

Pattern Grammar

A corpus-driven approach
to the lexical grammar of English

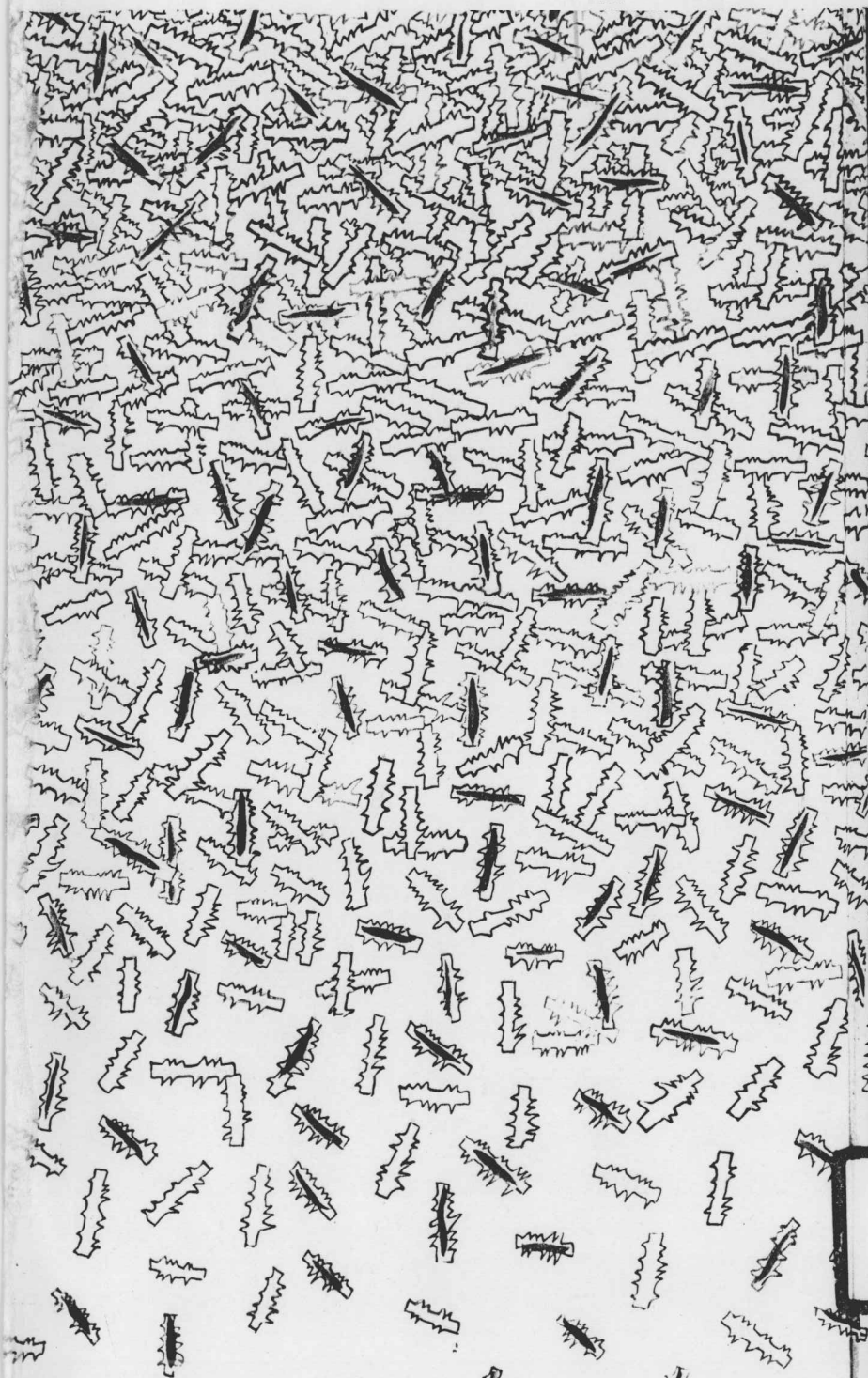
SUSAN HUNSTON

GILL FRANCIS

The University of Birmingham

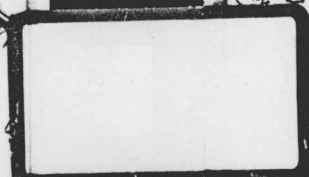
JOHN BENJAMINS PUBLISHING COMPANY

AMSTERDAM / PHILADELPHIA



SCL 4

Hunston/Francis Pattern Grammar



Susan Hunston

and Gill Francis
*Pattern
Grammar*

A corpus-driven approach
to the lexical grammar
of English

江苏工业学院图书馆
藏书章

Studies in Corpus Linguistics

SCL

Studies in Corpus Linguistics

Studies in Corpus Linguistics aims to provide insights into the way a corpus can be used, the type of findings that can be obtained, the possible applications of these findings as well as the theoretical changes that corpus work can bring into linguistics and language engineering. The main concern of SCL is to present findings based on, or related to, the cumulative effect of naturally occurring language and on the interpretation of frequency and distributional data.

General Editor

Elena Tognini-Bonelli

Consulting Editor

Wolfgang Teubert

Advisory Board

Michael Barlow (*Rice University, Houston*)

Robert de Beaugrande (*UAE*)

Douglas Biber (*North Arizona University*)

Wallace Chafe (*University of California*)

Stig Johansson (*Oslo University*)

M.A.K. Halliday (*University of Sydney*)

Graeme Kennedy (*Victoria University of Wellington*)

John Laffling (*Herriot Watt University, Edinburgh*)

Geoffrey Leech (*University of Lancaster*)

John Sinclair (*University of Birmingham*)

Piet van Sterkenburg (*Institute for Dutch Lexicology, Leiden*)

Michael Stubbs (*University of Trier*)

Jan Svartvik (*University of Lund*)

H-Z. Yang (*Jiao Tong University, Shanghai*)

Antonio Zampolli (*University of Pisa*)

Volume 4

Susan Hunston and Gill Francis

Pattern Grammar

A corpus-driven approach to the lexical grammar of English

Pattern Grammar

A corpus-driven approach
to the lexical grammar of English

SUSAN HUNSTON

GILL FRANCIS

The University of Birmingham

JOHN BENJAMINS PUBLISHING COMPANY

AMSTERDAM / PHILADELPHIA



TM

The paper used in this publication meets the minimum requirements of American National Standard for Information Sciences Permanence of Paper for Printed Library Materials, ANSI Z39.48-1984.

Cover design: Françoise Berserik
Cover illustration from original painting *Random Order*
by Lorenzo Pezzatini, Florence, 1996.

For Sam and Vivek

and for Elliott, Laura, Matthew and Toby

Library of Congress Cataloging-in-Publication Data

Hunston, Susan, 1953-

Pattern grammar : a corpus-driven approach to the lexical grammar of English / Susan Hunston, Gill Francis.

p. cm. -- (Studies in corpus linguistics, ISSN 1388-0373 ; v. 4)

Includes bibliographical references and index.

1. English language--Grammar, Generative. 2. English language--Discourse analysis. I. Francis, Gill. II. Title. III. Series.

PE1106.H453 1999

425--dc21

99-043483

ISBN 90 272 2273 8 (Eur.) / 1 55619 398 X (US) (Hb; alk. paper)

ISBN 90 272 2274 6 (Eur.) / 1 55619 399 8 (US) (Pb; alk. paper)

© 2000 John Benjamins B.V.

No part of this book may be reproduced in any form, by print, photoprint, microfilm, or any other means, without written permission from the publisher.

John Benjamins Publishing Co. · P.O.Box 75577 · 1070 AN AMSTERDAM · The Netherlands
John Benjamins North America · P.O.Box 27519 · Philadelphia PA 19118-0519 · USA

Table of Contents

Acknowledgements	xiii
CHAPTER 1	
A Short History of Patterns	1
1.1 About this book	1
1.2 Hornby: A Guide to Patterns and Usage in English	3
1.2.1 <i>Introduction</i>	3
1.2.2 <i>Verb patterns</i>	4
1.2.3 <i>Noun and adjective patterns</i>	6
1.2.4 <i>Conclusion</i>	7
1.3 Lexical phrases	7
1.3.1 <i>The lexicographical perspective</i>	8
1.3.2 <i>Language teaching</i>	10
1.3.3 <i>Psycholinguistics</i>	11
1.3.4 <i>Lexical phrases and a pattern grammar</i>	13
1.4 Sinclair: Corpus Concordance Collocation	14
1.4.1 <i>Corpus-driven language description</i>	14
1.4.2 <i>Sense and structure</i>	20
1.4.3 <i>The idiom principle</i>	21
1.4.4 <i>Units of meaning</i>	24
1.4.5 <i>Frames</i>	25
1.4.6 <i>Towards a lexico-grammar</i>	28
1.5 Francis: A corpus-driven grammar	29
1.5.1 <i>Meaning and pattern revisited</i>	29
1.5.2 <i>Lexis and grammar revisited</i>	30
1.5.3 <i>Towards a methodology</i>	31
1.6 COBUILD: the grammar patterns series	32
1.6.1 <i>Introduction</i>	32
1.6.2 <i>Grammar in the Collins COBUILD English Dictionary</i>	32
1.6.3 <i>The grammar patterns series</i>	35
1.7 Conclusion to Chapter 1	36

CHAPTER 2	
What a pattern is	37
2.1 A word and its patterns	37
2.2 A pattern and its words	43
2.3 The representation of patterns	44
2.4 What's in a pattern?	49
2.5 What kinds of pattern are there?	51
2.5.1 <i>The patterns of verbs</i>	51
2.5.2 <i>The patterns of nouns</i>	56
2.5.3 <i>The patterns of adjectives</i>	58
2.6 Looking at patterns from all sides	58
2.7 Different forms of a pattern	59
CHAPTER 3	
Problems in identifying patterns	67
3.1 Which word does the pattern belong to?	68
3.2 When is a pattern not a pattern?	71
3.3 Do patterns over-generalise?	77
3.3.1 <i>Introduction</i>	77
3.3.2 <i>'It's been a privilege to know you'</i>	77
3.3.3 <i>'Adamant in her refusal'</i>	78
3.3.4 <i>'Her success as a designer'</i>	79
3.3.5 <i>'It's illogical to believe'</i>	79
3.3.6 <i>'She qualified as a doctor'</i>	80
CHAPTER 4	
Patterns and Meaning	83
4.1 Meaning groups: some examples	83
4.2 Creativity in pattern use	95
4.2.1 <i>Introduction</i>	95
4.2.2 <i>Pattern and analogy</i>	96
4.2.3 <i>Patterns with less restricted lexis: N that and it v-link ADJ of n to-inf</i>	98
4.2.4 <i>Patterns with a meaning: V way prep/adv</i>	100
4.2.5 <i>Are there restrictions in a pattern?: V n into -ing</i>	102
4.2.6 <i>Exploiting the prosody of patterns: V n into -ing and V n as n</i>	104
4.3 Summary of Chapter 4	107

CHAPTER 5	
More on pattern and meaning	109
5.1 From meaning to pattern: lexis and function	109
5.1.1 <i>Using a 'meaning finder'</i>	109
5.1.2 <i>Further examples of notional groups</i>	111
5.1.3 <i>Notional groups in teaching languages</i>	118
5.2 Mapping meaning on to pattern	123
5.2.1 <i>Patterns and roles</i>	123
5.2.2 <i>The patterns of difficult</i>	130
5.2.3 <i>The patterns of difficulty</i>	136
5.2.4 <i>Summary of Section 5.2</i>	141
5.3 Semantic word groups and their pattern distribution	142
5.4 Conclusion to Chapter 5	146
Appendix to Chapter 5	146
CHAPTER 6	
Pattern and structure	151
6.1 Relating pattern to structure	151
6.1.1 <i>Introduction</i>	151
6.1.2 <i>Objects, Complements, Adjuncts and Object Complements</i>	152
6.1.3 <i>Patterns with finite clauses</i>	155
6.1.4 <i>'Extraposed' subjects</i>	156
6.1.5 <i>Structure and Meaning</i>	157
6.2 Problems with prepositional phrases	160
6.2.1 <i>Example 1: V for n/-ing</i>	161
6.2.2 <i>Example 2: V from n</i>	162
6.2.3 <i>Example 3: V n as n</i>	167
6.2.4 <i>Example 4: V n to n</i>	168
6.3 Problems with phase	169
6.3.1 <i>V to-inf</i>	170
6.3.2 <i>Verbs followed by prepositions</i>	175
6.4 Conclusion to Chapter 6	176
CHAPTER 7	
Word class and pattern	179
7.1 The word class as pattern set	179
7.2 Some 'new' word classes	181
7.2.1 <i>Classes of nouns</i>	181
7.2.2 <i>Ergative verbs</i>	183
7.2.3 <i>Shell nouns</i>	185

7.2.4	<i>Evaluative adjectives</i>	188
7.2.5	<i>Reciprocal adjectives and nouns</i>	191
7.3	Some problematic word classes	192
7.4	Words without classes	195
7.5	Are word classes necessary?	197

CHAPTER 8

Text and Pattern	199
8.1 Patterns in running text	199
8.2 Pattern flow	207
8.2.1 <i>Patterns as linear grammar</i>	207
8.2.2 <i>Representing patterns linearly</i>	209
8.2.3 <i>Pattern flow</i>	211
8.3 Pattern configurations	215
8.4 Collocation and clause collocation	225
8.4.1 <i>Introduction</i>	225
8.4.2 <i>Subordinating conjunctions</i>	225
8.4.3 <i>Clause collocation</i>	226
8.4.4 <i>Collocation</i>	230
8.5 The theory of a linear grammar	235
8.5.1 <i>The problem of embedded clauses</i>	235
8.5.2 <i>The problem of 'there'</i>	237
8.5.3 <i>Linearity and prospection: Brazil's 'A Grammar of Speech'</i>	240
8.5.4 <i>Halliday and the metaphor of metalanguage</i>	244

CHAPTER 9

Summing Up	247
9.1 Patterns and phraseologies	247
9.2 Questions of theory	249
9.2.1 <i>Is theory necessary?</i>	249
9.2.2 <i>Lexis and grammar</i>	250
9.2.3 <i>Pattern and meaning</i>	255
9.2.4 <i>What kind of theory?</i>	259
9.3 Questions of application to pedagogy	261
9.3.1 <i>The 'COBUILD' approach</i>	261
9.3.2 <i>Towards a pedagogic reference grammar</i>	262
9.3.3 <i>'Noticing' and using patterns</i>	265
9.3.4 <i>Do patterns matter?</i>	268
9.4 Conclusion	271

References	275
Appendix	283
Name Index	285
Subject Index	289

Acknowledgements

Much of the research for this book was undertaken as part of the project to publish the two 'pattern grammars': Collins COBUILD Grammar Patterns 1: Verbs; and Collins COBUILD Grammar Patterns 2: Nouns and Adjectives. The authors wish to acknowledge the outstanding contribution of Elizabeth Manning, co-author of those books, and of Professor John Sinclair, their instigator. Corpus examples in this book are taken from the Bank of English. The authors would like to thank HarperCollins publishers and The University of Birmingham for permission to use and quote from this corpus.

CHAPTER 1

A Short History of Patterns

1.1 About this book

This book is about the patterns that are associated with particular lexical items and that are observable from investigation of an electronically-stored corpus of written and spoken texts. The concept can be illustrated by this short extract from an article written by a lecturer in Philosophy on the subject of how Philosophy should be taught:

- (1) *[1] Philosophy is different from many other disciplines [2] in that learning about it is as much a matter of developing skills (in reasoning and argument) as it is a matter of learning a body of information. [3] In this sense there are no definitive 'answers' to many philosophical problems: [4] becoming a philosopher is a matter of becoming able to reason coherently and relevantly about philosophical issues. [5] Consequently, valuable contact time with lecturers is best spent actually 'doing philosophy', [6] and that means actively thinking and talking about it.*

There are several aspects of the grammar of this extract that a linguist may wish to draw attention to. For example, clauses 3, 5 and 6 each begin with something that summarises or 'encapsulates' (Sinclair 1995) the preceding clause or clauses: *In this sense; Consequently; that*, and that the grammar of each clause is therefore heavily dependent on the organisation of meaning in the text as a whole. Alternatively, we might make the observation that many of the processes indicated by this extract — learning about philosophy, becoming a philosopher, students interacting with lecturers — are expressed in the text as nouns (the Subjects of clauses) rather than as verbs, and that the relation of the grammar to the world is therefore metaphoric rather than congruent (Halliday 1994).

The aspect of the extract that this book focuses on, however, is the interaction between the particular lexical items in it and the grammatical patterns that they form a part of. A clear example is the noun *matter*, which appears three times in the extract. Each time it follows the indefinite article *a* and is followed

by *of* and a clause beginning with an 'ing' form: *a matter of developing skills; a matter of learning a body of information; a matter of becoming able to reason coherently and relevantly about philosophical issues*. It is clear that there is little point in treating *matter*, in the sense in which it is used here, as a single lexical item that can be slotted into a general grammar of English. Rather, the word comes, as it were, with its attendant phraseology, which in this case consists of '*a ___ of -ing*'. This phraseology is the grammar pattern belonging to the word *matter*.

The noun *matter* is far from unusual in having particular patterns or phraseologies associated with it. All words, in fact, can be described in this way. Let us take the verbs in the extract above:

- | | | |
|-----|----------|--|
| [1] | is | different from many other disciplines |
| [2] | is | ... a matter of... |
| [3] | are | no definitive answers... |
| [4] | is | a matter of... |
| [5] | is spent | actually doing philosophy |
| [6] | means | actively thinking and talking about it |

From this we may note that the verb *to be* is followed by a noun group (*a matter of, no definitive answers*) or by an adjective group (*different from...*); the passive verb *be spent* is followed by an '-ing' clause; as is the verb *to mean*. Investigation of a corpus will tell us whether these uses are typical or not. For example, here are ten concordance lines illustrating a typical use of *be spent*:

of this man whose early career was spent teaching at Harvard Business
 evening many valuable minutes were spent recounting the non-story of the
 in hard to find and empty days are spent wandering the streets or riding
 much of her time at the college is spent sitting through paperwork, Carolin
 job properly. 'Much of my time is spent making copious notes on what actio
 iles of border. Most of the day is spent riding along the riverbank. This
 out, son - a third of your life is spent sleeping, a third in daily
 'early ancestors' waking lives was spent chasing or being chased by various
 Much of the next 12 months will be spent celebrating or decrying the Spanis
 nd the rest of Bradford's life was spent restoring it. He planted well over

In each case, the verb is followed by an '-ing' clause, and is preceded by a noun group indicating a period of time.

A grammatical description of the verbs in this extract, then, needs to take into account their complementation patterns, that is, the kind of group or clause that may follow them, just as a description of *matter* needs to take into account its phraseology. The same could be said for the adjectives in the extract (*different from, able to*) — in fact, all words can be described in terms of their patterns. Patterns can be observed, intuitively, in a single text. This intuition is based on our previous experience of language: we know that we say *a matter of learning a body of information* rather than *the matter of learning a body of information* and

that, therefore, the article *a* is important to the pattern associated with *matter*. Intuition is not always a reliable guide, however, and it is advantageous to have a corpus to reveal what is typical patterning and what is unusual. The corpus is a concrete replacement for the rather vague 'previous experience of language'.

Briefly, then, a pattern is a phraseology frequently associated with (a sense of) a word, particularly in terms of the prepositions, groups, and clauses that follow the word. Patterns and lexis are mutually dependent, in that each pattern occurs with a restricted set of lexical items, and each lexical item occurs with a restricted set of patterns. In addition, patterns are closely associated with meaning, firstly because in many cases different senses of words are distinguished by their typical occurrence in different patterns; and secondly because words which share a given pattern tend also to share an aspect of meaning. The purpose of this book is to describe patterns and their association with meaning in more detail, and to discuss some of the theoretical issues arising out of this approach to grammar.

Chapter 1 sets the work in context, starting with the work by Hornby on patterns and usage and the growing interest in 'fixed phrases' by both lexicographers and language teachers. The immediate inspirations for corpus-driven grammar — Sinclair (1991) and Francis (1993) — are then discussed. In Chapters 2 and 3 we discuss the concept of 'pattern' in detail, and in Chapters 4 and 5 we give several examples of the association between pattern and meaning.

The second part of the book discusses various issues that arise in relation to a pattern-based approach to grammar. Chapter 6 takes a theoretical perspective and considers the relationship between patterns and the traditional structural analysis of clauses. Another comparison between patterns and traditional grammar is found in Chapter 7, which considers the notion of 'word class' in the light of our work on patterns. Chapter 8 takes the work in yet another direction by applying patterns to the analysis of running text. In Chapter 9 we consider some of the implications of this work for theories of grammar and for language teaching.

1.2 Hornby: A Guide to Patterns and Usage in English

1.2.1 Introduction

The closest forerunner (in concept, though not in time) of the work to be described in this book is Hornby's *A Guide to Patterns and Usage in English*, published in 1954. In the preface to this book, Hornby sets out his agenda: to give practical guidance to the language learner on usage rather than to expound analysis:

Analysis is helpful, but the learner is, or should be, more concerned with sentence-building. For this he needs to know the patterns of English sentences and to be told which words enter into which patterns. (Hornby 1954: v)

He points out that analogy is not an infallible guide to sentence patterns, especially with regard to verbs:

[The learner] may suppose that because he has heard and seen 'I intend (want, propose) to come', he may say or write 'I suggest to come', that because he has heard or seen 'Please tell me the meaning', 'Please show me the way', he can say or write 'Please explain me this sentence'. (Hornby 1954: v)

Hornby's book is not restricted to verbs, however, and he advocates learning pattern along with meaning for nouns and adjectives as well: "When [the learner] learns the meanings of the adjective *anxious*, he should also learn its patterns: 'anxious about his son's health', 'anxious for news', 'anxious (=eager) to start'." (Hornby 1954: vi).

Although he appears to make a sharp distinction between pattern and meaning — "A knowledge of how to put words together is as important as, perhaps more important than, a knowledge of their meanings" (Hornby 1954: v) — the description of *anxious* quoted above implicitly links the two.

The bulk of *A Guide to Patterns and Usage in English* is devoted to the description of 25 verb patterns, 4 noun patterns and 3 adjective patterns, as well as sections on adverbs, 'time and tense', indefinite pronouns and determinatives. A long final section details the way that various concepts can be expressed. The approach to grammar detailed in the book also informed the first three editions of the *Oxford Advanced Learner's Dictionary* (1948, 1963, 1974).

1.2.2 Verb patterns

Hornby's radical attention to usage rather than to analysis — to encoding rather than to decoding — leads him to describe pattern rather than structure. He does distinguish between transitive and intransitive verbs, and comments on which part(s) of the pattern constitute(s) the Object, but makes no further attempt at analysis. The headings to his tables, therefore, indicate the pattern (e.g. *Subject + Verb + to-infinitive*) rather than the elements of structure. In some cases, however, he distinguishes between examples of a single pattern that have different structures. For example, he comments with respect to *He likes his coffee strong* (verb followed by noun and adjective) that "[t]he combination of (pr)noun and adjective is the object of the verb" (Hornby 1954: 33). This is contrasted with *Can you push the door open*, which has the comment "...the

adjective denotes a state or condition that results from the action indicated by the verb". Although an analysis of this is not given, the implication is that the noun and the adjective comprise separate elements of the clause rather than a single element. Hornby's sparse use of functional categories occasionally leads to some curious analyses. In examples of a verb followed by two noun groups, as in *They made Newton President of the Royal Society*, the second noun group is termed a 'predicative adjunct' rather than an Object Complement. However, Hornby's stated aim is to describe usage, not to enter into the complexities of analysis: our discussion in Chapter 6 indicates our reasons for sympathising with this aim.

Hornby's descriptions of patterns are not entirely based on surface realisation, however. Possible transformations also play a role. For example, we have seen above how he deals with the pattern 'verb+noun+adjective'. In that pattern, however, he considers only those examples where the alternative 'verb+noun+to-be+adjective' (as in *Most people suppose him (to be) innocent*) is not possible. Another pattern (Verb Pattern 4, Hornby 1954: 22–25) deals with examples such as *They proved him wrong* (verb+noun+adjective) and *I have always found Smith to be friendly* (verb+noun+to-be+adjective), as well as *They knew the man to have been a spy* (verb+noun+to-be+noun). The same pattern accounts for examples such as *We all consider it wrong to cheat in examinations*, which could be considered a transformation of *We all consider [the act of] cheating in examinations to be wrong*. Arguably, Hornby groups together these particular patterns because he identifies a common meaning in the verbs with these patterns (verbs such as *believe, consider, declare, feel, find, guess, know, judge, prove, suppose* and *think*) and wishes to deal with them together.

As noted above, Hornby advocates that learners be told "which words enter into which patterns", and he attempts to do that by giving lists of the most common verbs used in each pattern. For example, for Verb Pattern 3 (verb followed by a noun and a to-infinitive, as in *We can't allow them to do that* and *They warned me not to be late*) he provides the following list:

advise, allow, ask, (can't) bear, beg, cause, challenge, choose, command, compel, dare (=challenge), decide, determine, encourage, entreat, expect, force, get, give (someone to understand...), hate, help, implore, instruct, intend, invite, lead (=cause), leave, like, love, mean (=intend), oblige, order, permit, persuade, prefer, prepare, press (=urge), promise, remind, request, teach, tell, tempt, trouble, urge, want, warn, wish (Hornby 1954: 21).

Using a corpus of course allows such a list to be made much more comprehensive (Francis et al. 1996 lists no fewer than 219 verbs and phrasal verbs with this pattern). However, it is more pertinent to note Hornby's concern with meaning

and pattern in this list, shown by his indication that some of the verbs in it occur with this pattern only when they have a particular meaning. In some cases, this concern is made more central. Verb Pattern 5, for instance (verb followed by noun and bare infinitive) is divided into verbs of perception (*feel, hear, listen to, look at, notice, observe, perceive, see, watch*) and others indicating permission, obligation and awareness: (*bid, help, know, let, make*).

The main omission in Hornby's work is that he does not deal with the large numbers of verbs that are associated with particular prepositions (although the item-by-item coding in the *Oxford Advanced Learner's Dictionary* does show the prepositions used with each verb). Instead of dealing with the prepositions separately, he puts them all together in Verb Pattern 18 (*We congratulated him on his success; They accused her of taking the money; What prevented you from coming earlier; I must remind him about it; Compare the copy with the original*) and Verb Pattern 24 (*You can rely upon that man; He succeeded in solving the problem; They all longed for the holidays; He consented to the proposal; She complained of the heat.*) In some cases, he gives a prepositional phrase as a variant of another pattern. For example, in describing the pattern 'verb followed by two noun groups', as in *They elected Mr Grey chairman*, he notes that *choose* and *elect* often have *as* or *for* before the second noun group, as in *They elected Mr Grey as chairman*.

1.2.3 Noun and adjective patterns

Hornby identifies only a handful of noun and adjective patterns, as follows:

Noun patterns

- noun + to-infinitive e.g. *Anne's desire to please her mother-in-law*
- noun + preposition + noun e.g. *a specialist in chest diseases; our anxiety for news*
- noun + that-clause e.g. *a hope that you would soon be well again*
- noun (+ preposition) + conjunctive + phrase or clause e.g. *the knowledge of how it should be done*

Adjective patterns

- adjective + to-infinitive e.g. *You were unwise to accept his offer*
- adjective + preposition + noun e.g. *Are you afraid of the dog?*
- adjective (+preposition) + clause or phrase e.g. *She was not aware that her husband earned £10 a week.*

Each of these patterns covers quite a range of examples. For instance, 'adjective + to-infinitive' includes patterns with introductory *it* and a prepositional phrase, such as *It's kind of you to say so*.

Hornby explicitly relates these patterns to the similar or contrasting verb patterns. For example, he points out that *another attempt to climb the mountain* can be seen as analogous to *They will attempt to climb the mountain*, but that whereas the noun *discussion* is followed by a preposition such as *on*, its cognate verb *discuss* is followed by a noun without an intervening preposition. Thus, learners can be encouraged to link together patterns that show similarities, but need to be warned against making false analogies.

1.2.4 Conclusion

It would be difficult to overestimate Hornby's achievement in *A Guide to Patterns and Usage in English*. The amount of detailed observation in the book is impressive, and the priority given to pattern over structure represents a radical reinterpretation of grammar from the point of view of the learner rather than the academic. It is perhaps an indication of the unusual quality of Hornby's work that it could be superseded only when technology gave us electronic corpora that allow the details missing from Hornby's classifications to be fleshed out.

1.3 Lexical phrases

Perhaps one reason for the comparative neglect of Hornby's work in language description is that he deliberately blurred the distinction between lexis and grammar, whereas theories since the fifties have tended to prioritise either one or the other. Both structuralism and Chomsky's work (largely) treated grammar as a system independent of lexis. On the other side of the coin, since the 1970s there has been an increasing emphasis, in both theoretical and applied linguistics, on lexis rather than grammar as the central principle of language. In this section we look at work in the area of lexis which has added to the perception of the phraseological nature of language.

It is by now a truism that a large amount of language encountered is not constructed from 'basic' structures and a lexicon, but occurs in sequences of morphemes that are more or less fixed in form. These sequences are called, variously, "lexical phrases" (Nattinger and DeCarrico 1989, 1992), "composites" (Cowie, following Mitchell in Cowie 1988), "gambits" (Keller in Cowie 1988), "routine formulae" (Coulmas in Cowie 1988), "phrasemes" (Melčuk 1988, 1995), "prefabricated routines and patterns" (Krashen 1981), "sentence stems" (Pawley and Syder 1983), "formulae" (Peters 1983), and "formulaic language" (Weinert 1995; Wray 1999). The research in this aspect of language comes from three

areas: lexicography, language teaching, and psycholinguistics. Common concerns are: the frequency and therefore importance of lexical phrases, the varying degrees to which lexical phrases are open to variation in wording, the functions of lexical phrases, and the importance of lexical phrases to a model of language that gives lexis and grammar equal priority.

1.3.1 The lexicographical perspective

Phrases of any kind pose a problem for the lexicographer in that they do not fit comfortably into the alphabetical headword list of the traditional dictionary. There is a time-honoured concern with 'idioms', that is, phrases which cannot be analysed or transformed according to normal syntactic rules (e.g. *He kicked the bucket* but not *They kicked the buckets* or *The bucket was kicked (by him)*), and whose meaning cannot be derived from their component parts (the meaning of *kick the bucket* cannot be derived from a knowledge of the meaning of *kick* and of *bucket*). However, as Melčuk (1995: 167) points out, idioms are only one small part of the total set of phrases which are, to some extent, 'fixed'. Moon (1992), for example, distinguishes between three types of so-called 'fixed expressions':

- a. 'anomalous collocations', which include examples such as *by and large* or *through thick and thin*, which cannot be analysed according to the normal rules governing English, in that a preposition (*by*) and an adjective (*large*) are not normally able to be coordinated, and adjectives such as *thick* and *thin* cannot normally occur as the compleitive of a preposition. Also included in this category are examples such as *kith and kin* in which one of the components 'is fossilised within that particular collocation': *kith*, for example, is found only in this phrase.
- b. 'formulae' such as proverbs, slogans, quotations, gambits, and closed-set turns, as in *You've never had it so good* and *Shut your mouth*. These items are in no way anomalous with respect to the language as a whole.
- c. 'fossilised or frozen metaphors': the 'pure idioms' such as *skate on thin ice* or *spill the beans*. These items are anomalous only in the sense that they cannot be manipulated grammatically, thus each part of the idiom (*skate*, *thin*, *ice*, *spill*, *beans*) is not treated by speakers as a separate lexical item, but as part of a phrase.

Melčuk (1988, 1995) proposes a complex set of what he calls 'non-free phrases' or 'phrasemes', each of which is 'fixed' in a particular way. According to Melčuk, a phrase is free "if and only if all its semantic and syntactic properties are completely determined by the respective properties of its constituent lexemes (and by the general rules of syntax)" (1988: 169). All other phrases are non-free.

He distinguishes between types of phraseme using formulae based on how transparent the meaning of the phrase is. In true idioms such as *shoot the breeze*, the meaning is not derivable from the constituents of the phrase; in collocations, or 'semiphrasemes' such as *crack a joke*, the meaning of one constituent (*joke*) is derivable from the general lexicon, but the meaning of the other constituent (*crack*) is determined by this particular collocation; in 'quasi-phrasemes' such as *start a family*, the meaning of the whole includes the meaning of both *start* and *family*, but includes also a further meaning: 'have one's first child'. In addition, he notes the existence of what he calls 'pragmatemes', that is, phrases which are transparent in meaning, but which are fixed in the sense that by convention one wording is consistently chosen over other possible alternatives in any given situation. He cites as an example the conventional phrase *Best before [date]*, used on food containers, which is consistently chosen in preference to, say, *To be consumed before...* or *Don't use after...*, each of which would be roughly equivalent in meaning.

For lexicographers, then, there appear to be two main questions with respect to lexical phrases: how fixed are they? and what is their relation to the grammar and lexicon of the language? This presupposes what Melčuk, with his distinction between 'free' and 'non-free' phrases, makes explicit: that lexical phrases compose only a part of the language as a whole, leaving the rest of the language to be described in other ways. The oddness of this becomes apparent when we consider Melčuk's category of 'collocation'. Collocation is to be accounted for in terms of non-free phrases only when the meaning of one of the items is tied to its co-occurrence with the other item. Thus *crack a joke* is a non-free phrase but *tell a joke* and *make a joke*, presumably, are free phrases. Yet *crack*, *tell* and *make* all collocate with *joke*. Among the significant collocates of the noun *joke(s)* in the Bank of English, for example, are the following (figures show the t-score which indicates how significant this collocate is for the node-word *joke*; see p. 231):

<i>telling</i>	14.6
<i>make</i>	13.9
<i>tell</i>	13.6
<i>making</i>	13.1
<i>told</i>	12.2
<i>made</i>	12.1
<i>cracking</i>	10.5
<i>cracked</i>	7.9
<i>crack</i>	7.8
<i>makes</i>	7.2

For the speaker wishing to talk about jokes, there is a limited set of verbs available. Neither *crack*, *make* nor *tell* represents a free choice: all are constrained by collocation, and in each case the precise meaning of the verb is determined by that collocation.

1.3.2 Language teaching

The second perspective from which lexical phrases have been investigated is that of language teaching and learning. Writers in this field include Pawley and Syder (1983), Nattinger and DeCarrico (1989, 1992), and more recently, Lewis (1993) and Willis (1990).

Researchers who are interested in language teaching place importance upon lexical phrases because of their frequency and their importance to a 'nativelike' production of the language. Pawley and Syder (1983: 191) argue that "fluent and idiomatic control of a language rests to a considerable extent on knowledge of a body of 'sentence stems' which are 'institutionalized' or 'lexicalized'". The relative fixedness of the phrases, they suggest, allows speakers to concentrate on other aspects of discourse, and thus to achieve the fluency that we associate with native speakers. They define lexicalised sentence stems thus:

"A lexicalized sentence stem is a unit of clause length or longer whose grammatical form and lexical content is wholly or largely fixed; its fixed elements form a standard label for a culturally recognized concept, a term in the language (Pawley and Syder 1983: 191).

Nattinger and DeCarrico (1989: 118) offer a similar definition of what they call lexical phrases: "These phrases are patterned sequences, usually consisting of a syntactic frame that contains slots for various fillers, and run the gamut from completely fixed, unvarying phrases to phrases that are highly variable". In a later work they add that lexical phrases are fixed in their functional application, as well as in their form (Nattinger and DeCarrico 1992: 11).

For researchers in the field of language teaching, then, lexical phrases are important because they should allow language learners to produce language that is phraseologically similar to that of native speakers and to produce language without undue hesitation or disfluency. (This assumes, of course, that learners wish to sound similar to native speakers: this will be discussed further in Chapter 9.) Cowie (1992: 10) comments that

It is impossible to perform at a level acceptable to native users, in writing or speech, without controlling an appropriate range of multiword units. Moreover, the demands of creative expression in the foreign language rests, as it does for

native speakers and writers, on prior knowledge of a repertoire of such expressions.

Lexical phrases are typically said to occupy a position 'between' lexis and syntax. Nattinger and DeCarrico (1989: 118) assert:

multi-word lexical phenomena ... exist somewhere between the traditional poles of lexicon and syntax. They are similar to lexicon in being treated as a unit, yet most of them consist of more than one word, and many of them can at the same time be derived from the regular rules of syntax, just like other sentences.

Pawley and Syder (1983: 217) suggest that they need to be described both as individual items (as lexical items) and as if they were created from rules of grammar:

If the native speaker knows certain linguistic forms in two ways, both as lexical units and as products of syntactic rules, then the grammarian is obliged to describe both kinds of knowledge; anything less would be incomplete.

We will return to the relationship between lexis and grammar in the discussion of the work of psycholinguists below.

1.3.3 Psycholinguistics

The concern of psycholinguists is how expert speakers of a language store and retrieve the language system, and how learners (of a first or second language) acquire the language. It is argued that lexical phrases play an important role in both processes, though there is substantial disagreement about this (see Weinert 1995; Wray 1999 for comprehensive reviews). Nattinger and DeCarrico (1989: 132) speculate that second language learners may acquire phrases which then provide the evidence for the learners' analysis of the language:

Lexical phrases may also provide the raw material itself for language acquisition... Later, on analogy with many similar phrases, they [learners] break these chunks down into sentence frames that contain slots for various fillers...

Krashen, on the other hand, suggests that the acquisition of prefabricated routines (such as *how are you*) and prefabricated patterns (such as *down with _____* or *that's a _____*) proceeds independently of what he calls the 'creative construction process', and that routines and patterns do not necessarily feed into the more important creative language. He concludes that "[t]he use of routines and patterns is certainly a part of language, but it is probably not a large part" (Krashen 1981: 98) and that "[t]he available evidence indicates that routines and patterns are essentially and fundamentally different from creative language" (Krashen 1981: 99).

Peters, from the point of view of first language development, disagrees (and

see Nattinger and DeCarrico 1992: 24–29 for a review of work in this area). Contrasting her own work with that of Krashen, she proposes “that formulaic speech ... is merely a facet of creative language” (Peters 1983: 4). She argues that children begin by acquiring phrases, rather than words or structure, and that these are then analysed into a system. She points out that, for the child, phrases constitute more meaningful data than words or syntactic systems:

It is not a dictionary of morphemes that the child is exposed to, but rather an intermittent stream of speech sounds containing chunks, often longer than a single word, that recur with varying frequency. It is out of this stream of unknown meaning and structure that the child must attempt to capture some pieces in order to determine their meaning and to preserve them for future use (Peters 1983: 5).

For some writers (e.g. Langacker 1987 cited in Weinert 1995) the distinction between formula and creativity is not a dichotomy but a continuum. Weinert (1995: 198) points out that this view leads to a theory of language production as comprising several distinct components, in place of the more conventional view that a single explanation may be found for it: “The view of language as a formulaic-creative continuum suggests that the units of knowledge and production may vary, including fixed formulas, mini-grammars, and general rules”.

The issue relates to language storage as well as to language acquisition. Peters argues that adult speakers of a language store the language in phrases, as well as in the form of words and syntactic rules. Becker (1975: 72, cited in Nattinger and DeCarrico 1989: 119) concurs, in a passage that is similar in essence to Francis’ suggestion for how language is encoded (see Section 1.5):

[the frequency of lexical phrases in performed speech implies] that the process of speaking is *Compositional*: We start with the information we wish to express or evoke, and we haul out of our phrasal lexicon some patterns that can provide the major elements of this expression. Then the problem is to stitch these phrases together into something roughly grammatical, to fill in the blanks with the particulars of the case at hand, to modify the phrases if need be, and if all else fails to generate phrases from scratch to smooth over the transitions and fill in any remaining conceptual holes.

Nattinger (1988: 75; see also Wray and Perkins 2000) refers to this approach as a more general theory of language performance:

Many theories of language performance suggest that vocabulary is stored redundantly, not only as individual morphemes, but also as parts of phrases, or even as longer memorized chunks of speech, and that it is oftentimes retrieved from memory as these preassembled chunks (Bolinger 1975).

Peters (1983: 90) concurs that ‘dual storage’, that is, having phrases available in memory as single (lexical) items, as well as the syntactic rules that allow them to be created, leads to redundancy in language knowledge. This in turn, she suggests, implies a fluidity between lexis and syntax:

...[T]here is considerable redundancy in the storage of both lexical and syntactic information. The relation between syntax and lexicon may therefore be more fluid than is usually supposed: Under some circumstances an expression may be retrieved from the lexicon as a single unit; under others it may be constructed from partially assembled pieces in the lexicon, requiring somewhat more syntactic processing; under yet other circumstances it may be constructed de novo from morphemes. Syntax and lexicon are thus seen to be complementary in a dynamic and redundant way. The same information may be present in both, in different forms...

This argument sounds somewhat similar to Sinclair’s assertion that “[t]he evidence now becoming available casts grave doubts on the wisdom of postulating separate domains of lexis and syntax” (Sinclair 1991: 104; see Section 1.4 for full discussion). It is worth pointing out, however, that the two positions are different. Peters does not disagree with the distinction between syntactic rules, which offer an abstract, very productive, system for the production of all possible sentences in a language, and the lexicon of that language. She simply argues that the speaker’s mind may store certain items both lexically (as single items) and syntactically (as the product of the operation of rules). The ‘fluidity’ she identifies refers to how the speaker encodes on different occasions, rather than to the description of the language itself. Sinclair’s position is more radical, and relates to the description of the language rather than to how speakers might encode. He argues that syntactic rules account for only a minimal part of the grammar of a language, and that the more important part is composed of the phraseological constraints upon individual lexical items. Thus, syntax is not a system independent of lexis: lexis and syntax must, ultimately, be described together.

1.3.4 Lexical phrases and a pattern grammar

It may seem that this discussion of lexical phrases has strayed a long way from the central concern of a pattern approach to grammar. There are two reasons for this digression. Firstly, the work on lexical phrases, much of which took place before language corpora were commonplace, in a sense prefigures Sinclair’s work on collocation and the ‘idiom principle’ (see Section 1.4). The availability of corpora allows us to identify with some certainty the frequently-occurring sequences of items that the lexicographers, language teachers, and psycholinguists

discussed in this section could identify only through intuition (see J. Willis 1997 for a discussion of the unreliability of intuition in this respect). In addition, placing the lexical item and its patterning centre-stage, as it were, breaks down the distinction which the concern for lexical phrases maintained: the distinction between what Melčuk called free and non-free phrases. The work of Sinclair and other corpus linguists suggests that all language is patterned, that there is no such thing as a free phrase, and that ultimately, the study of lexical phrases can simply be subsumed into a more general description of language.

Secondly, and more specifically, it is important to note that the grammar patterns discussed in the rest of this book are in a sense examples of lexical phrases. Although none of the writers we have discussed here mention phrases that we would call the product of grammar patterns, phrases such as *it is hard to believe that...*, *be interested in...*, *the fact that...* or *apologise to ... for ...ing* surely come within the remit of lexical phrases. Writers on lexical phrases and on grammar patterns, it might be argued, seek to account for the some of the same evidence in different ways.

It would be wrong to suppose, however, that grammar patterns are simply a special case of lexical phrases. Collections of lexical phrases are, ultimately, fairly random lists of phrases, organised either according to their relative fixedness, or to their function (discourse organising, opinion-giving and so on), or to one of their core words (see Pawley and Syder 1983 for examples of phrases with *think*). They are an attempt to account for a portion only of the lexicon. Grammar patterns, on the other hand, constitute an attempt to describe the whole of the language (or rather, all the frequently-occurring items in the language) in a principled way, and the lists of words collected in a given pattern are not random. The two approaches are far apart theoretically and in terms of language description in general.

1.4 Sinclair: Corpus Concordance Collocation

1.4.1 Corpus-driven language description

This section reviews the work of Sinclair, largely as it appears in the book *Corpus Concordance Collocation* (1991). This book is about more than it seems; from a handful of deceptively simple examples Sinclair sets out an agenda for a radical departure in the description of English.

Sinclair's work differs from that of other linguists in that he prioritises a method, or group of methods, and a kind of data rather than a theory. This

approach is what has come to be known as corpus linguistics: a way of investigating language by observing large amounts of naturally-occurring, electronically-stored discourse, using software which selects, sorts, matches, counts and calculates. The data that is the basis for this differs from that used with other methods of linguistic investigation in five respects:

- The data is authentic;
- The data is not selected on linguistic grounds;
- There is a lot of data;
- The data is systematically organised;
- The data is not annotated in terms of existing theories.

Each of these features may be stated as a principle, and is discussed in turn below.

The data is authentic The observation of actually-occurring language may be contrasted with introspection and intuition as ways of obtaining information about how language works. As Sinclair points out (1991:39): "It has been fashionable among grammarians for many years now to introspect and to trust their intuitions about structure...". He is, however, critical of the reliability of such intuition as an accurate reflection of language in use:

...the contrast exposed between the impressions of language detail noted by people, and the evidence compiled objectively from texts is huge and systematic. It leads one to suppose that human intuition about language is highly specific, and not at all a good guide to what actually happens when the same people actually use the language (Sinclair 1991:4).

As a result of this, Sinclair (1991:4) argues that "Their [intuition-led grammarians] study has ... been more about intuition than about language" and states what has become a commonplace: to find out about the language that people actually use, one must observe the language that people actually use.

Francis and Sinclair (1994: 197) quote the identification of ergative verbs as an example of the superiority of authentic data over intuition. They argue that it is not possible to tell whether a given verb is used ergatively or not simply by consulting intuition, and they cite the verb *clarify* as an instance: the authors could not determine through intuition whether this was an ergative verb or not. Consultation of a corpus of authentic English solved the problem, as *clarify* was found being used transitively in examples such as *She clarified the situation* and intransitively in examples such as *The situation clarified*, and could thus be classified as an ergative verb.

The data is not selected on linguistic grounds. Sinclair is by no means unique in advocating a reliance on authentic language as data. Halliday and other

proponents of Systemic-Functional Grammar, for example, typically derive their analytical categories from instances of actually-occurring discourse. Indeed, it would scarcely be possible to propose a grammar as social-semiotic (Halliday 1978) without taking as its starting-point language in actual use.

For most grammarians and other linguists, however, data is selected because it illustrates a particular language point. The grammarian cites observed instances of language in use that have caught his/her eye or ear. Sinclair (1991: 100) comments that: "This method is likely to highlight the unusual in English and perhaps miss some of the regular, humdrum patterns."

In other words, where instances of language are selected for analysis precisely because they strike the linguist as interesting, they are likely to exemplify the unusual rather than the mundane. Of course all language data is selected, but the texts that go into a corpus are chosen because they play a particular social role, rather than because they demonstrate a peculiarity of usage (see, for example, Renouf 1987).

There is a lot of data. Perhaps what is most striking about the data that Sinclair, and other corpus linguists, uses is its quantity. Even the earliest corpora consisted of around 1 million words of running text (Leech 1991: 10), far more than most linguists use as data. The corpus that Sinclair describes in *Corpus Concordance Collocation* consisted of just over seven million words; the current (1997) Bank of English corpus consists of over 300 million words. As a corpus gets bigger, it is possible to describe more and more accurately items of less and less frequency. For this reason, no corpus is really big enough, and no corpus could be too big (provided that other issues, such as spread of sources, are also taken into account; see Leech 1991). Limitations of size are imposed by storage considerations, and by the problems of devising software that can search, sort and count very large number of items quickly, but these are problems of hardware and software, not problems of language description.

Sinclair's argument that 'bigger is always better' relies on the assertion that in this case quantity is also quality. He says (1991: 4): "... the ability to examine large text corpora in a systematic manner allows access to a quality of evidence that has not been available before", and observes dryly that "[t]he language looks rather different when you look at a lot of it at once" (1991: 100). But why should quantity be quality? The difference between looking at a lot of data and a little, is that when a lot of data is examined, conclusions as to frequency can be drawn. Sinclair (1991: 4) draws attention to two observations about language which cannot be made without recourse to frequency information: firstly that some sequences of words co-occur surprisingly often, given that every utterance or written sentence spontaneously produced is unique; secondly, and in contrast,

that even so-called fixed expressions demonstrate surprising amounts of variability (cf. Moon 1994; 1998).

These observations are not peripheral to the serious business of grammatical description but challenge its very heart. Sinclair contradicts the 'received wisdom' that a small corpus is sufficient for doing grammar (an argument later propounded by Carter and McCarthy 1995: 143) by arguing that "[t]he new evidence suggests that grammatical generalizations do not rest on a rigid foundation, but are the accumulation of the patterns of hundreds of individual words and phrases" (1991: 100).

Sinclair's choice of the word 'generalizations' is significant here, in that it contrasts with those approaches to grammar that expound an abstract 'langue' or system that in some way underlies actual instances of language use. For Sinclair, it appears, there is no system setting the parameters of what may be said or written, only a set of generalisations capturing the essence of what has been said or written. Systemic-Functional terminology might be borrowed to rephrase this: language is not a system that is realised in actual instances, but a set of actual instances that may be regarded as construing an approximate and ever-changing system. Such a construal stems from the interpretation of hundreds of observations, made possible by the sheer amount of data available.

One of the outcomes of using large quantities of data is that some of it may be discarded, in the sense that instances of word-play or language that is strange because it is being used in strange circumstances, are deliberately ignored in terms of the general description of the language (Sinclair 1991: 99), though they may form the focus of studies of a different kind (e.g. Louw 1993, 1997). This is a different approach from that of many grammars, which concentrates on what is possible, not on what is frequent. Obviously, the view of what is of minor importance changes as the corpus gets bigger, and might differ according to the perceived audience for the description, but the point about a very large corpus is that it enables the observer to see what is 'central and typical' (Hanks 1987) and distinguish that from the less frequent usage. Sinclair (1987: 108) distinguishes between the typical and the possible thus:

For example, it is significant if, in several hundred instances of the base form of a verb, none signal the imperative. This is not to say that in the ethereal world of theoretical, school or traditional grammar the imperative of that verb is 'impossible'.

The data is systematically organised. Data alone is not enough, however. The greater the amount of data, the greater the need for organisation. Faced with a corpus containing hundreds of millions of words and no systematic organisation,