types in newspaper English.....	260
6.2.3	Styles of some British national newspapers.....	261
6.3	Insights into problems of description	264
6.3.1	The subject/complement overlap.....	264
6.3.2	Distinctions between -en participles and adjectives.....	265
6.3.3	The problem of sequential verbs.....	267

<b>6.4</b>	<b>Future work based on this thesis</b>	<b>268</b>
6.4.1	Further work based on the present data.....	268
6.4.2	Future work based on studies reported in this thesis.....	269
 <b><u>Bibliography</u></b>		<b>271</b>
 <b><u>Appendices</u></b>		
<b>Appendix 1a</b>	<b>Sample output from an Extract enquiry (pilot database)</b>	<b>278</b>
<b>Appendix 1b</b>	<b>Results of numerical enquiries (pilot study)</b>	<b>279</b>
<b>Appendix 1c</b>	<b>Sample output from an Extract enquiry (main database)</b>	<b>281</b>
<b>Appendix 1d</b>	<b>Examples of sentence contexts (main database)</b>	<b>282</b>
<b>Appendix 2</b>	<b>Results of numerical enquiries (main study)</b>	<b>293</b>
<b>Appendix 5</b>	<b>Sequential verbs in main data and supplement</b>	<b>295</b>

Chapter 1  
Introduction

1.1 The Thesis

1.1.1 'Newspaper English' and 'style'

'The style of newspaper English' may sound an uncontroversial phrase, but it takes little effort to question whether the underlying assumptions of such a phrase can be accepted.

'Newspaper English' (specifically in this case, British English) may or may not constitute a coherent variety of English but even academic studies often take its coherence for granted (see, for example, Rydén (1975)). Indeed, in casual conversation, remarking upon the latest excesses of the Sun or some pompous language in the Times, we usually assume that there is an identifiable variety of English associated with the press.

The examples in the last sentence may contain a clue as to one aspect of the language of national daily newspapers in Britain since we tend to remark on different features of language for different newspapers. There may, therefore, be different varieties for different groups of papers which correspond to the 'quality' and 'popular' press. One possible conclusion is that the formality of a paper's language, partly derived from its position on the quality-popular range, is the only determinant of its style. In other words, a popular paper using a casual

style of English will use language which is not significantly different from casual style in other contexts. Although the present work does not have space to test such a conclusion, I choose not to accept it as fundamental to my argument. Instead, I accept the contrary conclusion; that the popular image of 'newspaper language' as being to some extent coherent is correct.

However, the construction of this thesis also relies upon the assumption that within the rather wide register of 'newspaper language', there will be some variation according to types of text within papers, and according to the different papers themselves. Other writers recognise the flexible nature of terms such as 'register'. Zwicky and Zwicky (1982), for example, define 'register' as associated with specific contexts and functions, but allow the size to range from restricted varieties such as newspaper headlines to broader varieties such as 'the language of football'. Huddleston et al. (1968) also accept the usefulness of looking at their corpus as subdivided into three 'strata':

"Even taking into account the very considerable differences between texts in the same stratum, it was still possible to find a variety of features for which there were clear differences between the different strata taken as wholes." [p609]

The present study uses a carefully constructed corpus of data to make a partial syntactic description of 'the language of newspapers' as well as comparing more

restricted registers within the same data such as 'the language of popular papers' and 'the language of editorials'.

Another controversial term in the title of the present work is 'style'. I have already made an assumption about its meaning, namely that one can identify features of language associated with different registers of English and call them 'stylistic'. Even if this is accepted without question, a study with restricted time and limited space requires decisions to be made about which features of English are to be studied. Unlike Osgood (1960), I do not consider that stylistics is more concerned with structural than lexical choices. It seems possible to make stylistic choices at all levels of linguistic analysis, including lexical choices. However, the emphasis in the present work is upon structural choices, focussed on the verb phrase but including information about the context of the verb phrase. The reasons for this choice are explained in section 1.2 where the development of the main study out of a number of observations and background studies is explained.

The previous paragraph also made an assumption about the definition of 'style' when it mentioned 'choices'. Most writers agree that style involves an element of choice or freedom. Winter (1969) says:

"A style may be said to be characterised by a pattern of recurrent selections from the inventory of optional features of a language."

Although this is acceptable as a definition of style, it makes the notion of choice seem more straightforward than it is. Usually there are not clear-cut choices at every grammatical boundary for which one can identify the 'newspaper' version or the 'Dickens' version or the 'Jane Smith' version. Most structural choices rely upon a number of other choices and it is impossible to say which is primary. For example, the following data item from the main database for this study could have been written differently:

e.g. Dr Owen rocked the Alliance earlier this summer by publicly refusing to fight the next election without a clear commitment on nuclear defence.

[7 Sun Art]

If one is interested in the 'choice' of active verb phrases rather than passive ones, rewriting the example shows clearly that such a choice has implications for surrounding clauses:

e.g. The Alliance was rocked by Dr Owen this summer when he (or 'who') publicly refused to...

[rewritten - 7 Sun Art]

The subordinate clause with 'refuse' as its verb can no longer occur as an adverbial of manner to the main clause since Dr Owen is not the subject. Instead, the subordinate clause has to be rewritten as a time adverbial or a relative clause. Thus voice is not what Wells (1960) calls a "pure dimension of style, that is; it is not a variable

which can vary without variation in the other basic factors of style". There are probably very few features of language which could be described as 'pure' in this way; they would not be the interesting ones. Sampson and Haigh (1987) make a similar point about the difference between the styles of prose types in English, claiming that though certain constructions are typical of different genres, they are rare in absolute terms. They add:

"the really significant differences between the prose genres lie in quite small differences in mean values of topological properties of grammatical structure, which have a large cumulative effect because of the recursive nature of grammatical hierarchy but which are too small to allow anything to be predicted from individual productions." [p218]

The use of statistics in the study of style has been resisted by some, particularly those concerned with literary style, but it has now gained wide acceptance as one way of representing the general choices (Sampson and Haigh's "cumulative effect") made by a writer or speaker without being dogmatic about the choices in particular examples. Zwicky and Zwicky (1982) illustrate this "statistical preference" for some features over others in the following way:

"as when menus favour modified nouns over unmodified ones (ice-cold shrimp nestled on fresh lettuce rather than shrimp on lettuce)"

One further question that worries some linguists is whether the style of any variety of language can be established only after the identification of a set of

norms. Osgood (1960), for example, states that the student of style is interested in deviations from norms and adds:

"although the norms have to be determined before deviations from them can be noted and interpreted."

The problem with this is partly practical. Establishing the 'norms' would take many linguists many years work and it is unclear when the opportunity for the study of 'deviations' would arise. Another problem, of course, is whether we can justify the use of such a notion as 'norm' for such a wide-ranging and variable entity as the English language. This problem is particularly difficult when the study of statistical preferences of style is involved. In studying dialects one can have recourse to the notion of Standard English which, though evasive, can be agreed up to a point. What body of English, however, could form the basis of a description of 'normal style' in English?

A related problem, identified by Huddleston et al. (1968) and Dolezel (1969), is the verification of statistical results when "for all practical purposes text populations are open sets (uncountable)". For the purposes of the present study, I accept Dolezel's further statement about the status of stylistic description:

"Therefore all conclusions about the properties and nature of style characteristics, about the speaker type and text types, and about stylistic differences are to be considered hypotheses that will be confirmed or refuted by the accumulation of vast empirical material in the future."

### 1.1.2 A corpus-based study

This thesis is based on data which is described in Section 1.5 of the Introduction. The present section outlines the reasons for using data as a basis of description rather than relying on elicited data or on the introspective data of the researcher.

Corpus-based studies, which limit themselves to describing all and only the features of language present in the data, have been popular since linguistics established its status as a science. Some linguists produced descriptions based entirely on a corpus (e.g. Fries 1952) whilst others ignored this trend and based their descriptions "entirely on the language used by the author" (Palmer 1974). As Palmer points out, there are limitations on the usefulness of corpora:

"i. Even with an enormous amount of data, some forms will not occur."

"ii. In any recorded text, there are bound to be a number of anomalous forms."

The present study does not share the assumption underlying Palmer's first criticism of corpora which is that the linguist is aiming to describe areas of English structure exhaustively. In fact, this study has the aim of establishing whether there are any patterns of usage for certain areas of syntax which are typical of the data used. The fact that some forms do not occur will be considered potentially significant in stylistic terms. A

counter-argument to Palmer's second criticism is that in certain areas of study, anomalous forms are more likely to occur in data created by the linguist than in naturally occurring data. Svartvik (1966) points out this possibility in connection with the passive construction when he refers to W.S. Allen's use of the phrase 'grotesque curiosities' (Allen 1959) to describe some of the passives automatically derived by rule from actives.

e.g. Girls are liked by John

The close relationship generally acknowledged to exist between some active and passive sentence pairs may lead the unwary linguist to create and describe forms which would be highly unlikely to occur in any genuine linguistic context. Such forms may be more misleading than the anomalous forms created by performance errors which worry Palmer. Svartvik, who bases his work on a series of corpora, assumes that passive sentences recorded from actual speech and writing are likely to be normal and natural uses of the passive.

Proponents of both types of data (corpus / intuitive) acknowledge the drawbacks of their chosen method as well as those of the rejected method. Since there are clearly strengths and weaknesses in both approaches, it may be sensible to argue for an approach which includes both kinds of data. Huddleston (1971), for example, emphasises the importance of 'text' as the primary data of linguists

but insists that their studies should not be limited to a corpus alone:

"the description will be of interest only to the extent that the grammatical categories established have validity beyond the text."

Huddleston is describing a 'data-based' study here, where a corpus is supplemented by both incidental but naturally occurring data and by intuitive responses and judgements. There have been some studies of English which combine various types of data in this way; one fairly comprehensive example is *A Grammar of Contemporary English* (Quirk et al 1972). This work is well known for being based on material collected by the Survey of English Usage and in the Preface, the writers describe their data in the following way:

"we have augmented our own experience as speakers and teachers of the language with research on corpora of contemporary English and on data from elicitation tests"

The research on collected material is represented here as a supplement to intuitive data. One of the problems of such an approach is that there is a danger of circularity when collected data is referred to selectively rather than being described exhaustively. The researcher must have decided in advance what to look for and may simply not see some structures. However, Quirk (1984) indicates that the Survey of English in fact works from the data and uses elicited material as a supplement:

"there is a purposive relation between corpus and

elicitation. Not only do we turn to elicitation for information on features that chance to be of fairly rare occurrence: we do so when corpus instances reveal a variation that needs to be explained."

Although I accept the usefulness of elicitation techniques and of intuitive data in supporting data-based studies, the present study is limited to a description of a corpus. There are, however, two ways in which this study has implications extending beyond the corpus itself. First, the stylistic study of a number of syntactic and semantic features (chapter 2) not only describes the database itself, but also stands as a representative study of a wider body of data: newspaper English. Although I cannot claim that the findings of chapter 2 are necessarily true of all newspaper data, they are general enough to be used to compare with studies of different registers. Secondly, the corpus has been used to provide data on issues of semantico-syntactic description relevant to the language as a whole (chapters 3 to 5). In this function, the data replace intuition with concrete examples and represent English generally. These chapters also use some data which are not in the original corpus. This extra material is used to supplement the main corpus where there are not enough examples of a particular feature. It is taken from the wider collection of material forming the context of the main database.

The final point to be made here about data-based studies, is a criticism of their use of examples. Quirk et. al.

(1972) is typical of the kind of study which has been based on attested data, but fails to use the data for exemplification:

"For reasons of simplicity and economic presentation, however, illustrative examples from our basic material are seldom given without being adapted and edited;"

While their justification is understandable, I fear that there is a danger of obscuring the distinction between attested and invented examples and of consequently neutralising the advantages of using real data. It is particularly important for the reader to be aware of the status of examples; whether they are invented, attested but edited, or attested and quoted in full. This thesis quotes examples from the data in full in most instances. A few examples are rewritten, invented or cut short, but such editing is always made clear.

### 1.1.3 A descriptive study

As the previous section has made clear, this thesis is primarily descriptive and adopts no specific theoretical position at the outset. The previous section argued that for methodology an eclectic approach is most fruitful; here a similar point is made about the relevance of linguistic theories.

One of the consequences of basing a description on specific data is that the researcher is made very conscious of the dangers involved in formulating theories

of language without reference to attested data. If we consider, for example, the well-known semantic equivalence between active and passive pairs of sentences which was the departure point for Chomsky's original formulation of transformational grammar, it is clear that for many active examples from attested data, it would be impossible to construct passive equivalents and vice versa<sup>1</sup>. When the researcher is dealing with real data, both the immediate linguistic context (i.e. phrase, clause or sentence) and the wider context (i.e. paragraph, chapter, text) are at least as important for description and interpretation of the original as the kind of relationships between structures hypothesised by a theory.

This thesis, then, is influenced by various theories wherever they seem to provide some insight into the structure and patterning of the data. Quirk et al (1972) make the same point about the influence of "current or recently formulated theories" on their own, more ambitious description:

"Each of those (theories) propounded from the time of de Saussure and Jespersen onwards has its undoubted merits, and several (...) have contriuted very great stimulus to us as to other grammarians. None, however, seems yet adequate to account for all linguistic phenomena, and recent trends suggest that our own compromise position is a fair reflection of the way in which the major theories are responding to influence from others."

---

<sup>1</sup> Note that this point is made about individual examples, many containing a verb which, in other contexts, would happily passivise.

#### 1.1.4 A selective study

The background to the choice of newspapers as the data for this thesis is explained in section 1.2 but this section gives an outline of the aspects of structure and semantics which are the main focus of attention.

Even with the rather small corpus which is the main data for this study, it would be a huge task to describe all syntactic or semantic aspects of the data. The description presented here is therefore restricted to a number of linguistic features focussed primarily on the verb phrase, but also considers features of the context of verb phrases such as their main argument, the subject, and their clausal contexts.

As section 1.2 explains, the passive construction was one of the starting points of my work and this interest widened into a more general consideration of the verb phrase. As work progressed it became clear that my aim of considering the whole clause context as relevant to the description of the verb phrase was impractical in a work this size. I therefore chose two areas of potential interest; the overall structure and function of the clause containing each verb phrase, and the structure and semantics of the subject attached to each verb phrase. As chapter 6 (Conclusion) points out, each restriction imposed on the scope of this thesis indicates an area of future study. It would, for example, be interesting to

compare the structure and semantics of objects with those of subjects. But such a comparison would be to change the primary focus of the work from the verb phrase to the noun phrase and/or the clause.

#### 1.1.5 Structure of the thesis

The remainder of chapter 1 explains the background to the work and describes the pilot study.

Chapter 2 contains a stylistic description of the data. Features of the verb phrase and its context are described statistically for the whole data, for three different text types in the corpus, and for the eight different papers used in the corpus. Some significant differences in style between these sub-sections of the data are noted and tentative explanations are given for some of the differences.

Chapters 3 - 5 are more exploratory in method and take up some of the descriptive problems noticed during the pilot study in order to investigate the gap between models of description and attested data. Many such gaps were noticed in the pilot study, but restrictions of time and space limited further work to the following areas:

Chapter 3 explores the environment of the copula BE. In particular, the relationship between NP subjects and NP complements.

Chapter 4 investigates the borderline between -en

participles and adjectives, and encounters problems of identification and description.

Chapter 5 discusses verb phrase sequences<sup>2</sup> and questions the grammatical status of 'sequential' (or 'catenative') verbs, their semantic relationships and ordering priorities among them.

The Conclusion (chapter 6) summarises the findings of chapters 2 to 5 and delineates areas of future study indicated by this thesis.

## 1.2 Background Studies

The pilot study, which experimented with both data and methodology, is described in section 1.3. below. The present section introduces some of the premises underlying the thesis by describing the observations and preliminary studies which were its starting point.

### 1.2.1 Passive verb phrases in Mechanical Engineering texts

One observation which contributed to the methodology and subject matter of this thesis concerns the frequency of passive verb phrases in texts written by Mechanical Engineering students and in the textbooks that they read.

---

<sup>2</sup> Note that at this stage I make no assumptions about the nature of the relationship between so-called 'catenative' verbs and the non-finite verb phrases/clauses they govern. The use of the term 'sequence' is not intended to imply any particular analysis until the whole subject is discussed in chapter 5.

During a short course given to students of Mechanical Engineering to improve their written English, I noticed that there was a preponderance of passive verb phrases in their work. As a measure of the normality (or otherwise) of this feature, I compared the proportion of passive verb phrases in one student's design report with three other examples of written English. These were:

- a. Two pages of a magazine called 'Chartered Mechanical Engineer' which the students read regularly.
- b. The introductory chapter of a student textbook called 'System Modelling and Control' (Schwarzenbach and Gill 1978) which the students also claimed to have read.
- c. The opening chapter of a textbook in a different discipline at a comparable level. 'Introduction to Language Pathology' (Crystal 1982) was chosen to represent a fairly technical subject since I expected it to also show a high proportion of passive forms.

Table 1.1 shows the proportion of passive verb phrases in each extract as a percentage of the total number of verb phrases in the extract:

Table 1.1 Percentage of passive VPs in 4 written extracts

	pass VPs	total VPs	% pas
Student's report	119	381	31
Chartered Mech. Eng.	32	101	32
System Modelling	129	383	31
Language Pathology	99	653	15

Although the samples involved in this exercise were small, the similarity in percentage of passive forms among the three Mechanical Engineering texts was striking. I had expected the magazine to show a lower proportion of passives than the academic text as it has a more casual

style generally. I had also expected the two academic texts to resemble each other in this feature, but found that the difference of discipline had more effect on this aspect of style than difference of text-type.

One of the implications of the above for the present study was that even fairly simple statistical studies can be revealing in matters of style. Chapter 2 is based on this assumption.

The observations of Mechanical Engineering texts also suggested that there may be some correlation of stylistic features between texts written and read by the same people. This was one motivation for including letters written by readers of newspapers in the data for the main study.

### 1.2.2 Passive VPs in letters to the press and other text-types

Since the passive construction is a marker of fairly formal style, I decided to continue my investigation with more examples of written English. Newspapers were chosen for the reasons given in section 1.3 and because there is an interesting relationship between people who write to the press and the newspaper in which their letters are published. The correspondent (here meaning letter writer, not journalist) is presumably trying to write in the kind of style which the editor will publish. It may be assumed, then, that some care is taken even in letters written in a casual style and sent to the popular press.

Restricting the stylistic inquiry to the passive construction, I considered newspaper language from two angles. First, a range of newspapers were considered to see whether the popular image of a split between broadsheets ('quality') and tabloids ('popular') was reflected in style of English. Secondly, different text-types within newspapers were compared to see whether this line of enquiry should be pursued.

Table 1.2 shows the proportion of passive verb phrases in the letters published in five different newspapers on 15th March 1984:

Table 1.2 Percentage of passive VPs in letters to five newspapers

	pass VPs	total VPs	% pass
Times	69	311	22
Guardian	37	280	13
Mirror	7	25	28
Express	6	96	6
Sun	1	68	1

These figures, though taken from small samples, showed differences between the newspapers large enough to make further investigation seem worthwhile. It was particularly noticeable that there were no apparent groupings of serious and popular newspapers for this feature of style. The relationship between such groups is more complex than anticipated and is explored in chapter 2.

The second part of this observation exercise was to obtain figures for the passive construction from three different

kinds of text in two newspapers. The texts considered were the letters page, the editorial and a news article from The Guardian and The Times of 28th March 1984. The percentages are given in Table 1.3:

Table 1.3 Percentage of passive VPs in three text-types

	letters	editorial	article (% pass VPs)
Times	17	11	13
Guardian	13	11	23

Any conclusions had to be tentative at this stage, but the low frequency of passive verb phrases in editorials indicated that there may be some significant differences between text-types within newspapers.

### 1.2.3 Influence of background studies on the thesis

The observations reported above helped to settle a number of issues of data and methodology for this thesis. This section enumerates these decisions.

I had, from the beginning, intended to use attested data in my work. For stylistic studies it is clearly necessary to do so but the danger remains of prejudging what is likely to occur by only searching for particular features. As explained in section 1.1, the method used was to investigate a corpus of data exhaustively, but for a restricted area of syntax and related semantic features. This general restriction, imposed because of the size of the work, is not selective enough to result in the kind of prejudging I wished to avoid.

The results of the simple statistical comparisons in these background studies were interesting enough to indicate that larger studies of this type may show some stylistic features and relationships between types of data that are unpredictable from casual observation. I was careful, however, to use statistical comparison only as an indicator of general trends within the data. Although some linguists use such numerical measures to predict linguistic features in a precise way, I was concerned with gross differences between texts which might correspond with the public's intuitions about the language of different newspaper types and texts.

Another result of the background studies was the decision to base the thesis on the language of some of the main British daily newspapers. It seemed that the dimensions of newspaper (from the serious-popular range) and text-type (letters, editorials and articles) would be worthy of detailed study.

Finally, I decided that the features of language investigated should be centred on the verb phrase since it was the focus of my original interest in the passive construction. However, the general aim of using linguistic context to the full made this decision a matter of perspective rather than restriction.

In order to try out the methodology, in particular the storage and retrieval of data using computer database

software, and to discover some of the most fruitful areas of investigation, I set up a pilot study described in the following section (1.3).

### 1.3 The pilot study

#### 1.3.1 The Data

The data collected for the pilot study consists of the letters page from each of seven national daily newspapers on five consecutive days from 19th November 1984 to 23rd November 1984 inclusive. The newspapers represented are; The Times, The Guardian, The Telegraph, The Daily Mail, The Mirror, The Express and The Sun.

Each verb phrase in the data was to be coded with a number of features describing aspects of its structure, form and context. As the process of coding and entering the data into computer storage is very lengthy, I decided that for the pilot study a total of 1000 verb phrases would be sufficient. In fact 800 verb phrases only were coded in the time allocated for this stage but they were sufficient for trying out the intended methods of retrieval and manipulation of data. The verb phrases were not evenly spread across the range of newspapers collected so the pilot study does not contain any comparisons between single newspapers, or groups of newspapers. The pilot database therefore functions as a single corpus which is partly representative of newspaper English.

### 1.3.2 Computer entry, storage and retrieval

The use of a computer as an efficient way of storing, retrieving and cross-classifying data was envisaged early in the work for this thesis. The Amdahl computer of Leeds University has a database package called Extract (Duke and Screeton 1982) with which I was already familiar and this was chosen as the means of storing the pilot data. Another package called Inform, which was written to be used with Extract, was used to enter the data into suitable computer files.

Inform allows the user to create a 'form' on the screen which is then 'filled in' with the data. Each field of information associated with a data item is named on the screen and any special conventions of coding are included as a reminder to the user. Once I had decided which features of verb phrases were to be coded, an Inform file was created, containing instructions about the number of fields required for each data item and their names. The data was then input with all problems of coding being noted as the work progressed. Some of the coding conventions were changed during this process whilst other problems had to remain unsolved until they could be investigated in the main part of this thesis.

The resulting database was in a format recognisable by the Extract package. This package was designed as a bibliographic program for easy sorting of book and article

references on the basis of author, title, year or any other information which has been supplied. Despite this rather specialised purpose, the package is very flexible since the number and names of all the fields of information can be decided by the user. Once the data has been input, the package allows enquiries to be made which extract subsets of data, show the number of data items in any subset and can print out any combination of the fields associated with each data item in alphabetical order (see Appendix 1a for examples of Extract output). Whilst inputting was in progress, I made a number of practice enquiries to ensure that each record was structurally correct (i.e. had the right number of fields) and to discover which lines of enquiry would lead to interesting studies.

One practical problem which was solved by another system of computer storage was the need to refer to complete sentence contexts when investigating sets of verb phrases. It would have been very time-consuming to include the sentence as a field for each data item (i.e. each verb phrase) since many of the sentences are very long and contain up to 6 verb phrases. I therefore decided to type each sentence only once in a separate file with reference numbers corresponding to those entered in the pilot database.

### 1.3.3 Choice of features to be coded

The eighteen fields entered into the database for each verb phrase are considered here in turn. It should be

emphasised that the categories used to characterise features of verb phrases and their context were chosen as provisional categories; the most interesting cases were expected to be found where these categories broke down.

Apart from the reference codes in fields 2 and 18, the fields refer to four different aspects of the data. Fields 1 and 3-7 inclusive characterise the verb phrase<sup>3</sup> itself, fields 8-10 inclusive refer to the clause context of the verb phrase and fields 11-15 refer to two of the arguments of the verb phrase; the subject and the agent. These arguments were chosen because I wished to compare noun phrases as subjects with those functioning as agents. Finally, fields 16 and 17 identify borderline cases and obligatory cases of the passive construction respectively.

1. **VP** This field contains the verb phrase in exactly the form in which it occurs in the original context.

e.g. should raise [4/Te/19]<sup>4</sup>

2. **Ref** Each data item was given a unique reference consisting of a number, the name of the newspaper and the date of the newspaper.

e.g. [4/Te/19] = the 4th verb phrase to be found in the Telegraph from 19th November 1984.

3. **Mv** This field contains the citation form (infinitive

---

<sup>3</sup> It should be made clear at this point that the term 'verb phrase' in this study is not intended to include object or complement as it does in some work on generative grammar.

<sup>4</sup> Each example in this section is given with its database reference. The format of the reference is explained in the next paragraph.

without 'to') of the main verb from the current verb phrase.

e.g. RAISE

[4/Te/19]

4. **Simple / Finite** This field contains two subfields: The first is coded S (simple) for a verb phrase consisting of a main verb alone and C (complex) if it also contains one or more auxiliary verbs. The second subfield is coded F (finite) for finite verb phrases and NF (non-finite) for non-finite verb phrases.

5. **VP sequences** This field was used to identify those verb phrases connected by sequential ('catenative') verbs such as 'want', 'tell', 'begin' and 'make'. All verb phrases in a sequence had the reference codes of the other connected verb phrases entered in field 5.

6. **Polarity** This field identifies a verb phrase as positive (Pos), negative (N) and/or interrogative (Q).

7. **VP structure** The structure of the verb phrase was coded in terms of its constituent parts (modal, perfective, progressive, passive, main verb) and their forms (participles ing and en, 3rd person present s, other present forms o, infinitive i, past ed).

8. **Clause type** Here the clause surrounding the verb phrase is classified as main, adverbial, nominal, comparative, postmodifying or unidentified (?). These clause types are defined as in Quirk et al. (1986) except for the 'postmodifying' group which requires some explanation here. Since I was concerned with the clause contexts surrounding verb phrases and their main

constituents, it seemed unnecessary to code all the different types of clause that postmodify nouns and adjectives. I therefore decided to group all of these clause types together for present purposes. This group, then, comprises relative clauses, non-finite clauses which postmodify nouns, appositive clauses and clauses complementing adjectives.

9. **Clause function** In this field the subordinate clauses are classified according to their function within the sentence as one of; Subject, Object, Complement, Adverbial, Adjective modifier, Noun modifier - or unidentified (?).

10. **Clause structure** This field contains the skeletal structure of the clause, using function labels (SVOCA).

e.g. SVC that is a just reward for the loyalty of  
the staff [10/G/19]

11. **Subject type** In this field, four subfields show the following information about the subject associated with the current verb phrase:

- a. animate, abstract, inanimate or human
- b. male, female or neither/not known
- c. 1st, 2nd or 3rd person
- d. single, plural, uncountable or collective

e.g. ab n 3 s [98/Ti/19]  
(once) the strike (was..)

Note that in subfield a, the term 'animate' is a shortened version of 'animate and non-human'. It has this meaning throughout the thesis. The subject was marked 'not obvious' (n/o) if it had no physical realisation and its

identity was not clear from the context.

12. **Subject words** As a simple measure of complexity, the number of words constituting the subject was entered in this field, with an additional code if the subject contained coordination and/or subordination.

e.g. 7 the main effect of the Belcher initiative  
(will be...) [5/Te/19]

13. **Subject form** This field shows whether the subject occurs as a full form or as one of the pro-forms; personal pronouns (pn), relative pronouns, demonstrative pronouns, impersonal pronouns. If a subject was not realised, it was coded 'n/r'.

14. **Agent type** Four subfields characterise any agent following a passive verb phrase in the same way as for the subject (field 11):

- a. animate, abstract, inanimate or human
- b. male, female, neutral/not known
- c. 1st, 2nd or 3rd person
- d. single, plural, uncountable or collective

Alternatively, this field was marked 'not realised' (n/r) for passive verb phrases with no following agent, or 'not applicable' (n/a) for active verb phrases.

15. **Agent words** The agent, like the subject in field 12, was also coded according to the number of words and the presence of coordination or subordination.

16. **Adjective / participle** Some apparently passive verb phrases were marked as belonging to a set whose precise character was in doubt since they could also be analysed

as containing the copula BE and a participial adjective. The coding for this field was as follows:

clear passive	<u>Ven</u>
clear adjective	<u>Aj</u>
probably passive	<u>Ven?</u>
probably adjective	<u>Aj?</u>

17. **Passive / active** This field was intended to show how great a proportion of verb phrases which could be constructed with a passive auxiliary, were in fact passive. Active verb phrases were coded as either potentially passive (pass) or not (notp) and passive verb phrases were coded as either potentially active (act) or not (nota). This was regarded as being more than a matter of identifying transitive verb phrases since it is well-known that there are some transitive verbs which do not occur in the passive as well as a few which always occur as passive forms:

e.g. I was born in Germany (from Winter 1965)

However, it was very difficult to establish, for any particular example, whether there was an acceptable passive (or active) form since any change in the verb phrase involved changes in other clause elements and even in other clauses (see section 1.1.1 for a discussion of 'pure' features of style). This field was abandoned in the main study (see section 1.3.4 for more discussion of its problems).

18. **Sentence context** The final field contains a numerical reference to the sentence in which the current

verb phrase appeared.

#### 1.3.4 Practical problems of coding

The features described in section 1.3.3 were set up as carefully as possible on the basis of the preliminary observations but with the recognition that there would be some problems when coding larger quantities of data. The problems which arose during the coding of the 800 pilot data items fall into two groups. The first group consists of examples which caused practical problems for the coding and revealed discrepancies between the language expected and 'real' language data. The second group revealed more serious flaws in the formal categories I was trying to use and represent real problems of description. This section describes the first group of problems. The problems in the second group are taken up in section 1.3.5.

One of the practical problems which arose was the question of how to characterise verb phrases which were partly or wholly ellipted. The solution was to enclose the ellipted part of the verb phrase in brackets in field 1 and any ellipted main verbs were also bracketed in field 3.

e.g. "unless it is clear that we have" [43/Ti/19]  
 Mv: (GIVE) VP: have (given)

There is no space in this study for the deeper question of how we know when a form is elided; only clear cases were coded as described above.

The identification of verb phrases occurring in sequence caused a few problems, particularly where an intervening noun phrase obscured the connection:

e.g. compels me to advise of... [30/Te/19]

The data was checked through again carefully for examples which may have been missed.

One problem of assigning clause types in field 8 was how to classify comment clauses which interrupt other clauses, and are to that extent subordinate, but which are structurally independent.

e.g. a rusty German helmet, complete with bullet holes, which had lain there for 30 years, or to be more accurate, since November 1944.  
[295/Ti/19]

These cases were marked '?' since there was no available term for them, but it has since been suggested to me that they clearly perform a distinct function which some (e.g. Leech) have labelled 'interpolation'. There were very few examples of this kind in my data.

Difficulties of describing clause structure in field 10 arose when a clause element was interrupted by another element. In the following example, the prepositional phrase postmodifying the object is postposed after an adverbial:

e.g. to open (pay negotiations (with the working miners) (on productivity bonuses and other matters))

The coding solution was to repeat the symbol for the clause element which had been split (OAO in the above example).

#### 1.3.5 Problems of formal categories

The problems associated with the use of formal categories described in this section were not resolved during the pilot study. In the main database used in this thesis, some features had the basis of their coding changed from formal to semantic categories. These changes are described in section 1.5 of this chapter. The problems themselves are illustrated here.

There were a number of problems with field 6 which was set up on the assumption that all verb phrases could be satisfactorily classified as positive or negative on the one hand, and declarative or interrogative on the other. Although such a classification is formally possible, there were numerous examples where formal classification seemed to contradict semantic force.

For example, one kind of structure which opposes grammatical form and semantic force is the rhetorical question:

e.g. Are we not all supposed to be...?

[47/Te/19]

There were also many examples where the verb was not negated, but the clause as a whole was negated by some

other means.

e.g. no experienced manager believes .. [52/Ti/19]

Here the subject is formally negated, but there are examples which are more difficult to classify since they do not contain a directly negative term but convey negative meaning. Some comparative terms can be used in this way.

e.g. ... is less suitable .. [201/Ti/19]

The small number of examples in the data made the topic of negation an unlikely choice for further study in this thesis. There have, however, been many other studies of negation; most famous, perhaps, is Klima (1964), but more recently, for example, by Tottie (1980 and 1982).

The field characterising agents of passive verb phrases (field 14) also caused problems for coding. There were a number of examples of 'quasi-agents' which were introduced by prepositions other than 'by' [e.g. with ... 346/Ti/19]. They were only identifiable semantically since other similar prepositional phrases could not be interpreted as agents. These were excluded from consideration in the pilot study, but the main study was changed to include them in the coding scheme.

The coding of subject type (field 11), though not entirely based on formal categories, was originally intended as a

straightforward record of features of the subject. The first subfield might seem to entail a simple decision as to the semantic class of the subject, but frequent use of devices such as metonymy makes this decision more complex.

e.g. Are 'a nation' and 'the government' abstract or human?

e.g. Is 'a letter' (correspondence) abstract or inanimate?

e.g. Are 'Jersey and Guernsey' and 'the islands' abstract (the community), human (the inhabitants) or inanimate (the land surrounded by sea)?

There was a practical problem associated with coordinated subjects and with subjects made up of phrases in apposition since they often had different characteristics and no provision had been made in the coding scheme for multiple subjects. A similar, but more serious, problem arises when a simple subject refers to two or more referents with different characteristics.

e.g. Both groups have ... (i.e. the Church and the miners)

This example could support the view that, in this context, 'the Church' is intended as a plural, human noun on the grounds that coordinated nouns tend to have the same characteristics and 'miners' is unambiguously plural and human. This issue is related to that of the difference between notional and formal concord which is discussed in chapter 3. However, the present work has no space for a detailed discussion of this point.

All of these problems underline the fact that we tend to categorise nouns out of context as though the resulting classes, whether envisaged as formal or semantic, attached to the nouns themselves. The opportunity presented by this thesis to attempt to classify numerous nouns from real language contexts revealed that the classification of the referent is often contradicted by that of the noun itself. As the examples above show, it is usually the classification of the referent that is most relevant to the meaning of the clause as a whole.

A final example of the difference between analysing words in context and in isolation can be illustrated by field 17 in which I was trying to characterise verbs as 'able to occur in the passive (or active) voice'. This field was intended to give more accurate information about the proportion of verbs chosen to occur in the passive in any text. In fact, the problems of classifying words out of context made this aim impossible to fulfil. Clearly, I was not interested in whether a word-form in another sense could passivise, but excluding this possibility in principle does not in practice dispense with interference from other senses during the coding process. It also ignores completely the question of deciding when two senses are indeed separate. Assuming that a word-sense can be identified in each case, there remains the problem of whether to classify the verb as 'able to passivise' in its present context or in some contexts. The following

illustration contains GIVE as the main verb. Out of context it would normally be considered as able to passivise in two ways, transposing either the direct or the indirect object into subject position:

e.g. I gave him the clock  
 He was given the clock  
 The clock was given to him  
 [invented examples]

In one of the pilot data examples, where the verb is being used metaphorically, only one of these passive constructions would be possible:

e.g. gave me a clip round the ear" [22/M/19]  
 I was given a clip round the ear  
 \*a clip round the ear was given to me

It is very difficult to envisage some other verb phrases as passivising within their context without changing a number of adjacent clauses. For the pilot study, I decided to judge whether the verb-sense could be made passive (or active) in some context. Since it was clearly not likely to improve accuracy in the measurement of passive choice, this field was abandoned in the main study.

### 1.3.6 Enquiries from the data - numerical

Despite the problems of analysis explained in the previous section, there were some enquiries which could be made without modifying the database. These enquiries fall into two sets; the first set were designed to produce largely numerical results and provide a stylistic profile of the

data, the second set were exploratory enquiries, intended to extract the problem cases and present them in groups for closer inspection.

In the pilot study, numerical enquiries were intended to identify general areas of interest for the main study as well as give a stylistic description of features of the data. Since the pilot database was not divided internally along the dimensions of newspaper and text-type identified in section 1.2.4, the results of numerical enquiries are presented in this section simply as rounded percentages<sup>5</sup>, out of a total of 800 data items unless otherwise indicated. Significant differences between newspaper and text-type are revealed in the main study.

The numerical enquiries were first made on individual fields and subfields. The results of these enquiries have two main uses. First, they provide a general picture of the proportions of different linguistic features present in one kind of data (letters to the national press) which could be compared with other types of English text. Secondly, such information can be used as a profile of the whole data against which subsets of the data can be compared to see whether they behave 'normally' for this kind of data.

---

<sup>5</sup> All percentages in this thesis are rounded to the nearest integer. In the main study they are provided for interest, but the burden of assigning significance is left to statistical testing.

The contents of the tables below are made clear in their titles; discussion of the contents follows each table.

Table 1.4 Percentage of simple/complex and finite/non-finite verb phrases in 800 pilot data items (field 4)

simple	67	finite	73
complex	33	non-finite	27

It is perhaps surprising, given the attention to verb phrase structure in grammar books, that only one third of all verb phrases in this data are complex (i.e. have one or more auxiliaries). The majority of the verb phrases were, as expected, finite; the proportion could be compared with other types of data.

Table 1.5 Percentage of positive, negative and interrogative forms in 800 data items (field 6)

	declar.	interrog.
positive	91	2
negative	6	1

A majority of positive statements would be expected in any data, but I had expected to find a higher number of negative (rhetorical) questions and negative statements in this data, which often concerns controversial topics. Tottie (1983) found, however, that negation was more scarce in the written than the spoken language:

"I found that if I wanted to have samples of negation of equal size from spoken and written English, I had to examine more than twice as much written material as spoken material."

Table 1.6 Percentage of 800 VPs containing different auxiliaries (field 7)

	<u>% of 800</u>
modal verbs	15
auxiliary BE (prog)	3
auxiliary HAVE	8
auxiliary BE (pass)	10

[Note that some of the auxiliaries will cooccur in the verb phrases and the percentages in Table 1.6. do not, therefore, add up to 100%.]

Table 1.7 Percentage of 800 VPs occurring in different clause types (field 8)

	<u>% of 800</u>
main clause	40
adverbial clause	17
postmodifying clause	19
noun clause	15
doubtful cases (?)	10

The results in Table 1.7 would be slightly altered if the problem cases (10%) were assigned to existing classes. However, many of them are non-finite clauses following sequential verbs and are therefore rather difficult to place unless they are clearly similar in function to noun phrase objects (see chapter 5).

Some individual fields were not amenable to numerical enquiries as a result of the way in which they had been coded. For example, field 10, which coded the structure of clauses in terms of clause elements (S, V, O, C, A), contained all of the optional adverbials. There were too many possible structure patterns to identify any trends in

the data. For the main study, optional adverbials were excluded when they could be identified so that only the basic structures (SV, SVO etc.) and a few variations (e.g. OSV) were used.

Tables 1.8 Subject type (field 11) subfields 1 and 4

	% of 800		% of 800
human	54	single	60
abstract	34	plural	25
animate	1	uncountable	0
inanimate	5	collective	9

Although it does not fall within the scope of this thesis, it would be interesting to find out whether these figures are similar to those for other types of data.

Table 1.9 shows how the 'human' group of nouns break down into sex identified and unidentified (or mixed) groups:

Table 1.9 Percentage of 'human' nouns in male, female and unidentified groups (field 11, subfield 2)

	% human subjects
male	31
female	11
unidentified	58

I had expected a higher number of male than female subjects, but the difference is greater than anticipated. Other types of data would presumably show different proportions of these groups. The breakdown of male and female subjects into 1st, 2nd and 3rd person (field 11, subfield 3) also shows striking differences between the sexes:

Table 1.10 Percentage of sex-identified subjects in female and male groups x 1st, 2nd and 3rd person groups

	male (% 184)	female (% 184)
1st person	21	23
2nd person	0	0
3rd person	52	4

The figures in table 1.10 illustrate the fact that many nouns and pronouns in the pilot data were coded as 'female' or 'male' despite the fact that the words themselves (e.g. 'I', 'the teacher') were not linguistically marked for sex. When the referent of such (pro)nouns could be identified as female or male, it was coded as such. The interpretation of table 1.10, therefore, is that a similar proportion of female and male sex-identified referents are referred to by 1st person pronouns, but far more of the 3rd person references are to men. If women did not write to the press about themselves, their sex would have received only 4% of the sex-identified references. While these figures reflect the well-known fact that men are more prominent in public life than women, the size of the discrepancy as represented by table 1.10 is striking.

Table 1.11 Percentage of subjects occurring as full forms or as pronominal forms (field 13)

	% of 800
full form	34
personal pronoun	23
impersonal pronoun	6
relative pronoun	5
demonstrative pronoun	3
subject not realised	28

There were a large number of unrealised subjects in the pilot data. The following example is typical of the majority of data items with unrealised subjects in having a non-finite verb phrase:

e.g. One solution is  $\phi$  to allow all road users to keep much the same speed. [5 Te 19]

This example has no clear referent for the unrealised subject although many such subjects are omitted because their referents are entirely predictable from the preceding context:

e.g. the Government is covering up and  $\phi$  attempting to justify a wicked action. [130 Ti 19]

After all of the useful enquiries had been made on individual fields, enquiries were made on combinations of two fields, to see whether there were any strong correlations between features. The full set of results can be found in Appendix 1b; only the most interesting findings are discussed here.

Table 1.12 Complexity x finiteness in 800 data items (field 4)

	finite	non-finite
simple	41	26
complex	33	0

It is interesting to note that there were no cases of non-finite complex verb phrases in the pilot data. This can be used to illustrate the distorting effect of many grammar

texts; they often fail to indicate, even in general terms, the frequency of occurrence of many of the structures they describe. The resulting picture of English, which gives equal weight to all features, is unrepresentative of real language data.

Table 1.13 Complexity and clause type in 800 data items (fields 4 and 8)

	simple	complex	
main clause	61	39	(% of 316 MClS)
adverbial clause	72	28	(% of 135 ACls)
noun clause	60	39	(% of 120 NClS)
postmod clause	71	29	(% of 151 PClS)
doubtful cases (?)	82	18	(% of 78 ?ClS)

If the doubtful examples were allocated to the other classes, the picture could change, but otherwise these figures indicate that main clauses and noun clauses have a lower proportion of simple verb phrases than the data as a whole (66%). Adverbial and postmodifying clauses, on the other hand, have a higher proportion of simple verb phrases.

Table 1.14 Finiteness and clause type in 800 data items (fields 4 and 8)

	finite	non-finite	
main clause	99	1	(% of 316 MClS)
adverbial clause	66	33	(% of 135 ACls)
noun clause	71	29	(% of 120 NClS)
postmod clause	50	50	(% of 151 PClS)
doubtful cases	29	71	(% of 78 ?ClS)

As in Table 1.13, allocation of the doubtful cases could alter the proportions in Table 1.14. However, it is still

possible to compare the overall figures for finiteness (66% finite, 33% non-finite) with the figures in this table which show that the overall figure hides considerable variation. Main clauses were expected to be almost 100% finite and one might therefore have anticipated that all of the other classes would show a slightly lower proportion of finite verb phrases than the average 66%. Table 1.14 contradicts this expectation as noun clauses show a higher-than-average and doubtful cases a lower-than-average proportion of finite verb phrases. Many of the latter group (doubtful cases) are examples of non-initial verb phrases in sequence which are always non-finite.

Enquiries on two fields comparing subject and verb phrase features produced few signs of unexpected patterns leading to further study. The proportions in each set of figures were almost all similar to the figures for the whole of the data. There was no indication of features having a strong tendency to cooccur. The only exceptions to this were highly predictable relationships between features, such as the fact that all relative pronouns would occur in relative clauses. As an illustration of the results of this section, Table 1.15 shows the percentages relating subject type (field 11) to clause type (field 8):

Table 1.15 Subject and clause type in 800 data items (fields 11 and 8)

	% total	% MC1	% AC1	% NC1	% PC1
human	54	53	58	67	42
inanimate	5	5	6	1	8
abstract	34	39	27	28	38
animate	1	0	1	0	1
single	60	77	59	46	46
plural	25	17	21	31	37
uncountable	0	0	0	0	0
collective	9	4	10	19	1

None of the clause types seemed to differ very much from the data as a whole as far as proportions of subject type were concerned. However, statistical tests were to be used in the main study in order to discover significant differences not apparent from simple percentages.

### 1.3.7 Enquiries from the data - exploratory

After all the numerical enquiries had been made, a number of exploratory enquiries were used to extract all the examples relevant to the following topics:

- the passive voice
- the borderline between adjectives and participles
- clause types
- clause structures
- verb phrase sequences
- the copula BE and its environment

These topics were chosen because they all presented problems of description at the coding stage which were clearly more than notational problems. Many of the enquiries involved extracting those examples which contained a '?' in the relevant field.

Lack of time restricted detailed investigation to three of these topics; the environment of the copula 'BE', adjective / participle borderline and verb phrase sequences. These three topics are discussed below in sections 1.3.8 to 1.3.10.

#### 1.3.8 The environment of copula BE

All items which had BE as their main verb (field 3) were extracted from the database in order to consider the syntactic environment of the copula BE. I also wanted to collect examples of 'BE + -en form' which had been analysed as 'BE copula + complement' rather than as the passive form; 'BE auxiliary + past participle'. The latter examples would be used in considering the borderline between adjectives and participles (see section 1.3.9).

The number of examples containing BE as their main verb was 149, approximately 19% of the total. I examined the sentence contexts for each example extracted and attempted to identify the clause element following BE as belonging to one of the following classes: noun phrase, noun clause, adjective phrase, prepositional phrase, adverbial phrase. The percentages for each class are found in Table 1.16:

Table 1.16 Clause elements following BE copula

	% of 149
noun phrase	42
noun clause	11
adjective phrase	38
prepositional phrase	7
adverbial phrase	2

Although all of the examples were allocated to one of these classes, not all of the classifications were made easily. Some of the problem cases are discussed below.

The first problem occurs when the subject complement is realised by a noun phrase; in some examples there appears to be a kind of indeterminacy between the functions of subject and (subject) complement. These grammatical functions are normally regarded as differing in a number of ways, summarised here from Quirk et al. (1985). The form of the subject is almost always nominal (phrase or clause) whereas the complement may be either nominal or adjectival. Their normal positions also contrast since the subject occurs most often before the verb while the complement follows the verb. Syntactically, the subject is the element which determines the number (and person, where relevant) of the verb as well as having concord of number with any following (subject) complement. Semantically, the subject is associated with 'given' information and usually represents the theme of the clause whereas the (subject) complement introduces new information to identify or characterise the subject.

The problem of an indeterminacy between the functions of subject and complement arises because most of the distinctions outlined above are not absolute. For example, complements may sometimes be adjectival but in an individual example which contains a nominal complement, the form of the subject and complement are not fundamentally different.

Similarly, there may be a statistical tendency for subjects to precede the verb and complements to follow, but even a small number of exceptions to this general rule may be enough to cause doubt about a particular example in context. One distinction which cannot be doubted is the fact that it is the subject, not the complement, which determines the number of the verb. However, these clause elements also have concord of number between them (in most cases) and difference of number alone cannot, therefore, help to identify the subject in any individual case. The final distinction mentioned above is the semantic distinction between given and new information. It is difficult to discern whether these terms attach to the grammatical functions themselves, or to the positions in the clause which they normally adopt. Some of the examples discussed below (chapter 3) pose this question more explicitly.

When the main verb 'BE' is both preceded and followed by a noun phrase, therefore, the usual analysis would be to assume the first to be the subject and the second to be

the complement. Quirk et al. (1985) distinguish two subtypes of semantic role for complements: identification and characterisation. These can both contribute new information about the subject [CGEL 10.20 p741]. The following example shows that these two main features of the subject (i.e. position and information value) do not necessarily cooccur:

e.g. the victims will be the miners and power workers  
[30/G/19]

If we consider their information value, either of the noun phrases in this example could be interpreted as containing 'given' information; in fact the following noun phrase is the most likely candidate since the word 'victim' is being attributed to people who are already known to be miners and power workers. For the interpretation to assign the 'given' label to 'the victims', one would have to assume that the people concerned are already seen as victims, but are being given the tasks of miners and power workers. This example, therefore, seems to assign 'givenness' to the complement not, as is usual, to the subject. The apparent conflict between word order and information value in this example led me to consider the other 61 examples which had noun phrases (or clauses) following the verb phrase.

All of the other 61 examples made a clear distinction between subject and complement. Many subjects were pronominal as in:

e.g. I am a northerner [13/G/19]

Others were 'dummy' subjects such as 'here' or 'there':

e.g. There is no obstacle to... [15/G/19]

The difference between adverbial 'here' and 'there' and these 'dummy' subjects would normally be clear, since the word order is fixed for the latter:

e.g. \*no obstacle is there to...

Where the preceding noun phrase was complex, there was always a difference in complexity or definiteness between its structure and that of the following noun phrase. Subject noun phrases were more likely to be definite than complements and noun clauses occurred more following the verb phrase than in initial position:

e.g. The main effect of the Belcher initiative will  
be to add another string to the bow ...  
[5/Te/19]

Those articles which occurred in noun phrase complements in this data were all indefinite, except for the example quoted above [30/G/19] which was the original cause of the problem.

On the basis of this investigation, I decided that the apparent indeterminacy between the functions of subject and complement was worthy of further study. One possibility was that the problem examples showed a kind of splitting of the various functions of the subject as seen

in some languages (see, for example, Schachter 1977). Chapter 3 continues this investigation using examples from the main database.

Other problems for classification of contexts following BE included those cases whose complements were assigned to the adjective class, but where the -en form could also have been a participle. There were 8 such examples in the pilot data and these are discussed fully in section 1.3.9.

There were also 4 examples of adjectival complements ending in '-ing', which could represent the participle in progressive verb phrases, rather than complements following the copula BE.

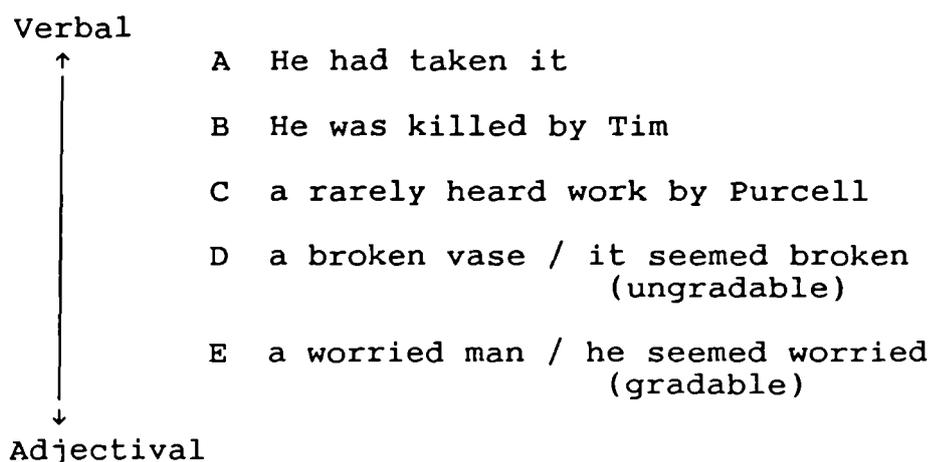
e.g. It is worrying that' ... [1/Te/19]

However, the small number of such cases did not invite further investigation at this stage, particularly since they all passed the test of intensification (i.e. they could all be premodified by 'very').

### 1.3.9 The borderline between adjectives and participles

The problem of identifying -en and -ing forms as members of either the adjective or the verb class is one which many writers mention but few confront by attempting to provide tests or guidelines. Kilby (1984) and Huddleston (1985) are exceptions to this generalisation (see also Johansson 1986 and Sampson 1985). Huddleston (1985) incorporates the notion of unclear boundaries between word

classes as a basic element of his descriptive grammar and also devotes a chapter (chapter 9) to some specific problems including the adjective/participle borderline. He provides examples ranging from clear adjective to clear verb with "at least three intermediate positions":



Kilby (1984) also compiles a list of criteria to distinguish between the two classes. However, he proceeds to show their ineffectiveness in classifying a number of examples. In chapter 5 he demonstrates that the -en form is essentially variable with regard to the word classes 'adjective' and 'verb'. He begins by setting up 8 criteria to define adjectives and 3 to indicate the presence of a participle. These are set out below, using Kilby's words and numbering:

- 2.a. Adjectives have comparative and superlative forms
- b. Adjectives can be modified by 'very', 'rather', etc.
- c. Adjectives occur both after a form of 'be' and before a noun within a noun phrase
- d. Adjectives occur as the complement of verbs such as 'prove', 'seem', 'become', 'feel'
- e. Adjectives often have a negative form in 'un-'
- f. Adjectives may be turned into adverbs by the addition of '-ly'

- g. Adjectives may often be turned into nouns by the addition of '-ness' or an equivalent suffix
  - h. Adjectives are conjoinable with other adjectives
- 3.a. Verbs may be modified by 'much' rather than 'very'
  - b. Verbs occur in the context 'NP saw/heard/had/ NP'
  - c. Verbs modifying nouns occur after the noun

The use of modal verbs ('can' and 'may') and 'often' show that Kilby is aware that these criteria apply in different combinations to different adjectives. He makes this point explicitly, and gives examples of non-participial adjectives that clearly fail some of the above tests of adjectival status:

e.g. \*unred, \*bluey, \*ruralness, \*very absolute

After discussing each of these criteria in turn, Kilby's next step is to pick, arbitrarily, 26 participles and mark them in a table as being positive (+), negative (-) or doubtful (?) for each of his 11 criteria. The resulting table successfully shows that the division between adjective-like and verb-like participles is not at all clear-cut. Table 1.17 gives two examples from Kilby's table:

Table 1.17 The adjective/participle borderline (from Kilby 1984)

	2a	2b	2c	2d	2e	2f	2g	2h	3a	3b
allowed	-	-	-	-	-	-	-	-	-	-
finished	-	-	+	+	+	-	-	+	?	+

Kilby ends his investigation of this "messy" area of

description by showing that English speakers disagree quite widely over the marking of his criteria. He presented 13 linguistics students with a questionnaire asking them to mark each of the 26 participles according to whether they could occur in three frames:

that seems ----  
 it became ----  
 it proved ----

He presents the results for the first frame and concludes that, "any attempt to read some neat hierarchical order into the results seems doomed to failure." (p94) He discusses the difficulties of methodology, in particular the amount of data needed to discover any possible sociolinguistic variation in the use of such structures:

"it seems to me that we need to reconcile ourselves to the possibility that certain phenomena simply yield no highly recognisable pattern when the judgements of different speakers are sampled." (p97)

Since Kilby did not appear to use any naturally occurring data for his work on participles, I decided that a useful first step in my investigation would be to try out his criteria on my data. I extracted from the pilot data all examples of -en forms occurring alone as verb phrases, all examples of -en forms following BE which had been classed as adjectives in the coding stage and all the cases where an -en form had been classed as verbal, but was marked as doubtful (?) in the database. These three groups form the basis of the following discussion.

It seemed as though most of the examples occurring in non-

finite clauses were near the verbal end of the classificatory scale (e.g. 'given' [192/G/19] in Table 1.18). The main exception was 'committed' [53/Te/19] which had most of the adjectival features and neither of the verbal ones<sup>6</sup>:

Table 1.18 -en forms from pilot data tested by Kilby's criteria

	2a	2b	2c	2d	2e	2f	2g	2h	3a	3b
given	-	-	-	-	-	-	-	-	-	+
committed	+	+	+	+	+	-	?	+	-	-
covered	?	-	+	?	+	-	-	+	?	+

Those words, like 'covered' [212/Te/19], which show features of both classes, or like Kilby's 'allowed' which show features of neither, cause the greatest problems for a classification system of this kind.

Apart from the non-finite forms mentioned above, there were two categories of -en forms following some form of BE in the pilot data. The first category contains those forms which I classified as verbal, but with a marker of doubt (?) in field 16 to show that I was unsure about their verbal status. The second category contains the -en forms recovered from the BE copula enquiry (see section 1.3.8). These forms had initially been classified as adjectival. Table 1.19 shows some of the pilot data from these categories tested against Kilby's criteria:

---

<sup>6</sup> Although Kilby mentions three features of participles, he only includes two in his table.



'contained could mean; a. be inside something  
or b. controlled, restrained

For these two examples, the relevant sense in its original context is sense a, but it would be very easy when classifying the forms out of context, to mistake the senses and fill in a '+' for tests 2a, b, c, d, and h. If informants were unaware of the danger, mistakes would be even more likely since they could invent contexts such as:

it is very dated	he is very contained
a dated expression	contained violence
it seems dated	he seems contained
dated and obsolete	angry but contained

The abundance of multiply ambiguous word forms in English is just one of the reasons why it is very difficult to classify words in isolation using sets of criteria. An improvement on Kilby's process of classification would be to try to identify which sense (whether polysemous or homonymous) of a word-form is being discussed. One could argue that it would be even better to classify words only when the context (or a number of contexts) has been considered. The effect of context will be discussed below and in Chapter 4.

After testing the pilot data against Kilby's criteria, I noticed that there was a strong correlation between the intuitive groupings taken from my database, and the marking of the criteria. For example, almost all of the word-forms which had been classed as adjectival in the database pass most of the adjective tests 2a-2h.

Whilst agreeing with Kilby that we should not expect sets of criteria, used with words in isolation, to provide an absolute division between adjective and participle classes, unlike Kilby I do not wish to leave the burden of explanation to sociolinguistic variation alone. There seemed to be some consistency in the intuitive analysis and this sense of confidence may be a result of looking at the words in context. I therefore examined both the immediately preceding and immediately following context for all examples from the pilot data.

With only one exception, the non-finite -en forms occur in postmodifiers to a noun and therefore appear to belong more to the verb than to the adjective class:

e.g. the total amount paid to a person for living  
 expenses [90 G 19]

Although many adjectives can occur in this position (e.g. 'president elect'), it is often stylistically marked for them. As Kilby points out (p89), a postmodifying position is less restricted for participles. The exception was 'served', which seemed to be functioning as a complement and was therefore more adjectival than verbal. It was also premodified by 'best' which would usually follow, rather than precede, a verbal participle. Many of the non-finite examples also had following prepositional phrases which clearly belonged to one of the major adverbial classes:

e.g. given to the select committee (A place)

The function, rather than the form of these prepositional phrases suggests that the -en forms are verbal and not adjectives postmodified by prepositional phrases.

The -en forms occurring in finite clauses were considered in two groups; those initially coded as participles (verbal) and those initially coded as adjectives. With one exception, the verbal group directly followed some form of the verb BE:

e.g. is made up of women

Such a context does not, in itself, indicate that these examples are verbal since adjectives can clearly also occur in this context. However, a smaller proportion of the adjectival examples actually occur immediately after BE in the pilot data. The only exception, 'controlled', has an adverb of manner, 'rigidly', before it. This supports the analysis of 'controlled' as verbal, since such adverbs do not normally premodify adjectives.

The preceding context of the second, adjectival, group shows 4 cases where a 'normal' adjective premodifier precedes the -en form. These are, 'excited' and 'stunned' which occur with an intensifier, 'so'; and 'experienced' and 'shocked' which both occur in comparative constructions:

e.g. at least as experienced in the affairs of Unesco  
the most shocked

Of the other examples, 3 occur immediately after the verb BE which was noted above as a common preceding context for verbal -en forms, but which cannot be used as conclusive evidence in individual cases. The final example is 'well-advised' which, as a compound not taking verbal inflections ("he \*well-advised me"), must join the adjective class.

The following contexts of the two groups of -en forms show no obvious differences. Both groups display a similar variety of following contexts as can be seen in Table 1.20:

Table 1.20 Distribution of following contexts of -en forms

	verbal	adjectival
non-finite clauses	3	3
finite clauses	1	1
prepositional phrases	6	2
no following context	0	2

If we compare the following contexts for both groups with those for 'true' (i.e. non-participial) adjectives, we find that the same range appears:

Table 1.21 Following contexts of 'true' adjectives

non-finite clause	e.g. necessary to point out
finite clause	e.g. sure that ...
prepositional phrase	e.g. most loyal to the government
no following context	e.g. even worse
comparative phrase	e.g. as true today as it was then

The following contexts for the 'true'<sup>7</sup> passives, on the other hand, show a wider range including complements, objects and adverbials:

Table 1.22 Following contexts of passive -en forms

non-finite clause	7
finite clause	1
prepositional phrase	11
no context	15
comparative	3
adverbial	26
complement	2
object	1

Since the passive examples all followed the verb BE, there can be no useful analysis of preceding context. The preceding context of the 'true' adjective group, however, included 7 examples of intensification and two comparative constructions.

The study reported in this section was intended to be exploratory, but some conclusions can be drawn from even this early stage of investigation. Consideration of preceding and following contexts indicates that context plays a large part in the interpretation of -en forms as either verbal participles or as participial adjectives. Unambiguous signs of an adjective seem to be the presence of an intensifying adverb or a comparative construction. Other following contexts are ambiguous with respect to the class of the -en form. The most common sign of a passive

---

<sup>7</sup> In this context, 'true' passives consist of those examples over which there was no hesitation in assigning them the label 'passive' during the coding phase of the pilot study.

participle in this data was the presence of a following adverbial which was distinguished from other prepositional phrases only semantically; by being a clear member of a major adverbial group such as 'place', 'manner', 'time'. Following contexts functioning as complements or objects were also a clear sign of the verbal status of an -en form. Cases where the -en form immediately follows BE were also generally interpreted as verbal. The examples where verbal status was unsure seem to be those cases which have no clear indication of adjectival status, such as intensification, and whose following context is of a type occurring after either -en form; non-finite clauses, finite clauses, prepositional phrases or no context.

Chapter 4 takes up the topic of this section and investigates the main database for further evidence of contextual involvement in distinguishing between participial adjectives and passive participles. The interaction between syntactic and semantic features is particularly considered.

#### 1.3.10 Verb phrases in sequence

Palmer (1974) used the term 'catenative' for those verbs which, whilst being closely connected to the following verb, showed few of the well-known characteristics of auxiliaries: their use in negation, inversion with the subject, avoidance of repetition (code) and emphatic affirmation. For example, the verb 'continue' dictates the

form of its following verb, 'defend' in this extract:

e.g. he continued to defend the horrors of  
vivisection [155 G 19]

To this extent it seems to behave like the modals and auxiliaries which also impose form on following verbs. However, it cannot, in modern English at least, be exchanged with the subject to form an acceptable question:

e.g. \*continued he to defend the horrors of  
vivisection? [rewritten 155 G 19]

In his classification of these 'catenative' verbs, Palmer produces eight classes which have semantic headings although he claims that they are based on formal similarities between the members of each class. The classes and their 'formal' features are summarised here:

futurity	possible occurrence of adverbials of future time
reporting	alternative 'that' construction possible passivisation of main <u>and</u> subordinate clause
perception	contrast of bare infinitive and Ving marks aspect
attitude	most object-like construction possessive alternative
process	heterogeneous class Ving forms not belonging to other classes
effort	Vi forms not handled elsewhere
causative	small set of common verbs
needing	marginal catenative

It is clear from this summary that only the first four of these classes can be called 'formal', and within these classes there is some variation. For all except the last two classes, the semantic similarities seem to be more convincing than the suggested 'formal' similarities.

The original reason for investigating this feature of the pilot data was in order to decide whether each sequence of verb phrases should be counted as a single, complex verb phrase or as two or more verb phrases. Quirk et al. (1985) go into the question of the independence of such verb phrases in some detail. They perceive a scale between the clearest auxiliary verbs and the clearest main verbs, postulating four intermediate groups between these extremes: marginal modals, modal idioms, semi-auxiliaries and catenatives. The basis of their analysis will be discussed in chapter 5. On the question of the number of verb phrases in any text, Quirk et al. (1985) claim that one cannot resolve the question of whether any string of verbs should be analysed as one verb phrase or as two:

"There is no clear answer to this question, since the gradient relating auxiliary to main verb functions implies a comparable gradient between a single verb phrase analysis and a multiple verb phrase analysis."

[CGEL 3.57 p154]

Returning to Palmer's classification, the problems noted above showed that this type of verb phrase sequence needed detailed research. It would not have mattered that his classification was not neat or elegant if his classes had some intuitive appeal or enlightening potential. Since Palmer did not use any naturally occurring data, it may be

that he overlooked some aspects of sequential verbs<sup>8</sup>. The present section, and Chapter 5, provide an opportunity to consider some genuine examples of verb phrase sequences and seek patterns of behaviour in this data.

There were 61 verb phrase sequences, 56 consisting of 2 verb phrases (as defined in section 1.2.1), 4 consisting of 3 verb phrases and 1 consisting of 4 verb phrases. At this point, Palmer's method of considering only sequences of two verb phrases was followed, but the longer examples were included by dividing them into sequences of two verb phrases. The 67 sequences thus produced can be grouped according to their meaning into the following categories:

permission	e.g. to <u>allow</u> sb. to keep, overtake, increase etc.
persuasion	e.g. <u>influencing</u> sb. not to take up smoking
forcing	e.g. <u>compels</u> me to advise of ...
process	e.g. <u>started</u> barking
ability	e.g. should <u>seek</u> to brush
appearance	e.g. <u>appear</u> to want, represent etc.
intention	e.g. <u>are prepared</u> to accept
movement	e.g. <u>went</u> to get
unclassified	e.g. <u>designed</u> to give, <u>forgot</u> to hold

Some of the unclassified examples would presumably form separate groups in a larger sample. Even in a small database like this, some of the semantic groupings show formal similarities. The first three groups, for example, all contain a noun phrase between the verb phrases. The other groups had only a single example of such a

---

<sup>8</sup> I am using the term 'sequential' in preference to Palmer's 'catenative' as a way of distancing myself from his approach. The term also implies the possibility of ordering priorities in long sequences; this implication seems to me to be absent from Palmer's term.

structure.

The complete database entries for each of the verb phrases in sequence were scrutinised to see whether there were any features among those coded which characterise sequential verb phrases as a group.

The first field investigated was 'subject type' (field 11) and it was immediately clear that a large majority of subjects (89 out of 112) occurring with verb phrase sequences were human. The figures are given in Table 1.23 with the proportions of each subject type in the total pilot data as a comparison:

Table 1.23 Subject types occurring with VP sequences

	VP seq (% of 112)	total data (% of 800)
human	79	54
abstract	12	35
animate	2	1
inanimate	1	5
not obvious	6	5

The proportions of human and abstract subject types occurring with verb phrases in sequence are clearly not normal when compared with the overall figures. The real number of human subjects may be even higher since some of the 'not obvious' subjects, though their referents are unidentifiable, are clearly human:

e.g. is to allow all road users to keep much the same  
speed [104,5/Te/19]

The subject of 'allow' here is presumably human, though it has a vague reference in the context.

The second field investigated for signs of correlation with verb phrase sequences was clause type (field 8). The verb phrases were divided into two groups; those occurring first in a sequence and those occurring second. The figures for these groups, and for the whole database, are shown in Table 1.24:

Table 1.24 VPs in sequences and their clause types

	1st VPs (% of 52)	2nd VPs (% of 60)	total VPs (% of 800)
main clause	35	0	40
noun clause	17	41	15
relative clause	23	3	19
adverbial clause	18	5	16
doubtful cases	6	50	10

The two groups of sequential verb phrases have very different distributions of clause types and should clearly be treated separately from this angle. As expected, the figures for the first group resemble those for the overall data, showing that they are 'normal' with respect to this feature. Although not tested statistically at this stage, the percentages are roughly comparable. Another general feature that supports this view that 1st verb phrases are 'normal' is the proportion of finite to non-finite verb phrases in this group. The percentages are shown in Table 1.25:

Table 1.25 Percentage of finite and non-finite VPs in sequences

	1st VPs (% of 52)	total VPs (% of 800)
finite	79	73
non-finite	21	27

The second verb phrases in two VP sequences, however, show a very different pattern, not only in being 100% non-finite, which is a well-known feature of such sequences, but also in having a preponderance of NCl clause types (41%) and of doubtful clause types (50%). The size of this group indicates a dissatisfaction with the practice of labelling all such clauses 'nominal' on the grounds that some of them are similar to ordinary NP objects. This problem and the question of whether second (and subsequent) verb phrases should anyway invariably be considered to be separate clauses is addressed in Chapter 5.

The final characteristic of verb phrase sequences to be discussed in this section is the ordering of verb phrases in sequences containing more than two verb phrases. There are only 5 such sequences in the pilot data, but they were investigated on the assumption that Palmer (1974) was wrong to disregard them:

"For simplicity we shall deal only with complex phrases involving two verb forms (...) In longer complex phrases there would be successive 'layers' of subordination, but the grammatical relations between each pair are the same." (p167)

Many African languages and English-related pidgins and creoles have verb sequences which are comparable to these English ones, and there are often semantic tendencies to be observed in the ordering of the verbs. It may be the case then, that there are some ordering priorities among the sequential verbs in English which Palmer's method of considering only two verb phrase sequences, as well as using only introspective evidence would miss entirely. The 5 longer sequences in the pilot data are presented here in full with the 'main' verb of each verb phrase underlined:

- i. we want to stop tobacco companies from sponsoring sport
- ii. in persuading young people to start smoking
- iii. it would be designed to allow the working miner to substantially increase his take-home pay
- iv. the BBC producer (...) asked me not to give my own views but to summarise the issues and let the audience make their judgement
- v. you now appear to want the NUM/NCB negotiations to resume to negotiate for a settlement

Although there are not enough examples here to draw any firm conclusions, it was considered worthwhile to pursue this investigation in the main part of the thesis to see, for example, whether 'intention' verbs like 'want' tend to precede 'process' verbs like 'resume' or 'stop'. This was foreseen, not as a grammatical ordering rule, but as a tendency to order sequential verbs in particular ways which reflect socio-cultural requirements or restrictions on the kinds of statements which are made.

## 1.4 Literature

### 1.4.1 Corpus linguistics

Modern linguistics, in its efforts to reject traditional, subjective thinking and to emerge as a 'science', has often (though not invariably) laid emphasis on the importance of using attested data rather than intuitive and introspective data for analysis. There have, therefore, been corpus-based studies since the days before computers were commonplace tools of the discipline.

Since computer storage and retrieval have made the task of studying corpora easier, studies of computer-based corpora have been more common. Three large projects based on computer corpora are particularly worthy of mention.

First, the Survey of English Usage, based at University College London, has produced a variety of grammars which are partly based on corpus research. These include Quirk et al (1972), Quirk and Greenbaum (1973), Leech and Svartvik (1975) and Quirk et al (1985). The main aim of the project was to provide a comprehensive description of the English language and the use of corpora was accepted as a natural part of this process. The published grammars themselves, therefore, do not contain lengthy descriptions of the corpora, or their method of collection and exploitation.

Other works based on the Survey of English Usage, however,

specifically mention the corpus-based origin of their data. Tottie (1982), for example, reports a study of the types and extent of negation in texts from the Survey.

The second project to be mentioned in connection with computer corpora is the project which set up the Brown corpus of American English. This corpus, containing a million words of the written language of 1961 is significant in that it attempted to include a balance of written styles so that it was representative of the written language in America at that time (see Francis 1964 for details of its contents). The third project was set up in response to the Brown corpus. The LOB corpus of British English was the result of a collaboration between the Universities of Lancaster, Oslo and Bergen and it aimed to match the Brown corpus in everything except the provenance of the material, which was to be British (details of the LOB corpus can be found in Johansson et al 1978). These two corpora provided a unique opportunity to study aspects of the grammar and lexis of American and British written English and were established as a general resource for linguistic research.

One further development of the Brown and LOB corpora which is very useful for grammatical research is the grammatical tagging of each word. The Brown corpus has a tagged version produced at Brown University and there is also a tagged version of the LOB corpus (details of the latter can be seen in Johansson et al. 1986). More sophisticated

analysis of the corpora has also been undertaken. This high-level parsing was applied to a small part of the Brown corpus in a project based at Gothenburg (for details, see Ellegård 1978) and is currently being turned into "a more accessible and useful research resource" by the SUSANNE project based at Leeds University (details in Sampson (forthcoming)). Two 'analysed' (parsed) versions of the LOB corpus are currently available, and known as the 'Lancaster-Leeds Treebank' and the 'Parsed LOB Corpus' respectively. Sampson (forthcoming) describes a number of other analysed corpora, most of which are derived from the Brown and LOB corpora.

Since the Brown and LOB corpora have been widely available there have been many studies on subsets of the data. Sometimes these are focussed on linguistic features or areas. For example, Rissanen (1980) examines the placement of 'only' in examples from the Brown corpus and Johansson (1986) investigates the order of adverbial particles and objects in the LOB corpus. Other studies may combine a grammatical focus with a contrastive study of registers or 'genres'. Gustafsson's study of the Brown corpus (1982), for example, focusses on the fronting of adverbials, but also compares four 'genres': journalism, science, popular fiction and literary essays.

Many other works (notably unpublished PhD theses) are also based on corpora which differ from the Brown and LOB corpora in being collected for a specific study with well-

defined focusses. Examples of this type of study are Gustafsson (1975), which describes 'Syntactic Properties of English Law Language' and Varantola (1984) which has both a grammatical focus (noun phrase structure) and is based on a specific variety of English: Engineering English. Huddleston (1971) also bases his wide-ranging syntactic study on a corpus which is composed of scientific texts in English.

A small number of studies have shared the 'register focus' of the present study; newspaper English. Most of these (e.g. Straumann 1935, Mårdh 1980) have been concerned with the linguistic features of headlines, but a few have taken other parts of newspapers as their data: Rydén (1975), for example, takes the linguistic focus of 'noun-name collocations' and Fries (1987) is concerned with the relationship between full reports and front page summaries.

There is, as far as I know, no work closely resembling the present study which is based on a corpus of texts contrasting two dimensions of newspaper language: text-type and newspaper-type.

#### 1.4.2 Syntax and semantics

Although the focus of the present study is syntactic, it will already have become clear that I do not therefore consider that semantic matters must be excluded. In fact, there are a number of problems raised in this thesis which

are made simpler by the consideration of semantic features alongside syntactic ones.

Modern linguistics made early efforts to reject semantic definitions of grammatical categories in order to establish a more scientific and rigorous discipline in which the results of analysis could be duplicated by any researcher using the same material and methods. Palmer (1976) argues for 'formal grammar' in which semantic information is not used as the basis of categorisation:

"If, then, the grammatical categories are given semantic definitions, the definitions are circular. An excellent example is the definition so often found in grammar books of a noun as 'a word used for naming anything'. The difficulty is that we have no way of establishing what 'anything' may be." (p118)

Palmer's second argument is that grammatical and semantic categories do not often coincide. Leech (1981), however, in Jacobson's (1986) words:

"proceeds to make some laudatory attempts to outline certain general processes that relate syntactic and semantic structures" (p7)

The debate about the autonomy of syntax and semantics was also continued by those encountering difficulties in putting transformational-generative grammar (TG) into practice, notably Fillmore (1968) and McCawley (1968). For both of these writers, the level of 'deep structure' had to be semantic as well as syntactic.

However, the debate about the relationship between these

two 'levels' of language is no longer a simple opposition between those who exclude semantic information from grammatical analysis and those who do not.

Like many writers responding to Chomsky's 'Aspects of the Theory of Syntax' (1965), Liefcrink (1973) was concerned to reintroduce semantics into structural analysis. He proposed that deep structure consists of much more abstract categories than were put forward by Chomsky, and that these categories were both syntactic and semantic rather than syntactic with semantic interpretation rules as in the standard TG theory. Liefcrink's approach is based upon the premise that words do not (and cannot) have meaning independently of the sentences in which they occur and his 'semantico-syntactic categories' are therefore always associated with the sentence. Examples of his categories are: declarative, causative, locative, attributive.

In contrast to Liefcrink's approach, Miller (1985) argues strongly that though intimately connected via a set of rules, syntax and semantics should not be combined into "one large monolithic system". The reasons for this approach are partly based on the failure of previous attempts to combine the two 'levels':

"What has happened hitherto (....) is that the syntactic analysis suffers" (p193)

Miller's work is basically generative, but as in Lyons

(1966) the notional definitions of parts of speech found in traditional grammar are partly reinstated. Thus it is argued that since many of the forms occurring in noun positions denote concrete objects, other forms whose reference is not essentially object-like are seen as physical entities when they occur in noun positions. A similar argument is used by Miller in discussing syntax:

"the facts of syntax and morphology are not to be regarded as accidental but as systematic and indicative of semantic structure" (p211)

This part of Miller's argument suggests that when semantic information is taken into account, apparent exceptions may be seen to 'fit' the rule after all. A similar idea is found in Green (1974) who also explores the relationship between syntax and semantics from a TG standpoint. She states that her original purpose was:

"to reanalyze several syntactic phenomena thought to be exceptional, in order to demonstrate the value of assuming that there are reasons for the behaviour of most items believed to be exceptions." (p10)

This is echoed by the sentiment indicated in Quirk (1984) when he expresses:

"our adherence to the postulate - advanced above all by Bolinger - that there is no linguistic variation without linguistic reason" (p186)

Green proposes that syntactic rules which appear to be irregular, exceptional or idiosyncratic are actually as regular as other rules when semantic composition is the

basis of their operation. Green tests her position by discussing the problems caused to generative grammar by 'dative movement' and claims to have discovered not a single rule but several independent alternations which were nevertheless similar to each other.

Two articles which explore specific parts of the relationship between syntax and semantics are Declerck (1986) and Schachter (1977). Declerck argues that there is a difference between formal definiteness and semantic definiteness and that linguistic studies often fail to take this into account. This position could be extended to many grammatical areas; one of the problems of coding discussed in section 1.3.5 arose because of the difference between formal and semantic plurality, animateness and abstractness. The coding was set up to recognise formal categories, but it soon became clear that the semantic categories were often more revealing.

Schachter's article discusses a division of the function of the grammatical relation of 'subject' into 'reference-related' and 'role-related' properties. His argument is based upon the division of these functions in various Philippine languages, Tagalog in particular. These languages split the syntactic properties of the subject between two clause elements: 'topic' and 'actor'. The topic has features such as the ability to be postmodified by a relative clause which is a feature primarily carried by the subject in English, while the actor has in common

with English subjects features such as the tendency to occur first in a clause (although word order is not always fixed in Tagalog). 'Topic' and 'actor' are much more semantically based notions than 'subject' and to this extent Schachter is arguing for more semantically based grammar.

#### 1.4.3 Categorical Indeterminacy

The descriptive problems discussed in chapters 3 and 4 both concern the nature of indeterminacy in an area of syntax and semantics. Chapter 3 is mainly concerned with the boundary between subject and complement and Chapter 4 with that between adjective and verb (-en participle). There has been increasing recognition, in recent linguistic work, that categorial indeterminacy should be incorporated as part of any description of English. Quirk et al. (1985), for example, acknowledge the existence of 'gradience'<sup>9</sup> between categories and they define it in the following way:

"A gradient is a scale which relates two categories of description (for example two word classes) in terms of degrees of similarity and contrast." (p90)

The resulting description includes gradience as one of its basic descriptive tools.

More extensive investigations of the nature of

---

<sup>9</sup> See Bolinger (1961) for a full description of gradience

categorisation by human beings have been carried out by Rosch (1978), Coates (1983) and Lakoff (1987). Rosch proposes that people use categories which are not full of equally 'good' members, but that some members are better examples of the category than others. The better examples are known as 'prototypes' of the category, although Rosch is careful to point out that what is really being referred to are "judgements of degree of prototypicality" (Rosch 1978:40). Prototypicality may constitute a gradation within a category where worse members are nevertheless still members of that category. Alternatively, the degrees of prototypicality may shade evenly across the boundary between two adjacent categories and constitute the kind of gradience usually referred to by linguists.<sup>10</sup>

Lakoff (1987) reviews the work of Rosch and others who have contributed to the demise of "the classical view that categories are based on shared properties" (p.5), replacing it with a new theory of categorisation based on prototypes.

As well as exploring the principles upon which human categorisation is based, Lakoff makes three case studies of examples of human categories; two of these are linguistic examples. The longest case study concerns existential and deictic there-constructions and aims to demonstrate that syntactic constructions occur in radial

---

<sup>10</sup> See Lakoff's discussion of Rosch's work (1987:45).

categories which have a central group of members (prototypes) and other, non-central, groups which are related to the central group.

Coates (1983) draws on the work of Zadeh to incorporate his "fuzzy sets theory" into the description of linguistic categories. In particular Coates is concerned to describe the different meanings of the modal auxiliaries and their relationship to each other. Her dissatisfaction with the model of gradience usually employed by linguists is that her data did not show two clear extremes covering most of the examples with a small number of examples spread along the cline between the extremes. Instead, Coates found only one clear extreme and most of her examples clustered between the end-points of the cline. Her version of a fuzzy set consists of a 'core' (the modal meaning learnt first by children / the stereotypical meaning), the 'skirt' and the 'periphery': "The latter often has the qualities of an emergent category (...) because it is often possible to define peripheral examples by contrast with the core." (Coates 1983:13). Coates' description of a fuzzy set as applied to linguistics has many similarities with the cognitive categories described by Lakoff (1987), although their aims, and therefore their emphases differ. Coates distinguishes three types of indeterminacy in her data, based on her initial description of a fuzzy set. First she describes 'gradience' as the graded membership of a category.

Secondly she identifies ambiguity as an inability to decide to which of two discrete categories an example belongs. Finally, she uses the term 'merger' to describe examples where there are two possible categories and neither is excluded by the context. In these cases the distinction between categories is neutralised and the example can be understood without any need for a decision as to category membership. An example of 'merger' from Coates (1983:17):

" A: Newcastle Brown is a jolly good beer

B: Is it?

A: Well it ought to be at that price.

Here it is not clear whether the speaker is referring to the maker's obligation to provide good beer (Root OUGHT), or whether he is making a logical assumption - 'it costs a lot, therefore it is good' (Epistemic OUGHT)>"

The notions of merger and prototypicality will be used later in this thesis (sections 3.3 and 4.5) to describe some of the problem examples investigated.

#### 1.4.4 Borderlines: subject/complement

One of the 'borderlines' chosen for study in the present work is that between the grammatical relations of subject and (nominal) subject complement. The possibility of any overlap between these clause functions has not arisen, as far as I am aware, in the main descriptive and theoretical work on English. One of the reasons is that the term 'subject complement' is used to refer to the complements (both adjectival and nominal) of clauses containing the copula 'BE'. Many theories and descriptions based on

English would consider this to be a sub-category of a more general 'complement' function (often including what I would call 'object') and the possibility of this larger category having an indistinct border with the subject category does not arise. Another reason is that the 'overlap' which I suggest exists between these functions is only evident in a small number of individual examples. Any work based on intuitive data, or generalising away from attested data would not necessarily notice this problem.

One work that has some relevance for this topic is Huddleston (1971). In his book, Huddleston makes a subcategorisation of the uses of main verb 'BE' into intensive, extensive intransitive (with following adjuncts) and extensive transitive. The last category, which he labels 'equative', contains examples where the verb has the effect of identifying two referents with each other:

e.g. the result was a failure

This example from Halliday (1967) is supposed to be ambiguous between an intensive interpretation (the result was unsuccessful) and an equative interpretation (a failure resulted). Huddleston claims that the grammatical difference between the two is that equative uses of the verb allow the subject and complement (which he labels 'object') to be reversed. The 'reversability' of a clause

is also used in the present study to test the proximity of the subject and complement functions in examples where their forms do not differ in definiteness and complexity.

#### 1.4.5 Borderlines: adjective/verb

As already mentioned in section 1.3.9, the problem of establishing whether -en forms (and -ing forms) are adjectival or verbal has not been widely treated in the literature, although it is often mentioned. In his rather short chapter on the passive voice, Palmer (1974) discusses what he calls the 'statal passives', but dismisses the problem by claiming that the status of Ven forms acting as adjectives "is in no real doubt". His arguments for their adjectival nature are:

- i. that they occur with 'already' and a present tense verb
- ii. that they may be coordinated with adjectives
- iii. that they occur with verbs like 'look' or 'stay'

These criteria are subject to the same kind of criticism as those used by Kilby (1984), namely that it is often difficult to make decisions on acceptability in such tests. Tense, as implied in test i, however, seems to have some relevance for interpretation in individual cases (see chapter 4). In chapter 4 I also argue that the attempts to establish the adjectival or verbal status of an -en form permanently<sup>11</sup> may be misplaced in a grammatical

---

<sup>11</sup> The use of the word 'permanently' is not intended to imply that Palmer (and others mentioned below) do not recognise that languages change through time. It is used to contrast with my argument (in chapter 4) that word-senses may be temporary

description.

Kilby, whose work was discussed in section 1.3.9 above, is mainly concerned to illustrate that a 'battery' of tests such as those proposed by Palmer is inadequate to make the borderline between participle and adjective clear. He provides no alternative procedures, but suggests that sociolinguistic variation, which (he explains) would be extremely difficult to test, is responsible for the 'messiness' of the data.

Quirk et al (1972) discuss the 'clear' cases of participial adjectives including those where there is no corresponding verb (e.g. 'unexpected', 'talented'), those where there is no predicative use (e.g. 'escaped', 'departed') and those where the corresponding verb has a different meaning (e.g. 'calculating', 'relieved'). They also mention the borderline cases and, like Palmer, are concerned with the 'permanent' categorisation of word-forms into word-classes:

"For both participle forms, modification by the intensifier 'very' is an explicit indication that the forms have achieved adjective status" (p.244)

Another point they make is that the occurrence of an object (with -ing forms) or an agentive by-phrase (with -en forms) makes explicit the verbal force of the participle. However, they also show that it is becoming

---

members of a particular word-class in individual contexts.

increasingly acceptable for -en forms to be premodified by 'very' as well as being followed by an agent:

"e.g.       ?The man was very offended by the  
              policeman"

Other writers (e.g. Johansson et al 1986 and Sampson 1985) are concerned with this borderline for the practical purpose of tagging large corpora of data. Johansson et al, for example, explain their procedures for assigning verbal and adjectival tags to -en forms, but the practical aim of constructing coding guidelines which are reasonably straightforward for a number of researchers to apply creates some anomalies that are discussed in chapter 4.

A special group of -en forms are discussed by Siegel (1973) who is not as certain as Quirk et al. that what she calls the 'unpassives' (e.g. 'unexpected', 'unemployed') are clearly adjectival. There is a difference between these forms and the other examples that Quirk et al. cite as having no corresponding verb. The unpassives do not have corresponding verbs with the -un prefix: \*uneducate, \*unemploy, \*unexpected. However, they all have corresponding positive verbs (educate, employ, expect) unlike examples such as 'talented' or 'downhearted' (from Quirk et al.) which have no corresponding verb at all. Siegel is attempting to fit the unpassives into a TG framework and concludes that they are passives in deep structure and have the un- prefix added after the participle has been relabelled 'adjective'. For the purposes of the present

study, even Siegel's conclusions support the interpretation of the unpassives as surface adjectives.

#### 1.4.6 Borderlines: full/auxiliary verbs

The status of verb phrases following what I have called 'sequential' verbs has been discussed at length and from a number of different points of view.

Quirk et al. (1972) classify all of the verb phrase sequences as different types of verb complementation. They make formal subcategorisations according to the form of the second verb phrase (to- infinitive, bare infinitive, -ing participle or -en participle) and according to whether there is a subject which intervenes between the two verb phrases. In the more recent grammar from the Survey of English Usage, Quirk et al. (1985) make a similar classification, although some of the types of complement are shown to occur with lists of verbs analysed into unlabelled semantic classes.

In Palmer's treatment (1974) of these examples, he sets up a separate class of verbs which he calls 'catenatives' (after Twaddell 1965). His approach has been examined in section 1.3.10, but it should be noted here that his concern is with the non-final verb phrases in sequences, whereas Quirk et al. were primarily concerned with non-initial verb phrases. The emphasis, therefore is on the class of verbs occurring with following non-finite clauses in Palmer, but with the non-finite clauses themselves in

Quirk et al.

Quirk (1965) is, however, concerned with a class of verbs similar to Palmer's catenatives when he uses them to illustrate their gradience in terms of the constructions in which they may appear. He uses seven frames such as 'He X to come every day' and 'He would X to come every day' to show that if a verb occurs in certain frames, this fact presupposes its occurrence in other frames. Quirk goes on to illustrate his notion that some sub-classes of words are in 'serial relationship'; they have a kind of 'rolling' overlap between them in terms of the types of context in which they occur. This idea is followed through in an extensive corpus-based study of verb complementation by Andersson (1985) who analyses the data into 20 groups, defined partly syntactically and partly semantically. Andersson's groups form a circular overlapping set and seem neatly to share some of the features of neighbouring groups.

Matthews (1981) is concerned with various grammatical problems including the status of non-finite clauses following sequential verbs and postulates the existence of a number of 'fused constructions' to explain their structure. Matthews suggests that just as linguists accept that categories have both clear, central members and peripheral, unclear members, "it is not surprising that there should also be cases of marginal subordination, in which it is undecidable whether a smaller clause is



verb phrases are construed as a 'double-barrelled' unit (e.g. [kept visiting]) or as two separate units (e.g. [kept] [visiting]). Matthews' argument is that since no single tree-diagram can capture the whole set of dependency relations such bracketing will inevitably be incomplete.

The distinction between analysing the second verb phrase in a sequence as a verb complement or as a sentential complement is not taken so lightly by Borsley (1984). He uses evidence from Welsh to argue that the 'control' verbs (e.g. 'try') and the 'raising' verbs (e.g. 'seem') which are treated as taking a sentential complement in the extended standard theory are better analysed as taking a verb phrase complement in accordance with the Lexical Functional grammar of Bresnan (see Kaplan and Bresnan 1982).

Foley and Olson (1985) make a more fundamental attack on the notion of the clause and argue that it should not be taken for granted that the clause can retain its traditional definition as a single predicate simple sentence. Their examples which cause problems for this definition of the clause all contain serial verbs (similar to sequential verbs) and come from the Kwa languages of West Africa and the Papuan languages of Papua New Guinea. Their description, which is based on a valency treatment of verbs and their contexts, is claimed to provide a unified account of both 'traditional' clauses and clauses

containing serial verb constructions. They conclude:

"Much more detailed work in these languages and others will be necessary to extend the idea of the layered structure of the clause. We hope we have demonstrated its importance in accounting for a large number of diverse facts in several unrelated groups of languages." (p57)

### 1.5 The main database

After the pilot study was finished and its implications for future work considered, I began to set up a larger database to provide the basis for my main study. The framework of the database and the computer software used are the same as for the pilot study. Some details of coding and features were changed; these changes are described in the remainder of section 1.5.

#### 1.5.1 Data collected and coded

The data collected for the main part of this thesis came from eight national daily newspapers published on the five days between 15th and 19th November 1986. The newspapers chosen were: The Times, The Guardian, The Daily Telegraph, The Mail, Today, The Mirror, The Express and The Sun. The choice of papers was intended to represent the range from 'quality' to 'popular' and were considered to be those recognised as belonging to the mainstream.

From each of the 40 newspapers collected, I extracted the letters page, the editorial and one news article. The news articles were intended to be on the same subject for any one day, but on two days there were no suitable articles

appearing in all eight papers. For these days I had to choose articles which were similar in subject matter.

The data extracted from the newspapers were too extensive for the detailed coding which I wanted to adopt for the computerised corpus of material. I therefore coded the first 100 verb phrases/clauses from each of the categories. The main computer database, then, consists of 2400 data items, identified as verb phrases but coded for contextual features relating to clause structure, clause function and subject. These 2400 items are distributed as shown in Table 1.26:

Table 1.26 Distribution of data in the main database

	article	editorial	letters	tot.
Times	100	100	100	300
Guardian	100	100	100	300
Telegraph	100	100	100	300
Mail	100	100	100	300
Today	100	100	100	300
Mirror	100	100	100	300
Express	100	100	100	300
Sun	100	100	100	300
total	800	800	800	2400

The reason for this carefully constructed and balanced corpus was that I wanted to use the detailed coding of the corpus in a number of different ways. First, it was intended to represent newspaper English generally. Although some rather specialised types of excerpt were not included (e.g. sports and weather reports), I felt that there was some justification in describing the data

stylistically as a whole so that it could be compared with other styles in the future. The second way in which the data was to be used was to describe the different styles of two cross-cutting sets of sub-databases: text-type (article, editorial and letters) and newspaper-type (Times to Sun). Both of these stylistic aims are fulfilled in chapter 2. Finally, the data was used as a corpus to supply examples of descriptive problems which had arisen during the pilot study and seemed to be of relevance to the description of English generally. The result of these investigations can be seen in chapters 3, 4 and 5.

#### 1.5.2 Features coded

The features of each verb phrase to be coded for the main database were based on those used in the pilot study. Some changes were made, both in which features were coded and in how the codes were applied. The 21 features coded in the main database are described below. Where there were changes made on the basis of the pilot study, these changes are indicated.

1. VP This field contains the verb phrase in exactly the form in which it occurs in the original context.

e.g. don't need [33 S Let]

2. Verb This field contains the citation form (bare infinitive) of the main verb

e.g. need [33 S Let]

3. **Sense** In the pilot study there was a feeling that it would be useful, in some cases, to have an indication of which sense of the word was being used in the verb phrase. Field 3 provided this disambiguation, usually in the form of synonyms or near-synonyms:

e.g. invent (verb: make up) [20 Ti Art]

4. **Ref** The reference field in the main database contained the number of the verb phrase (1 to 100), the number of the sentence in which it occurred, the code for the paper and the kind of text it came from:

e.g. [50/22 To Ed] = the 50th verb phrase in the 22nd sentence of Today's editorial.

All references given in the remainder of this thesis are based on this lay-out, although in many cases only the sentence number, paper and text are mentioned.

5. **Finite** Each verb phrase was coded as finite or non-finite according to the form of its first verb.

6. **Polarity** This field identifies a verb phrase as positive (P), negative (N) and/or interrogative (?). The problem of negative force, discussed in connection with the pilot study, was alleviated by the use of a special code (semN) followed by an indication of the word(s) which supply the semantic force:

e.g. semN never [26 Mi Let]  
on condition the Teddies were never taken out of  
the window

Although limits of time and space mean that this topic is not taken up in the present work, I hope to return to these data items in the future.

7. **VP struc** I was unhappy about the accessibility of the information about VP structure in the pilot study since it had been coded in the same field as verb phrase form. I therefore decided to adapt Svartvik's (1966) schema of 16 verb phrase types, adding an extra code for verb phrases containing auxiliary 'do'. The structures are as follows:

- |     |                                 |                               |
|-----|---------------------------------|-------------------------------|
| 1.  | Main verb                       | (Mv)                          |
| 2.  | Modal + main verb               | (Mod + Mv)                    |
| 3.  | Perfective + main verb          | (Have + Mv)                   |
| 4.  | Progressive + main verb         | (BE1 + Mv)                    |
| 5.  | Passive + main verb             | (BE2 + Mv)                    |
| 6.  | Modal + perf + main verb        | (Mod + Have + Mv)             |
| 7.  | Modal + prog + main verb        | (Mod + BE1 + Mv)              |
| 8.  | Modal + pass + main verb        | (Mod + BE2 + Mv)              |
| 9.  | Perf + prog + main verb         | (Have + BE1 + Mv)             |
| 10. | Perf + pass + main verb         | (Have + BE2 + Mv)             |
| 11. | Prog + pass + main verb         | (BE1 + BE2 + Mv)              |
| 12. | Modal + perf + prog + main verb | (Mod + Have + BE1 + Mv)       |
| 13. | Modal + perf + pass + main verb | (Mod + Have + BE2 + Mv)       |
| 14. | Modal + prog + pass + main verb | (Mod + BE1 + BE2 + Mv)        |
| 15. | Perf + prog + pass + main verb  | (Have + BE1 + BE2 + Mv)       |
| 16. | Modal + perf + prog + pass + Mv | (Mod + Have + BE1 + BE2 + Mv) |

8. **VP form** This field coded the form of the initial verb in the verb phrase. The codes used were i, ing and en for the three non-finite forms and s, o and ed for the 3rd person present tense, all other persons (present tense) and past tense respectively. Forms of the verb 'be' not conforming to this pattern were assigned their own codes: am (m), are (r), was (was), were (wer).

9. **Cl type** The clause surrounding the verb phrase was coded according to its type. The categories used were: main, adverbial, nominal, postmodifying, comparative and 'doubtful', the last category being used for examples which were difficult to classify and which might prove interesting at a later stage. The 'doubtful' category contained a majority of examples which were non-initial in verb phrase sequences, but which did not seem to correspond to object noun clauses:

e.g. the nuclear shield began to become the cheaper  
option [14 Te Art]

Other clause types are as described for the pilot data in section 1.3.3.

10. **Cl function** The function of some clause types is stable (e.g. relative clauses always post-modify in a noun phrase), but since there is not a one-to-one relationship between clause function and form as they are usually recognised, I decided to code the function separately. The codes used in this field were:

Subject, Object or Complement (for noun clauses)  
Adverbial (adverbial clauses)  
Noun modifier (relative clauses, non-finite  
postmodifying clauses and appositive noun  
clauses)  
Adjective modifier (adjective complement clauses)  
Verb complement (non-initial verb phrases in  
sequences)

11. **Cl struc** The structure of each clause context was coded in terms of the constituents Subject, Verb, Object, Complement and Adverbial. Optional adverbials were omitted so that basic structures could be compared.

e.g. It is a dunce's dream      SVC      [10 S Ed]

12. **Subj type** The subject of each clause context was characterised in the same way as in the pilot study. The four subfields showed whether the subject was:

animate, abstract, inanimate or human  
male, female or neither/not known  
1st, 2nd or 3rd person  
single, plural, uncountable or collective

As in the pilot study, the subject was marked 'not obvious' (n/o) if it had no realisation and no clear referent from the context.

13. **Subj sem** In the pilot study (see section 1.3.5) I had noticed that there was frequently a mismatch between the type of referent and the word(s) representing the referent. For example, a word may be literally inanimate, but actually refer to a person or people. In order to collect examples of such mismatches for future research, I included field 13 which has the same kind of information as field 12, but is only filled in if the word and referent differ in their characterisation.

14. **Subj form** The form of the subject was recorded in field 14, using a slightly enlarged set of codes from those used in the pilot study:

**full** - any noun phrase which consisted of at least a non-pronominal noun.  
**pn** - personal pronouns  
**PN** - names  
**dem** - demonstrative pronouns  
**imp** - impersonal pronouns 'it' and 'there'  
**rel** - relative pronouns  
**Q** - 'question' words  
**wh** - wh- words not used for interrogatives  
**NCl** - noun clauses  
**n/r** - 'not realised'

15. **Subj struc** The structure of the subject was recorded in terms of its premodification (determiners, enumerators, adjectives, nouns) and its postmodification (prepositional phrases, postmodifying clauses).

e.g. The huge international operation [14 Mi Art]  
def aj aj hn

16. **Agent** Any by- phrase agents occurring after passive verb phrases were entered in full in this field in order to save time looking up their sentence contexts at a later stage.

17. **Agt type** The same four subfields were coded as for subject type (field 12). The default code in this case, however, was 'not applicable' (n/a) since most verb phrases were not passive. The code 'not realised' (n/r) was also needed for passives with no obvious agent.

18. **Agt sem** As in the case of subjects, any agents whose referents need to be coded differently from field 17 are given a second coding here.

19. **Agt form** The form of realised agents was coded as for subjects (field 14) but, as in field 17, the default code for active verb phrases was 'not applicable' (n/a).

20. **Agt struc** The structure of agents was also coded as for subjects (field 15). The default 'n/a' code was used for active verb phrases.

21. **Special Interest** This field was used to identify examples falling in two categories of 'special interest' which were to be investigated in detail in the main study.

These were the borderline between adjectives and -en participles (v/aj) and verb phrases occurring in sequence.

### 1.5.3 Extracting information from the database

As in the pilot study, the database was used to supply two kinds of information. For the stylistic study reported in chapter 2, the main information needed was statistical. Using the facilities provided by the Extract database package, I was able to discover the numbers of data items in any category defined by one or more features in the 21 fields described above. For example, the number of examples containing a nominal clause and a human subject could be extracted by an enquiry of the following form:

```
? clt == 'N' and sty @ 'h'
```

This enquiry asks for all examples where the clause type (clt) is equal to (==) nominal (N) and where the subject type (sty) contains (@) a subfield which is human (h). The first response from the computer is to show how many data items match the description in the enquiry. For statistical information the enquiry is finished at this point.

In order to extract data for more exploratory investigations such as those in chapters 3 to 5, the next stage was to make the computer format the information in the 21 fields in an accessible way. I worked out a general format which would be suitable for the majority of my enquiries and stored this in a file which was used for most of the remaining enquiries (see Appendix 1c for an

example of the formatted data).

After these enquiries had been made, much of the investigation made use of the printout produced and a separate file containing complete sentence contexts for each item (extracts from the latter file are reproduced in Appendix 1d).

## Chapter 2

### Aspects of styles in the corpus

This chapter describes the main database in terms of the features coded for each data item which are described in section 1.5 of the Introduction. The description is in three parts: a description of the database as a whole, a description to compare the three text-types (articles, editorials and letters) with each other, and a description comparing individual newspapers and groups of newspapers with each other.

#### 2.1 Description of the whole database

One of the most straightforward ways of exploiting the data described in section 1.5 is to examine the features which were encoded in the database and present them as a partial stylistic description of the data. Such a description can, of course, have no explanatory value but stands as a contribution to the investigation of style more generally, since the figures given below can easily be compared for other types of data.

The aspects of the data described in this section can be used in at least two ways. In the present work, there is little space for comparison with other types of data although this is an obvious application of this information. The most important function of the general stylistic study for this thesis is to provide a standard

against which descriptions of sub-groups of data can be compared. These 'standards' can be seen in use in sections 2.2 and 2.3 of the current chapter and sporadically in chapters 3, 4 and 5.

The statistical test used throughout this chapter is the chi-squared ( $X^2$ ) test which gives the level of probability of the deviation observed in the data being due to chance alone (Mulholland and Jones 1975)<sup>1</sup>. Although this test can be used for purposes of prediction, particularly by looking for high levels of similarity between samples, in the present work it is used simply to identify areas of significant difference. The  $X^2$  values in this chapter are considered for significance at the 0.05 level which means that if the differences were due to chance then the observed figures would only be expected to occur in 5% of possible cases. In fact, many of the values of  $X^2$  are also significant at the 0.001 level, but the 0.05 level is used as the standard for the tests in this work.

In this first section, the  $X^2$  tests are used rarely; the main purpose of the figures given below is to establish some facts about the style of the whole database, not to compare it with other data. However, there are a few examples of parts of the data being compared with each other and one example of the main database being compared

---

<sup>1</sup> See Chapter 15 of Huddleston et al. (1968) for a fuller explanation of the use of chi-squared tests in linguistic research.

with the pilot database. These comparisons are suitable for statistical analysis and the observed figures were tested against an expected figure for each population (pilot/main data or two sub-databases). Since all of the tests in this section compared the figures for just two samples, there was one degree of freedom in each case.

### 2.1.1 Main verbs and the copula 'BE'

The verb 'BE', when it occurs as a main verb, is studied in detail in Chapter 3, so it was considered useful to have an overall impression of the frequency with which this verb occurs in the data:

BE	455	19% of total data items (2400)
----	-----	--------------------------------

The frequency of 'BE' is striking when it is compared with the frequency of other verb-forms. There are 729 different verb-forms represented in the data, although the number of verb senses is probably higher; there was not time to analyse the polysemy of each verb-form during the initial analysis of the data. However, the frequency count for all of the data shows that no other verb-form is nearly as frequent as the copula 'BE'. Table 2.1 shows all of the verb-forms occurring 10 times or more in the data; only 31 out of 729 occur even this frequently.

Table 2.1 Frequency of verb-forms

verb-form	frequency	verb-form	frequency
be	455	leave	17
become	14	make	44
believe	13	put	10
claim	16	replace	10
come	16	save	10
deny	11	say	81
find	19	see	14
get	16	show	15
give	22	take	18
go	27	tell	21
have	56	think	12
hear	13	try	11
injure	10	use	23
kill	14	want	16
know	22	warn	10
		win	12

At this point, the figures simply serve to show that the copula 'BE' occurs very frequently in the data for this thesis. It would, perhaps, be interesting stylistically to see whether similar frequency counts occur in other types of data, or whether this proportion of occurrences is typical of English as a whole<sup>2</sup>. However, within the current work the high frequency of 'BE' does at least justify devoting a chapter to the study of its environment (Chapter 3).

### 2.1.2 Finiteness

A majority of finite verb phrases would be expected in most types of data, but the exact proportion of finite to

---

<sup>2</sup> Note that a superficially similar figure of 22.8% occurrences of BE was obtained from the scientific corpus investigated by Huddleston et al. (1968) p687.

non-finite verb phrases could not have been predicted:

Table 2.2 Finite and non-finite VPs

	number	% (of 2400)
finite	1763	73
non-finite	637	27

There may be some correlation between these figures and those for main and subordinate clauses. A written style which had less subordination, for example, would be expected to have a lower number of non-finite verb phrases. This hypothesis must await further comparative study for which there is no space in this work.

The division of non-finite verb phrases (total 637) into infinitive (i), progressive (ing) or passive (en) was coded in field 8 of the data. The figures and percentages for each type of non-finite VP are given in Table 2.3 below, and will be used in later sections of this chapter for comparative purposes:

Table 2.3 Non-finite verb phrase classes

	number	% (of 637)
ing	224	35
i	344	54
en	69	11

The finite verb phrases were also coded according to the form of their first verb in field 8. Again for comparative purposes, both within this work and for future studies, the figures and percentages (of 1763 finite VPs) for these

groups are given:

**Table 2.4 Finite verb phrase classes**

	number	% (of 1763)
base form	499	28
3rd person form	466	26
past form	495	28
am	4	0
are	87	5
was	156	9
were	56	3

### 2.1.3 Verb phrase types

One of the motivations for this study was a sense of dissatisfaction with the way that grammatical descriptions fail to match everyday linguistic experience. In particular, I was unhappy about the amount of attention given to the structure of the verb phrase as consisting of four possible auxiliary positions plus a main verb, when casual observation (see section 1.3.10) led me to believe that the sequences of verb phrases which are discussed in Chapter 5 were at least as common as the verb phrases with three or four auxiliaries.

The numbers of each verb phrase type (1 to 17 as described in section 1.5) support the notion that verb phrases in naturally occurring data are most likely to consist of a main verb alone, or a main verb with one auxiliary. The figures for the 17 categories are given below in table 2.5 with percentages (of 2400):

Table 2.5 VP structures

	number	% (of 2400)
1 Main verb (Mv)	1694	71
2 Modal (Mod) + Mv	217	9
3 Have + Mv	127	5
4 BE1 + Mv	45	2
5 BE2 + Mv	158	7
6 Mod + Have + Mv	32	1
7 Mod + BE1 + Mv	4	0
8 Mod + BE2 + Mv	50	2
9 Have + BE1 + Mv	3	0
10 Have + BE2 + Mv	30	1
11 BE1 + BE2 + Mv	4	0
12 Mod + Have + BE1 + Mv	0	0
13 Mod + Have + BE2 + Mv	4	0
14 Mod + BE1 + BE2 + Mv	0	0
15 Have + BE1 + BE2 + Mv	0	0
16 Mod + Have + BE1 + BE2 + Mv	0	0
17 Do + Mv	33	1

The figures in this section are supported by those of Huddleston et al. (1968) who found that the majority of their verb phrases were in the simple present tense (p643).

Although it is clearly possible to find, and indeed invent, acceptable English sentences containing a verb phrase with a structure like numbers 11 to 16, the point here is that there has been an unwarranted emphasis in both pedagogical and research work on these rather rare structures, whilst verb phrase sequences, which occur relatively frequently in this data (see chapter 5 below), have been largely ignored.

#### 2.1.4 Passive verb phrases

The number of passive verb phrases in the data, and how they are distributed between different structure types, is

shown below in Table 2.6:

Table 2.6 Distribution of passive verb phrases

	number	% (of 315)
1 non-finite Mv; -en form	69	22
5 BE2 + Mv	158	50
8 Mod + BE2 + Mv	50	16
10 Have + BE2 + Mv	30	10
11 BE1 + BE2 + Mv	4	1
13 Mod + Have + BE2 + Mv	4	1
14 Mod + BE1 + BE2 + Mv	0	0
15 Have + BE1 + BE2 + Mv	0	0
16 Mod + Have + BE 1 + BE 2 + Mv	0	0

The above table shows that the structure BE2 + Mv (auxiliary verb BE) is by far the most common passive structure, occurring in just over 50% of the passive examples. The total number of passive verb phrases, 315, represent 13% of the whole data (2400 VPs) which is not significantly different from the proportion found in the first exploratory study of passives in letters to newspapers (see section 1.2.3). The number of passives for this exploratory data was 197 out of a total of 1383 verb phrases; 14% overall. An expected number of passive verb phrases was calculated for each database and the hypothesis that the observed figures were not significantly different was confirmed by a  $X^2$  test ( $X^2 = 0.81$ ). These figures appear to be much lower than those for scientific and technical English. Huddleston et al. (1968), examining scientific written English, found 26.3% passives (p687) and my background investigation of Mechanical Engineering texts found over 30% of passive

verb phrases .

Since most writers agree that there is an important difference between passive verb phrases with a realised agent and those without an agent, it may be interesting for future study to know what proportion of the total passive verb phrases these two groups represent. The figures are given below in Table 2.7:

Table 2.7 Passive VPs with/without agents

	+ agt	- agt
1	17	52
5	35	123
8	7	43
10	9	21
11	0	4
13	1	3
total	69	246
% (of 315)	22	78

The large percentage of passive VPs without agents is probably not particular to the style represented in the present data since it is comparable to the 80% of agentless passives found by Svartvik in a widely-based corpus (ranging from scientific to literary texts) and the 70% found in literary works by Jespersen (1933).<sup>3</sup>

---

<sup>3</sup> Stein (1979) points out that in the small number of passive verb phrases that have an agent realised (15-20%), the agent phrase is probably "not that optional after all" (p126).

### 2.1.5 Verb phrase sequences

As mentioned above in connection with verb phrase structure, the proportion of verb phrase sequences in the data under consideration is as high as many other subgroups of verb phrases. In Table 2.8, for example, the number of verb phrases occurring in sequence is compared with those which contain modal verbs, those which contain the progressive auxiliary (BE1), those which contain the perfective auxiliary (Have), and the passive verb phrases.

Table 2.8 Subgroups of verb phrases as a proportion of data

	number	% (of 2400)
modals	307	13
BE1	52	2
Have	196	8
pass	315	13
VPs in sequence	373	16

The verb phrases which occur in sequences are examined in detail in Chapter 5. The present section aims simply to establish the group as large enough to be worthy of further investigation both in future comparative studies and within the present work.

### 2.1.6 Clause types and functions

It was noted in the Introduction (section 1.3.3) that the categories of clause type and clause function were in need of further study, and that a working set of categories was set up, based on observable differences of physical form

and position in order to input the data in some retrievable way.

Here the aim is to show, in general, the proportions of main clauses and of main classes of subordinate clauses, although the latter may not correspond exactly to the categories in other works.

The classes of subordinate clause represented in table 2.9 below include some which were set up on the basis of their apparently similar functions. These include the 'noun appositive' class, which consists of those clauses not fitting into the relative clause class, but which nevertheless follow and modify in some way, a noun (see Quirk et al. 1972, section 13.16):

e.g. a mutual acceptance on the part of the  
superpowers that their relations are too  
important to ... [18 T Ed]

The 'verb complement' class contains all the verb phrases occurring as second or subsequent phrases in a sequence. The 'adjective complement' class has some precedent in the literature (see Quirk et al. 1972, section 12.35) and it will not be described here.

Table 2.9 Proportions of clause types in the data

	number	% (of 2400)
main	1081	45
noun	289	12
adverbial	348	15
relative	259	11
verb complement	195	8
noun complement	127	5
adjective complement	76	3
problem cases	24	1

It was stated in the discussion above that there is no clear dividing line between the class membership of a clause and its function. It is therefore difficult to base any firm conclusions on data drawn independently from fields 9 (clause type) and 10 (clause function). However, there may be some comparative value in presenting the proportions of noun clauses which were found to have the functions of subject, object and complement:

Table 2.10 Functions of noun clauses

	number	% (of 289)
subject	26	9
object	221	76
complement	34	12

[Nb. The total number of noun clauses represented here does not match the number given in Table 2.9 because there are a few examples (8) in the data of noun clauses functioning as modifiers within noun phrases.]

#### 2.1.7 Subject type

As mentioned in the Introduction (section 1.3.5), there were a number of problems involved in setting up the fields associated with subjects, particularly field 12 which is labelled 'subject type' in this study. The coding of this field does, however, allow us to draw some conclusions from the data.

The first sub-field of field 12 concerned the animacy or otherwise of the referent of a subject. As might be expected in data which largely concerns the affairs of human social and political life, there were high

proportions of human and abstract subjects and very few inanimate or animate (non-human) subjects:

Table 2.11 Subject type - animate/inanimate/human/abstract

	number	% (of 2201)
human	1101	50
abstract	950	43
inanimate	142	6
animate	8	0

[Nb. There were 199 subjects whose referent could not be identified sufficiently for such categories to be applied.]

The second subfield of field 12 identifies the subjects as female or male, and as such represents a breakdown of the human category. Table 2.12 shows the proportion of human subjects identified by sex:

Table 2.12 Human subjects - sex groups

	number	% (of 1101)
female	106	10
male	415	38
unidentified	580	53

Although it is not surprising to find more subjects identified as male than as female, the proportions in this table may be of comparative interest when other types of data are considered.

The third sub-field in field 12 shows whether the subject is in the first, second or third person. As one would expect in this kind of data, the majority of subjects are

third person, as shown in Table 2.13:

Table 2.13 Subjects - person

	number	% (of 2400)
1st	168	7
2nd	17	1
3rd	2016	84

One of the observations made during the pilot study was that there seemed to be a high coincidence of female and 1st person features<sup>4</sup> in the subjects of the pilot data (see section 1.3.6 of the Introduction). The coincidence of these features was investigated for the main database, but the proportions of male and female subjects which occurred in 1st and 3rd person form were not significantly different:

Table 2.14 Subjects - sex and person

	male	% (of 415)	female	(% of 106)
3rd person	398	96	102	96
1st person	16	4	4	4

The main difference between the two databases is that the pilot data all came from letters pages in newspapers, whereas the present data contain an equal balance of examples from letters, news articles and editorials. The high proportion of 1st person, female subjects seems, therefore, to be a feature of letters and therefore of the

---

<sup>4</sup> As explained in chapter 1, the sex coding was given on the basis of known referents, not only for words which are marked for sex in any context.

'public' rather than being typical of newspaper English generally.

The fourth subfield of field 12 recorded the 'number' of each subject in the data and the figures are given in Table 2.15:

Table 2.15 Subjects - number

	number	% (of 2400)
singular	1330	55
plural	635	26
uncountable	137	6
collective	99	4

#### 2.1.8 Agent type

There were only 69 agents (by-phrase and equivalents) realised in the data; a total of 246 passive verb phrases were marked as having no agent whose referent was identifiable. Unlike the coding of subjects, whose referents are usually clear even when there is no subject form realised, agents which are not realised are usually unidentifiable as well. This reflects the use of the passive structure specifically to avoid mentioning the agent.

Table 2.16 Agents - animate/inanimate/human/abstract

	number	% (of 69)
human	37	54
animate	0	0
inanimate	1	1
abstract	31	45

With so few examples of human agents to consider, it is probably not useful to split the human category further into male and female subgroups. However, it is interesting to note the apparently consistent proportion of human subjects and human agents which are identified as male. The figures for agents are given in Table 2.17:

Table 2.17 Human agents - sex

	number	% (of 36)
female	0	0
male	14	39
unidentified	22	61

Although there can be no statistical significance attached to numbers of this size, the proportion of male subjects (38%) was very similar to that of male agents (39%). This may be a feature which is worth comparing with other texts in the future.

The third subfield in field 17 codes the person of any realised agent. As I had expected, all of the agents (69) were in the 3rd person in the data. 2nd and 3rd person agents, seem to be restricted to the spoken language ('This house was built by me').

The figures for the last subfield of field 17, number, when taken as a percentage of realised agents, are comparable to those for subjects. The figures for agents can be seen in table 2.18. The percentages for subjects are included for comparison:

Table 2.18 Agents - number

	number	% (of 69) subjects (%)	
singular	36	52	55
plural	23	33	26
uncountable	5	8	6
collective	5	7	4

2.1.9 Subjects and agents - form

The figures for field 14, subject form, and field 19, agent form, are given together in Table 2.19 to facilitate comparison:

Table 2.19 Subjects and agents - form

	subject	% (1619)	agent	% (69)
full	705	44	45	70
pronoun	437	27	0	0
proper noun	204	13	18	26
demonstrative	37	2	0	0
impersonal	90	6	0	0
relative	114	7	0	0
question/wh	13	1	0	0
noun clause	19	1	3	4

[Note that, since subjects and agents have different reasons for being 'not realised', the percentages here are calculated from the number of realised examples, 1619 for subjects and 69 for agents.]

Although a competent speaker of English can envisage situations where almost any of these forms could occur as agents (e.g. 'He was killed by them'), except for the impersonal pronouns such as 'there' or 'it' and noun clauses, the fact is that in the present data at least there are agents in only a limited set of forms. A comparison with other types of data may show this to be a feature of newspaper style. In the present data, subject

forms range over a wider set of realisations.

Fields 15 and 20, which recorded the structures of subjects and agents respectively, have entries which are too diverse to use for statistical analysis but which would provide data for an interesting future study of this topic. However, some simple features can be extracted from the data to give some idea of the character of subjects and agents in the data. The features examined in this way were definiteness (the presence of definite or indefinite articles) and modification (the presence of premodification other than articles and of postmodification) and the figures are given in tables 2.20 and 2.21 below:

Table 2.20 Subjects and agents - definiteness

	subject	% (of 1619)	agent	% (of 69)
indefinite	80	5	8	12
definite	367	23	13	19

The significant difference between subjects and agents here is that there is a higher proportion of indefinite agents than indefinite subjects, although the number of agents is so small that this figure may be unrepresentative. In a  $X^2$  test, taking the average of the proportions for subjects and agents as the expected figure,  $X^2$  had the value 11.9 for indefinite articles and 0.35 for definite articles. Only the first value was significant. A more complex definition of 'definiteness'

(for example, including names and demonstratives) might have altered the latter result. The significant difference in the number of indefinite articles may indicate that the choice of passive voice is sometimes influenced by a more general ordering tendency, placing definite NPs before indefinite ones, which also affects the form of subjects as compared to subject complements (see chapter 3).

**Table 2.21 Subjects and agents - modification**

	subj	% 2201	agt	% 69	
premod.	565	26	27	39	$X^2 = 4.62$
postmod.	49	2	2	3	$X^2 = 0.18$
pre- and postmod	178	8	20	29	$X^2 = 31.55$
unmodified	1409	64	20	29	$X^2 = 12.62$

The figures in table 2.21 show that the number of subjects and agents which are postmodified do not differ significantly. However, the three other categories produce values of  $X^2$  which are significant. Subjects are more likely to be unmodified than agents which tend to occur with premodification or both pre- and postmodification. The fact that agents seem to be modified more often than subjects may be one of the factors affecting the choice of a passive construction since English favours a late position in the clause for long or complex elements.

Although the coding of subject form (field 14) and agent form (field 19) included the features of apposition and conjunction, there were not enough examples in the data from which to draw any firm conclusions. The only

implication in these figures (given in table 2.22) is that the slightly higher number of appositive structures in the agents (11%) may show that the passive construction has been chosen to avoid the clumsiness of a very long subject, allowing the agent instead to carry the complexity. This hypothesis would need further study in order to establish its accuracy.

Table 2.22 Subjects and agents - apposition and conjunction

	subject	% of 2201	agent	% of 66
apposition	52	2	7	11
conjunction	35	1	2	3
app and conj	7	0	0	0

## 2.2 Comparative description by text-type

### 2.2.1 Features showing no difference between text-type

Many of the features coded in the database showed no significant difference between the three types of text investigated. Some of the figures are given below both to illustrate the uniformity of the data when cross-sectioned in this way and to provide a comparison for those features which differ along the dimension of newspaper type as opposed to text-type.

The  $X^2$  tests used in sections 2.2.1 and 2.2.2 calculated the deviation of the figures for articles, editorials and letters from the expected figure which was, in most cases, one third of the total figure for the feature concerned.

For example, the observed number of finite verb phrases in the total data was 1763. Assuming the sub-databases are not significantly different from each other, the expected number of finite verb phrases in each one would be  $1763 \div 3 = 587.67$ . This calculation was made simple by the fact that the main database was set up with equal numbers of data items in each of the text-type x newspaper categories. Since the  $X^2$  tests in this section compare three sub-databases, there were two degrees of freedom in each case.

There was no significant difference between the text-types in the proportion of finite verb phrases:

Table 2.23 Finite verb phrases in three text-types

	total	art	ed	let
finite	1763	566	604	593
%	73	71	76	74

The value of  $X^2$  for the figures in table 2.23 was 1.3, which was not significant at the 0.05 level.

Most of the numbers in field 7, verb phrase structure, are too small to be considered statistically. The figures are therefore given (in table 2.24) for structure 1 (Mv) alone which shows no significant difference across text-type. Structures 2 and 3, which do show a difference, are discussed in section 2.2.2.

Table 2.24 Verb phrase structures in three text-types

	total	art	ed	let
1 Mv only	1694	587	533	574
%	71	73	67	72

The figures in table 2.24 revealed a  $X^2$  value of 2.81, which is not significant at the 0.05 level.

The number of verb phrases which form part of sequences (field 21) does not seem to differ between these text-types, as the figures in Table 2.25 show:

Table 2.25 Verb phrases in sequence in three text-types

	total	art	ed	let
VPs in seq	373	128	130	115
%	16	16	16	14

The impression that the figures in table 2.25 do not differ significantly was confirmed by a  $X^2$  value of 1.07. The distribution of clause types (field 9) across text-types is also fairly uniform as can be seen from Table 2.26:

Table 2.26 Clause types in three text-types

	total	art	ed	let	
main	1081	351	378	352	$X^2 = 1.3$
%	45	44	47	44	
noun	289	106	89	94	$X^2 = 1.59$
%	12	13	11	12	
postmod	462	146	161	155	$X^2 = 0.74$
%	19	18	20	19	
verb comp	195	68	68	59	$X^2 = 0.83$
%	8	9	9	7	

The only clause type which differed between the three text-types was the adverbial clause. The figures for this clause type are given and discussed in section 2.2.2. The other clause types did not differ significantly between text-types; the values for  $X^2$  given alongside table 2.26 are not significant.

As explained above (section 1.5), the clause functions as set up in the data reflect in a fairly straightforward way the clause types just described. The figures (which can be seen in Appendix 2) therefore show the same kind of uniformity as those for clause types. However, there were some differences found between the text-types regarding the distribution of noun clauses among the functions of subject, object and complement. These are discussed in section 2.2.2.

Some differences of subject type (field 12) found between the text-types are also discussed in section 2.2.2 below. However, one term of the fourth sub-field, plural, was found not to differ between text-types:

Table 2.27 Number of subjects in three text-types

	total	art	ed	let
plural	635	217	190	228
‡	26	27	24	29

Although the other terms of this sub-field were significantly different, a  $X^2$  test on the figures in table 2.27 produces a value of 3.61, which is not significant.

The figures for some of the subject form categories (field 14) were too small to use for statistical analysis. There were, however, roughly equivalent proportions of fully worded subjects in the three text-types as shown in Table 2.28:

Table 2.28 Full subjects in three text-types

	total	art	ed	let
full subject	700	252	234	214
%	29	32	29	27

The figures in table 2.28 were not significantly different at the 0.05 level ( $X^2 = 3.1$ ). Some other subject forms, notably personal pronouns and proper noun phrases, were found to differ slightly between text-types and these are discussed in section 2.2.2.

Because the number of realised agents in the whole database was so low (69 out of 315 passives), there can be no useful analysis of sub-sections of the data on agent types and forms. However, the distribution of passive forms, both with and without agents, was investigated and since some significant differences were found, is dealt with in the next section.

### 2.2.2 Features which differ between text-types

The overall frequency of the copula verb 'BE' in the data was given as 19% in section 2.1. Although this was shown to be considerably higher than the frequency of any other

verb, the total figure conceals a difference between the three text-types under consideration here. As seen in Table 2.29, the news articles in the data contain a lower proportion of copula 'BE' verbs than the other two text-types:

Table 2.29 Occurrence of copula 'BE' in three text-types

	total	art	ed	let
BE	455	74	203	178
%	19	9	26	22

The percentages in table 2.29 show a clear difference between the text-types which is supported by the very high value of  $X^2$  at 61.72, significant at the 0.001 level. It would appear that in this data at least, news articles make fewer attributions than either letters or editorials. This may be a reflection of the kind of objective style expected in news reporting, which would avoid the use of any construction that gives the impression of personal judgement or opinion. Table 2.30 shows the total figures for each text-type broken down into the four main types of complement (see chapter 3 for more on the complementation of 'BE'):

Table 2.30 Complementation of 'BE' in three text-types

	total	art	ed	let	
Noun phrase	200	39	85	78	$X^2 = 0.82$
%	44	53	42	44	
Adjective phrase	169	27	72	66	$X^2 = 0.16$
%	37	36	35	37	
Prep phrase	36	3	16	17	$X^2 = 1.99$
%	8	4	8	9	
Noun clause	29	3	17	9	$X^2 = 2.39$
%	6	4	8	5	

[Nb. The percentages in each case represent the proportion of the total number of copula verb phrases in each type of data which are; 455 (Total data), 74 (Articles), 203 (Editorials) and 178 (Letters) respectively.]

Although the total number of 'BE' copula verbs differed significantly between the three text-types, the figures in table 2.30 show that the text-types do not differ in their proportion of BE complements. None of the values of  $X^2$  given in table 2.30 is significant at the 0.05 level.

There is one slight difference (not statistically significant) between the articles and the other text-types which may support the claim that article writers aim for an impression of objectivity. This is the fact that a higher proportion of those copula 'BE' verb phrases which appear in articles are complemented by a noun phrase rather than any other kind of complement. It can be argued that such complementation, exemplified below, apparently attributes to the subject something more verifiable than do adjectival or prepositional complements.

The first example shows an uncontroversial noun phrase complement following the copula verb 'BE':

e.g. It was the fourth explosion in a week in the city..... [12 S Art]

In contrast, note the relative subjectivity of the prepositional and adjectival complements in the next two examples:

e.g. ...strongman Dr Owen (-) is in complete command of the SDP. [5 S Art]

In this example, Dr Owen's ability to control the direction of the SDP is stated as fact, although it can only be the considered opinion of the writer. The next example is more obviously giving the opinion of the writer:

e.g. It would be absurd if someone living on the sale of inherited assets were entitled to benefit when a hard-working manager was not. [3 Te Let]

Clearly these classes of complement are not related directly to the 'objective' truth or falsehood of a proposition, but the impression they give to the reader can be partly contrived by the use of one type of complement rather than another. For example, the noun phrase can be used to give a sentence an air of authority which belies its subjective nature:

e.g. (The Liberals have not won an election for 70 years.) They are electoral poison. [47 S Ed]

Without looking more closely at the data, which is not the purpose of this mainly statistical chapter, the conclusion must simply be that article writers use fewer attributive forms generally, but a higher proportion of noun phrase complements in particular, which may indicate an effort to appear objective in their reporting.

The next field investigated which revealed differences between text-types was verb phrase structure (field 7). As mentioned in section 2.2.1, structures 2 (modal + main verb) and 3 (have + main verb) appeared to differ between text-types. These structures were therefore investigated with other related structures (containing modal verbs or have) in order to test the differences for statistical significance.

The classes of verb phrase structure which contain a modal verb are presented below with the number of occurrences in the total data and in each of the three text-types. In the bottom line, the total figures are given as a percentage of the total number of verb phrases in each database; 2400 for total data and 800 for each text-type:

Table 2.31 Occurrence of modal verbs in three text-types

	total	art	ed	let
2 Mod + Mv	217	40	103	74
6 Mod + Have + Mv	32	4	18	10
7 Mod + BE1 + Mv	4	2	1	1
8 Mod + BE2 + Mv	50	16	10	24
13 Mod + Have + BE2 + Mv	4	0	2	2
Total	307	62	134	111
% (of 2400/800)	13	8	17	14

A  $X^2$  test on the total figure for each text-type produces a significant value for  $X^2$  of 26.43.

As with the frequency of occurrence of the copula 'BE' discussed above, the articles tend to have fewer modal verbs than either editorials, which have a higher than average number of modals, or the letters, which fall in the middle and have a proportion close to the average for the total data. A tentative explanation (given below) is related to that given for the different frequency of copula 'BE' and concerns the apparent objectivity of news article style.

The modal auxiliaries may be described as introducing an element of doubt into a proposition. They indicate to the reader/hearer that there is some qualification of the verbal meaning. This is not generally compatible with the journalist's aim of appearing to report accurately, truthfully and objectively, so it is not surprising that the articles should contain fewer modals than other text-types. The converse of this explanation can be applied to the high proportion (17%) of modals in the data from editorials. The newspaper reader expects to read the judgements and opinions of the Editor in the editorial and modals are ideal both for protecting the Editor from later charges ("I only said it might...") and for allowing projections of what could or should be done or have been done in a given situation. Some examples from the data follow:

e.g. America could respond to such a lead. So could Britain. [30, 31 M Ed]

e.g. There may be, in fact, a distinction between appearance and reality in this instance. [12 Ti Ed]

A pattern is beginning to emerge from the figures in this section which shows the articles as having a more 'objective' and 'factual' style (fewer attributions using 'BE', less use of modals) and editorials as having an openly more 'subjective' style (more use of attributions and more modals). The letters in each case conform more closely to the average (measured over the total data) than the other two text-types; 22% copula 'BE' (average 19%) and 14% modals (average 13%). However, this pattern is not repeated by the figures for verb phrases containing the auxiliary 'have':

Table 2.32 Occurrence of auxiliary 'have' in three text-types

	total	art	ed	let
3 Have + Mv	127	39	57	31
6 Mod + Have + Mv	32	4	18	10
9 Have + BE1 + Mv	3	1	0	2
10 Have + BE2 + Mv	30	10	15	5
13 Mod + Have + BE2 + Mv	4	0	2	2
Total VPs with 'have'	196	54	92	50
% (of 2400)	8	7	12	6

Although the pattern described above is not upheld here, there are differences between the three text-types and the value of  $\chi^2$  (16.45) showed them to be significant. It is not easy to explain the occurrence of 'have' in terms of

objectivity or the papers wanting to appear objective. It may simply be the case that the use of perfective constructions reflects directly the kind of statements typically made in editorials:

e.g. Ronald Reagan has now taken command of the crusade to save the most powerful nation on earth from the chemical and biological plague that threatens to rot it from within

[25 M Ed]

The pattern found in figures for modals and BE copula is, however, repeated when we consider the number of passive verb phrases in each text-type. The figures are given in Table 2.33:

Table 2.33 Passive verb phrases in three text-types

	total	art	ed	let
5 BE2 + Mv	158	82	30	46
8 Mod + BE2 + Mv	50	16	10	24
10 Have + BE2 + Mv	30	10	15	5
11 BE1 + BE2 + Mv	4	0	4	0
13 Mod + Have + BE2 + Mv	4	0	2	2
Total	246	108	61	77
% (of 2400/800)	10	14	8	10

As mentioned above, these figures confirm the general difference noted between article style and editorial style. A  $X^2$  test on the total figures produced the value 13.93 which is significant at the 0.001 level. The articles have a higher than average number of passives which may reflect the fact that the passive construction is often associated with formality and with factual reporting as a result of having been used extensively in

scientific and technical writing.<sup>5</sup> The editorials, however, are not constrained to appear impartial, indeed they are often expected to be partisan, and the passive construction, which gives an impression of rationality and objectivity, occurs less often in this text-type. The letters again fall between the other text-types and have a proportion of passive verb phrases close to the average for the total data.

The example below shows the passive being used as part of a formal, informative style:

e.g The document will also be debated at the Liberal assembly in Eastbourne next week. [3 Te Art]

The last feature of the verb phrase coded in the data which shows some difference between the three text-types under consideration is that of the three kinds of non-finite verb phrase. It was stated in section 2.2.1 that there was no difference among text-types in the proportion of their verb phrases which were non-finite. However, although they are not significantly different, the pattern being established in the present section, where letters

---

<sup>5</sup> As noted in Section 2.1.4, Huddleston et al. (1968) found about 26% of passive VPs in their data. They also found that the most formal scientific writing, the 'high stratum' in their data showed the highest proportion of passives without agents:

"In the high stratum, in particular, there is a convention that authors should efface themselves by the use of the passive, thus stressing the objectivity of their research." (p646)

have close to average figures for some features while the other text-types are ranged either side of the norm, is upheld even in the overall figures for finite verb phrases (and therefore for non-finite verb phrases) which were given in Table 2.23. Within these figures, the text-types also differ according to the non-finite form concerned as shown in Table 2.34 below:

Table 2.34 Types of non-finite verb phrase in three text-types

	total	art	ed	let	
ing	224	96	51	77	$X^2 = 7.19$
%	35	41	26	37	
i	344	114	122	108	$X^2 = 3.80$
%	54	49	62	52	
en	69	24	23	22	$X^2 = 0.23$
%	11	10	12	11	

[Nb. In the above table, the percentages represent the non-finite forms as a proportion of the total number of non-finite forms in each text-type. These are 637 for all the data, 234 for articles, 196 for editorials and 207 for letters.]

Only the present participle form (ing) of non-finite verb phrase was found to occur in significantly different proportions in the three text-types. The value of  $X^2$  for the 'ing' form was 7.19 which is significant at the 0.05 level. The other values of  $X^2$  in table 2.33 are not significant at this level.

There is no easy way to explain why the articles should have a higher than average proportion of -ing forms and a lower than average proportion of infinitive and -en forms.

The non-finite forms do not reflect in any obvious way the objectivity of style which was discussed in connection with the copula 'BE' and the modals. However, the fact that the style of the letters yet again falls between that of the articles and that of the editorials suggests that a closer investigation of the present participle and its semantics in particular may be worthwhile.

The single clause feature which shows a significant difference between the three text-types is the function of noun clauses (field 10). As table 2.34 shows, the letters fall between the other text-types yet again and have figures which are close to average for the whole data. The articles, on the other hand, have a very high proportion of their noun clauses functioning as objects whilst the editorials have a higher than average proportion of subject and complement noun clauses:

Table 2.35 Noun clause functions in three text-types

	total	art	ed	let	
subject	26	1	16	9	$X^2 = 15.65$
%	9	1	18	10	
object	221	101	48	72	$X^2 = 10.82$
%	79	95	54	77	
complement	34	3	20	11	$X^2 = 15.87$
%	12	3	22	12	

[Nb. The percentages in Table 2.35 represent the proportion of the total number of noun clauses in each text-type; 281 in the total data, 106 in articles, 89 in editorials and 94 in letters.]

The difference between the three text-types was

significant for noun clauses functioning as subject, object and complement. To interpret these figures in the context of the style of articles and editorials one must consider the different effect of using noun clauses as subjects, objects and complements. Some examples from the data illustrate such effects:

e.g. Last night, Dr Owen welcomed the conference decision and said he was confident that the Liberal Conference next week would also support the compromise. [12 M Art]

This example (starting 'he was confident..') is typical of the kind of noun clause object appearing in the articles; a number of the examples in the data are used to report indirectly the speech of participants in the news story in question. Another example of this straightforward use of noun clause objects<sup>6</sup> follows:

e.g. Mr John Cartwright, spokesman on defence, agreed that the 1985 policy was not engraved on tablets of stone but would be interpreted with flexibility. [10 Ti Art]

Other kinds of noun clause object (those not reporting speech) do not seem to occur very frequently in the articles. They do, however, occur in the other text-types. The following example comes from a Times Editorial:

e.g. The simultaneous transfer of Nicholas Daniloff and Gennady Zakharov to the custody of their respective ambassadors has temporarily defused what threatened to become a new explosion of East-West acrimony. [1 T Ed]

---

<sup>6</sup> Note that not all writers agree that clauses following verbs of 'saying' are objects. See, for example, Munro (1982).

The fact that there are a high number of noun clauses functioning as complements in the editorial data is probably a result of the more general tendency to use attributive constructions (often involving the copula 'BE') which has been discussed above. The formality of the style is not obviously affected by the use of long complement noun clauses, whereas the use of attributive constructions may imply personal judgement or opinion, particularly when the copula is modified by a modal as in the following example:

e.g. Its only purpose can be to tame the broadcasters  
before a general election gets under way.  
[2 Mi Ed]

The use of noun clauses in subject position is perhaps the most complicated of the three functions to explain stylistically. The data contain examples which have a variety of different effects. One of the simplest is where the noun clause functions as the subject of a passive verb phrase whose agent is left vague for purposes of persuasion by the writer. This is primarily a use of the passive construction and is illustrated by the following example:

e.g. What she says must be treated seriously.  
[3 To Ed]

One possible effect of having a noun clause in subject position is that the resulting subject will be rather long and unwieldy. This may be one of the reasons why article

writers in the data seem to avoid such constructions. English is well-known for the feeling of discomfort invoked in a reader/hearer by long and complex subjects. However, the long and complex subject can be used precisely to create that feeling in the reader by the style of writing found in editorials, where the writer wishes to persuade the reader of her/his point of view. One example from the data illustrates this effect:

e.g. The fact that he may have been turned into an addict by an unscrupulous seller is an important factor, but doesn't justify execution.  
[8 To Ed]

Although the proposition of this sentence partly upholds the importance of the drug pushers' guilt in creating addicts, the construction tends to 'rush over' this point by placing it in subject position and, as in most written English sentences, leads toward the focus at the end. Another example has the effect of focussing attention on the end of the sentence and, like the above example, emphasises a comparison between two features:

e.g. To put off such a fundamental decision about the safety of our country is as frightening as it is irresponsible.  
[12 Ex Ed]

There is also a symbolic reflection of meaning in the use of this long subject of the kind more often associated with poetic style. The effect of the long subject 'puts off' (delays) the main verb phrase in a way that reflects the 'putting off a decision' referred to in the noun clause itself.

Another symbolic effect is observed in the next example, where the long noun clause subject produces a discomfort in the reader similar to that hypothesised as afflicting Dr Owen:

e.g. Listening to his party give David Steel a near four minute standing ovation while he gets stabbed in the back may prove to be too much for Dr Owen. [36 Ex Ed]

The final example of subject noun clauses in this section illustrates another persuasive use of such constructions. The proposition of the subject in the following sentence can really only be an assumption, but its realisation as a subject noun clause puts it into a form which cannot easily be questioned: it is taken for granted that something is irritating the Government:

e.g. But what irritates the Government is that we have also exposed the devastation which the cruel cuts in spending have caused to the NHS. [28 Mi Ed]

By using a noun clause in complement function, this sentence also takes for granted both the process of 'exposing' and the further embedded proposition that 'cruel cuts in spending have caused devastation to the NHS'. The only proposition which is put forward openly, and can therefore be questioned, is the one linking the subject and the complement.

As mentioned in section 2.2.1, all of the subfields of field 12, 'subject type', except for plural number of

subfield 4, showed some differences between the text-types. Apart from the figures for animate subjects which were too small to be of interest, the figures in the first subfield continue the pattern noticed in the other fields discussed above, namely that the letters fall nearest the norm and the articles and editorials are ranged either side of the norm. The figures and percentages (of total data items for each text-type: 800) are given in table 2.36:

Table 2.36 Subject type in three text-types

	total	art	ed	let	
inanimate % (of 2400/800)	142 6	71 9	30 4	41 5	$X^2 = 19.03$
human % (of 2400/800)	1101 46	420 53	293 37	389 49	$X^2 = 23.89$
abstract % (of 2400/800)	950 40	265 33	403 50	283 35	$X^2 = 35.55$

A detailed explanation for these figures would rely on a comprehensive analysis of the referents of the subjects in each sub-database. However, the higher than average number of abstract subjects in editorials and of human subjects in articles is probably related to the kind of proposition made in each text-type. The editorials often make rather general statements whose subjects contain abstract nouns such as 'deterrent', 'difference', 'event', 'oppression' (all taken from Today Editorials) and there are also a high number of impersonal subjects in this sub-database. An example from Today:

e.g. There is absolutely no evidence of that.  
[10 To Ed]

News articles, on the other hand, are more likely to contain specific references to individuals, or groups of people such as 'CND supporters', 'Jo Binns', 'the two men' (examples from Today Articles) and even the non-human subjects are more likely to be inanimate than those in editorials:

e.g. One abandoned package... [28 To Art]  
The document... [3 Te Art]

Differences between text-types in subfields 2 and 3 of field 12 do not conform to the pattern described a number of times above in which the letters have figures close to the average for the total data. In fact, the editorials are closest to the overall average of number of identified male and female subjects in subfield 2, as can be seen in Table 2.37 below:

Table 2.37 Sex of subjects in three text-types

	total	art	ed	let	
human	1101	420	293	389	
male	415	241	97	77	$x^2 = 77.89$
% (of human)	38	57	33	20	
female	106	28	28	50	$x^2 = 8.03$
% (of human)	10	7	10	13	

The figures for both male and female subjects differ significantly. The high proportion of male subjects in articles is probably partly a reflection of the high

profile of men in public life, but may also be attributed to the unconscious but widespread tendency to consider male activities newsworthy while ignoring areas of life which are primarily female. The slightly higher than average proportion of female subjects identified in the letters may be connected to the high number of first person subjects occurring in this sub-database. A correlation between first person and female subjects was found in the pilot study, which was based on data from letters alone. However, no such correlation appeared to exist for the main data when considered as a whole. The figures for person are given below:

Table 2.38 Person of subjects in three text-types

	total	art	ed	let	
1st person % (of 2201)	168 8	0 0	20 3	148 20	$X^2 = 238.55$
2nd person % (of 2201)	17 1	0 0	2 0	15 2	$X^2 = 24.3$
3rd person % (of 2201)	2016 91	756 100	704 97	556 77	$X^2 = 24.11$

A  $X^2$  test was hardly necessary to confirm that the figures in table 2.38 were significantly different for the three text-types. All three values of  $X^2$  are significant at the more stringent 0.001 level.

It is not surprising to find that the articles, which are supposed to avoid personal judgements and opinions, use no first or second person pronouns. The editorials and

letters, which are the two sides of a supposed interaction between the newspaper and its readership, use a small number of first and second person forms.

The final difference between text-types which will be discussed in this section is found in field 14, subject form, where the proportion of personal pronouns and of proper nouns showed a significant difference between the text-types. The personal pronouns were not considered as a uniform group for the purposes of this section since 1st and 2nd person pronouns are likely to be predicted by the nature of the communication itself, whereas a 3rd person pronoun is in some cases freely chosen from a number of options including noun phrases of differing complexity. The letters, for example, have already been shown to have a higher than average number of 1st person pronouns which reflect their usual purpose of expressing the attitudes of the newspaper's readers. The proportion of 3rd person pronouns in articles was close to average for the total data, while the figure for editorials was well above average and that for letters well below:

Table 2.39 Personal pronouns as subject in three text-types

	total	art	ed	let	
3rd person	252	96	118	38	$X^2 = 40.67$
%	11	12	15	5	

The three text-types were shown to be significantly different in their proportion of 3rd person pronouns.

Contrary to my expectations, the number of 3rd person pronouns was higher than average in the articles. I had expected that the 'formal' style of the articles would lead the writers to avoid the potential ambiguity of reference represented by these pronouns. In fact, the editorials had the highest proportion of 3rd person pronouns and this may reflect the kind of discussion of individuals which they often contain:

e.g. What she (Mrs Reagan) says must be treated seriously. But she is wrong. In the American context, she has a superficially attractive argument. [3-5 To Ed]

In contrast with the editorials, the letters show a very low proportion of 3rd person pronouns as well as a low proportion of proper nouns as shown in table 2.40:

Table 2.40 Proper nouns as subject in three text-types

	total	art	ed	let	
proper nouns	204	90	86	28	$\chi^2 = 23.02$
%	9	11	11	4	

The difference between text-types seen in table 2.40 is also significant at the 0.001 level. The lower number of proper nouns in letters might have been explained by arguing that writers of letters to the press are less careful in avoiding the potential ambiguity of pronouns than professional journalists. However, since there were also fewer 3rd person pronouns in the letters than in the other text-types, the figures for proper nouns and 3rd

person pronouns probably reflect the high number of 1st person pronouns already discussed.

### 2.3 Comparative description by newspaper

The final section of this chapter considers the data when it is cross-sectioned to compare the styles of eight different newspapers. There are therefore eight sub-databases consisting of 300 data items (clauses) each. With the relatively small numbers being considered here, it is difficult to be sure that any differences noted could be generalised to larger bodies of similar data. For this reason only the most striking differences between newspapers are discussed. Another result of the small numbers being considered here is that it is more difficult to be sure when there are no differences between sub-databases. There are therefore no illustrations of uniformity through the data as given in section 2.2.1. All the figures extracted from the data can, however, be seen in Appendix 2.

#### 2.3.1 The copula 'BE'

The figures from some fields show a general uniformity with only one or two newspapers being significantly different for that feature. This is the pattern of occurrence of copula 'BE' verbs found for each sub-database as table 2.41 shows:

Table 2.41 Copula 'BE' in eight newspapers

	number	% (of 300)
Times	61	20
Telegraph	62	20
Guardian	67	22
Today	58	19
Mail	62	21
Mirror	54	18
Express	53	18
Sun	38	13
Total	455	19 (% of 2400)

Although these figures are very close, the Sun has a particularly low proportion of 'BE' verbs and the Guardian a slightly higher proportion than the other newspapers. There seems to be a tendency for the 'quality' papers (Times, Telegraph and Guardian in the present data) to have figures slightly above average and for the 'popular' papers (Mirror, Express and Sun) to have figures slightly below average. In order to test the hypothesis that some of the features investigated will differ according to newspaper type rather than individual newspapers, the  $X^2$  test was run both on the figures for individual papers and on the combined figures for three groups of papers. These groups were set up on the basis of observations from the pilot study and from considering the percentages worked out for the tables in this section some of which indicated a quality group, a popular group and a central group consisting of the Daily Mail and Today. The result of the first  $X^2$  test was that there was no significant difference between the papers when they were regarded individually

( $\chi^2 = 9.72$ ). When the papers were considered in three groups, the value of  $\chi^2$  was significant ( $\chi^2 = 6.39$ ).

One of the tendencies noted in this cross-sectioning of the data is that the papers considered to be in the centre of the quality - popular range, namely the Daily Mail and Today, fluctuate in their features so that they vary between seeming to belong to the quality group, the popular group and the middle ground. In the case of copula 'BE' verbs, the Mail is closer to the serious group while Today has a figure close to the average for the whole data.

The tendency for the quality papers to have more 'BE' verbs than the popular papers is not immediately reconciled with the pattern noted in section 2.2 where the articles had the lowest proportion of 'BE' verbs and the editorials the highest proportion. It was suggested there that the use of attributive constructions may be typical of a subjective style whilst the appearance of objectivity aimed at by article writers might lead to them avoiding such constructions. Because the popular press has a reputation for its subjective style of reporting, one would have expected the figures for copula 'BE' to be higher for this group. One hypothesis to explain this apparent discrepancy is that the difference lies in the distinction between overt and covert subjectivity. For example, the popular papers as a whole may use other, more covert strategies to convey their opinions whereas the

editorials as a group are not trying to hide their subjective nature.

### 2.3.2 Modal auxiliaries

Two other features of the verb phrase were found to differ between text-types in the previous section, and they also show some differences between papers. They are the incidence of modal auxiliaries and the proportion of passives in each sub-database. In the case of the modals the percentages seemed to show a difference between the quality papers and the popular papers. Although the newspapers would appear to split into two groups on the basis of the figures for modals given in table 2.41, the tables in this section as a whole show Today and the Mail as fluctuating between the extremes. I therefore decided to continue testing statistically as though there were three groups as described above in connection with the copula 'BE'. The  $X^2$  test was run both on the individual papers and the quality, central and popular groups.

Table 2.42 Modal auxiliaries in eight newspapers

	number	% (of 300)
Times	53	18
Telegraph	56	19
Guardian	44	15
Today	43	14
Mail	32	11
Mirror	23	8
Express	33	11
Sun	23	8
Total	307	13 (% of 2400)

The values of  $\chi^2$  for the individual papers and the three groups were 29.18 and 23.84 respectively, both significant at the 0.001 level. The three quality papers (Times, Telegraph and Guardian) have higher than average occurrence of modals in their verb phrases and the popular group have figures considerably below the average for the whole data. As with the copula, Today is close to the average. The Daily Mail, however, changes its allegiance and has a figure closer to those of the popular papers. The reason for the difference between quality and popular papers in their frequency of modals may be related to their desire for impact or accuracy. Adding a modal auxiliary to a verb phrase may diminish the impact but increase the accuracy as in the rewriting of an example below:

e.g. And they (his key words) are certain to be  
seized upon by Tory chairman Norman Tebbit  
[5 Ex Art]

they may be seized upon by Tory chairman Norman  
Tebbit [rewritten]

Omitting a modal, on the other hand, may reverse this process and increase the impact whilst committing the writer to a stronger statement:

e.g. That calculation may be wrong. [34 Ti Ed]

That calculation is wrong. [rewritten]

The modal, then, as well as being more evident in persuasive language (e.g. editorials) than in factual

reporting (e.g. articles) may also be important to the quality papers for avoiding the kind of inaccuracy which the sensationalism of the popular press often produces.

### 2.3.3 Passive verb phrases

Table 2.43 shows the figures for all of the passive structures<sup>7</sup> occurring in each of the newspapers:

Table 2.43 Passive verb phrases in eight newspapers

	number	% (of 300)
Times	40	13
Telegraph	47	16
Guardian	15	5
Today	41	14
Mail	17	6
Mirror	27	9
Express	36	12
Sun	23	8
Total	246	10

From the figures in table 2.43 it seems as though the three groups of papers which were shown to behave differently with respect to copula 'BE' and modal auxiliaries do not cohere in the same way as far as the passive is concerned. The quality papers have two figures which are above average for the total data and one (Guardian) well below average. The popular papers in contrast have two below-average figures and one above-average (Express). The group labelled 'central' in this section has one figure above (Today) and one below (Mail)

---

<sup>7</sup> Note that these figures do not include the non-finite passives.

the average. As expected, the  $X^2$  test produced a highly significant value for  $X^2$  (32.31) at the 0.001 level when run on the figures for individual papers. The value of  $X^2$  (1.65) when the three groups were tested was not significant as a result of the internal variation within the groups which was described above. Here we can only hypothesise that the passive has a highly stigmatised reputation as being typical of formal and objective styles and that the writers of Guardian texts, including the letter writers, use the passive sparingly in a rather extreme, and probably unconscious attempt to avoid its over-use. The converse may be true of the surprisingly high proportion of passive forms in the Express database. The writers may be aiming for what they see (or hope the reader will see) as an objective and authoritative style. More definite conclusions cannot be drawn from these figures alone.

#### 2.3.4 Verb phrase sequences

The number of verb phrases forming part of a VP sequence did not differ significantly across the three text-types investigated in section 2.2, but shows some differences between newspapers. The figures are given in table 2.44:

Table 2.44 Sequential VPs in eight newspapers

	number	% (of 300)
Times	42	14
Telegraph	53	18
Guardian	49	16
Today	46	15
Mail	42	14
Mirror	43	14
Express	33	11
Sun	65	22
Total	373	16 (% of 2400)

These figures were surprising because I had expected the papers to have different proportions of verb phrase sequences according to their groups, such sequences often being associated with the spoken or casual style of English. In fact, neither the individual papers nor the three groups were found to be significantly different when the figures for verb phrase sequences were tested statistically. The values of  $X^2$  for these tests were 13.43 and 0.43 respectively, neither being significant at the 0.05 level. It is clear, however, that the Express and the Sun have figures which differ considerably from the others. A detailed examination of the verb phrase sequences from the main database can be found in Chapter 5 where the formality of such sequences is considered. The explanation for the difference between numbers of sequences in the Express and the Sun may, however, be of the kind suggested for the passive figures above. This is the phenomenon whereby a newspaper aspires to a style normally associated with papers from a very different group. Here, one could suggest that the Express

unconsciously avoids the kind of construction which is readily associated with its stereotypical style and found in abundance in the Sun.

### 2.3.5 Subordinate clauses

One prediction that most linguists would make about the styles of quality and popular type newspapers is that there will be more clause subordination in the quality group. This prediction is upheld by the figures from field 9, clause type:

Table 2.45 Subordinate clauses in eight newspapers

	number	% (of 300)
Times	185	62
Telegraph	186	62
Guardian	182	61
Today	155	52
Mail	171	57
Mirror	139	46
Express	155	52
Sun	146	49
Total	1319	55 (% of 2400)

The figures in table 2.45 indicate that there is a fairly sharp split between the three quality papers and the other papers when subordinate clauses are counted. The tests showed that the differences were significant whether the papers were taken individually or in three groups. The values of  $X^2$  were 14.41 for individual papers and 13.11 for the groups, both significant at the 0.05 level.

The three quality papers are clearly separated from the others by this feature, although the Mail also has a

higher than average figure. It is interesting to consider the reasons why some features of 'formal' style correlate so clearly with popularly perceived groups of newspapers while others encourage the kind of over-correction noted for passive constructions and VP sequences. It may be the case that the use of subordinate clauses seriously complicates the reader's job in unravelling the sense of a text and is not as salient a feature for the average reader as the passive construction or VP sequences.<sup>8</sup>

Within the class of subordinate clauses the small figures do not allow much interpretation. Three of the classes, however, were thought to be worth investigating; the figures for noun clauses, adverbial clauses and relative clauses are given in table 2.46:

**Table 2.46 Three classes of subordinate clause in eight newspapers**

	NCl	% NCl	RCl	% RCl	ACl	% ACl
Times	42	23	30	16	58	31
Telegraph	45	24	47	25	34	18
Guardian	42	23	33	18	37	20
Today	29	19	32	21	43	28
Mail	39	23	26	15	54	32
Mirror	36	26	25	18	42	30
Express	33	21	32	21	43	28
Sun	23	16	34	23	37	25
Total	289	22	259	20	348	26
8 papers	$X^2 = 4.96$		$X^2 = 7.37$		$X^2 = 11.69$	
3 groups	$X^2 = 0.87$		$X^2 = 0.82$		$X^2 = 3.66$	

<sup>8</sup> Note, however, that Huddleston et al. (1968) found that their 'high stratum' of scientific English contained the highest number of simple sentences. (p647)

The figures in table 2.46 show that there is no straightforward relationship between type of newspaper and proportion of different clause types. The  $\chi^2$  tests revealed no significant differences between the papers, whether they were taken individually or in groups.

### 2.3.6 Subject type

Field 12, subject type, demonstrates a fairly clear relationship between features of the subject and groups of newspapers. The first subfield shows the quality and popular papers as having different proportions of inanimate, human and abstract subject referents. The figures for animate referents were too small to be of use here.

Table 2.47 Subject types in eight newspapers

	inan	% inan	human	% hum	abs	% abs
Times	7	2	103	34	158	53
Telegraph	7	2	109	36	159	53
Guardian	2	1	119	40	151	50
Today	19	6	133	44	115	38
Mail	25	8	145	48	98	33
Mirror	39	13	150	50	95	32
Express	27	9	142	47	111	37
Sun	16	5	201	67	63	21
Total	142	6	1101	46	950	40
8 papers	$\chi^2 = 60.48$		$\chi^2 = 48.17$		$\chi^2 = 70.33$	
3 groups	$\chi^2 = 43.62$		$\chi^2 = 31.81$		$\chi^2 = 58.95$	

All of the  $\chi^2$  tests produced significant values for  $\chi^2$  at the 0.001 level. This feature, then, is shown to be important both in the styles of individual papers and of

groups of papers. It should be noted that this use of the word 'style' does not describe the limited sense where a straightforward choice is assumed in every case (see chapter 1 for a discussion of the term 'style'). The feature of subject type is close to the borderline between content and style since it is presumably the wider choice to report human stories that produces such high proportions as 67% human subject referents in the Sun. However, there are many intermediate levels of choice which would be difficult to allocate to discrete style and content categories and since the features coded by this subfield are often taken to be on the borderline between grammar and semantics, they may also be on the borderline between content and form.

The second subfield of field 12 codes the human subjects according to their sex, if identifiable. Table 2.48 gives the figures for female and male subjects in each newspaper and shows them as a percentage of the total sex-identified subjects for that paper. One clear conclusion to draw from this table is that all of the papers identify a much higher proportion of their subjects as male than as female.

Table 2.48 Female and male subjects in eight newspapers

	female	% fem	male	% male
Times	2	5	35	95
Telegraph	1	2	53	98
Guardian	3	6	46	94
Today	11	17	55	83
Mail	28	33	57	67
Mirror	14	19	58	81
Express	11	13	40	87
Sun	36	47	71	53
Total	106	20	415	80
8 papers	$X^2 = 54.28$		$X^2 = 13.87$	
3 groups	$X^2 = 24.28$		$X^2 = 6.2$	

From these figures one can see that there is a difference in the proportions of female and male sex-identified subjects in the groups of newspapers. The value of  $X^2$  found for the individual papers and the groups was significant for the female subjects at the 0.05 level, but not for the male subjects. The popular papers identify a higher proportion of female subjects than the quality papers in this data and the central group of papers (Today and Mail) resemble the popular group for this feature. There may be a number of reasons why these papers mention women more than the quality papers. The popular papers have a reputation for their contribution to the various stereotypes of women as sex-objects, housewives and mothers and this may be one factor in the higher proportion of female subjects, although it should be remembered that the news stories were chosen as occurring in all of the papers concerned and cannot be their subject

matter alone be held responsible for the difference between the groups of papers. Another contributing factor may be the number of women featuring in the anecdotal letters which are typically written to the popular press:

e.g. Mum went to the chemist to buy Dad some deodorant to use after jogging.

[36 S Let]

The very low frequency of subjects identified as female in the quality papers may be a result of the kinds of subjects which these papers typically treat, including politics, financial affairs, business and industry all of which are most concerned with, and run by, men. A more detailed analysis of the sex-identified subjects would require a close analysis of a large number of examples. Such a study is not undertaken in the present thesis because of the small number of examples available from the main database and because it is not central to the argument of the thesis.

#### 2.3.7 Subject form

The last feature of the subject to show any significant differences between newspapers is the occurrence of personal pronouns as a proportion of all subjects coded in field 14, subject form (see table 2.49).

Table 2.49 Personal pronouns as subject in eight newspapers

	prons	% (of 300)
Times	44	15
Telegraph	31	10
Guardian	46	15
Today	59	20
Mail	57	19
Mirror	68	23
Express	56	19
Sun	76	25
Total data	437	18 (% of 2400)
8 papers	$X^2 = 25.77$	
3 groups	$X^2 = 19.6$	

There is a wide difference between the quality and the popular papers with the central group again having figures which reflect the popular rather than the quality group. Both the  $X^2$  test on individual newspapers and that on the three groups showed that the difference was significant at the 0.001 level. The quality papers show a lower than average proportion of pronouns which may indicate a concern to avoid any ambiguity of reference. Alternatively, these figures may result from a difference in the depth of reporting of the two groups. If the popular papers contained less detail, and therefore fewer 'characters' in each text, there would be less chance of pronouns having ambiguous reference.<sup>9</sup>

---

9 It should be noted that the figures for the whole data fit into the pattern of pronoun (and name) distribution recorded in Quirk et al. (1985). Table 17.124 (p1251) presents figures for four types of text: informal speech, fiction, serious talk and scientific writing. There appear to be some mistakes in the lower four boxes of the first column of this table, the figures for which should be obtained by adding the two columns labelled

One interesting feature of the other categories of subject form coded in field 14 is that the pattern noted in Table 2.49 is not repeated elsewhere. One might have expected that the quality papers would reflect their lack of pronouns in a higher proportion of fully worded subjects and/or proper nouns, but this is only partly the case. As

---

'All'. The lower boxes of the first column would then look like this:

2212
1980
2431
2803
2088
2511
1167
1769

From these re-calculated figures, it was possible to calculate the percentage of subject noun phrases which are realised by names or personal pronouns. In this way I obtained figures which can be compared with the joint percentage of personal pronouns and names in my data. The figures (calculated from Quirk et al.) are as follows:

	names and pronouns
% of all NPs	47.2
% of all subj NPs	73.7
% of subj NPs in informal speech	87.7
% of subj NPs in fiction	79.8
% of subj NPs in serious talk	70.8
% of subj NPs in scientific writing	39.3

At about 27%, the proportion of names and pronouns in my data is even lower than that found in the scientific texts of the Survey of English Usage. It therefore appears to place Newspaper English toward the formal end of the formality gradient for this feature. None of these facts, however, invalidate the observations of internal differences of pronoun distribution found in my data.

the figures in Table 2.50 show, the Times and the Telegraph have a higher than average frequency of fully worded subjects, but the Guardian's low number of pronouns is compensated not by non-pronominal (full) subjects but by a relatively high number of impersonal subjects:

Table 2.50 Proper nouns, impersonal and full subjects in eight papers

	PN	% PN	full	% full	imp	% imp
Times	24	8	96	32	9	3
Telegraph	24	8	108	36	12	4
Guardian	37	12	70	23	20	7
Today	26	9	93	13	12	13
Mail	17	6	78	11	14	16
Mirror	20	7	93	13	5	6
Express	25	8	89	13	13	14
Sun	31	10	78	11	5	6
Total data	204	9	700	29	90	4
8 papers	$X^2 = 10.59$		$X^2 = 11.91$		$X^2 = 15.24$	
3 groups	$X^2 = 2.20$		$X^2 = 0.62$		$X^2 = 5.53$	

Unlike the case of personal pronouns discussed above, the figures for proper nouns, impersonal subjects and fully worded subjects were not found to be significantly different for the three groups of papers. When they were tested for differences between individual papers, only the impersonal pronouns were found to differ significantly among the papers. A larger sample would be needed to make a thorough investigation of the interaction of these different subject forms which do not appear to stand in a simple relationship with newspaper type.

### Chapter 3

#### The copula 'BE' and its environment

After completing the stylistic and statistical survey of the main database described in Chapter 2, I changed the emphasis of the study in two ways. Firstly, the focus was now on the problems involved in description rather than on description itself. From the first stages of the pilot study, I had realised that the areas of interest in the data were at least twofold. The data provided insights into the differences between different styles of newspaper English (chapter 2) and in addition the process of describing it highlighted problems which had to be confronted. Some of these problems form the subject matter of chapters 3, 4 and 5. The second change in emphasis was that the database was no longer being used as the basis of a contrastive study of newspapers and newspaper text-types. Instead, it was regarded as a sample of newspaper English which also represents written English more generally. The choice of which descriptive problems should be studied in depth was made partly on the basis of the results in Chapter 2 and partly on the basis of the pilot study. The features chosen are all difficult to describe in a purely formal way and share the characteristic of occupying the border area between syntax and semantics. Although the observations made in these chapters cannot necessarily be generalised to English as a whole, I hope that the discussion of descriptive, and at times

theoretical, problems will contribute to the debate on these problems and provide illuminating examples from attested data.

The first feature to be investigated in detail was the copula 'BE' and its surrounding context. The pilot study had raised the question of the difference, apart from position, between noun phrase subjects and noun phrase complements and it seemed worthwhile to consider this issue by analysing the large number of examples in the main database.

The first step was to categorise the following contexts of the copular verb (see section 3.1). This categorisation was useful for the thesis in two ways. First, it helped to locate both 'true' adjectives and participle forms analysed as adjectives. These are discussed in the examination of the participle/ adjective borderline which can be seen in Chapter 4. Secondly, it provides data about the forms of complements associated with the verb 'BE' in one kind of written English.

The second step was to investigate the forms of all the NP complements and NP subjects, comparing aspects of their structure in general terms and also comparing the subjects and complements in pairs linked by the copula (sections 3.2 and 3.3).

### 3.1 Types of following context

The first stage in examining the following context of the copula 'BE' was to extract from the database all examples which had the verb 'BE' specified as the main verb. There were 455 such examples out of the total of 2400 verb phrases. The output was formatted to show all the fields associated with each example so that there was easy access to all the information (see Appendix 1e for examples).

The sentence contexts for each example were consulted and in each case the kind of clause element following the verb was noted. On the basis of the pilot study, I expected to find examples from the following categories:

#### noun phrase

e.g. Whether such sense pertains at the Liberal  
assembly in Eastbourne next week is another  
matter  
[16 Te Ed]

#### adjective phrase

e.g. The Liberals are content to postpone the  
question of replacing Polaris  
[18 Ti Art]

#### prepositional phrase

e.g. this would be on seven different floors  
[21 Te Let]

#### noun clause

e.g. whose primary concern was that President Reagan  
should meet Mr Gorbachov  
[30 Ti Ed]

#### adverbial clause

e.g. This is where the blame should be laid  
[7 M Let]

adverbial phrase

e.g. A summit may not always be just ahead  
[39 Ti Ed]

I also allowed for a 'problem' category to mark those cases which were not immediately identified as belonging to the above categories. The problems which arose were all eventually assimilated into one of the other categories. The examples over which I hesitated in this way can be illustrated by:

e.g. ...members of the RIBA have not been the most  
vocal in opposition to the idea of...  
[25 Te Let]

In the above example, the head word of the complement is an adjective, but the phrase was nevertheless classed as a noun phrase.

As noted in Chapter 2, there was a significant difference between the proportion of main verbs realised by 'BE' in the articles (9%) as compared with the editorials (26%) and the letters (22%). The categories of complement were investigated for these sub-databases as well as for the total data but although I had found differences of quantity between the three text-types, they showed no significant difference of kind. The figures for each type of following context in the main and pilot databases are given in table 3.1 below, and they are also given as percentages of the total number of 'BE' verbs in each database:

Table 3.1 Following context of copula 'BE'

	pilot	% (139)	main	% (455)
NP	62	42	200	44
AjP	57	38	169	37
PP	10	7	36	8
NCl	17	11	28	6
?	0	0	15	3
ACl	0	0	5	1
AvP	3	2	3	1

Many linguists make a distinction between the grammatical functions of 'complement' (adjective and noun phrases) and 'adverbial' or 'adjunct' (prepositional and adverbial phrases) following BE.

complements:

e.g. She is an engineer.  
He is very tired.

adjuncts:

e.g. They are in the garden.  
They are inside. [invented]

Although there are clearly basic differences in the kind of information given about the subject by these types of phrase, there are areas of overlap in semantic terms, as exemplified below:

e.g. John is a pauper.  
John is penniless.  
John is out of money. [invented]

Those prepositional phrases which have similar semantic force to adjectives and nouns in this way tend to be either idiomatic or at least used metaphorically. However, the point remains that there is an unclear boundary between the forms which fulfil the functions of complement

and adjunct.<sup>1</sup> A more satisfactory approach would be to regard all following contexts of the verb 'BE' as complements, sub-dividing this larger group into sub-groups partly on the basis of form and partly on the basis of meaning. Huddleston (1971), for example, makes a distinction between the intensive use of 'BE' which is followed by an 'attribute' and what he calls 'equative' uses of the copula which is followed by an 'object', but he adds:

"However, the difference is not 'signalled' by the contrast between Adj Phr's and NP's - the former are normally attributes, but the latter occur freely as either attributes or objects."

Huddleston's divisions are partly formal (adjectival phrases can only be attributive) and partly semantic (nominal phrases can occur as attributes or objects).

We will return to Huddleston's analysis in the discussion of following NPs (see section 3.2). As an initial stage in my analysis it was considered useful to categorise following contexts on the basis of form but not taking the division into complements and adjuncts as basic.

The figures in table 3.1 show that a large majority of examples containing the main verb 'BE' have adjectival or nominal complements following the verb phrase. This

---

<sup>1</sup> See Quirk et al. (1985), section 10.11 (p732) for discussion of the gradience between adjunct and complement.

tendency may be typical of the data rather than of English more generally; in particular it was surprising to find so few adverbial complements to the copula.

### 3.2 NP complements

The pilot study had indicated that examples with NPs following the verb 'BE' may be of interest in the discussion of the borderline between subject and complement functions. The investigation of NP complements<sup>2</sup> occupies the remainder of the present chapter.

Having identified those clauses which contained BE followed by a noun phrase complement, I investigated their structure and the structure of the subjects associated with them. For comparison, I also looked at the structure of all the subjects preceding BE (regardless of complement type) and of all the subjects in the database.

Some well-known features of NP complement structure, as compared with NP subject structure, had been supported by the pilot study. These were the tendency for noun phrases occurring late in the clause, such as complements, to be more complex and more indefinite than earlier noun phrases, such as subjects. On the topic of complexity,

---

<sup>2</sup> The term 'complement' is being used in a very general way here to indicate any following context of the copula. As implied in section 3.1, I consider that either the whole group of following contexts or sub-groups divided according to form are the only valid bases for an investigation. Here I am concerned with those complements which have a noun phrase form.

Bock (1982) produced evidence to show that it is easier to process structurally simple items first and structurally complex items later in the clause, since the simpler items are processed automatically and the complex items need controlled processing. Givón (1979) made a study of definiteness in English fictional and newspaper texts and found that a large majority (94%) of indefinite direct objects were preceded by definite subjects.

One of my intentions in this section was to investigate the database to see whether it conformed to these observations about definiteness and complexity. A related, but potentially more interesting aim was to look at examples where these general differences of structure did not occur. The interest in such examples rests in the fact that the distinction between the functions of subject and complement begins to blur when differences of structure disappear.

### 3.2.1 The problem

One of the most interesting problems that arose during the pilot study was the case of the main verb 'BE' whose following NP complement could just as easily have been its subject. The example was:

e.g. the victims will be the miners and power  
workers [30/G/19]

As noted in the Introduction (section 1.3.8) this example was shown to be the only case in the pilot data where the

possibility of reversing the subject and complement noun phrases arose.

If the function and meaning of the copula 'BE' is sometimes likened to that of the equals sign in mathematics, where the two sides of the equation are in some sense balanced, this kind of example is close to supporting that simile.

As Huddleston (1971) points out, there are some uses of the verb 'BE' which are 'equative' in this way and differ from the more common intensive use of 'BE'. The latter attributes some characteristic to the subject rather than emphasising the equivalence of the referents of subject and complement.

If this 'equality' between subject and complement were widespread, the distinction between these grammatical functions would begin to break down; indeed it would rest on the tendency for subjects to precede verbs and for complements to follow them. But word order is not a wholly reliable indication of grammatical function, as the following example, with its adjectival complement before the verb and subject following, shows:

e.g. more fundamental by far are the arguments that  
 any nuclear arms are militarily useless, a waste  
 of human resources and by their nature a threat  
 to the survival of humanity. [11 G Let]

In the majority of cases (180 out of 200 = 90%), however, there seems to be an imbalance between subject and

complement which makes NP subjects and NP complements distinguishable in context if not in theory. In the pilot data, the subjects and complements differed in definiteness and/or complexity in all but one case. In the remainder of this chapter, the main database is investigated to see whether there is a general difference of definiteness and complexity between the subjects and complements and to consider whether individual subject-complement pairs uphold the distinction between subject and complement functions.

### 3.2.2 The investigation

Nearly half of the examples containing main verb 'BE' had NP complements (200 out of 455) and their subjects and complements were examined to see what differences of structure were evident.

A large majority of the subjects (134 out of 200) were not realised as full noun phrases, but were divided among the other subject categories of field 14 as shown in table 3.2. The equivalent figures for subjects of all copula BE verbs and all subjects in the data are included for comparison.

Table 3.2 Subjects not realised as full NPs or noun clauses

	BE + NP	% 200	tot BE	% 455	tot data	% 2400
impers pn	59	30	78	17	90	4
n/r	22	11	35	8	781	33
pers pn	30	15	74	16	441	18
dem pn	13	7	21	5	38	2
proper nouns	8	4	41	9	275	11
rel pn	2	1	14	3	114	5

As the figures in table 3.2 show, there are some differences in realisation between subjects preceding BE + NP, subjects preceding all BE copula verbs and subjects in the total data. The total data group, of course, includes the BE group which in turn contains the BE + NP group. It would not be valid, therefore, to perform statistical tests on smaller and smaller subdatabases treating them as though they were independent. However, the differences are both clear and relatively simple to explain. Subjects preceding the copula when it is complemented by a noun phrase are seen to be realised as impersonal pronouns more often (30%) than the subjects in the other groups (17% and 4%). The data reveals that the constructions 'It is...' and 'There is...' are very common:

e.g. It would be a poor nose which didn't smell a rat  
[10 Mi Ed]

The 'impersonal' subjects differ from each other in the following way: 'it' sets up an intensive relationship between the complement head ('a poor nose') and the relative clause ('which didn't smell a rat'). 'There', on the other hand, simply establishes the existence of the complement (exemplifying what Huddleston (1971) calls an 'extensive intransitive' use of the verb 'BE'). However, they share the property of influencing the structure of the clause towards end-weight and end-focus. In the example above the impersonal pronoun enables a long and

highly modified noun phrase to occur toward the end of the clause.

The BE + NP constructions also seem to attract more than the average number of demonstrative pronouns as subject:

e.g. That, at least, is the appearance. [11 Ti Ed]

This construction is probably most often used, as in the example above, to place the emphasis on new information (see section 3.2.6 for a discussion of end-focus). Other differences observed from table 3.2 show the BE + NP examples as having fewer personal and relative pronouns and fewer proper nouns than the other groups. These differences are also explained by the need to place both emphasis and complexity toward the end of English clauses:

e.g. The mastermind behind the gang, claimed counsel,  
was Peter Dye, who used a company, Golden Falcon  
Ltd, with offices "believe it or not" in Great  
Scotland Yard. [32 Te Art]

The subject in this example is quite long, but the complement is even longer and more complex; it would be difficult to understand had the subject and complement exchanged roles.

Although any English speaker can imagine examples where the main verb 'BE' is followed by a complement drawn from table 3.2 (e.g. "the main point is this" or "it was him"), there were no such examples in the present data, with the exception of a few proper nouns which occurred as

complements. The lack of pronominal complements may be a feature of the particular data being analysed here, or of the written language more generally since the above examples would seem 'normal' in the spoken language. This absence of pronouns in clause-final complements may also represent a general difference between subjects and complements. The finding that complements are more likely to be realised by 'heavy' and subjects by 'light' noun phrases supports the conclusions of Aarts (1971), who investigated subject and non-subject noun phrases in extracts from the files of the Survey of English Usage. On the 'light' side, his findings were that personal pronouns were typical of 'subjectness' for all of his data. Other pronouns were typical subjects for scientific writing and formal English whilst names were typical subjects for light fiction. The 'heavy' exponents, i.e. noun phrases pre-or postmodified, were shown by Aarts to be typical of non-subject functions. The latter finding is upheld by my data and discussed in section 3.2.3 below.

Allan (1987) discusses a number of hierarchies affecting the ordering of noun phrases. He is mainly concerned with the ordering of conjuncts, but implies that ordering among noun phrases more generally, such as between subject-object and subject-complement, would adhere to the same rules. Allan includes the hierarchy of 'definiticity' which dictates that definite noun phrases have priority (= left placing) over indefinite noun phrases:

e.g. The £29 million given out by the Arts Council  
 is a paltry sum in comparison to the amount  
 spent on weapons each year [21 To Let]

Allan also discusses the 'formal' hierarchies which ensure that structurally simple noun phrases have priority over those which are structurally complex. There are also, according to Allan, priorities between the hierarchies, definiticity having priority over the formal hierarchies. This relationship between hierarchies is supported by the present data and discussed later in this section.

To compare the full NP subjects with NP complements, I counted the number having definite or indefinite articles and the number containing other premodification (enumerators, quantifiers, adjectives or nouns) or postmodification (adjectives, prepositional phrases or postmodifying clauses). The figures for definite and indefinite articles are given in Table 3.3:

Table 3.3 Definiteness of subjects (+ BE + NP) and NP complements

	subject	% (200)	NP comp	% (200)	
indef	1	0	70	35	$X^2 = 67.06$
def	33	17	41	20	$X^2 = 0.86$

Although these figures do not show the direct relationship between individual subjects and complements (see sections 3.2.3 to 3.2.6), they indicate the well-known tendency for NP subjects to contain few indefinite articles and for NP complements to contain a significantly higher number of

indefinite articles. It could be argued that this difference can be explained by the different uses (or meanings) of the indefinite article in subjects and complements respectively. In many cases, for example, complements containing indefinite articles have a descriptive function known as 'non-referring'. Subjects containing an indefinite article, in contrast, are often assumed to have a definite referent as in 'A man phoned you today'. However, indefinite subjects of copular clauses tend to have general reference to a class which is similar to the 'non-referring' use mentioned in connection with complements. An example of such a subject might be: 'An oak is a beautiful tree'.<sup>3</sup> Differences of use of the indefinite article may not, therefore, be the main explanation for their different frequency in subjects and complements.

The  $X^2$  test comparing NP subjects and NP complements for definiteness produced a significant value for indefinite articles. The value of  $X^2$  for definite articles, on the other hand, was not significant at the 0.05 level. This similarity in the occurrence of definite articles between NP subjects and NP complements was surprising, since the difference in occurrence of indefinite articles tends to lead to the assumption that the converse will apply in the case of definite articles.

---

<sup>3</sup> I am grateful to G. Sampson for discussing this point with me and contributing this example.

When the figures in table 3.3 are compared with those for subjects in the total data, the number of definite articles for both groups is seen to be close to the average ( $447/2201 = 19\%$ ). The number of indefinite articles in complements, however, is considerably higher than the average in all subjects ( $80/2201 = 4\%$ ) and the subjects occurring with those complements also differ from the average in having 0% indefinite articles.

On a stylistic note, the proportions of subjects and complements containing definite articles in the present data are lower than those found by Varantola (1984) in both general English and engineering English. In that study, about 30% of cases contained a definite article which compares with an average of 19% in my data. Söderlind (1962 - reported in Varantola) found that styles which aimed at brevity (including newspaper style) often chose a form with zero article and that, in a familiar context, such zero forms may have specific reference for the reader. Varantola's data on indefinite articles showed an average of 18% of cases containing indefinite articles. This compares with 35% of NP complements and 4% of subjects in my data. Since Varantola's data is not divided between subjects, complements and other NPs, it may be that her percentage conceals variation similar to that found in my data.

The modification of noun phrases as subject and as

complement was investigated statistically by considering each example and assigning it to one of the four categories; premodified, postmodified, pre- and post-modified and unmodified. The figures for these categories are presented in table 3.4:

Table 3.4 Modification of subjects (+ BE + NP) and NP complements

	S +NP	% 200	NP comp	% 200	
premod	18	9	60	30	$X^2 = 22.62$
postmod	17	9	65	33	$X^2 = 28.1$
pre-/post-mod	9	5	55	28	$X^2 = 33.06$
unmodified	156	78	26	13	$X^2 = 92.86$

The extremely high values for  $X^2$  in table 3.4 (all significant at the 0.001 level) confirm the striking differences in modification between subjects occurring with noun phrase complements and the noun phrase complements themselves. A large majority of the subjects (78%) are unmodified whereas a total of 91% of the NP complements have modification of some kind. If the figures for these 200 subjects are compared with those for the total number of realised subjects, we can see that subjects followed by copula BE and a noun phrase complement are not typical of the data as a whole:

Table 3.5 Modification of subject (+ BE + NP) compared with all subjects

	S +NP	% 200	tot S	% 2201	
premod	18	9	565	26	$X^2 = 21.65$
postmod	17	9	49	2	$X^2 = 35.26$
pre-/post-mod	9	5	178	8	$X^2 = 0.03$
unmodified	156	78	1409	64	$X^2 = 6.11$

Although the figures for both pre-/post-modified subjects and unmodified subjects do not differ between the two groups (values for  $X^2$  are not significant), the figures for premodification and postmodification individually differ significantly. In other words, subjects followed by the copula and a noun phrase complement are more likely to be postmodified and less likely to be premodified than subjects as a whole. Neither group of subjects, however, has as much modification of any kind as the group of noun phrase complements. This finding is not surprising in view of the large number of studies referring to the tendency for 'light' elements to occur earlier and 'heavy' elements later, in the clause (see, for example, Aarts 1971, Halliday and Hasan 1976 and Allan 1987). One study, by Yngve (1961), illustrates how the short term memory of people constrains the grammar of English from too much left-branching phrase structure which would result in complex items at the beginning of clauses while simple items occur toward the end.

Having established that my data conformed to the general pattern of light elements preceding heavy ones, the direct relationship between subjects and complements as they occur in context was investigated. First, I checked every data item containing a noun phrase complement to ensure that the basic hypothesis was correct; that no complement would be more definite than its subject and no subject more modified than its complement. The majority of the

examples (180/200) were easy to identify as supporting this hypothesis. They are illustrated by the following example:

e.g. They were the work of Arab guerillas seeking freedom for a Lebanese terrorist jailed in France for four years. [16 S Art]

Most of the items were similar to this example and had a pronominal subject followed by a complex complement. A few examples, like the one given below, show that an indefinite article will occur in the complement even though the subject is longer and more complex than the complement:

e.g. The recent demand for greater sensitivity in treating prisoners is a complete joke. [36 To Let]

This example supports the claim in Allan (1987) that the definiticity hierarchy takes precedence over the formal hierarchies.

After setting aside the cases where the hypothesis was clearly supported, I looked at the group of NP complements which contain definite articles or proper nouns to see whether they would compare with the example from the pilot study quoted above, where the subject and complement seemed to be interchangeable. The hypothesis was that since there is a general difference of definiteness and complexity between NP subjects and NP complements, any complements containing definite articles would be expected

to cooccur with subjects at least as definite as them and probably less complex. This means that the subjects would also contain definite articles (with countable nouns) or proper nouns which have a specific reference and are therefore semantically definite. Where both the definiteness and complexity of subject and complement are comparable, I expected the distinction between the syntactic functions of 'subject' and 'complement' to be least clear.

There were twenty examples in the main database which contained a complement with definite article. They were investigated to see whether any of the examples had subjects and complements which were not only equally definite but also equally complex in structure. They are discussed below in groups of cases which appear to have similar properties distinguishing between their subjects and complements.

### 3.2.3 End-weight

The data contained a number of examples whose subjects and complements could be judged equally definite. That is, they contained a definite article, a possessive pronoun or a proper noun with specific reference. Some of these, however, had complements which were rather complex and were therefore unlikely to be interpreted as subjects. The well-documented preference in English (and other languages) for long and complex elements to occur toward

the end of clauses as well as the conventional positioning of subject before the verb and complement after, preclude the possibility of reversing the identities of subject and complement in such examples.

The first three examples given below share the feature of having a proper noun as head of their complements. In each case, the distinction between subject and complement might have been obscured if the proper noun had not been postmodified by a relative clause, a prepositional phrase and a noun phrase in apposition respectively. This fairly extensive postmodification, I would suggest, overrides the equal definiteness observed between subject and complement and makes the assignment of grammatical functions unambiguous.

e.g. The mastermind behind the gang, claimed counsel, was Peter Dye, who used a company, Golden Falcon Ltd, with offices "believe it or not" in Great Scotland Yard. [32 Te Art]

e.g. It is alleged that the brains behind the operation was businessman Paul Dye, 41, of The Poynings, Richings Park, Iver Bucks. [42 To Art]

e.g. His right-hand man, it was alleged, was Clive Williamson, 29, an electrician of Northolt, Middlesex. [31 M Art]

The examples above all correspond to what Huddleston (1971) calls the 'equative' use of the verb 'BE' for which he proposes the arguments 'identifier' and 'identified'. These are like deep cases since they refer to the semantic function of the arguments and can be associated with

either the subject or the following NP (which Huddleston calls 'object' in this use of the verb, but which is part of the large class of 'complement' in this study). These examples all have the 'identifier-as-subject' format which is the unmarked version, according to Huddleston. If the subject and complement had their positions reversed, they would have the marked 'identified-as-subject' format:

e.g. Counsel claimed that Peter Dye, who used a company, Golden Falcon Ltd, with offices "believe it or not" in Great Scotland Yard, was the mastermind behind the gang.  
[rewritten - 32 Te Art]

In fact, as Huddleston points out, this format is virtually indistinguishable from (if not identical to) the intensive use of the verb 'BE' which attributes a characteristic to the subject rather than equating two NP referents. However, the most obvious effect of rewriting examples in this way is that the subject of the subordinate clause in the new version is over-long and complex. It may also be significant that the 'BE' clauses in these examples are all subordinate to the main clause. More examples are needed to test the hypothesis that violation of end-weight within clauses is less acceptable in subordinate than in main clauses.

The following example is similar to the above, but its complexity is found not in postmodification but in the conjunction of three proper nouns as equal heads of the NP complement:

e.g. The other alleged couriers are Ahmed from Pakistan, Graham Ellis from Whitton, Middlesex, and David Millard, 37, from Peterborough.  
[47 To Art]

In all of the examples discussed so far, one could argue that the information structure of the clauses is also important, the focus on new information in each case falling on the names of the accused. The reversed versions would only be suitable in a context where the people concerned had already been mentioned or were part of shared knowledge. Nevertheless, the argument here is not simply whether the subject and complement phrases can be reversed physically but whether there is a point where the functional distinction between subject and complement disappears.

The set of examples where end-weight seems influential in the ordering of NPs includes one with a rather long appositive noun phrase which emphasises the descriptive nature of complements:

e.g. The latest victim of its vendetta is the Monocled Mutineer, a brilliantly acted, beautifully written but violent drama about a British army mutiny in 1917. [3 Mi Ed]

Without this appositive phrase, the example would be easily reversed with only a minor difference of focus. The length of the phrase, and the extent of adjectival premodification, however, emphasise the attributive<sup>4</sup> nature

---

<sup>4</sup> I am using 'attributive' in a non-technical sense.

of the phrase.

In contrast, the following example has both premodification and postmodification in its complement, but the conjoined superlative adjective phrases seem to serve equally well as indicators of a descriptive complement:

e.g. Khomeini's Iran is possibly the cruellest and  
the most fanatical regime in the world today  
[25 To Ed]

As well as a postmodifying prepositional phrase, the complement in the next example has an adjective, 'vocal', as its head. The presence of adjectives, whether as heads or as premodifiers, seems to be one of the factors influencing the assigning of NPs to subject and complement categories. Perhaps the existence of adjectival complements in general (e.g. She is sad) would incline us to analyse such highly descriptive phrases as complements:

e.g. members of the RIBA have not been the most vocal  
in opposition to the idea of a move back to  
Portland Place [25 Te Let]

The examples in this section have shown that in many cases where the subject and complement are equally definite, the complexity of the complement distinguishes it from the subject.<sup>5</sup> Noun phrases containing adjectives, as well as

---

<sup>5</sup> Clearly the usual ordering priorities of subject before the verb and complement following also have a great influence, but here I am concerned with other, less obvious, distinctions.

being more complex than their corresponding subjects, also influence their interpretation as complements so that it would be difficult to rewrite the examples with subject and complement reversed.

#### 3.2.4 Negation and number

The first example in this section shows that negation of the verb 'BE' can have the effect of 'unbalancing the scales' in the same way as end-weighted complements. In testing the 'equality' of subject and complement in copula clauses by reversing these elements the presence of a negative verb changes the sense of the clause quite dramatically:

e.g. that Gadaffi is not the only ogre in the world  
[30 To Ed]

the only ogre in the world is not Gadaffi  
(reversed)

This example illustrates the important relationship between given information and the first main clause element. In the rewritten version, it is taken for granted that there is only one ogre and the question is one of his, or her, identity. The original version takes for granted the equating of Gadaffi and "ogre", but questions whether there are other ogres.

The other negative example has a difference of number between the subject and the complement which makes a reversal of the clause unacceptable (in written English)

as well as different in its focus:

e.g. Constant bloodthirsty remakes are not the sign  
of artistic integrity [34 To Let]

?the sign of artistic integrity is not constant  
bloodthirsty remakes (reversed)

There is a difference between these versions in the assumptions they make. The original takes for granted that the reader would agree with 'constant bloodthirsty remakes' as a description whereas the rewritten version takes the existence of a 'sign of artistic integrity' as given. However, the main interest of this example lies in the apparent violation of number concord between subject and complement. Since the 'rules' of number concord state that concord exists between the subject and the verb (Quirk et al 1972), it is certain that a reader would identify "constant bloodthirsty remakes" as the subject in this example, the complement being in the singular. The, for me, unacceptable attempt at reversal suggests that subject-complement concord of number (as opposed to subject-verb concord) may be more easily violated when the subject is plural and the complement singular or uncountable. A famous example of this kind occurs in the Bible and is also quoted in one of the data items:

e.g. The wages of sin are death [29 Te Ed]

In the Grammar of Contemporary English (Quirk et al. 1972), number concord between subject and complement is said to arise "naturally from the denotative equivalence

of subject and subject complement." The authors explain exceptions to this concord by claiming that complements in such examples, though nominal in form, are adjectival in function. This is a very difficult argument to uphold since most nouns seem adjectival when they occur in a position which is shared with adjectives. It is, for example, mainly because we know that the word 'poison' occurs in other positions more typical of the noun class as a whole, that we identify it as a noun even in positions considered central to the adjective function, such as after BE and premodifying nouns:

e.g. They are electoral poison                    [47 S Ed]  
       (compare: They are poisonous)            (invented)

e.g. poison pen            (compare: big pen)    (invented)

In order to see whether subject-complement concord is violated equally with singular and with plural subjects, I went through the 200 data items which have NP complements, recording the number of each subject and corresponding complement. The numbers of items showing each combination of subject and complement are shown in table 3.6:

Table 3.6 Number concord / discord in subjects and complements

subject	complement	number	
sing	sing	103	] — concord
plural	plural	17	
uncount	uncount	13	
sing	uncount	23	
uncount	sing	6	
impers	any	29	
plural	uncount	4	] — discord
uncount	plural	0	
plural	sing	5	
sing	plural	0	

As expected, there was a large number of examples (189) which displayed only the usual concord between subject and complement. Although many writers fail to make it clear, it seems reasonable to count mixtures of singular and uncountable NPs as having normal concord since they both typically occur with a singular verb. Notice, however, that the combination of singular subject with uncountable complement is more common, in this data, than the reverse. The last four combinations, in contrast, mix plural NPs with singular or uncountable NPs and there is no obvious reason why they should not all be equally unacceptable to the English speaker since they all violate number concord between subject and complement. The complete absence of combinations of singular or uncountable subject with plural complement contradicts this expectation and reinforces my dissatisfaction with the attempted reversal of the example above.

It would be unwise to make any far-reaching

generalisations from such a small number of examples, but there seems to be a tendency, in these data at least, for any violation of subject-complement concord to have a plural subject and non-plural complement. Three illustrations from the data follow; the first has a plural subject and uncountable complement and the second and third have a plural subject and singular complement:

e.g. Cartwright's claims to be putting forward a  
politically realistic alternative are hogwash  
[4 G Let]

e.g. but her hands and arms were a real give away  
[7 Ex Let]

e.g. the cranks who get money from the Arts Council  
run by Sir William Rees-Mogg are a very small  
percentage of the recipients of grants  
[23 To Let]

It may be significant that the majority of the examples in the data which displayed subject-complement discord (7 out of 9) were from the Letters text-type. This might indicate that it is a feature of comparatively casual style. It also seems likely that research on the spoken language would discover examples of non-plural subjects occurring with plural complements. The following invented examples<sup>6</sup> sound quite acceptable when spoken:

e.g. That's them  
Here's the crayons

Another contrastive study, perhaps using spoken data,

---

<sup>6</sup> This point and these examples were contributed by Loreto Todd.

would be needed in order to establish the patterns of usage in other styles and media.

### 3.2.5 Anaphora

Some examples of potential subject-complement equality are particularly resistant to reversal as a result of having subjects which refer back to their immediately preceding context. This is a particular kind of difference in information value because the subject is not only 'given', but is given in context and the complement represents the new information brought in by the writer:

e.g. Yesterday's bombing was the eleventh this year  
[18 Ex Art]

This example refers to a bombing which was first mentioned five sentences previously:

e.g. A bomb blasted a crowded waiting room at police headquarters in the heart of Paris yesterday  
[13 Ex Art]

Since the intervening sentences mention other bombings in such phrases as "the fourth bombing" and "further attacks", the writer has to identify the attack to which this sentence refers unambiguously. "Yesterday's bombing" achieves this aim and being completely 'given' information, is assigned to the subject function.

The straightforward given/new contrast as seen in the example above is exploited for the writer's persuasive purpose in the following:

e.g. I realised this 'circus' would be the power  
 behind the Labour throne [20 M Let]

The 'circus' mentioned in this example refers back to the previous sentence which mentioned "the TUC conference". The writer's aim of making the reader unconsciously accept one attribute (TUC = circus) as given, whilst being presented with a less controversial one (TUC = power behind Labour) as the ostensible main point of the clause is served by this use of information structure since the reader is conditioned to expect more 'important' information in the later stages of a clause, in this case the complement. As the rewritten version shows, a reversal of subject and complement in this example would make explicit the underlying aim of the original version:

e.g. I realised that the power behind the Labour  
 throne would be this TUC 'circus'.  
 [reversed - 20 M Let]

### 3.2.6 End-focus

There were only six examples in the data, out of a total of 200 having NP complements, which displayed a convincing balance between their subjects and complements. These are listed below. In none of the examples is there any striking difference of complexity between subject and complement and they all have definite reference:

e.g. France has been the main target [35 Mi Ed]

e.g. The Union Jack is the national flag  
 [23 S Ed]

e.g. Heathrow was the main hand-over point

[16 Mi Art]

- e.g. why a man of such achievement should feel  
resignation to be the only course [27 Te Let]
- e.g. the main objectors have been the architectural  
historians and journalists [25 Te Let]
- e.g. The controversy over the Monocled Mutineer is  
the latest example of the BBC's incompetence  
[24 Ex Ed]

In all of the above examples, it is possible to imagine a reversed version which does not significantly change the sense of the clause as a whole. However, there is always a change of emphasis when clause elements are moved and the difference between these examples and those discussed under other headings is that the change here would be least disruptive to the sense.

### 3.3 Conclusions about NP subjects and NP complements

Using the main database, this chapter has confirmed a number of well-known features of noun phrases which depend upon their position in the clause. It also discovered a number of interesting aspects of the relationship between noun phrase subjects and noun phrase complements.

Those features which were confirmed include the following:

A. Complements tend not to be realised by pronouns or proper nouns, and data from other studies (Aarts 1971) suggest that this is not a feature of newspaper English alone but may be more typical of written than spoken language.

B. Complements contain a higher number of indefinite articles than subjects.

C. Complements are much more heavily modified than subjects. They are more likely to be premodified or postmodified or both than subjects.

D. In any individual clause containing copula BE followed by a noun phrase complement, the subject will be at least as definite as the complement. There were no examples of indefinite subjects with definite complements. If there is a conflict between definiteness and complexity, definiteness wins. There was one example where the subject was clearly more complex than the complement, but the latter was indefinite.

There were a number of discoveries made whilst I was working on this topic which are not, to my knowledge, reported in other work.

E. The rules of number concord between subject and complement seem to be more complicated than is normally recognised. There is a tendency, at least in casual style, to mix plural subjects with non-plural complements, but not to mix non-plural subjects with plural complements. I also found that the violation of number concord and the negation of the verb are both unambiguous markers of the distinction between subject and complement.

F. The structures of subjects preceding BE + NP complement are not typical of the structures of subjects

in general. In particular, they are even less likely than subjects in general to contain an indefinite article. They are also less likely to contain premodification but more likely to contain postmodification than subjects as a whole. These observations may be specific to newspaper English, since other studies (notably Varantola 1984) show different proportions of definite and indefinite articles and of pre-and post-modification.

G. Although complements were, unsurprisingly, found to contain more indefinite articles than subjects, they had the same proportion of definite articles as NP subjects occurring with NP complements and as NP subjects generally. It would be interesting to collect more data on NP structure, possibly including objects, in order to establish whether some features (such as definiteness) are more stable than others in their frequency of occurrence irrespective of their function.

H. The indeterminacy between the grammatical functions of subject and complement has been narrowed down to one small area of overlap. In practice, most NP complements are distinguished from their subjects by complexity and/or definiteness. However, the few examples where the noun phrases preceding and following the verb BE are equal in definiteness and complexity illustrate the notion of 'merger' introduced by Coates (1983).<sup>7</sup> The merger in this

---

<sup>7</sup> See section 1.4.3 for more discussion of this type of indeterminacy.

case means that subject and complement share similar functions and, to use Huddleston's terminology, are simultaneously 'identifiers' and 'identified'. It is in such cases that the normal positions of subject before the verb and complement after the verb lose their power to distinguish between the functions which have merged. The reversal of a clause such as 'The Union Jack is the national flag' results in no more than a small shift of focus from one NP to the other: 'The national flag is the Union Jack'.

Chapter 4The borderline between adjectives and participles

The second descriptive problem of the data to be studied in detail was the borderline between the adjective class and -en participles. This topic had been investigated at some length in the pilot study and produced the general conclusion that the context of individual examples was useful in assigning them to the verbal or adjectival class. Clearly there is no absolute boundary between these word classes; various writers have illustrated the gradience between them.<sup>1</sup>

My dissatisfaction with Kilby's attempt to define this borderline (Kilby 1984) is based upon his aim of characterising -en forms as adjectival or verbal irrespective of their context. This tendency, which is present in other treatments of the same topic (e.g. Huddleston 1984, Johansson 1986 and Quirk et al 1972), causes Kilby to confuse two different kinds of information. First, there is the information about the actual form and context of the -en form and secondly, there is the information about other possible forms and contexts of the -en form. For example, the criterion relating to the occurrence of intensifiers before

---

<sup>1</sup> See, for example, Huddleston (1984) pp320-324 and Quirk et al. (1985) pp414-416.

adjectival -en forms may be applied by looking at the data to see whether there is, in fact, such an intensifier:

e.g. I was so frightened I spent the next day at my  
mother's. [29 S Let]

It may, on the other hand, be applied by trying to insert an intensifier to see whether the clause remains acceptable. This is a much more difficult test to perform and, as Kilby shows with his informant test, opinions on acceptability range almost as widely as the number of informants questioned. In some cases the unacceptability may be very clear:

e.g. ...in a world where the streets were \*very paved  
with gold. [18 Te Let]

In other cases, a decision on acceptability may rely on the tester's ability to imagine a convincing situation in which a slightly unusual clause might be produced:

e.g. The gang's activities over two and a half years  
were ?very/?highly geared to obtaining high  
prices from... [29 Ex Art]

Johansson et al. (1986) had a very practical aim in confronting the problem of the borderline between the -en participle and adjectives. This aim was the grammatical tagging of the LOB corpus of English. The Users' Manual discusses all situations in which -en forms can occur, but recognises that: "The principal problems arise with -ed forms after BE". The criteria used to assign verbal and adjectival tags to such examples are described, but at

times they unconsciously contradict each other. For example, the verbal tag is applied with active verbs, even when the meaning is stative:

e.g. " (9) The chimney was closed and the hearth recess cleaned of its soot..."  
 "(10) Only Henry's head nodded, his eyes were closed and his breathing loud and heavy..."  
 [Johansson 1986]

The first criterion for adjectival status, however, also claims examples which "denote a state" and in my opinion should reclaim the second of the 'closed' examples above for the adjectival group. The concern with permanently assigning an -en form to one or other part of speech is revealed in a comment on the second 'closed' example:

"The structure of coordination in (10) shows that closed is similar to an adjective. Nevertheless, this form and other 'stative' -ed forms of action verbs have not achieved clear adjective status, as shown by their inability to accept intensification by very: \*very closed, very built etc."

Johansson et al. can be criticised on the grounds that many 'true' adjectives, specifically the non-gradable adjectives, cannot occur after 'very' (e.g. 'female') and it is therefore not a criterion to be used rigidly. This extract can also be faulted for moving away from considering items in context. Although it is generally accepted that some participial forms 'become' adjectives over a period of time, it seems to me that the more interesting question for anyone concerned with describing data rather than writing dictionaries is to consider the

status of individual -en forms in context. One implication of this position is that the same -en form, in the same sense, may be verbal in some contexts and adjectival in others.

The present chapter sets out to investigate examples of -en forms following BE in order to establish how far the formal context of an example can indicate the verbal or adjectival status of such a form. Having considered the formal context, semantic influences on word class assignment are discussed.

#### 4.1 Decisions arising from the pilot study

Section 1.3.9. of the Introduction considered the borderline between adjectives and participles, with particular reference to the past participle. In that section, I sought to discover whether there was more certainty about the adjectival or verbal status of an -en form<sup>2</sup> when it was considered as an individual instance in context, rather than as a lexeme with a number of theoretically possible contexts.

One of the conclusions of section 1.3.9. was that there were some features of context which would always incline

---

<sup>2</sup> To avoid confusion, the phrase '-en form' is used in this chapter to mean a word in the form of a past participle whose verbal or adjectival status has not yet been determined. Because it is the passive form which can look formally identical to copula + adjective, the verbal -en forms are referred to as passive participles.

the reader/hearer towards interpreting the -en form as either verbal or adjectival. It is well-known, for example, that comparative forms are concrete evidence of an adjective:

e.g. the Soviet Union is more concerned about...  
[28/T Ed]

The -en form, on the other hand, is usually considered to be passive if there is a following object or complement:

e.g. Flanagan was given a suspended six-month  
prison sentence [75/36 Mi Art]

Other contextual features seemed to have no more than a tendency to be interpreted as verbal or adjectival. These included features such as adverbials following the -en form, which in the pilot data were usually taken to indicate a passive form:

e.g. the theories embodied in her "Daughter of  
Time" [69 Te 19]

Whilst some results were obtained from the pilot study, there were a few problems with the method used in investigating this topic. I did not wish to repeat the process of coding doubtful examples as either potentially verbal or potentially adjectival; these were left as one larger group of doubtful cases which were suspected of being more adjectival than verbal. The reason for this decision was that the -en forms themselves are in a verbal form and any doubt over their classification must mean

that they are potential adjectives. Another change that was made in the main study was to keep finite and non-finite cases together, since most of the non-finite examples had a clearly recognisable elided subject and auxiliary:

e.g. a Soviet UN employee (who is) charged with  
 espionage [86/38 Ex Art]

In the pilot study, the context investigated included only elements immediately preceding and immediately following the -en form and took no account of which clause the context items belonged to. For the main study I decided to broaden the notion of context to include all preceding and following context, but restricting this context to the clause containing the -en form concerned. Note that this restriction to a single clause allows for following clauses to be included if they are subordinate to the clause concerned.

e.g. is convinced [that when the real issues..]  
 [4 Ex Ed]

The noun clause complementing 'convinced' in the above example would be included in its context. Verb phrases in sequence, whether they preceded or followed the verb phrase in question, were also included since it can be argued that they are not in separate clauses at all (see Chapter 5):

e.g. with whom the SDP is supposed to be allied  
 [3 G Let]

Another change that was made in the present study was to divide the context into three rather than two areas. Assuming that all examples contain a form of the verb BE (in some cases ellipted) and an -en form, the context can be divided into that preceding BE, that occurring between BE (or an auxiliary other than BE) and the -en form, and that following the -en form. The three contexts considered were labelled 'preceding', 'intervening' and 'following'.

#### 4.2 Method of investigation

Three groups of examples were investigated for this chapter. The first group were examples classed as unambiguously passive in the coding stage of the research and they were easily retrieved from the data by the passive identification in the 'special interest' field (no. 21). The group of doubtful cases were also easily retrieved since they too had a marking (Ven/aj) in field 21. The third group considered were the authentic adjectives; those which were not in a participial form. These were not directly available from the database since adjectives had not been anticipated as being an area of special interest. However, those following the verb BE had already been identified during the investigation of the copula reported in Chapter 3 and were therefore available from the printout used in that part of the study.

Each group of examples was considered in turn, using the file containing full sentence contexts (see Appendix 1d)

to characterise preceding, intervening and following contexts for each example:

e.g. [One's future happiness] will [of necessity] be  
 built [upon the ruins of another family's sorrow  
 and despair.] [39/11 M Let]

In the above example, the preceding context is the subject, the intervening context is the adverbial 'of necessity' and the following context is another adverbial. The intention was to consider the three groups of examples in exactly the same way, but it was difficult to use the same terms to characterise the context of clear adjectives and clear passives, even when they were formally identical.

e.g. People are too kind [to the Guardian]  
 [92/56 Mi Ed]  
 vital resuscitation was given [to two other  
 survivors] [90/43 M Art]

In these examples, the prepositional phrase seems to be a complement to the adjective ('kind') and more of an independent, though necessary, adverbial following the passive ('given'). In Johansson (1986) one of the criteria for identifying adjectival forms was to distinguish between "prepositions which typically occur after adjectives" and those which occur after verbs. The distinction is illustrated with neat contrasts such as: 'pleased with' (adj) / 'pleased by' (verb), but the 'adjectival' preposition is found later on the same page in an example of a verbal form: "the poet is concerned

with an audience". This difficulty of deciding between syntactic and semantic criteria is discussed further in section 4.4. For the purposes of this part of the study, I decided to classify contexts together when they were both syntactically and semantically similar.

Having characterised the three kinds of context for each example, I counted the number of examples containing each different context for the three groups. The results are shown in Tables 4.1, 4.2 and 4.3 in the next section. Because some of the contexts were not comparable across the groups, separate tables are used to show percentages for the sub-group of context types which were comparable. Tables 4.4, 4.6 and 4.8 contain the results of these comparisons for preceding, intervening and following contexts respectively.

One other perspective was taken on this data. It was considered useful to calculate the proportion of all passive and -en forms (with different kinds of context) which were assigned to the doubtful group. These proportions are presented as percentages in Tables 4.5, 4.7 and 4.9.

### 4.3 Results and analysis

Tables 4.1, 4.2 and 4.3 show the range of items occurring in the contexts of three groups<sup>3</sup> of examples: passive verb

---

<sup>3</sup> Note that the 'group' containing -en forms of doubtful verbal status is not homogeneous, but contains examples which range along a gradient between participle

phrases, BE copula + adjectives and BE copula + -en forms whose class membership is in doubt. The tables display the figures for preceding, intervening and following contexts respectively. The whole clause context of each example was considered:

Table 4.1 Preceding context of three groups of examples

	passive	adjective	-en form
subject	216	120	50
adverbial	9	2	0
VP sequence	16	3	2
no context	69	2	5
other (prep/conj)	5	0	0

The preceding context of all three groups is dominated by simple subjects:

e.g. Mrs Thatcher is convinced that...

[8/4 Ex Ed]

The passive group also has a number of examples with no preceding context which can mostly be classed as elided subjects:

e.g. allegations (which were) made by her  
colleagues [69/28 Ex Let]

This feature illustrates a striking difference between the passive and adjective groups; that 22% (69/315) of the passive examples had an ellipted subject and auxiliary whereas less than 2% (2/57) of the adjectives occurred in an equivalent position with no preceding context at all.

---

and adjective. For convenience, however, they are referred to as a 'group' in this chapter.

The two examples of this kind are also extremely different to the reduced relative clauses containing -en forms:

e.g.  $\phi$  to be fair to Dr Owen,  
[60/24 Te Ed]

Unlike in the passive examples, there is no clearly identifiable ellipited subject in the above case and the verb has not been omitted. Of course, there are probably some examples in the database with adjectives which are closer counterparts to the reduced relative clause (e.g. vegetables suitable for frying). Such examples were unfortunately not extracted from the data because adjectives had not been coded in the database. In contrast with the adjective group the -en group contained a number of examples with no preceding context. Other preceding contexts for the passive group include a number of sequential verb phrases<sup>4</sup> and a few adverbials (with subjects as well in most cases). These are exemplified below:

e.g. have their stripes and badges taken off their  
sleeve. [7/1 Mi Let]

This year a new exam, the General Certificate  
of Secondary Education, is being introduced  
to.... [1/1 S Ed]

Table 4.2 describes the intervening contexts discovered in the data between the first auxiliary of the verb phrase and the -en form, passive participle or adjective:

---

<sup>4</sup> See chapter 5 for a discussion of whether such sequences of VPs consist of one or more clauses.

Table 4.2 Intervening context of three groups of examples

	passive	adjective	-en form
adjuncts	33	13	5
quantifying adverb	0	26	5
comparative adverb	0	10	2

As the above table shows, each of the groups of examples has some degree of interruption between the verb and the -en form or adjective. The differences between the groups are considered below in conjunction with tables 4.4 to 4.9 showing the figures as percentages. However, it is clear that the passive groups favours intervening adjuncts rather than adverbs which modify the -en form itself. An example of the former:

e.g. The two leaders' efforts {to...} was strongly  
criticised last night by... [80/26 G Art]

Intervening contexts are more varied for the adjective and -en form groups; they divide into those having an adjunct function, as in the passive example above, and those which directly modify the adjective or -en form. The example below shows an intervening intensifier, but comparative adverbs were also found in the data.

e.g. even the Liberals, (...), are extremely  
unenthusiastic [8/3 G Let]

The next table (4.3) shows the following contexts of the three groups of examples studied:

Table 4.3 Following context of three groups of examples

	passive	adjective	-en form
adverbial	153	13	7
agentive adv	60	0	1
VP sequence	25	34	18
object	9	0	0
complement	7	0	3
no context	61	56	0
complementary phr.	0	26	16

Following contexts seemed to differ more noticeably among the groups than preceding or intervening contexts. The passive group had predominantly adjuncts in its following contexts:

e.g. millions of pounds worth of drugs were smuggled  
through European airports ....  
[60/29 M Art]

The adjective group, however, had no dominant following context but a range of following contexts which occurred fairly frequently. It was common, for example, to find adjectives with no following context or with a following clause:

e.g. it is obvious that the positive acceptance of  
....  
[62/17 G Let]

The -en form group had following contexts dominated by sequential verb phrases and complementary phrases, the latter illustrated below:

e.g. The serial is only slightly related to the  
truth  
[4 Mi Ed]

In addition to the dominant following adjunct, the passive examples contain the same range as the adjectives and a small number of direct objects and complements. The objects occur with ditransitive verbs in the passive form. No objects or complements are found in the adjective group and this may be an indication that the small number of objects and complements in the -en forms group should be classed as passive (see section 4.4 for discussion of these examples).

Although Tables 4.1 - 4.3 give some indication of the character of contexts for these three groups, the calculation of percentages for contexts which were comparable across the groups was the next stage of analysis. Tables 4.4, 4.6 and 4.8 below contain these figures.

It should be noted here that there is a basic circularity in the method of investigation employed in this chapter. In considering the best approach to the problem of the adjective/participle borderline, I decided that to compare the contexts of all -en forms regardless of their coding with the contexts of all true adjectives would be a long and possibly fruitless process. The final decision was to make a shorter study using an element of intuitive knowledge. The process had begun in the coding stage of the project when I had allocated all -en forms which were not immediately seen to be passive forms to a 'doubtful' category. There were, presumably, some unconscious principles on which the categorisation was made and part

of the aim of this chapter was to uncover such intuitive criteria. A second aim of this investigation was to find out whether the doubtful category shared context types with the true adjectives.

Table 4.4 presents comparable figures (percentages) for preceding contexts in each of the three groups:

**Table 4.4 Percentages of preceding context types for 3 groups**

	passive	adjective	-en form
subject	69	94	96
VP sequence	5	2	4
adverbial	3	2	0

From this table it can be seen that there are similarities between the adjective and -en groups which are not shared by the passive group. Although all three groups have a large majority of cases which have a simple subject (i.e. a subject alone) preceding the verb, the proportion in the passive group is significantly lower than in the other groups. The figures for simple subjects from table 4.1 were tested for difference and produced a value for  $X^2$  of 8.78, which is significant at the 0.05 level. The other percentages in this table are based on rather low figures, but also show a similarity between the adjective and -en groups. Both of these groups have proportionally fewer preceding sequential verb phrases and fewer, or no adverbials.

The comparison of preceding contexts is not

straightforward. For example, although the adjective and -en groups share a similar proportion of simple subjects preceding the verb, the passive group also has a large majority of examples with such a preceding context. This cannot, therefore, be a very compelling criterion for judging an -en form to be adjective-like. Indeed, the proportion of all -en forms with a simple subject which were put in the doubtful category during coding was 19%. The remaining 81% were classed as unambiguous passives. This distribution of examples containing an -en form and a simple subject between the passive and the doubtful groups is not significantly different from the overall distribution of -en forms. A  $\chi^2$  test on the observed distribution of examples with simple subjects compared with that expected from the average distribution, produced a value for  $\chi^2$  of 2.45, not significant at the 0.05 level.

The other two entries in Table 4.4 concern the occurrence of preceding sequential verb phrases and preceding adverbials. From the table, it seems that the presence of a preceding adjunct acts in favour of a passive classification during the coding process. This conclusion is upheld by considering the figures from the viewpoint of distribution between passive and -en groups as seen in Table 4.5.

Table 4.5 Distribution of -en forms between passive and -en groups (preceding context)

	total	subject	VP seq	adjunct
passive	85	81	84	100
-en forms	15	19	16	0

The proportions of preceding subjects, as mentioned above, are similar to the overall proportions. The figures for sequential verb phrases are even closer to the average. For adjuncts, however, a higher than average number of examples were classed as unambiguously passive. This means that the occurrence of adjuncts in preceding contexts may have been partly responsible for the decision to grant passive status to some examples.

Table 4.6 shows the proportion of each type of intervening context as they occur in the three groups of examples:

Table 4.6 Percentages of intervening context types for 3 groups

	passive	adjective	-en form
adverbial	10	17	12
quant adv	0	20	9
compar adv	0	8	7

Unlike the figures in Table 4.4, the columns in Table 4.6 do not add up to 100% because not all of the examples had an intervening context. Some conclusions can still be drawn from these percentages. If the figures for the passive and -en groups are converted to show the distribution of -en forms among these two groups (see table 4.7), we can see that the intervening adjuncts had

little effect on decisions; only an average number of -en forms with intervening adverbials were assigned to the doubtful group. A  $X^2$  test on the figures (from table 4.2) produced a figure for  $X^2$  of 0.14, not significant at the 0.05 level. On the other hand, the occurrence of an intervening intensifying adverb or comparative adverb was consistently interpreted as being a sign of potential adjective status; no such examples were classified as passive:

Table 4.7 Distribution of -en forms between passive and -en groups (intervening context)

	total	adjunct	adverb
passive	85	87	0
-en forms	15	13	100

It seems from this data that whilst an intervening modifying adverb is a sign of an adjective, an intervening adverbial is neither an indicator of passive status nor of adjectival status.

The next table (4.8) shows the proportion of comparable following contexts occurring in the three groups:

Table 4.8 Percentages of following context types for 3 groups

	passive	adjective	-en form
adjunct	49	10	14
VP sequence	8	27	30
complementary phrase	0	20	25

Here, as with the preceding contexts, there is a very clear similarity between the adjective and -en groups which is not shared by the passive group. The large percentage of passive examples which were followed by an adverbial (68%), and the comparatively low percentages for the two other groups (10% and 14%) implies that following adverbials are a fairly strong indicator of passive status. A  $\chi^2$  test on the figures from table 4.3 supported this conclusion;  $\chi^2$  was calculated as having the value 47.67, significant at the 0.001 level. Following sequential verb phrases and complementary phrases, however, seem to be indicative of adjective status. The latter are never found after clear passives and the figures for the former were found to be significantly different at the 0.001 level;  $\chi^2 = 31.66$ . As with the other context types, the distribution of -en forms between the clear passive group and the 'doubtful' group shows whether a particular context resulted in a higher-than-average or lower-than-average number of -en forms being classed as 'doubtful'. Table 4.9 shows this distribution for following contexts:

Table 4.9 Distribution of -en forms between passive and -en groups (following contexts)

	total	VP seq	adverbial
passive	85	45	96
-en forms	15	55	4

The percentages in the above table show clearly that

compared with the total number of forms classed as either a passive form or an -en form, those followed by adverbials are relatively more frequently assigned to the passive group. Conversely, those followed by a verb phrase in sequence are less frequently assigned to the passive group.

#### 4.4 Syntactic and semantic features of the borderline

In section 4.3, the contexts of all examples in the main database containing either BE + -en forms or BE + adjectives were described. The next stage in my investigation was to look more closely at the 57 examples which had been coded as 'doubtful' during the inputting of data.

My aim here was to see how far the general conclusions about context made in section 4.3 could be applied to individual examples. The expectation was that although one can characterise typical contexts of verbal and adjectival -en forms by analysing a corpus of data, it would be much more difficult to use such generalisations as criteria for assigning each occurrence of an -en form to the verbal or adjectival class. A further expectation was that semantic considerations would complement grammatical information to produce decisions in some cases.

The 57 examples in the 'doubtful' category represent -en forms which I suspected of being adjectival during the initial coding process. They were assigned to this

category impressionistically and their formal and semantic contexts were not consciously taken into account. On looking more closely at the examples, I saw that some of them were clearly marked as adjectives by their formal contexts.

A number of the examples (25/57) were marked as adjectives by contexts exclusive to that class such as intervening comparative or intensifying adverbs:

e.g. Your story mentions four women who are deeply  
committed to politics... [3 M Let]

As in the example above, many examples contain complementary prepositional phrases which are also typical of adjectives:

e.g. if someone living on the sale of inherited  
assets were entitled to benefit when... [3 Te Let]

One problem with these phrases which was mentioned in section 4.2, is that they are often formally indistinguishable from adjuncts which more frequently follow passive forms:

e.g. ...the centre of the city seemed to have been  
surrendered to the "toughs" [30 T Let]

However, the difference between the two functions of prepositional phrases is usually clear. Most prepositional phrases following passive forms can be identified as belonging to one of the major categories of adjunct, such

as 'time', 'manner' or 'place' and their prepositions have a meaning which can be isolated and sometimes paraphrased. The phrases complementing adjectives, on the other hand, are impossible to categorise in this way. Their prepositions are very closely linked with the adjective and often seem to have no identifiable meaning of their own.

The occurrence of a complementary phrase, therefore, is established by a combination of semantic and syntactic information but once it has been established, it can be taken as a clear marker of adjectives.

Another large group of examples (15/57) each contained a clause following the -en form. Most of the following contexts were 'to infinitive' clauses, but there were also two examples containing 'that-' clauses:

e.g. Mrs Thatcher is convinced that when the real  
issues are put before the public, she will win  
an election hat-trick. [4 Ex Ed]

The similarity of such examples to those containing adjectives followed by that- clauses is fairly convincing:

e.g. he is also well aware that the more Leftward-  
leaning tone of some of his potential  
political bedmates does not augur well for the  
happiness of the marriage. [24 Te Ed]

But there are also other, clearly passive examples which are similarly followed by that- clauses:

e.g. I was told that they were there to wave to  
passing children... [26 Mi Let]

The difference between the passive example and the other (adjectival and doubtful) examples seems to be that there is a necessary, though unidentified, agent associated with the action in the passive. The other examples lack this shadowy figure. A more formal difference can also be observed; the passive example has the verb in a past tense whereas the other examples have a present tense for the verb BE. In order to interpret the doubtful example as passive, one would need to change the verb to the past tense:

e.g. Mrs Thatcher was convinced that when the real issues are put before the public, she will win an election hat-trick. [4 Ex Ed - rewritten]

This rewritten version could still be interpreted adjectivally, but there is now the suggestion of an agent and the action (of 'convincing') is described as taking place in the past rather than attributing a static state of mind to Mrs Thatcher. The question of whether the 'that-' clause functions as an object, making the main clause passive, or whether it functions as a complement to an adjective relies mainly on subtle aspects of context which are more semantic than syntactic.

The group of following to- infinitive clauses which are more common than that- clauses in the 'doubtful' group are also reflected both in the passive and the adjectival groups. Three examples follow, from passive, adjectival and doubtful groups respectively:

e.g. Toplis is shown to be a petty crook, a con-man  
... [7 Mi Ed]

- e.g. Mr Kounba Balogun is not fit to be a leader of  
his community [31 To Ed]
- e.g. Dr Owen is determined to get an agreement of the  
precise criteria on which a decision....  
[9 G Art]

As with that- clauses, there seems to be a difference between the passive example and the other examples. This difference cannot reside in the tense since all three have a present tense verb. Similarly, there is no 'typical' distinction between human subjects before -en form adjectives and non-human subjects before passive forms (see Johansson 1986 p33) since the examples all have human subjects. In this case, as in other examples in the data, the distinction between static and dynamic meaning seems to be the most crucial distinguishing factor. Although the static-dynamic contrast is often mentioned in connection with the participle-adjective borderline (see Johansson 1986, Quirk et al 1972, Sampson 1985), it "is often far from easy to apply in practice" (Sampson 1985).

Although I have previously stated that tests hypothesising possible forms should be used with caution, some of them may help to capture the elusive static-dynamic distinction.

The tests which concern us here are of two kinds. There are those which serve to establish the polysemy of a particular word-form, thereby distinguishing the sense of a participial adjective from that of an active verb and there is the 'time test', already mentioned above, which

questions the appropriateness of the existing time reference (tense and/or aspect) for a dynamic interpretation of the -en form. One example of the latter test would require a past or perfect form of BE in order to be interpreted as dynamic:

e.g. if the Alliance leaders are now clear what is agreed the general public are likely to be much less so. [16 Ti Art]

The same form, however, in the same sense, can occur as a passive participle:

e.g. the need for a joint policy has now been agreed on both sides [14 G Art]

The combination of a perfect VP and the adjunct 'now' leave the reader in no doubt that 'agreed' is a passive participle. In order to obtain an adjectival interpretation this example would either need to contain a present tense:

e.g. the need for a joint policy is now agreed [rewritten - 14 G Art]

or it would have to omit the adjunct referring to present time:

e.g. the need for a joint policy has been agreed (in the past) [rewritten - 14 G Art]

Sampson (1985) alludes to the time dimension in his discussion of an example from the LOB corpus. The example is 'the knives which are now well rounded by wear' and Sampson chooses an adjectival analysis "since the latter

(i.e. a verbal analysis) would correlate with the interpretation according to which the knives are now being rounded, (...) whereas it is clear that the writers mean that the knives are now in a state resulting from rounding action over a long period in the past". Although he does not elaborate the topic of time, it is clearly crucial to the interpretation of some -en forms and in some cases the decision depends upon combinations of time adjuncts and VP forms.

The polysemy of a word-form is also important, in some cases, for distinguishing between passive and adjectival occurrences of an -en form:

e.g. But just how far the Alliance in general is  
 prepared to note those lessons will be clearer  
 after.... [27 Te Ed]

This example has a preceding adjunct, 'in general', and a following sequential verb 'to note'. Both of these features have been shown to be more typical of adjectives than of passive participles, but both could also occur with the passive. In trying to establish why the -en form 'prepared' seems to be undeniably adjectival, we have to take a semantic, more specifically a lexicological viewpoint. The word-form 'prepared' has at least two polysemous senses which are represented by the verb (e.g. 'The Alliance prepared to govern' - invented) and the derived adjective respectively. One type of evidence which can be used in support of this argument is that there are

a number of adjectival near-synonyms which could be used to paraphrase the quoted example.

e.g. But just how far the Alliance in general is  
willing/keen/happy to note....  
[rewritten - 27 Te Ed]

These synonyms do not work for a revised version of the verb, although it has a paraphrase in 'get ready'. A more difficult test to apply is to see whether it is possible to construct an active equivalent to the original example:

e.g. But just how far X is preparing the Alliance in  
general to note  
[rewritten - 27 Te Ed]

In some cases, where there is a lack of non-participial near-synonyms, this test is useful:

e.g. if any Libyans or Palestinians had been  
involved [20 To Ed]

e.g. if X had involved any Libyans or Palestinians  
[rewritten - 20 To Ed]

The active equivalent of this example, which would be acceptable if we could identify the subject, shows however that the original is adjectival since no outside agent is implied. The decision to identify 'involved' as adjectival in such examples illustrates the difference between my approach and that of Johansson (1986). He distinguishes between a similar example, 'he's involved in several of my business ventures...', which he classes as verbal, and the better established adjectival use of 'involved' in 'even

that was too involved and tortuous for some foreign readers...'. As I have already pointed out, the dictionary word class of a form, though interesting, can often be violated in context and is therefore not reliable as a criterion in borderline cases.

A final group of examples occurring in the data were those which seemed to contradict the clearest of the formal rules. Quirk et al (1972) suggest that there is an increasing tendency to use apparently clear markers of adjectives, such as intensifiers, with equally 'clear' markers of passive participles, such as following agentive by-phrases. The authors do not decide whether their example ("The man was very offended by the policeman") is, in fact, adjectival or verbal. For them, it is an example of indeterminacy. My data, however, included some examples which I felt were definitely adjectival, but which displayed contextual features overwhelmingly associated with the passive in the rest of the data. One of these examples contained an agentive by-phrase:

e.g. Abdallah is wanted by the Italians in connection  
with.. [26 Te Art]

When the past participle of a verb has such a specialised meaning and this meaning is rarely, if ever, attached to the active form of the verb, it is a good case for ascribing adjectival status to the -en form. In addition, the usual sense of the verb 'want' is never used in the passive (\*A bike is wanted by him) so the two senses are

separated by their different syntactic potential. The agentive phrase does not seem to influence the interpretation of this example, perhaps because it is part of an almost formulaic clause; X is wanted by Y.

The other apparently contradictory examples both contain adjectival -en forms followed by object complements, which I have previously argued (see section 4.3) are a clear indicator of verbal status. The examples both occur in the same sentence:

e.g. My last statement was dated August 18, the envelope was franked August 27... [34 Ex Let]

Although there is no problem constructing an active sentence using the verbs 'date' and 'frank', there does seem to be a difference in meaning which is not ascribable to the invention of an agent or the usual difference between active and passive transforms:

e.g. They dated my last statement August 18 and the Post Office franked the envelope August 27... [rewritten 34 Ex Let]

Despite the fact that someone must have performed these acts, these -en forms in the original version seem to be describing not an act, but a state; the state of the letter and envelope. Had the author intended to retain some verbal force in these potential passive verb phrases, she or he would presumably have used a past perfect form of verb phrase.

#### 4.5 The adjective/participle borderline - summary

There are a number of conclusions to be drawn from the discussion of examples on the borderline between the adjectival and verbal classes.

The first is that there are certain indicators in the formal context of most examples as to their complexion. These indicators are summarised here. The following conditions may apply to passive verb phrases:

- a. it is the verb in a reduced relative clause
- b. it has a following agentive -by phrase
- c. it has a preceding adjunct
- d. it has a following object or complement

The -en form tends also to be classed as passive under the following conditions:

- e. its preceding context contains a sequential verb phrase
- f. it has an intervening adjunct
- g. it has a following adjunct

An -en form is classed as unambiguously adjectival if:

- h. it has an intervening modifying adverb such as an intensifier or a comparative adverb
- i. it has a following complementary phrase

An example tends also to be classed as adjectival if:

- j. it has no more than a subject in its preceding context
- k. it has a following sequential verb phrase

Although context may be of some help in drawing the

boundary between adjectives and passive participles, this investigation has shown that semantic considerations are just as important in assigning these categories. Three types of semantic influence were discovered in the data:

First, it was suggested that some apparently formal criteria (e.g. the occurrence of complementary prepositional phrases) are based upon semantic categories. Secondly, it was shown that time, in the form of tense, aspect and time adjuncts is a very subtle marker of word class for -en forms. Thirdly, the polysemy of word-forms was shown to be important in establishing a distinction between verbal and adjectival uses of some -en forms.

The discovery that some examples of -en forms can be assigned to either the adjective or the verb class by consideration of their context does not deny the existence of gradience between these word classes. The lists of attributes usually cited to distinguish these word classes can now have contextual features added. However, these lists define not the boundaries between absolute categories but prototypical<sup>5</sup> adjectives and verbs. Rosch (1978) and Coates (1983) assume that the attributes are part of the machinery of gradience and do not investigate them further. The analysis of gradience in this chapter indicates that attributes are not all simple 'yes/no'

---

<sup>5</sup> See section 1.4.3 for a discussion of prototypicality as introduced by Rosch (1978).

features. Even when we restrict our consideration of the -en form to examples in context, as in this study, there remain distinctions between those attributes which guarantee membership of a category (see a-d and h-i above) and those which are more commonly found with members of one category than another (see e-g and j-k above). As in chapter 3, the overlap of two categories in this chapter can be described in terms of a 'merger' (see Coates 1983:16). In this case there are examples which are difficult to assign to either the verb or the adjective class because their context provides either conflicting evidence or no evidence at all. In many cases a decision is not crucial to the meaning of the clause as a whole.

## Chapter 5

### Verb phrases in sequence

#### 5.1 Verb phrases in sequence - a problem of description

The notion of sequential verb phrases was introduced in chapter 1 (1.3.10) and the term has been used subsequently in this thesis to refer to sequences of VPs which are formally related in that all non-initial VPs are in a non-finite form dictated by the previous verb phrase:

e.g. common sense seems to be absent [M Let]

These sequences first came to my attention at an early stage of the pilot study when I found some of them difficult to classify. The problem was a simple one; does a 'sequence' of two VPs count as one complex clause or two related, but simple clauses?

A straightforward answer to this question was not to be found in available treatments of the verb phrase in English. Palmer (1965), for example, spends about a quarter of his 'catenative' chapter on 'Problems of statement' in which he states:

"catenatives occur in complex rather than simple verb phrases, involving subordination, though it has been stressed that these complex phrases still share some of the characteristics of the simple phrase. As such they are clearly full verbs, not auxiliaries." (p168)

On the next page, however, in discussing those catenatives he calls 'subject complementation' verbs, Palmer acknowledges that the distinction is not so clear:

"Again it is clear that there are no clear-cut divisions between primary auxiliaries, modals and catenatives."

To summarise the problem, the sequential verbs (i.e. non-final verbs in sequences) have some of the properties of main verbs and some of the properties of auxiliaries. In common with auxiliaries they restrict the form of their following VP and have a close semantic link with this VP. A small subset (notably the 'aspectual' verbs: see section 5.3) of examples also passivise as though they were in simple verb phrases. Unlike auxiliaries, the sequential verbs are not significant in the structures known as 'negation', 'inversion', 'code' and 'emphatic affirmation'.

Many of them can be assigned tense, be negated or be passivised independently of their following VPs.

Other, more intuitive considerations are also indecisive about the status of sequential verbs. Such verbs seem to belong to a restricted set like modal auxiliaries, but the resulting class is much larger than the modal class and possibly even more indeterminate at its boundaries (section 5.4 discusses this issue). Another intuitive reaction is that examples where three or four verb phrases follow each other do not seem to be as complex as the analysis in terms of main verbs would suggest:

e.g. we want you to help us create an outspoken  
intolerance for drug use [Te Ed - supp]

The analysis which assigns these three verb phrases to different clauses, two of which are subordinate, may seem to over-complicate the data. This intuitive reaction is supported by evidence from child language acquisition studies which show that children learn to string sequential verbs together at a very early stage in syntactic development (see Limber 1973).

The status of sequential verbs as auxiliaries, main verbs, or something in-between is discussed again in section 5.3. First, however, the examples of sequential verb phrases found in the data are described in relation to other features coded in the database.

## 5.2 Sequential verb phrases and other features

This section is concerned with the relationship between verb phrases occurring in sequence and the other features coded in the data. Some areas of potential interest were indicated by the pilot study, and these were followed up for the main database.

The first field to be investigated was subject type (field 12). The pilot study had indicated that there may be some differences between the subjects occurring with sequential verbs and subjects as a whole. In particular there seemed to be more human and fewer abstract subjects with sequential verbs. The figures and percentages for the main database are shown in table 5.1:

Table 5.1 Subject type<sup>1</sup> of sequential verbs

	seq VPs	% (350)	tot VPs	% (2201)
human	205	59	1101	50
abstract	130	37	950	43
animate	0	0	8	0
inanimate	15	4	142	6

Although human subjects occurred more frequently with sequential verbs<sup>2</sup> than they did overall, the difference was smaller than in the pilot data. I tested the distribution of examples from the sequential verb sub-database among the categories human, abstract and 'other' against the distribution in the whole database using a chi-squared distribution to measure the discrepancy between observed and theoretical frequencies. The value of  $X^2$  (9.77) was significant at the 0.001 level. The high number of human subjects occurring in this group is not surprising when the semantics of the sequential verbs are considered (see section 5.5). Many of them are characterised by human features such as volition and intention.

The next field investigated was finiteness (field 5) which shows up a gross and well-known feature of such verb phrase sequences; that second and subsequent verb phrases are 100% non-finite. The number of finite and non-finite

---

<sup>1</sup> Many of the sequences of verb phrases contained no intervening NP and can therefore be assumed to have the same subject as the first VP. Intervening subjects of this kind are included in the figures for this table.

<sup>2</sup> From this point the term 'sequential verb' should be taken to refer to non-final verbs occurring in sequences.

verb phrases occurring as 1st verb phrases is more interesting and these figures are given in Table 5.2 with figures from the whole database for comparison:

Table 5.2 Finiteness in sequential VPs from main data

	1st VPs	% 1st	total	% total
finite	150	82	1763	73
non-finite	32	17	637	27

One tentative conclusion of the pilot study on this aspect of the data was that 1st verb phrases in sequences behaved 'normally' as far as finiteness was concerned. If we compare the figures in Table 5.2 with those from the pilot study (see Table 5.3 below), it seems that there is a consistent difference between the overall proportion of finite verb phrases and the proportion of 1st VPs which are finite:

Table 5.3 Finiteness in sequential VPs from pilot data

	% f.	% n.f.
1st VPs	79	21
Total data	73	27

In the main database the difference between total data and 1st verb phrases was tested using the chi-squared distribution and a value of  $X^2$  was obtained (6.46) which is significant at the 0.05 level. One explanation of this tendency is that the non-finite clauses following sequential verbs cause the observed percentage of non-finite verb phrases in the total data to be higher than in

the non-final VP data. The percentages for 1st VPs would, therefore, be typical of all verb phrases which do not have their finiteness determined for them by a preceding sequential verb. This explanation is supported by recalculating the percentages of finite and non-finite verb phrases in the total data, excluding those 197 VPs which follow a sequential verb. These percentages show that sequential verbs are distributed between the finite and non-finite categories in a frequency which is close to normal for 'free' verb phrases as a whole; finite VPs representing 80% and non-finite 20% of total 2203 cases.

The next field investigated in relation to VP sequences was field 9, clause type. In the pilot study, final VPs were found to have an 'abnormal' distribution among clause types when compared with the total data as would be expected from data which contains 100% non-finite subordinate clauses. More important, perhaps, is to establish whether the first verb phrases in sequences, which have no such absolute restrictions on their behaviour, show figures comparable to those for the total data.

**Table 5.4** Clause types in sequential VPs  
(1st in sequence)

	1st VPs	% 1st	total data	% total
MCl	85	47	1081	45
NCl	25	14	289	12
ACl	31	17	348	15
RCl	24	13	259	11
Ncomp	9	5	127	5
Ajcomp	6	3	76	3
Vcomp			195	8

A chi-squared test was performed on the figures in table 5.4 to establish whether the distribution between clause types of the sequential verb group of examples is similar to that of the data as a whole. The value of  $X^2$  (2.42) was not significant at the 0.05 level, confirming the finding of the pilot study that sequential verbs are not more likely to occur in any clause types more than other main verbs.

In summary, it has been established that verbs occurring non-finally in verb phrase sequences behave grammatically like most freely occurring verbs by occurring in clause classes in proportions comparable to those found for verbs as a whole. The lower-than-average number of non-finite sequential verbs was hypothesised as being closer to the true percentage of 'freely occurring' verbs which are non-finite. The only significant difference found between sequential verbs and the verb class generally was the higher proportion of human subjects occurring with sequential verbs.

### 5.3 The grammatical status of sequential verbs

Although many linguists have investigated and/or discussed the problem of sequential verbs in English (see, for example, Joos 1964, Palmer 1965, Rosenbaum 1967, Huddleston 1971, Quirk 1972 and Matthews 1981), there does not emerge from their work any consensus of opinion as to the grammatical status of the sequential verbs themselves

or their following non-finite verb phrases. These verbs and their consequences for sentence structures are, however, generally problematic for most models of grammar. The problem is summarised by Chomsky (1961) in the following way:

"In regard to them, our intuitions 'concerning the propriety of particular classifications' fail us."

Transformational linguists such as Rosenbaum (1967) and Huddleston (1971) concentrate on the dependent non-finite clauses and the difficulties involved in assigning appropriate deep structures to superficially similar pairs of examples such as:

e.g. He persuaded John to come.  
He expected John to come.

Other linguists, such as Quirk et al (1972), begin by classifying the non-finite dependent clauses and then proceed to classify groups of verbs according to the forms they dominate. Quirk et al. (1985) make distinctions, according to (morpho-)syntactic properties, between 'marginal modals' (dare, need) 'modal idioms' (had better), 'semi-auxiliaries' (be going to) and 'catenatives' (appear, fail) (see sections 3.40-49). In their examination of verb complementation (chapter 16.20-67) they also make some semantic subclassifications based on syntactic distribution (see, see for example, section 16.38: "Subjectless infinitive clause as direct object"). Palmer (1965) also starts by setting out the two

main formal criteria for categorising such verbs. The first is whether the verb can be followed by one or more of the four non-finite forms, i.e. bare infinitive, to-infinitive, -ing form or -en form. The second criterion is the presence or absence of a noun phrase between the first and second verb phrases as in:

e.g. Fred told Joe to run fast (invented)

Although I will later be arguing that these criteria are not very useful in establishing categories of sequential verbs, it may be of interest here to record the numbers of examples in my data which occurred in each category. There were 43 (24% of 182 sequences) examples containing intervening noun phrases. Table 5.5 shows the numbers of examples followed by each of the non-finite forms:

Table 5.5 Non-finite forms following sequential verbs

	number	% (197)
to- infinitive	162	82
bare infinitive	10	5
-ing form	19	10
-en form	6	3

The clear indication from this table is that, in the present data at least, the to- infinitive form is by far the most frequent non-finite form to follow sequential verbs.<sup>3</sup> Descriptions based on introspective data often give

---

<sup>3</sup> Note that in Andersson's data (1985) infinitive forms are also the most frequent (p.275).

almost equal emphasis to each of the four non-finite forms, and are to this extent unrepresentative of real language data. The same point can be made about the extensive 'testing' of groups of sequential verbs to see what combinations of aspect, tense and voice are possible after them (see Palmer section 7.1.3). In the main database, only 13% of non-finite forms following sequential verbs were marked for these features (7% passive, 5% perfective and 1% progressive). Another 'test' for the independence of the following clause, negation, only occurs in 2% (4 examples) of the same data.

It may be objected that I am not taking into account the fact that these are intended to be tests of possible constructions and not criteria for analysing actual data. However, there is widespread dissatisfaction with such tests of acceptability amongst those who have done more than consult their own intuitions (e.g. Kilby 1984). I therefore prefer to generalise from attested data wherever it seems reasonable to do so. The rare occurrence of features such as passive or negative non-finite verb phrases after sequential verbs, though important in establishing the independence of these clauses, makes them relatively uninteresting beyond this simple function.

Although Palmer (1965) uses some of these tests and criteria for his classification, he objects to one, the pseudo-clefting test on the following grounds:

"First, it is very difficult to draw any clear lines where pseudo-clefting is or is not possible. (...) Secondly, unless the pseudo-cleft test links up with other syntactic features of the catenatives (and it does not) it would seem to prove no more than that some verbs allow pseudo-clefting and others do not!" (p179)

I would extend this argument to the tests Palmer uses, since they do not form coherent sub-classes of sequential verbs. Palmer's own categorisation suffers from the problem of being over-complicated partly because it is descriptively thorough. Although it may be argued in some cases that 'commonsense simplicity' is not a good measure of the adequacy of a grammar (e.g. García 1967), it is certainly true that the most complex descriptions must be criticised on the grounds that they are almost as complex as the data being described. In the present case, a complex description would either have categories of only one member or (as in Palmer's description) it would have categories subdivided by recurring cross-classificatory features which undermine the significance of the basic distinctions.

There are a number of conclusions I have reached from other work on the subject of sequential verbs. The first is that sequential verbs are not grammatically similar to auxiliaries, except in so far as they impose a restriction (or restrictions) on the form of the following verb phrase. There are, however, some researchers who see a semantic similarity between the auxiliaries and sequential verbs in that they have fundamental meanings which can be

used to modify the meaning of almost any verb. García (1967), for example, discusses the class which he calls 'aspectual semi-auxiliaries' (including 'begin', 'continue' and 'stop') and makes a strong case for them to belong to the auxiliary class since they have no selectional restrictions and are therefore highly productive:

"Since selectional restrictions are precisely what the lexicon is made of, while freedom of distribution is the hallmark of grammatical items, it must be admitted that 'begin' and its peers make rather poor lexical items" (p861)

Joos (1964), despite making the decision to "(leave) them where I found them among the hundreds of verbs that govern infinitives", also makes the point that some of the sequential verbs are close to having little lexical meaning:

"we see a gradation - not precise, but still very clear - from one extreme of pregnant meaning and rare use to the other extreme of banal meaning and frequent use."

Palmer also uses the notion of gradience to capture the semantico-syntactic significance of the clause following a sequential verb. His range has the most auxiliary-like catenatives at one end, with subordinate clauses not equivalent to objects and the most lexical verbs at the other extreme, with following clauses which are undoubtedly object-like.

Andersson's (1985) investigation of verb complementation

leads him to construct a circular continuum of verb groups "to show how closely consecutive verb groups are knit together semantically and syntactically and how at the end (or beginning) of one group or subgroup we may already have one foot in the following verb group" (p267). Andersson's 20 groups are defined partly syntactically and partly semantically. For example, group 16 contains verbs of 'mental processes' and 'affirmation' which are followed by a to-infinitive and "normally or only occur in the passive" (p271). Andersson's is the most successful description of a large number of sequential verbs that I have seen. However, the detailed syntactic information is not matched by similar detail in the semantic part of the information on each group. The next step in my analysis, therefore, was to examine semantic aspects of the sequential verbs in my data.

#### 5.4 Restricted class of non-final VPs in sequences

One of the most important aspects of verb phrase sequences is the occurrence and distribution of individual verbs in such sequences. The non-final verbs in these sequences are drawn from a large, but possibly limited, set of verbs which have precisely the property of being followed by another verb in a non-finite form. The nearest to a definition of this class of verb to be found in Palmer (1965) is the following:

"These are those that combine with other verbal forms in complex phrases (2.1.4) with regular rules of co-occurrence."

He also describes the class as;

"(3) Catenatives KEEP, WANT, LIKE, SEE and many others."

In order to understand Palmer's definition fully, we have to appreciate what he means by the term 'complex phrase'. He uses this term to describe those sequences of verbs which have (or, presumably, could have) more than one marker of tense, negation and passive voice. The 'catenatives' (Palmer's term after Twaddell 1960) are therefore those verbs which occur in positions other than last in such complex phrases. This section is concerned with establishing the possible range of such a class of verbs.

For the present section (and for section 5.5) I investigated the whole of the data collected in the first stage - i.e. the whole of the letters page, editorial and article from each newspaper for each of the five days concerned. Although this data was not all exploited in the main study of 2400 examples, it was recognised from the beginning that such a natural extension of my database would be useful for areas of investigation that yielded small numbers of examples in the main database. This wider database allowed me to confirm findings from the computerised database and gave a large number of extra examples on which to base lexical description. In the following discussion, the computerised database is referred to as 'the data' and the rest of the material is

referred to as the 'supplement'.

Table 5.6 gives a summary of the number of different verb types occurring in non-final position of verb phrase sequences in both the data and the supplement. These verb types are also presented as a percentage of the number of sequence tokens. The complete list of these verbs and their distribution can be found in Appendix 5.

Table 5.6 Verbs occurring in non-final positions in sequences

	diff vbs (types)	VP seqs (tokens)	% (of seqs)
data	75	181	41
supplement	127	444	29

This table shows that the class of verbs which occur in non-final position in verb phrase sequences is limited. The number of different verbs seen as a percentage of the total number of verb phrase sequences decreases as the amount of data increases. This indicates that a greater proportion of the total set of sequential verbs has been 'found'. I would expect the proportion of verb phrase sequences having 'new' sequential verbs to reduce to near 0% as the amount of data increases. This hypothesis will have to await further investigation which may well show that the set of sequential verbs is not completely closed in English.

Joos (1964), discussing these verbs under the heading 'quasi-auxiliaries', makes an estimation of the number of

sequential verbs to be found in English, based on the numbers found in his data:

"Then the laws of vocabulary statistics suggest that perhaps a few hundred different verb-bases are freely used in this way, but certainly not over a thousand, from among the dozens of thousands of English verb-bases." (p20)

Joos also gives figures for sequential verbs followed by an infinitive form and he states that only 24 out of 75 'verb-bases' (32%) occurred once in the data. By comparison, 35 out of 75 (47%) verbs in the pilot data and 59 out of 127 (46%) in the main data occurred only once. The difference between Joos' figures and mine may be a result of his restriction to examples with following infinitive forms although a large majority (82%) of my examples also contained infinitives. Another explanation of the difference may be that different kinds of text are being investigated. Joos studied the structures to be found in a published account of a murder trial: 'The Trial of Dr. Adams' by Sybille Bedford. He quotes Henry Lee Smith who describes the language of the book as "unique in English writing. There is no other work to my knowledge that so carefully mirrors actual English speech". It remains for future research to establish whether different styles or media use different quantities of sequential verbs, more or less repetitively. The consistency of the 'single occurrence' figures for the data and the supplement, however, support the suggestion made above that the larger the database of newspaper English, the more sequential verbs are found repeated.

### 5.5 Non-final sequential verbs - a partial semantic analysis

In section 5.3, I rejected syntactic criteria as the basis for making a classification of sequential verbs. Restrictions on form of following clause, marking of tense, aspect and voice in the following clause and potential occurrence of an intervening noun phrase are different enough for individual verbs to be stated lexically for each one. Although Palmer (1965) is clearly correct in stating that this approach would not be "very illuminating", I find his categorisation even less illuminating than a list of verbs with their possible constructions. The size of the sequential verb class as estimated in the previous section, however, makes some sort of subclassification desirable. Other works often mention semantic groupings in passing. Some, like Palmer (1965), seem almost embarrassed by the semantic coherence of groupings which had started out syntactically:

"Although the headings are semantic, the basis of the verb classes is a formal one."

It may be partly linguistic fashion that makes a semantic classification seem more acceptable at this stage in the development of the subject. The rest of this section investigates the possibility of analysing the sequential verbs into sets of related lexical fields which primarily share semantic features, but which may also share syntactic properties. Such an analysis has one distinct

advantage over a purely syntactic analysis; it allows for the set of non-final sequential verbs to be expanded as new examples occur, or are found to occur, since many of these would fit into existing lexical fields. The restriction on the class could thus be said to be semantic rather than lexical and avoids the problem of establishing whether the class is actually closed in English.

This section then, outlines my view of lexical field theory and componential analysis and then attempts a partial semantic analysis<sup>4</sup> of the verbs occurring non-finally in verb sequences in the pilot database, the main database, and in the supplementary material described above in section 5.2.

Lexical field theory was first made popular by Trier (1934) and his version of the theory made the strong claim that the entire lexicon consists of a neat, interrelating structure of lexemes. Although many have since rejected this claim, the notion of lexical fields has been seen as intuitively satisfying for at least some of the data, and has therefore been used widely. The question of how to define a lexical field has been tackled in a number of different ways. Most writers are content to give "a generalised definition which would cover a number of different types of grouping" (Willis and Jeffries 1982).

---

<sup>4</sup> The practical application of these theories draws on the working methods (unpublished) of the Oxford University Press Lexical Research Unit which was in operation at Leeds University between 1980 and 1984.

Miller (1972), for example, writes that "a semantic domain is any set of words implied by an incomplete definition" and Lyons (1977) writes "the set of lexemes in any one language system which cover the conceptual area and by means of the relations of sense which hold between them, give structure to it, is a lexical field". Willis and Jeffries react to these definitions in the following way:

"However, both these definitions avoid the taxing question - which part of the definition does one start with? Do we group 'swim' with 'run' and 'dance' on the grounds that MOVE is part of its definition? (...) Or do we, alternatively, group 'swim' with 'sail', 'water-ski' and other WATER words or with its subordinates: 'crawl', 'breaststroke', 'butterfly' etc.? Basically there is no one correct way to organize a lexical field and how it is done will depend on what relationships you are hoping to show through such organisation"

This view of lexical fields is appropriate for the study of sequential verbs in two ways. First, many of the sequential verbs occurring in the data belong to wider lexical fields when they are analysed for reasons other than their sequential nature. For example, the verb 'oblige' is found among the following verbs:

bid	appoint	coerce
charge	decree	compel
command	dictate	constrain
direct	ordain	expect
enjoin	prescribe	insist
instruct	rule	oblige
order		require
stipulate		
tell		

Whilst some of these verbs would presumably be found non-finally in verb phrase sequences if the database were large enough, others (e.g. 'stipulate' and 'insist') occur

only with finite clause complements. Similarly, the placing of the field as a whole within a wider framework may differ according to the required emphasis. For example, the 'oblige' field is classed as one concerning subject volition (SUBJ WANTS OTHER TO ACT) in the present study where only sequential verbs are in focus. By contrast, the 'oblige' field fits into the range of LANGUAGE AND COMMUNICATION verbs when wider terms of reference are involved.

The preceding paragraph introduced the notion of componential analysis into the discussion; this technique of analysing the meaning of a word (or, in the case of polysemy, a word-sense) is often used in conjunction with lexical field analysis. Herbst et al. (1979) say:

"Interestingly, some of the scholars who have been explicitly concerned with semantic theory (-) have attempted to combine the theoretical framework of componential analysis and the study of lexical fields."  
(p80)

The word 'interestingly' sounds as though Herbst were surprised, but it seems wholly unsurprising that while componential analysis was first developed independently of field theory it should have been adopted by adherents of this theory. The barren nature of lexical field theory used alone is illustrated by asking a foreign learner of English to use a Thesaurus without the help of a dictionary. Although accomplished speakers of English can use a thesaurus by drawing on their knowledge of the

meaning differences and similarities, learners hoping to enrich their vocabularies need these similarities and differences explained. As Wierzbicka (1972) says:

"Replacing a word by its synonym or a number of more or less exact synonyms has nothing to do with defining. A definition is an act of analysis - it must replace one word with a group of words corresponding to particular components of its meaning." (p13)

The converse of the argument given above, that componential analysis is diminished if it is not used in a lexical field framework, is also true. Many components of meaning are important only in so far as they distinguish between two words with similar meanings. One example is Nida's definition of the verbs 'run' and 'walk' which he distinguishes by contrasting the continuous contact made with the ground in walking whilst running involves both feet leaving the ground for short intervals. If these verbs had been defined in isolation it would probably have been the speed of movement which was emphasised in the definition. By contrasting the words, one is made aware of the fact that we can walk very quickly and run slowly; the contrast must therefore lie elsewhere.

The beginning of this section established the background to the semantic analysis of sequential verbs occurring in the data. The remainder of the section describes the resulting analysis.

The list of verbs occurring in non-final positions in verb

phrase sequences was examined carefully for groups which seemed to share semantic features and the most widely shared features were taken as the major divisions in the whole group. The list was divided into three main fields on the basis of this preliminary examination and a fourth was added later to accommodate some of the 'basic' verbs. These fields are characterised below in terms of the semantic feature which they appear to share:

A S WANTS SOMETHING TO HAPPEN

B S ATTRIBUTES SOMETHING (ACTION OR QUALITY) TO SOMEONE

C VERBS REFERRING TO THE PROCESS OF AN ACTION OR STATE

D .This group is not semantically united except in containing verbs which appear to have little lexical meaning (see end of this section for more discussion).

It should be noted here that many of the verbs listed in fields A and B below can be used with non-personal subjects despite the clearly human components of 'wanting' and 'attributing' which form part of their definitions. I would argue that all such cases are examples of a particular metaphorical process by which non-human subjects are interpreted as acting in a human way. Some of these transferred uses of a verb are well-established:

e.g. The rain threatened to spoil the picnic.  
(invented)

The computer inspired me to write.  
(invented)

Other verbs are not known for functioning in this way, but

can do so given an appropriate context:

e.g. The black clouds defied me to take a walk.  
(invented)

Still other members of the fields below are unlikely to be used in this way, except in the most creative of contexts:

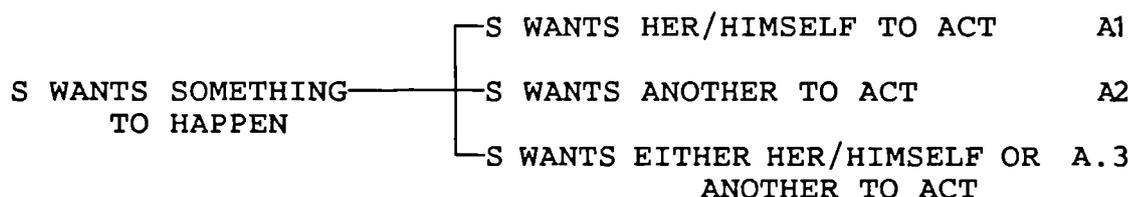
e.g. The sun urged me to lie down and rest.  
(invented)

Since all of the relevant examples in the data were preceded by a human subject (or subject whose referent is human), my approach is to characterise the verbs as requiring human subjects and supplying human qualities to those subjects whose referents are not, in fact, human.

Fields A to D are treated separately below, but it will become clear that the first two, in particular, have sub-fields containing similar semantic features and are to this extent related to each other. These kinds of cross-classification are common in lexical field analysis and serve to show that any two-dimensional classification is arbitrary since any of the shared features could have been chosen as the main lines of division between fields. However, the following description fulfils the criterion of being the simplest descriptively adequate analysis which caters for the data in question and is in this sense the 'best' analysis of my data.

The first lexical field introduced above (A) can be divided into another three sub-fields according to the

data:

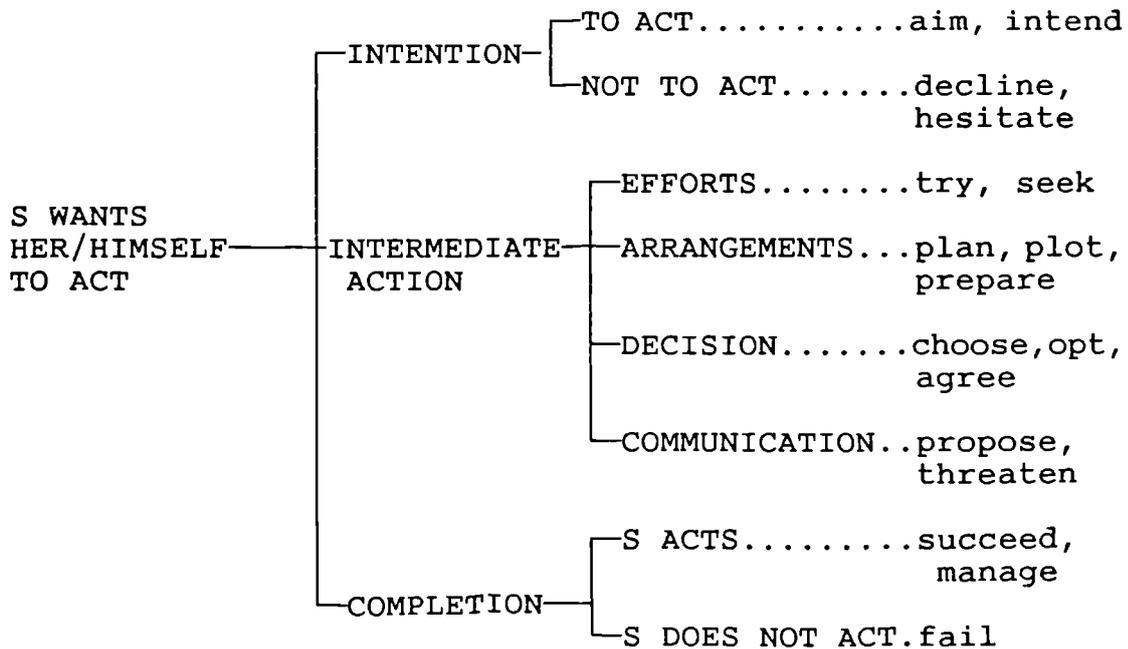


The wording of the above semantic features illustrates one of the difficulties of lexical description; the researcher is restricted to the circularity of using words to describe words. In this case, the verb WANT is being used in the semantic features as well as occurring as one of the verbs in the field.

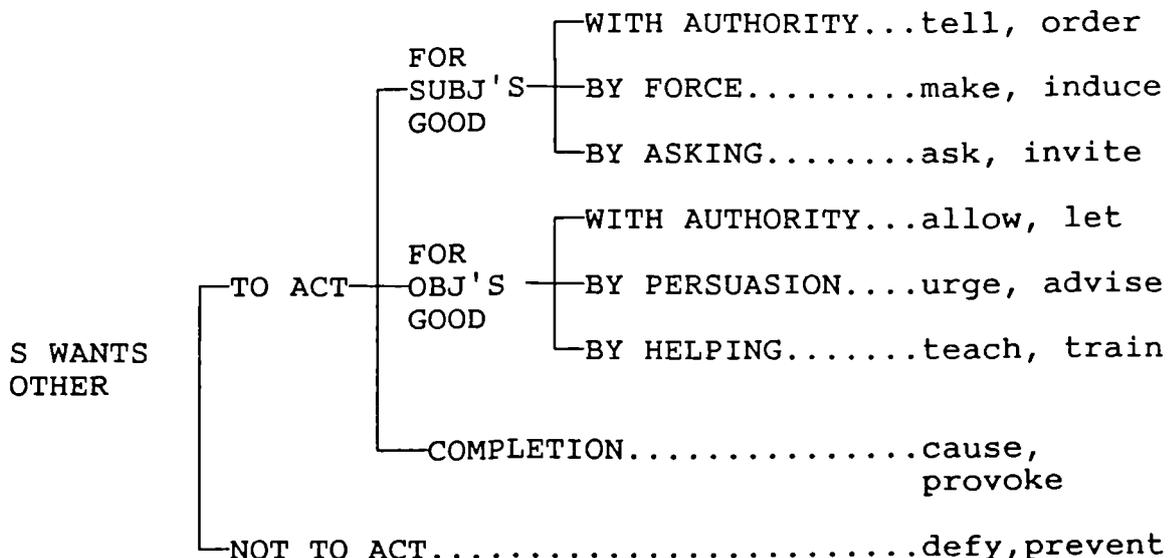
There have been various attempts, both theoretical (Wierzbicka) and practical (LDOCE) to avoid this circularity, but it is clear that the 'problem' is unavoidable once the description moves away from distinguishing rather general, barely related verbs toward the problems of distinguishing specific and closely-related verbs (Willis and Jeffries 1982). I consider this issue something of a pseudo-problem, created by the desire to appear scientific in one's methodology. Once the 'problem' is acknowledged as unavoidable, the researcher has to find the best way of incorporating circularity into her/his description. One solution is to make the circle as wide as possible; to use defining words far from the field under investigation, making sure that none of the words defined also appear in the field. However, there are many fields where even this approach is impossible or where the resulting description is awkward or unwieldy. In these

cases, there seems to be some kind of semantic 'primary' at work (see Wierzbicka) and one of the verbs described by the field has to be used to characterise it. In the fields under investigation here, WANT has been used to represent the notion of the Subject's wish/desire (also circular) that some action should (or should not) be taken.

The first sub-field above (A.1) divides as follows:

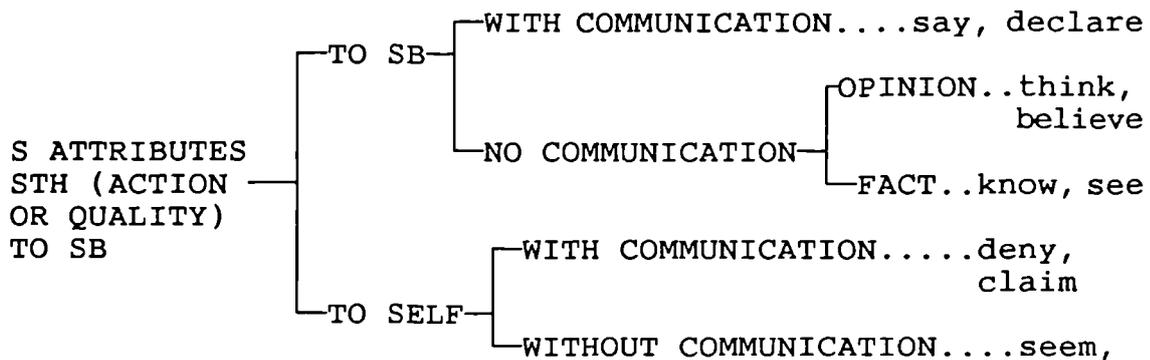


The second sub-field (A.2) divides in the following way:

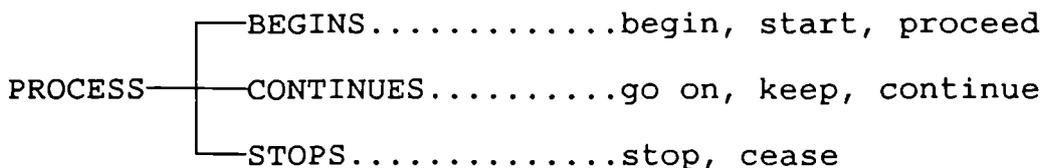


The third subfield (A3) consists of a single field and contains verbs such as 'like', 'love', 'want' and 'wish'.

The second main group of sequential verbs (group B) has the following divisions:



The third main group of sequential verbs found in my data (C) are those which are often referred to as 'process' verbs since they describe the beginning, middle or end of an action. These aspects of processes provide the main divisions of the group:



The preceding analysis represents most of the sequential verbs occurring in the data and in the supplement. There were, however, a number of verbs which did not seem to belong to these fields. Some of these were 'impersonal' verbs such as 'happen', 'prove' and 'turn out' which seem to form a group on their own. This group, in particular, share the characteristic of occurring with an impersonal

pronoun when followed by a finite clause, but allowing the logical subject of the following clause to be 'raised' to subject in the main clause when followed by an infinitive clause:

e.g. Ras Kuomba Balogun turns out to be a chum of  
Tottenham's wild man, Bernie Grant.

[Ex Ed]

It turns out that Ras Kuomba Balogun is a chum  
of Tottenham's wild man, Bernie Grant.

[rewritten]

Others, such as 'serve' and 'stand', which are also impersonal, seem rather idiomatic in their collocation potential:

e.g. The whole exercise served to reinforce the  
unspoken bargain

[Ti Ed]

Some of the verbs which do not fit into the fields above would presumably form their own fields if there was enough data to produce more of the set of sequential verbs. There are a few verbs, however (group D), which seem to be different from the exceptions already mentioned. They correspond to special uses of the verbs 'be', 'have', 'come', 'go' and 'get' when they are followed by non-finite verb phrases. Some of these verbs ('be to' and 'have to') are treated as straightforward examples of sequential verbs by Joos (1964). Others treat them as special individual cases as a result of their different overt and covert syntactic features. Palmer (1965), for example, decides to treat 'have to', but not 'be to' as a

full verb on the grounds that the former (but not the latter) can cooccur with other auxiliaries and tends to use 'do' for negative, interrogative, emphatic and 'code' forms. In the semantic treatment proposed in this chapter, the sequential uses of these verbs are grouped together on the grounds that they are in some way 'basic' in their meaning. The same group of verbs are part of what Liefrink (1973) calls the 'prime' verbs; although his reasoning is based on their frequent occurrence in phrasal verbs, he also implies that they are semantically less complex than other verbs:

"the combination (prime verb + complement) in these cases is felt to form some kind of semantic unit"

#### 5.6 The ordering of longer sequences

In the description of the pilot study in chapter 1 (section 1.3.10), the hypothesis was put forward that, as in many African languages and related pidgins and creoles, verb phrase sequences in English have ordering priorities among them which influence the order of two or more sequential verbs occurring in sequence. Any such ordering would clearly not be grammatical, since the logic of the relationship between the sequential verbs determines the order. Thus we can say both of the following sentences according to our intended meaning:

e.g. He started to want the meal	[invented]
He wanted to start the meal	[invented]

However, it was possible that there would be some signs of semantic ordering in the data and I decided to investigate the longer sequences for any signs of selectivity. There was not enough data in the pilot study to investigate this hypothesis and the main database also produced only 2 examples of sequences containing two or more sequential verbs. The data supplement (as described in section 5.4), however, produced 21 further examples of longer sequences, bringing the total for all available data (including examples from the pilot study) to 27.

Although 27 examples do not go far towards providing evidence for ordering priorities, any patterning which seems to emerge from them may provide the motivation for future work on this topic. I therefore decided to investigate the ordering of sequential verbs in these 27 examples using the semantic groupings established in section 5.4 as a guide.

Assuming that sequences with two near-synonyms such as "I stopped ceasing.." or "They wanted to wish to.." are the most unlikely combinations of sequential verbs, my hypothesis was that sequential verbs occurring together would tend to come from different lexical fields. I therefore coded the 27 examples of longer VP sequences according to the fields in which their verbs belonged. For this purpose I took the fields to be those which correspond to groups A1, A2, A3, B, C and D in my analysis.

Whilst I was studying these sequences, it became clear that the occurrence (or non-occurrence) of an intervening noun phrase may be significant for the ordering of the sequential verbs. This information was added to my coding for the 27 examples. Table 5.7 shows the combinations of lexical fields with intervening noun phrases in these examples:

Table 5.7 Combinations of sequential verbs in longer sequences

	1st verb	int NP	2nd verb	int NP
1.	A3		A2	NP
2.	A1		A2	NP
3.	B		A2	NP
4.	D		A2	NP
5.	A1		A2	NP
6.	A1		A2	NP
7.	C		B	NP
8.	B		D	NP
9.	A1		A2	NP
10.	B		A2	NP
11.	B		A3	NP
12.	A3		C	NP
13.	A2	NP	C	
14.	A2	NP	A2	
15.	A2	NP	A1	
16.	A2	NP	C	
17.	A2	NP	A1	
18.	A3	NP	A2	
19.	A2	NP	A2	
20.	A2	NP	C	
21.	A3	NP	C	
22.	A2	NP	A2	NP
23.	A2	NP	D	NP
24.	A3	NP	A2	NP
25.	D		A1	
26.	A3		B	
27.	D		C	

The most striking result of this coding exercise was the discovery that in sequences of three verb phrases (containing two sequential verbs and one 'other' verb) only a small minority of examples (3 out of 27 = 12%)

contain two intervening noun phrases and the same number of examples contain no noun phrases, apart from the original subject. In other words, three examples have the same subject for all three verbs and three examples have different subjects for each of the three verbs; the large majority have two verbs sharing a subject and the third with a different subject. Some examples will illustrate these three groups:

e.g. we have to try to hold objective discussion  
[Te Let]

we want you to help us create an outspoken  
intolerance for drug use [Te Ed]

I cannot order an education authority to stop  
circulating such a book [M Ed]

The first and second examples above represent those having only one subject and three subjects respectively. The reason why they are rare in this data may be that they are typical of styles not represented here. The first example in particular seems 'conversational', and may be more frequent in spoken English. Alternatively, it may be the case that we avoid such structures because they overload the memory in some way. Note, for example, that in casual speech (and in children's speech) we often treat the second and third verbs of a sequence having only one subject as though they were coordinated rather than subordinated. The first example above would sound quite natural rewritten in this way:

e.g. we have to try and hold objective discussion  
[rewritten - Te Let]

Explanations for the lack of structures such as these must await further research and, in particular, further data collection.

In continuing the search for ordering priorities among sequential verbs, we should note that there seems to be no absolute restriction on combinations of sequential verbs. As with many other areas of linguistics, even the 'unlikely' examples given above and containing two near-synonyms can be given a plausible context by an 'armchair linguist' with time and imagination. No individual example could have its sequential verbs reversed without a dramatic change in meaning:

e.g. I would prefer to trust the British Embassy in  
Japan to look after... [T Let]

I would trust the British Embassy in Japan to  
prefer to look after... [rewritten - T Let]

But most examples can be reversed without loss of grammaticality. The question is, therefore, whether there are any tendencies, in different styles of English or in English generally, to choose certain combinations of sequential verbs in preference to others. I have already hypothesised from my data that there is a tendency to combine a verb which is followed by a noun phrase with one that is not. I have also hypothesised that there will be few examples where both sequential verbs are from the same lexical field. There were only three examples in the data with both sequential verbs being coded alike. All of them

had two verbs from the field labelled A2. These pairs of verbs were, however, far from being near-synonyms. The pair 'allow - prevent' are opposite in meaning and the pairs 'call for - allow' and 'ask - help' are in separate main sections of the field.

Other ordering priorities are much more difficult to establish from such a small number of examples, but some directions for future study can be inferred. For example, verbs from field A3 occur only initially in the data, verbs in field C occur after verbs from A2 and D but before verbs from B. The only group which seems to have a virtually unrestricted combining potential is the group of A2 verbs. More data would, presumably, uphold some of these generalisations and disprove others.

## Chapter 6

### Conclusions

#### 6.1 Aims of the thesis

The first stage of this thesis was to collect a corpus of 'newspaper English' and store it in computerised form. The kind of computer storage used was different from some corpora in that it set out to exploit the advantages of database programming which allow the user flexible access to her or his data.

The first main aim of the research was to analyse the linguistic features coded in the database in three different ways: as though the corpus were a homogeneous and representative sample of newspaper English in general, by contrasting and comparing three kinds of text-type and by contrasting and comparing different newspapers.

The second main aim of the thesis was to use the data as the basis of a discussion of three problems of description encountered in the early stages of the research. The problems shared the characteristic of being questions of borderlines: between subjects and complements, between adjectives and participles and between full and auxiliary verbs.

## 6.2 Results of the thesis

### 6.2.1 The style of newspaper English

The result of considering the database as an example of 'newspaper English' was a description of the coded features presented mainly as percentages of the total data items. In view of the rather wide scope of the projected study, it was impossible to include any comparisons with data from other registers of English. These results are, however, available for comparison with work on other areas of English and could also be compared with similar studies of other corpora of newspaper English. Both kinds of comparison could be used to verify the figures obtained in this study and distinguish between those features which were typical of newspaper data and those which had a wider relevance.

### 6.2.2 Styles of different text-types in newspaper English

The database was then investigated for internal structuring, concentrating first on the three text-types which had been chosen for study: news articles, editorials and letters.

These text-types were shown to resemble each other in a number of the features which had been coded. These similarities included; the finiteness of their verb phrases, the proportion of main verbs with no auxiliaries, the number of verb phrases occurring in sequences, the

distribution of most clause types and the frequency of occurrence of plural subjects and complex subjects.

Differences between the text-types were more interesting since they revealed a recurrent pattern in which the letters text-type had figures falling closest to the average for the whole data while the figures for the other text-types were ranged on either side of the average. Some of the features conforming to this pattern were explained in terms of their apparent objectivity or air of authority. These were; the occurrence of copula 'BE', the use of modal auxiliaries and the use of the passive voice. In these cases it seemed as though the article writers were aiming at objectivity whilst the editorial writers were conforming to their expected function and giving opinions. Other features were not so easily explained in this way: the '-ing' participle and the distribution of noun clauses among the functions of subject, object and complement.

### 6.2.3 Styles of some British national newspapers

The data for this study was set up with two cross-cutting classifications. As well as the classification based on text-type, there was also a division into eight different newspapers from which the data were collected. My own understanding of the range of British newspapers led to the choice of a variety of papers which correspond to the popular image of a 'quality' and a 'popular' press. I was

therefore assuming that there would be some differences between the individual papers. However, in treating them separately for coding purposes, I made no assumptions about whether the popular groupings would be reflected in the linguistic features of the papers.

The results of the statistical analysis of coded features supported the idea that these eight newspapers were representative of groups, but there seemed to be three, rather than two groups. In addition to the quality and popular groups, there was a group of two papers each of which varied between holding a genuinely 'central' position between the quality and popular groups and having figures typical of one or other extreme.

This pattern of three groups was indicated for all of the features coded except two which were found to have no significant difference between the groups. Of these two (passive and VP sequences), the first showed a significant difference between newspapers and the second did not. Both, however, had figures which indicated that these may be stigmatised features of the language which are sometimes used by a writer aspiring to a more elevated style or, conversely, are avoided in a kind of linguistic 'inverse snobbery'.

All of the remaining features investigated conformed to the pattern described above. Most of these results were, if not predictable, at least explainable in terms of the

different aims of the groups of newspapers. For example, it was not surprising to find that subordinate clauses were much more common in the quality than in the popular group nor that there were more human subjects in the popular newspapers and more abstract subjects in the quality press. The unexpected results were that the quality papers had fewer identified female subjects than the popular papers and that the popular papers had fewer occurrences of copula 'BE' than the quality papers. The first of these results is explained, on reflection, by the contexts in which women feature in popular papers. The quality press, by comparison, reveals not bias, but accuracy in reporting public affairs. The second unexpected result, the lack of examples containing 'BE' in the popular papers, seemed to contradict the findings for text-types from which I had drawn the conclusion that the kind of attributions made by the copula were seen as subjective and therefore occurred more rarely in articles than in editorials. It was hypothesised, but not tested, that there may be a complex relationship between type of complement following 'BE' and the apparent authority of the clause.

The features as tested for individual newspapers were mostly used to confirm the significance of differences between groups of papers. There were, however, some features which did not differ between paper despite being significantly different for the groups. These features

(e.g. 'BE', VP sequences and male subjects) showed a cumulative difference when the groups of papers were tested which was too small to show up when the individual papers were tested.

### 6.3 Insights into problems of description

#### 6.3.1 The subject/complement overlap

Chapter 3 investigated what appeared to be an area of overlap between the grammatical functions of subject and complement. Much was already known about the observable differences between subjects and noun phrases occurring later in the clause. These overall differences such as the fact that NPs occurring late in the clause are more likely to be complex than than subjects and the fact that indefiniteness occurs later rather than earlier in the clause were confirmed in the present study. However, this study revealed a further aspect of these well-known facts which was that, for definiteness at least (and for complexity in most cases), individual complements will not be more definite (or less complex) than their corresponding subjects. There is an important difference here between the subject-complement and subject-agent relationships. The agents occurring in the data also showed an overall greater complexity and more frequent indefiniteness than the subjects, but individual cases were found to contradict this finding.

In addition to their different usual positions, the difference of complexity and of definiteness, the subjects and complements were unambiguously differentiated by the violation of subject-complement concord (which makes clear which NP is dictating the number of the verb), the occurrence of adjectives and adjective phrases in the complement (which are indicative of an attributive function) and, in some cases, the occurrence of a negative verb (which makes a reversed version not only different in focus, but also different in cognitive meaning). There were a small number of cases where the functions of subject and complement seemed to be 'merged' around an equative use of the verb BE. It was suggested that in these examples there was minimal difference of meaning when the NPs were reversed.

A further discovery of this chapter was the fact that subjects occurring with copula 'BE' when it is followed by a noun phrase are not typical of subjects as a whole. In particular, they are more often realised as impersonal pronouns, have even fewer indefinite articles and are more often simple (unmodified) than subjects in general.

### 6.3.2 Distinctions between -en participles and adjectives

Chapter 4 continued an investigation, begun in the pilot study, into the -en forms which were difficult to identify as belonging to the verbal or adjectival word-class. There were two kinds of conclusion drawn from this part of the

study. First, it was shown that some of the 'tests' which are often quoted as distinguishing between these word classes are much more satisfactory when used to test actual context rather than potential contexts based on introspective data. Some contexts were found to be unambiguous markers of passive (e.g. following objects or complements) and some were typical of passive forms rather than adjectives (e.g. intervening adjuncts). Similarly, some contexts were only found with adjectives (e.g. intervening intensifiers) and some were found to be more typical of adjectives than passives (e.g. following sequential verb phrases). A distinction was also made between the 'temporary' categorisation of an individual -en form as a participle or an adjective and the 'permanent' assignment of -en forms to these categories when they had proved themselves full members of the category. The notion of 'merger' (Coates 1983) was again invoked to describe the kind of overlap discovered in this chapter.

The second type of conclusion was that semantic considerations constitute a useful supplement to purely formal tests of grammatical status of -en forms. These conclusions included the finding that 'complementary phrases' are identifiable only by semantic means, that time (indicated by a combination of tense, aspect and adjuncts) is often conclusive in assigning individual examples to the verbal or adjectival groups and that the

identification of polysemous senses can also contribute to the distinction.

### 6.3.3 The problem of sequential verbs

The final part of this study was the investigation of sequential verbs, reported in chapter 5. These had begun as a descriptive inconvenience (how many verb phrases were involved?) but soon became a source of deeper interest.

One of the findings of my work on these verbs was that, apart from having a higher-than-average number of human subjects, they behaved similarly to other freely occurring verbs (in the proportion of finite/non-finite verbs and in their distribution among the clause types).

The debate on the grammatical status of these verbs has been long, complex and inconclusive. From my reading of other work on the subject it seemed that one had to accept that they were not auxiliary verbs, despite sharing some features of auxiliaries. It was, however, clear that some of them (e.g. sequential uses of 'be', 'have' and 'go') are semantically simpler than many full verbs. Another conclusion from reviewing other work was that the attempts to sub-classify these verbs on formal grounds were unsatisfactory.

The examples of sequences in my data suggested that the sequential verbs form a large, but restricted sub-class of full verbs which are more satisfactorily sub-classified

into fields of lexical items related by shared semantic features. It is probable that syntactic features would be shared, in many cases, by members of a field. The arrangement of verbs into lexical fields allows for the possible expansion of the class of sequential verbs, without a fundamental change in the description, when other near-synonyms are used in this way.

Longer sequences of verb phrases have been ignored by most treatments of this topic in English but discussion of 'serial verbs' in other languages led me to consider the possibility of there being ordering priorities among sequential verbs in English. One discovery of this section was that any sequence of two sequential verbs (followed by any full verb) will usually contain one intervening NP, either between the sequential verbs or between the second sequential verb and the 'other' full verb.

Another general restriction on the co-occurrence of sequential verbs was that they are usually from different lexical fields. This restriction may be part of a wider ordering tendency which did not show up in the small number of examples I was able to gather from my data.

#### 6.4 Future work based on this thesis

##### 6.4.1 Further work based on the present data

Many of the features coded in the database used for this study were not exploited to the full, and some were not

used at all. Future studies using the same material could include the following:

An investigation of the difference, in practice, between formal and notional aspects of subject type (e.g. the categories of human/abstract/inanimate/animate) on the same basis as the distinction between formal and notional number.

A study of the different ways of negating the clause as used in the data. Field 6 contains many examples where the verb phrase itself was not negated, but the force of the clause was felt to be negative.

The structure of clauses was coded in field 11 and these entries could form the basis of an interesting study of the range of basic clause types occurring in this kind of English.

Finally, there is a large amount of information coded in field 15, subject structure, which could be used to investigate different types of pre- and post-modification of the subject noun phrases in this data.

#### 6.4.2 Future work based on studies reported in this thesis

Perhaps more likely areas of future work are those based on the investigations reported here. Like all research, this thesis raises as many questions as it answers, thereby pointing to possibly fruitful future studies.

Those which I would like to continue, given the opportunity, are the following:

The comparison of the present database with other data representing different 'registers' of English. This would involve setting up similar databases or adapting the information available in existing ones.

Further study of the relationship between grammatical structures (such as the passive and complementation of 'BE') and intended or received impressions of 'objectivity' and 'authority'. Related to this topic would be the study of apparent aspirations to style not stereotypical of the context.

Finally, I would like to continue the investigation of sequential verbs from both a semantic and a syntactic basis. A particular aim is to collect and analyse large numbers of longer sequences in order to confirm or contradict the findings of chapter 5 and compare data from different styles and media in English.

Bibliography

- Aarts, F.G.A.M. 1971 "On the distribution of Noun-Phrase Types in Clause-Structure". *Lingua* 26.
- Aarts, J. and T. van den Heuvel 1984. "Linguistic and computational aspects of corpus research" In Aarts and Meijs (eds) 1984.
- Aarts, J. and W. Meijs (eds) 1984 *Corpus Linguistics*. Amsterdam: Rodopi.
- Allan, K. 1987. "Hierarchies and the choice of left conjuncts (with particular attention to English" *Journal of Linguistics* 23.
- Allen, W.S. 1959. *Living English Structure*. London.
- Andersson, E. 1985. *On verb complementation in written English*. Lund: CWK Gleerup.
- Bach, E. and R.T. Harms (eds) 1968. *Universals in Linguistic Theory*. New York
- Bock, M. 1982. Some Effects of Titles on Building and Recalling Text Structures. In *Discourse Processes* 3,4. 301-311.
- Bolinger, D. 1961. *Generality, Gradience and the all-or-none*. The Hague: Mouton
- Borsley, R.D. 1984. "VP complements: evidence from Welsh". *Journal of Linguistics*, 20.
- Bresnan, J. 1982. *The mental representation of grammatical relations*. Cambridge, Mass: MIT Press.
- Coates, J. 1983. *The semantics of the modal auxiliaries*. London: Croom Helm.
- Chomsky, N. 1965. *Aspects of the Theory of Syntax*. Cambridge, Mass: MIT Press.
- Crystal, D. 1982. *Introduction to Language Pathology*. Edward Arnold.
- Crystal, D. and D. Davy 1969. *Investigating English Style*. London: Longman
- Declerck, R. 1986. "Two notes on the theory of definiteness" *Journal of Linguistics*, 22.
- Doležel, L. 1969. "A framework for the statistical analysis of style" In Dolezel and Bailey (eds) 1969.

- Doležal, L. and R. Bailey (eds) 1969. *Statistics and Style*. New York: Elsevier.
- Duke, J. and T. Screeton 1982. *Extract User Manual*. University of Leeds Computing Service.
- Ellegård, A. 1978. *The Syntactic Structure of English Texts*. Gothenburg Studies in English 43. Stockholm: Almqvist and Wiksell
- Enkvist, N.E. 1964. "On defining style: An essay in applied linguistics" In Spencer (1964)
- Fillmore, C. 1968. "The Case for Case". In Bach and Harms (1968).
- Fodor, J.A. and J.J. Katz (eds) 1964. *The Structure of Language*. Englewood Cliffs, N.J.: Prentice-Hall.
- Foley, W.A. and M. Olson 1985. "Clausehood and Verb Serialisation". In Nichols, J. and A.C. Woodbury (1985).
- Francis, W.N. 1964. *Manual of Information to accompany a standard sample of present-day edited American English, for use with digital computers*. Providence, R.I.: Department of Linguistics, Brown University.
- Fries, C.C. 1952. *The Structure of English*. New York: Harcourt, Brace and World
- Fries, U. 1987. "Summaries in Newspapers: A Textlinguistic Investigation" In U. Fries (ed) *The Structure of Texts*. Tübingen: Gunter Narr Verlag
- García, E.C. 1967. "Auxiliaries and the Criterion of Simplicity" *Language* 43.
- Garside, R.G., G.N. Leech, and G.R. Sampson (eds) 1987. *The Computational Analysis of English* London: Longman.
- Gillon, B.S. 1987. "The readings of plural noun phrases in English." In *Linguistics and Philosophy* 10.
- Green, G. 1974. *Semantics and syntactic regularity*. Bloomington and London: Indiana U.P.
- Gustafsson, M. 1975. *Some syntactic properties of English law language*. Publications of the Department of English, No. 4. University of Turku.

- Gustafsson, M. 1982. "Fronting of adverbials in four genres of English." In Jacobson (ed) 1982.
- Halliday, M.A.K. 1967 and 1968. "Notes on transitivity and theme in English, Parts 1, 2 and 3". Journal of Linguistics 3 and 4.
- Halliday, M.A.K. and R. Hasan 1976. Cohesion in English. London: Longman
- Harris, Z.S. 1946. "From morpheme to utterance" Language 22.
- Herbst, T. 1979. "Some issues in German semantic studies." In Herbst, Heath and Dederding (eds) 1979.
- Herbst, T., D. Heath and H.M. Dederding (eds) 1979. Grimm's Grandchildren. Current Topics in German Linguistics. London: Longman.
- Huddleston, R.D. 1971. The sentence in written English: A syntactic study based on an analysis of scientific texts. Cambridge: C.U.P.
- Huddleston, R.D. 1984. Introduction to the Grammar of English. Cambridge: CUP.
- Huddleston, R., R. Hudson, E. Winter and A. Henrici. 1968. Sentence and clause in scientific English. London: Office of Scientific and Technical Information.
- Jacobson, S. 1980. Papers from the first Scandinavian Symposium on syntactic variation. Stockholm: Almqvist and Wiksell.
- Jacobson, S. 1982. Papers from the second Scandinavian Symposium on syntactic variation. Stockholm: Almqvist and Wiksell.
- Jacobson, S. 1986. "Synonymy and hyponymy in syntactic variation" In Jacobson (ed) 1986.
- Jacobson, S. (ed) 1986. Papers from the third Scandinavian Symposium on syntactic variation. Stockholm: Almqvist and Wiksell
- Jespersen, O. 1933. Essentials of English Grammar. London: Allen and Unwin.
- Johansson, S. 1986. "Some observations on the order of adverbial particles and objects in the LOB corpus." In Jacobson (ed).

- Johansson, S., G.N. Leech and H. Goodluck 1978. Manual of information to accompany the Lancaster-Oslo/Bergen corpus of British English, for use with digital computers. Department of English, University of Oslo, Oslo.
- Johansson, S. with E. Atwell, R. Garside and G. Leech 1986. The Tagged LOB Corpus. Users' Manual. Norwegian Computing Centre for the Humanities, Bergen.
- Joos, M. 1964. The English verb: form and meaning. Madison: The University of Wisconsin Press.
- Kaplan, R.M. and J. Bresnan. 1982. Lexical-functional grammar: A formal system for grammatical representation. In Bresnan 1982.
- Kilby, D. 1984. Descriptive syntax and the English verb. London: Croom Helm.
- Kittredge, R. and Lehrburger, J. (eds) 1982. Sublanguage. Studies of language in restricted domains. Berlin and New York.
- Klima, E.S. 1964. "Negation in English" In Fodor and Katz (eds) 1964.
- Kucera, H. and W.N. Francis 1967. Computational Analysis of Present-Day English. Providence, Rhode Island.
- Kytö, M., I. Ossi and M. Rissanen (eds) 1987. Proceedings of the Eighth International Conference on English Language Research on Computerized Corpora.
- Lakoff, G. 1987. Women, fire and dangerous things. Chicago: U. of Chicago Press.
- Langacker, R.W. 1987. Nouns and Verbs. Language 63,1.
- Leech, G.N. 1974 (1st edn) and 1981 (2nd edn) Semantics. Harmondsworth: Penguin.
- Leech, G.N. and J. Svartvik 1975. A Communicative Grammar of English. London: Longman
- Liefcrink, F. 1973. Semantico-syntax. London: Longman
- Limber, J. 1973. "The genesis of complex sentences" In T.E. Moore (ed) Cognitive development and the acquisition of language. New York: Academic Press.

- Lyons, J. 1966. "Towards a 'notional' theory of parts of speech" *Journal of Linguistics* 2.
- Lyons, J. 1977. *Semantics*. 2 vols. Cambridge: C.U.P.
- Mårdh, I. 1980. *Headlines*. On the grammar of English front page headlines. *Lund Studies in English* 58. Lund.
- Matthews, P.H. 1981. *Syntax*. Cambridge: C.U.P.
- McCawley, J.D. 1968. "The role of semantics in a grammar" In Bach and Harms (eds)
- Miller, J. 1985. *Semantics and syntax*. Cambridge: C.U.P.
- Mulholland, H. and C.R. Jones 1969. *Fundamentals of statistics*. London: Butterworths.
- Munro, P. 1982. "On the transitivity of 'say' verbs" In *Syntax and Semantics* 15 New York: Seminar Press.
- Nichols J. and A.C. Woodbury (eds) 1985. *Grammar inside and outside the clause*. Cambridge: C.U.P.
- Nida, E.A. 1975. *The Componential Analysis of Meaning*. The Hague: Mouton.
- Osgood, C.E. 1960. "Some effects of motivation on style of encoding." In Sebeok (ed) 1960.
- Palmer, F.R. 1974. *The English Verb*. London: Longman.
- Palmer, F.R. 1976. *Semantics. A new outline*. Cambridge: C.U.P.
- Quirk, R. 1965. "Descriptive statement and serial relationship" *Language* 41.
- Quirk, R. 1984. Recent work on adverbial realisation and position. In Aarts and Meijs (eds) 1984.
- Quirk, R. and S. Greenbaum 1973 *A University Grammar of English*. London: Longman
- Quirk, R., S. Greenbaum, G. Leech and J. Svartvik 1972. *A Grammar of Contemporary English*. London: Longman.
- Quirk, R., S. Greenbaum, G. Leech and J. Svartvik 1985. *A Comprehensive Grammar of the English Language*. London: Longman.
- Quirk, R. and J. Svartvik 1966. *Investigating linguistic acceptability*. The Hague: Mouton.

- Rissanen, M. 1979. "On the position of only in present-day written English." In Jacobson (ed) 1980.
- Rosch, E. 1978. Principles of categorization. In E. Rosch and R.B. Lloyd (eds) Cognition and Categorization. New Jersey: Lawrence Erlbaum Associates.
- Rosenbaum, P.S. 1967. The grammar of English predicate complement constructions. Cambridge, Mass: MIT Press.
- Rydén, M. 1975. "Noun-name collocations in British English Newspaper Language" Studia Neophilologica 47.
- Sampson, G. 1985. The 'parsing law' document produced by the Lancaster-Leeds automatic parsing research project. (unpublished)
- Sampson, G. forthcoming. Analysed Corpora of English: A consumer's guide. In M. Pennington and V. Stevens (eds) Computers in Applied Linguistics: An International Perspective.
- Sampson and Haigh, 1988. "Why are long sentences longer than short ones?" In Kytö, Ossi and Rissanen (eds).
- Schachter, P. 1977. "Reference-related and role-related properties of subjects" In Syntax and Semantics 8. P. Cole and J.M. Sadock (eds) New York: Seminar Press.
- Schwarzenbach, J. and K.F. Gill 1978. System modelling and control. London: Edward Arnold.
- Sebeok, T.A. (ed) 1960. Style in Language. Cambridge, Mass: MIT Press.
- Siegel, D. 1973. "Nonsources of unpassives" In Syntax and Semantics 12. J.P. Kimball (ed) New York: Seminar Press.
- Somers, H.L. 1983. An investigation into the application of the linguistic theories of valency and case to the automated processing of language. Unpublished PhD. thesis, UMIST.
- Spencer, J. 1964. Linguistics and style. London: O.U.P.
- Stein, G. 1979. Studies in the functions of the passive. Tübingen: Narr.
- Straumann, H. 1935. Newspaper Headlines Allen and Unwin.

- Sullivan, W.J. 1980. "Syntax and linguistic semantics in stratificational theory" In *Syntax and Semantics 13* E.A. Moravcsik and J.R. Firth (eds).
- Svartvik, J. 1966. *On voice in the English verb*. The Hague: Mouton.
- Tottie, G. 1980. "Negation and ambiguity" In Jacobson (ed) 1980
- Tottie, G. 1982. "The co-occurrence of negation, modality and 'mental verbs' in spoken and written English" In Jacobson (ed) 1982.
- Trier, J. 1934. *Das Spracheliche Feld. Eine Auseinandersetzung*. Neue Jahrbucher fur Wissenschaft and Jungendbildung 10.
- Twaddell, W. F. 1965 *The English Verb auxiliaries*. Brown: Brown U.P.
- Varantola, K. 1984. *On noun phrase structures in engineering English*. University of Turku.
- Wells, R. 1960. "Nominal and verbal style" In Sebeok (1960) ed.
- Wierzbicka, A. 1980. *Lingua Mentalis*. Australia: Academic Press.
- Willis, P. and L. Jeffries 1982. "Participant (case) roles and lexical analysis" Nottingham Linguistic Circular, 11.2.
- Willis, P. and L. Jeffries 1982. *An internal report to the O.U.P Lexical Research Unit, Leeds University*.
- Winter, W. 1965. "Transforms without kernels?" *Language* 41.
- Winter, W. 1969. "Styles as Dialects" In Dolezel and Bailey (eds) 1969.
- Yngve, V. 1972. "The Depth Hypothesis." In F. W. Householder (ed) *Syntactic Theory* 1. Harmondsworth: Penguin.
- Zwicky, A.M. and Zwicky, A.D. 1982 "Register as a dimension of linguistic variation". In Kittredge and Lehrburger (eds).

Appendix 1aSample Output from an Extract Enquiry (pilot database)

(Only one record is given here; the format is the same for all enquiries which extract all of the fields for each verb phrase.)

VP:	permitted	reference:	
12/Te/19			
main verb:	PERMIT	simple vp?:	S NF
attached vp:	13/Te/19	polarity:	POS
vp structure:	Mvben	spare field:	
clause type:	AjCl	subj type:	H 3
P			
clause function:	NM	subj wds:	4
clause structure:	P?	subj form:	full
agent type:	n/r	lex field:	SC
agent wds:	n/r	Ven/Aj/?:	Ven
pass/act?:	act		
context ref:	4		

Appendix 1bNumerical Enquiries - Results

1.	simple	532	2.	finite	587
	complex	266		non-finite	212
3.	positive	742	4.	modal	117
	negative	48		perfective	65
	question	18		progressive	27
	neg question	4		passive	80
				no auxiliary	533
5.	MCl	316	6.	O. no C.	365
	ACl	127		C. no O.	137
	NCl	120		O and C	15
	AjCl	151		no O or C	283
	CCl	8			
	?	78			
7.	animate	5		single	476
	inanimate	40		plural	203
	human	435		uncountable	1
	abstract	274		collective	72
	male	135			
	female	49			
	male and 1st	39		female and 1st	42
	male and 2nd	1		female and 2nd	0
	male and 3rd	95		female and 3rd	7
8.	full	273		relative	40
	demonstrative			pronoun	185
	impersonal		n/r		221
9.	simple & finite	325			
	complex & finite	261			
	simple & non-finite	207			
	complex & non-finite	5			
10.	simple & positive	517		complex & pos	223
	simple & negative	12		complex & neg	36
	simple & question	3		complex & q	13
11.	MCl & simple	192		MCl & complex	123
	ACl & simple	90		ACl & complex	37
	NCl & simple	72		NCl & complex	47
	AjCl & simple	107		AjCl & complex	44
	CCl & simple	7		CCl & complex	1

12.	Mcl & finite	314	MCl & non-f	1
	Acl & finite	82	ACl & non-f	45
	NCl & finite	85	NCl & non-f	35
	AjCl & finite	75	AjCl & non-f	76
	Ccl & finite	8	Ccl & non-f	0
13.	Atvp & MCl	18	14. neg & MCl	24
	Atvp & ACl	14	neg & ACl	5
	Atvp & NCl	33	neg & NCl	8
	Atvp & AjCl	15	neg & AjCl	7
	Atvp & CCl	1	neg & CCl	1
15.	NCl & subject	12		
	NCl & object	77		
	NCl & complement	18		
	NCl & ?	11		

	F	NF	AT	NEG	POS	Q	MCL	ACL	NCL	JCL	CCL
H	319	115	88	31	399	10	167	73	80	64	12
I	30	10	1	3	37	0	17	8	1	12	58
AB	225	48	16	11	257	6	125	33	34	58	3
AN	3	2	2	0	5	0	1	2	0	1	0
1	116	29	21	12	133	3	77	27	23	9	1
2	7	1	2	0	8	0	6	1	0	0	0
3	454	145	84	33	557	13	227	88	92	126	7
S	391	85	39	27	442	9	243	72	55	70	8
P	141	61	41	16	184	7	53	29	37	56	0
U	0	1	0	0	1	0	0	0	0	1	0
C	45	27	25	3	69	0	14	14	23	9	0
FULL	255	18	23	12	257	4	142	38	49	27	1
PN	181	4	17	15	166	5	104	35	28	12	3
DEM	23	0	2	5	17	1	17	3	1	0	2
REL	39	1	4	4	36	0	0	0	0	39	0
IMP	53	0	1	5	47	1	36	7	4	2	1
N/R	31	189	67	7	214	0	15	43	35	72	1

	FUL	PN	REL	DEN	IMP	
H	127	170	19	0	0	118
I	22	1	3	1	2	8
AB	118	13	17	22	51	52
AN	2	1	0	0	0	2
1	1	111	0	0	1	33
2	0	6	0	0	0	2
3	267	68	39	23	52	145
S	157	135	17	22	53	90
P	73	43	23	1	0	61
U	1	0	0	0	0	0
C	42	2	1	0	0	27

Appendix 1cSample Output from an Extract Enquiry (main database)

(Only one record is given here; the format is the same for all enquiries which extract all of the fields for each verb phrase.)

VP:	are put	reference:	8/4 Ex
Ed			
main verb:	PUT	sense:	n/n
FIN:	F	polarity:	POS
vp structure:	5	vp form:	r
clause type:	A	subj type:	ab n 3
p			
clause function:	A	sem subj ty:	n/a
clause structure:	spa	subj form:	full
agent:	n/r	subj struc:	def aj
hn			
agent typ:	n/o	agt form:	n/r
sem agt typ:	n/a	agt struc:	n/r
special int:	p		

Appendix 1dFull Sentence Contexts for Main Database

(restrictions of space allow for only 5 examples from each newspaper/ text-type)

Times / Article / 15-9-86

1. The SDP conference yesterday endorsed the tough stand on defence by Dr David Owen, the party leader.
2. It supported him and Mr David Steel, the Liberal leader, in their bid to settle the parties' differences on the replacement of Polaris by pursuing their quest for a minimum European deterrent.
3. Alliance leaders will now intensify their drive for agreement on a minimum European deterrent which Dr Owen claims could be a "turning point of British defence policy", and which will at the least prove a handy diversion in the nuclear policy clash which has seen Alliance support drop sharply in recent opinion polls.
4. After yet another coded attack on his leadership style had been beaten off by a margin of 4-1, a delighted Dr Owen declared: "I've got what I wanted - the freedom to go for the Labour Party's jugular on the crucial question of defence at the next election."
5. Party chiefs breathed a sigh of relief after the SDP conference, meeting at Harrogate, had successfully wriggled through its biggest problem of the week - using the defence debate to send conciliatory messages to the Liberals in advance of their conference next week without undermining their own leader.

Times / Editorial / 15-9-86

1. The simultaneous transfer of Nicholas Daniloff and Gennady Zakharov into the custody of their respective ambassadors has temporarily defused what threatened to become a new explosion of East-West acrimony.
2. It may also have saved the second Reagan-Gorbachov summit meeting.
3. But the calm in American-Soviet relations has been bought at a price, and that price may have been too high.

4. By agreeing to trade the transfer of the US journalist for the Soviet scientist, the United States has for the first time acknowledged the equivalence of the two cases.

5. No matter that the American Secretary of State, George Schultz, publicly rejected that interpretation.

Times / Letters / 15-9-86

1. In your editorial (September 9) on the breakthrough by the medical researchers at Sheffield Children's Hospital in discovering that an enzyme deficiency is present in some babies who die mysteriously of "cot deaths", you rightly point out that establishing neo-natal screening for all new-born babies would be extremely costly.

2. You conclude, however, that there is, nevertheless, a good case for proceeding with such tests.

3. We would question that conclusion and would argue that if funds were to become available for nationwide screening there are alternative, more cost-effective, uses for those resources in neo-natal and post-natal health care.

4. The maximum number of babies whose lives might be saved by identifying and treating MCAD (Medium Chain Acyl-coenzyme A dehydrogenase) deficiency is estimated at between 5 and 10 percent of the 1500 to 2000 "cot deaths" per year.

5. Let us say that 150 deaths might be prevented.

Telegraph / Article / 15-9-86

1. The Alliance would fail to gain public confidence if it appeared to be putting political interests before the proper defence of Britain, Mr John Cartwright, SDP Defence spokesman, warned the conference at Harrogate yesterday.

2. He was speaking in reply to a debate which ended with conference endorsing proposals in the Alliance joint commission document on defence despite reservations from some speakers.

3. The document will also be debated at the Liberal assembly in Eastbourne next week.

4. Mr Cartwright stressed that any agreed policy must carry both credibility and conviction with electors.

5. He said the two parties had already moved a long way towards achieving the sort of agreement needed on the "difficult issue" of a replacement for the Polaris nuclear weapon system.

Telegraph / Editorial / 15-9-86

1. The BBC serial about the mutiny at Etaples in 1917, "The Monocled Mutineer", has been disowned by the academic experts on the subject, by surviving eyewitnesses, and even by the BBC's own "historical adviser" to the scriptwriter and producer of the drama.

2. A total of \$3 million of licence-payers' money has been expended on what amounts to a distortion of a localised incident at a training camp into a "class war on the Western front", as the banner headline in the Radio Times proclaimed.

3. The shameless fabrication of events in a serial advertised as historical, not fictional, has been compounded by showing gratuitous violence, including brutal rape scenes, at nine o'clock on a Sunday evening.

4. Those involved in the making of this programme, including the managing director of BBC TV, Mr Bill Cotton, make no apology for any of this; Mr Cotton claims that it is "a play about the greater truth about the First World War".

5. It is not surprising that many members of the public, including a number of MPs, are angry about this state of affairs.

Telegraph / Letters / 15-9-86

1. Two aspects of the proposal for phasing out mortgage interest relief appear not to have received sufficient consideration.

2. To be equitable, the proposed housing benefit would have to be related to wealth and not simply to income.

3. It would be absurd if someone living on the sale of inherited assets were entitled to benefit when a hard-working manager was not.

4. Nor is it clear why benefit should be denied to someone like myself, buying my first house at the age of 35, at which time I had a reasonable income but no assets, having previously worked abroad as a missionary on an allowance-only basis.

5. Second, the gain to the Inland Revenue from ending mortgage interest relief would be nowhere near the five billion pounds which it is alleged to cost.

Guardian / Article / 15-9-86

1. The Social Democrats smoothed over their differences on Polaris yesterday to prepare the way for a comprehensive deal with the Liberals on Defence.
2. The SDP conference in Harrogate endorsed the Alliance commission report on defence which leaves open the question of replacing Polaris.
3. They said that it was essential to agree a detailed policy with the Liberals before the general election campaign.
4. In doing so, the SDP has moved some way from the position adopted by Dr David Owen, the party leader, in early summer when senior figures in both parties became alarmed at the likelihood of lasting, deep disagreement over the future of Britain's nuclear deterrent.
5. At the Liberal assembly in Eastbourne next week there will be strong opposition, probably from a substantial minority, against an agreement on a policy which does not rule out a replacement for Polaris but Mr David Steel, the party leader, is expected to win acceptance of the Alliance commission report.

Guardian / Editorial / 15-9-86

1. The delegates all held their yellow cards loyally in the air in the end.
2. But there was no mistaking the sense of unease which preceded the mostly unanimous votes.
3. The Social Democrats, debating defence policy at their Harrogate conference yesterday, know now that their relationship with the Liberals will be on the line when their Alliance partners debate the same issue at Eastbourne next week.
4. They know, too, that yesterday's debate was not just about the defence issue, important though that is.
5. It was implicitly about something much wider - the nature of the relationship with the Liberals - and, as some speakers dared to say, it was about Dr Owen's leadership of the party, too.

Guardian / Letters / 15-9-86

1. SDP defence spokesman John Cartwright makes much of the "credibility" of his idea of a Franco-British force de frappe.
2. In fact, it's a lot less credible than he would have us believe.
3. First, no British political party except for the SDP will swallow it: even the Liberals, with whom the SDP is supposed to be allied, are extremely unenthusiastic, and Labour and the Tories want nothing of a Franco-British bomb (though for different reasons).
4. Given that the SDP has no chance of forming a government on its own in the foreseeable future (and the Alliance has only a tiny chance), Cartwright's claims to be putting forward a politically realistic alternative are hogwash.
5. Second, a Franco-British nuclear force would face significant international opposition.

Mail / Article / 15-9-86

1. SDP members rallied round their leader yesterday with a neat sidestep over Alliance defence policy.
2. Dr David Owen is openly committed to wanting an eventual replacement for the Polaris nuclear system - something the Liberals have consistently rejected.
3. Yesterday SDP delegates at the annual conference in Harrogate made their support for Dr Owen overwhelmingly clear.
4. But they also left the door open for a pre-election deal with the Liberals by hinting that last year's conference decision on Polaris is not irrevocable.
5. The message came in the opening of the debate on the joint SDP-Liberal commission's report.

Mail / Editorial / 15-9-86

1. It is enough to make Britannia blush for very shame.
2. Contrary to the soothing public relations guff which heralds the annual Defence Review, the fighting power of our Royal Navy could become seriously reduced

in the coming years.

3. That is the sensational thrust of documents found in a secret file dumped on a Thames towpath near Reading.

4. If the file is genuine - as it would appear to be - then the security row over how it came to be missing should be no more than a squall compared to the political storm over what was in that file.

5. This newspaper has consistently argued that, under the Tories, Defence commitments threaten to outstrip cash estimates of what they will cost.

Mail / Letters / 15-9-86

1. Perhaps Mrs Thatcher's supposed reluctance to appoint women to her Cabinet is because she has learned from experience that their 'femaleness' will always distract attention from their effectiveness as politicians.

2. No matter how professional, clever or competent a woman is in any field, there are still those unable to see her in terms of ability, unlinked to physical appearance.

3. Your story mentions four women who are deeply committed to politics: 'Dark, vivacious Edwina Currie' and 'statuesque blonde' Marion Roe; not forgetting, of course, those two 'tall, cool blondes with devastating smiles' Virginia Bottomley and Ann Winterton.

4. I'm afraid it's going to be a long, hard climb, ladies.

5. I was bewildered to read the news story on the woman who stabbed her young son because she thought he was the Devil.

Today / Article / 15-9-86

1. David Owen saw off his SDP critics over defence yesterday.

2. His party virtually wrote him a blank cheque to complete an Alliance policy deal with David Steel.

3. The two men are planning more defence co-operation with France, especially over a successor to Britain's ageing Polaris fleet.

4. Dr Owen also wants to build more nuclear missile

submarines to provide a European nuclear umbrella beyond the year 2000.

5. But that is anathema to many Liberals and the delicate compromise the two leaders have stitched together is that no decision over replacing Polaris is needed yet.

Today / Editorial / 15-9-86

1. Nancy Reagan thinks that drug-pushers who are responsible for the deaths of any of their customers should themselves face the death penalty.
2. Whatever else one may feel about Mrs Reagan, her commitment to the anti-drugs campaign is beyond question.
3. What she says must be treated seriously.
4. But she is wrong.
5. In the American context, she has a superficially attractive argument.

Today / Letters / 15-9-86

1. Mrs Thatcher was shown in a cheerful mood when she opened the Nissan plant.
2. It is so nice to know that the Japanese giant is going to settle happily in this country, flood the place with their products and hit our own car production.
3. Not only that, but they are going to receive \$100 million in grants.
4. And the Prime Minister finds all of this thrilling.
5. There is just no sense in giving such a vast amount of British money to a foreign concern.

Mirror / Article / 15-9-86

1. The Social Democrats patched over their differences on defence yesterday in a bid for unity with their Liberal partners.
2. It was a triumph for SDP leader David Owen who warned his party to get ready for an early general election.
3. After the debate at the SDP's conference in Harrogate, Dr Owen said: "I am pretty pleased. The so-called challenge to my position has disappeared out of the window."
4. Dr Owen has upset some Liberals by insisting that in a European-based nuclear strategy there must be a

replacement for Polaris.

5. But the policy agreed yesterday skirted round the Polaris issue and endorsed a joint SDP-Liberal approach to other European leaders for closer collaboration on defence.

Mirror / Editorial / 15-9-86

1. The Tory Party is stepping up its campaign against the BBC.
2. Its only purpose can be to tame the broadcasters before a general election gets under way.
3. The latest victim of its vendetta is *The Monocled Mutineer*, a brilliantly acted, beautifully written but violent drama about a British army mutiny in 1917.
4. The serial is only slightly related to the truth.
5. Its author, Alan Bleasdale, has never pretended his work was other than fiction, even if BBC publicists foolishly did so.

Mirror / Letters / 15-9-86

1. Most people believe that "to chance your arm" was a tailor's expression (1889), taken up by soldiers (1899) to mean to risk a court martial, and have their stripes and badges "pretty pretties" taken off their sleeve.
2. That is not correct.
3. In 1492, two prominent families, the Ormonds and Kildares, were in the midst of a bitter feud.
4. Besieged by Gerald Fitzgerald, Earl of Kildare, Sir James Butler, Earl of Ormond, and his followers took refuge in the chapter house of St Patrick's Cathedral, Dublin, bolting themselves in.
5. As the siege wore on, the Earl of Kildare concluded that the feuding was foolish.

Express / Article / 15-9-86

1. The SDP made it clear yesterday they may be prepared to leave Britain without her own nuclear deterrent.
2. The position was revealed by MP Charles Kennedy, speaking on behalf of the party's powerful policy

committee.

3. He told their conference in Harrogate:

"Our commitment to replace Polaris should not be construed as replacement come what may. It is a judgement on the likely balance of future probabilities."

4. His key words came in a defence debate designed to smooth over serious differences between the SDP and their Liberal partners in the Alliance.

5. And they are certain to be seized upon by Tory chairman Norman Tebbit, who has already made much of the deep split on defence which threatens to shipwreck the whole Alliance campaign at the next election.

Express / Editorial / 15-9-86

1. It's the annual party conference season.

2. From Brighton to Bournemouth, Harrogate to Eastbourne, the great seduction of the voter is under way.

3. Politicians will be turning triple reverse somersaults to win over the electorate; it is time for the public to be on guard.

4. Mrs Thatcher is convinced that when the real issues are put before the public, she will win an historic election hat-trick.

5. She is right.

Express / Letters / 15-9-86

1. I awoke at five in the morning to find our bungalow flooded throughout to a depth of two inches.

2. The Fire Service told me their pumps would not work on such a small quantity.

3. The Southern Water Authority Emergency Service at Brighton asked 'Are you a Council property?'

4. Upon replying 'No, this is a private address,' I was told 'Sorry, we can't help you get a plumber in.'

5. This we managed to do and he arrived at between 5.30 and 6a.m.

Sun / Editorial / 15-9-86

1. This year a new exam, the General Certificate of Secondary Education, is being introduced to replace the old O-levels.
2. It is designed for pupils aged 16.
3. The exam has itself been tested at a Liverpool school, where pupils did specimen arithmetic papers.
4. The children found the questions absurdly easy.
5. All but three of 28 passed.

Sun / Letters / 15-9-86

1. My grandson had been learning about sex and babies at school.
2. He and his family were at a wedding and it was getting very late.
3. Seeing the bride and groom go up to bed at the hotel, my grandson said: 'Would you like me to go with you and show you how to make a baby?'
4. My four-year-old son Kevin had been playing with Douglas, who is a twin.
5. Next morning there came a knock at the door.

## Appendix 2

## Results of numerical enquiries (main database)

Feature	Tot	Art	Ed	Let	Ti	Te	G	To	Ma	Mi	Ex	S
BE	455	74	293	178	61	62	67	58	62	54	53	38
Fin	1763	566	604	593	213	216	212	227	221	232	223	222
Pos	2250	772	741	737	287	281	276	276	284	279	284	283
Neg	86	17	36	33	8	11	14	13	11	11	9	9
PQ	16	0	7	9	1	0	1	3	1	6	1	3
SemN	49	12	16	21	5	8	9	8	4	4	6	5
Vst 1	1694	587	533	574	199	196	223	198	226	216	209	227
2	217	40	103	74	34	34	35	28	25	20	22	19
3	127	39	57	31	17	12	14	13	20	22	17	12
4	45	13	16	16	2	4	3	5	5	9	5	12
5	158	82	30	46	19	24	9	30	11	21	27	17
6	32	4	18	10	6	3	5	6	4	0	7	1
7	4	2	1	1	0	1	1	1	0	0	1	0
8	50	16	10	24	12	18	3	6	3	3	2	3
9	3	1	0	2	0	0	0	0	1	0	0	2
10	30	10	15	5	8	5	3	3	1	2	6	2
11	4	0	4	0	0	0	0	0	2	1	0	1
12	0	0	0	0	0	0	0	0	0	0	0	0
13	4	0	2	2	1	0	0	2	0	0	1	0
14	0	0	0	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0	0	0
16	0	0	0	0	0	0	0	0	0	0	0	0
17	33	7	11	15	3	3	5	8	2	6	2	4
VFing	224	96	51	77	45	25	29	18	35	26	26	20
i	344	114	122	108	38	43	51	44	40	32	44	52
en	69	24	23	22	5	16	9	11	4	10	7	7
Siseq	373	128	130	115	42	53	49	46	42	43	33	65
Clt M	1081	351	378	352	115	114	118	145	129	161	145	154
A	348	125	93	130	58	34	37	43	54	42	43	37
N	289	106	89	94	42	45	42	29	39	36	33	23
Postm	462	146	161	155	60	76	73	55	52	39	57	50
Vcomp	195	68	68	59	19	27	26	24	22	22	20	35
CLF S	26	1	16	9	4	3	2	4	3	5	4	1
O	221	101	48	72	32	37	32	22	31	29	20	18
C	34	3	20	11	5	5	5	3	4	2	7	3
A	348	125	93	130	58	34	37	43	54	42	43	37
Nm	388	129	130	129	49	65	55	49	42	34	51	43
Am	79	19	31	29	14	11	18	9	10	5	7	5
?	222	72	85	65	24	31	33	24	27	22	23	38
Sty an	8	1	1	6	0	2	0	0	1	1	0	4
in	142	71	30	41	7	7	2	19	25	39	27	16
hu	1101	420	293	389	103	109	119	133	145	150	142	201
ab	950	265	403	283	158	159	151	115	98	95	111	63
m	415	241	97	77	35	53	46	55	57	58	40	71
f	106	28	28	50	2	1	3	11	28	14	11	36
1	168	0	20	148	18	9	8	13	23	28	30	39
2	17	0	2	15	3	0	2	6	0	4	2	0
3	2016	756	704	556	246	268	262	248	246	252	248	246

s	1330	505	424	400	109	185	163	165	180	179	157	191
p	635	217	190	228	91	69	70	71	64	88	96	86
u	137	17	55	64	32	13	16	26	23	11	14	1
c	99	19	55	25	35	10	22	4	2	6	13	7
Sfull	705	254	235	216	96	108	70	93	78	93	89	78
pn	437	96	140	201	44	31	46	59	57	68	56	76
PN	204	90	86	28	24	24	37	26	17	20	31	25
dem	37	4	16	17	6	5	4	6	3	7	2	4
imp	90	14	34	42	9	12	20	12	14	5	13	5
rel	114	44	39	31	16	20	18	12	18	6	12	12
Q	4	0	3	1	0	1	0	0	0	3	0	0
WH	9	0	8	1	2	1	1	1	0	2	2	0
NCl	19	1	13	5	2	3	0	2	3	5	3	1

Appendix 5Sequential verbs in main data and supplement

	data	supp.		data	supp.
accuse of		2	force		5
advise		2	forgive	1	
agree	2	1	get	1	1
aim		2	go	2	2
allege	4	2	going to	5	2
allow	7	22	go on	1	1
appear	7	4	go on to	2	4
arrange for		1	happen		1
ask	2	7	have to	16	43
attempt		2	have (+en)		5
avoid		1	help	4	13
awake	1		hesitate		1
be	11	25	hope	3	2
be able		3	imagine		1
be upto	2		incline		1
begin	3	5	induce		1
believe	2	1	insist	2	1
bother		1	inspire		1
bring oneself		2	intend	1	4
brainwash	1		invite		3
calculate		1	jolt into		1
call for/upon		2	keep	1	2
care		1	know	1	1
cause	1	1	learn		1
cease		1	leave	1	1
choose	2	8	let	1	5
claim	1	3	like	3	8
come		3	love		1
come to	1	2	lure		1
consider		2	make	2	5
conspire	1	1	manage	2	6
continue	2	14	mean		1
contrive	1		need	4	13
criticise for		1	oblige		1
dare	1		opt		1
decide	1	4	order	2	3
declare		1	permit		1
decline		2	persuade		2
defy		1	plan	2	2
deny		1	plot	1	1
deserve		2	prefer	1	2
design	2	3	prepare		1
discourage		1	prevent	1	3
do well	1	1	proceed		1
dream of		1	promise		2
enable	1		propose		3
encourage		7	prove	1	1

entail		1	provoke		1
expect	3	13	purport	1	
fail	4	10	recall		1
feel	1		recognise		1
fight		1	refuse	4	1
find	3	1	relegate		1
report	1	1	take	1	
result in		1	teach		1
risk		1	tell		3
said	2	2	tend	1	4
see	3	2	think	1	1
seek	1	4	threaten	4	1
seem	2	18	train	1	
send	1	2	trust		2
serve		2	try	9	24
show	2	2	turn out		2
start	1		urge	1	2
stand		5	volunteer		1
stop	2	2	wait	1	1
succeed	1	2	want	8	17
suppose		4	warn	2	2
suspect		1	wish	4	13