

# Corpus Approaches to Evaluation

Phraseology and Evaluative Language

Susan Hunston

# **Corpus Approaches to Evaluation**

# Routledge Advances in Corpus Linguistics

EDITED BY **TONY McENERY**, *Lancaster University UK*

**MICHAEL HOEY**, *Liverpool University, UK*

## **1. Swearing in English**

Bad Language, Purity and Power  
from 1586 to the Present  
Tony McENERY

## **2. Antonymy**

A Corpus-Based Perspective  
Steven Jones

## **3. Modelling Variation in Spoken and Written English**

David Y. W. Lee

## **4. The Linguistics of Political Argument**

The Spin-Doctor and the Wolf-Pack  
at the White House  
Alan Partington

## **5. Corpus Stylistics**

Speech, Writing and Thought  
Presentation in a Corpus of  
English Writing  
Elena Semino and Mick Short

## **6. Discourse Markers Across Languages**

A Contrastive Study of Second-Level  
Discourse Markers in Native and  
Non-Native Text with Implications for  
General and Pedagogic Lexicography  
Dirk Siepmann

## **7. Grammaticalization and English Complex Prepositions**

A Corpus-Based Study  
Sebastian Hoffman

## **8. Public Discourses of Gay Men**

Paul Baker

## **9. Semantic Prosody**

A Critical Evaluation  
Dominic Stewart

## **10. Corpus Assisted Discourse Studies on the Iraq Conflict**

Wording the War  
Edited by John Morley and Paul Bayley

## **11. Corpus-Based Contrastive Studies of English and Chinese**

Richard Xiao and Tony McENERY

## **12. The Discourse of Teaching Practice Feedback**

A Corpus-Based Investigation of  
Spoken and Written Modes  
Fiona Farr

## **13. Corpus Approaches to Evaluation**

Phraseology and Evaluative Language  
Susan Hunston

# **Corpus Approaches to Evaluation**

Phraseology and Evaluative Language

**Susan Hunston**



**Routledge**

Taylor & Francis Group

New York London

First published 2011  
by Routledge  
270 Madison Avenue, New York, NY 10016

Simultaneously published in the UK  
by Routledge  
2 Park Square, Milton Park, Abingdon, Oxon OX14 4RN

This edition published in the Taylor & Francis e-Library, 2010.

To purchase your own copy of this or any of Taylor & Francis or Routledge's collection of thousands of eBooks please go to [www.eBookstore.tandf.co.uk](http://www.eBookstore.tandf.co.uk).

*Routledge is an imprint of the Taylor & Francis Group, an informa business*

© 2011 Taylor & Francis

The right of Susan Hunston to be identified as author of this work has been asserted by her in accordance with sections 77 and 78 of the Copyright, Designs and Patents Act 1988.

Page 187. Figure 9.1 'Statement types' (From chapter 'Evaluation and the planes of discourse' by Susan Hunston) from "Evaluation in Text: authorial stance and the construction of discourse" edited by Susan Hunston and Geoff Thompson (2000 hb). Reproduced by permission of Oxford University Press.

All rights reserved. No part of this book may be reprinted or reproduced or utilised in any form or by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying and recording, or in any information storage or retrieval system, without permission in writing from the publishers.

**Trademark Notice:** Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

*Library of Congress Cataloging-in-Publication Data*

Hunston, Susan, 1953–

Corpus approaches to evaluation / Susan Hunston.

p. cm. — (Routledge advances in corpus linguistics ; 13)

Includes bibliographical references and index.

1. Corpora (Linguistics) I. Title.

P128.C68H86 2010

410.1'88—dc22

2010014464

ISBN 0-203-84168-9 Master e-book ISBN

ISBN13: 978-0-415-96202-5 (hbk)

ISBN13: 978-0-203-84168-6 (ebk)

*For Linda and for John*



# Contents

<i>List of Figures</i>	ix
<i>List of Tables</i>	xi
<b>1 Evaluative Language, Phraseology and Corpus Linguistics</b>	<b>1</b>
<b>2 Appraisal, Stance, Evaluation</b>	<b>10</b>
<b>3 Status in Written Texts and Multi-Modal Texts</b>	<b>25</b>
<b>4 Evaluation, Quantity and Meaning</b>	<b>50</b>
<b>5 Modal-Like Expressions</b>	<b>66</b>
<b>6 Corpus Approaches to Investigating Status</b>	<b>92</b>
<b>7 Grammar Patterns, Local Grammars and Evaluation</b>	<b>119</b>
<b>8 Phraseology, Intensity and Density</b>	<b>151</b>
<b>9 Conclusion</b>	<b>166</b>
<i>Appendix to Chapter 5</i>	173
<i>Notes</i>	183
<i>Bibliography</i>	187
<i>Index</i>	195



# Figures

2.1	Concordance lines for <i>trust</i> .	15
3.1	From Hunston (2000a: 187).	32
4.1	Examples of <i>terrible</i> .	57
4.2	<i>CAUSE</i> and <i>BRING ABOUT</i>	59
4.3	<i>a shred of</i> .	60
4.4	<i>PAY . . . price</i> .	61
5.1	Concordance lines for <i>decide wh</i> (from Hunston 2003b).	70
5.2	Concordance lines for <i>distinguish between</i> .	71
5.3	Concordance lines for <i>distinguishing between</i> .	72
6.1	Concordance lines for <i>assumption that</i> .	94
6.2	Phraseologies associated with the negative evaluation of assumptions.	100
6.3	adjective <i>to assume that</i> .	101
6.4	Writer's assumptions.	102
6.5	<i>we assume that</i> .	103
6.6	<i>we assume that</i> .	104
6.7	Concordance lines for <i>this discovery</i> .	107
6.8	Examples of <i>Vn as n</i> .	111
7.1(a)	Co-text: <i>RECOVER</i> is followed by a noun phrase ( <i>V n</i> ); also included here is the passive equivalent ( <i>be V-ed</i> ).	124

x *Figures*

7.1(b)	Co-text: <i>RECOVER</i> is followed by a prepositional phrase beginning with <i>from</i> ( <b>V from n</b> ).	124
7.1(c)	Co-text: <i>RECOVER</i> is followed by a noun phrase and a prepositional phrase beginning with <i>from</i> ( <b>V n from n</b> ).	124
7.1(d)	Co-text: <i>RECOVER</i> occurs at the end of a clause or is followed by a non-dependent prepositional phrase or adverb ( <b>V</b> ).	125
7.2(a)	Adjectives indicating emotion.	129
7.2(b)	Adjectives indicating human qualities.	129
7.2(c)	Adjectives indicating qualities of things.	129
7.2(d)	Adjectives indicating attitudes.	130
7.3	Examples of <b>ADJ about n</b> .	140
8.1	Intensifying sequences.	158

# Tables

3.1	Status in Example 3.5	29
3.2	Status in Example 3.6	31
3.3	A Classification of Source Types	35
3.4	'Engagement' Account of Example 3.16	45
3.5	'Status' Account of Example 3.16	45
3.6	'Clause Relations' Account of Example 3.16	46
3.7	'Engagement' Account of Example 3.17	48
3.8	'Status' Account of Example 3.17	48
5.1	Frequencies of Left Collocates before <i>distinguish</i> (Rounded Percentages)	74
5.2	Sequences, Patterns and Meanings with <i>to distinguish</i>	75
5.3	Frequencies of Left Collocates before <i>distinguishing between</i> (Rounded Percentages)	76
5.4	Modal-Like Elements Identified by Selecting <i>of + -ing form</i>	78
5.5	Rounded Percentage of Each Wordform in Four Lemmas	82
5.6	Left Collocates of Four Base Forms (200 Lines Each)	82
5.7	Relative Frequency of Wordforms for Four Lemmas	83
5.8	Modal Meaning and <i>whether</i>	84
5.9	Sentences with <i>for fear of</i> and Their Paraphrases	87
5.10	Summary of Uses of 10 Selected Verb Sequences	89

xii *Tables*

6.1	Frequencies of <i>fact</i> in the BNC	109
6.2	Frequencies of <i>hypothesis</i> in the BNC	109
7.1	Parsing the <b>V n n</b> Pattern	121
7.2	Parsing the <b>V n from n</b> Pattern	121
7.3	Adjectives in Two Patterns	132
7.4	Adjectives in Appraisal Types	137
7.5	Parsing the <b>V n as adj</b> , <b>V n n</b> and <b>V n adj</b> Patterns	141
7.6	Parsing Definitions	143
7.7	<i>It v-link ADJ -ing</i>	145
7.8	<i>It v-link ADJ for n to-inf</i>	145
7.9	<b>V it ADJ to-inf</b> (Example from FrameNet)	145
7.10	<b>V for n to-inf</b>	145
7.11	<b>V for n to-inf</b>	146
7.12	Using ‘Evaluation Limiter’	146
7.13	Using ‘Evaluation Limiter’	146
7.14	Parsing <b>v-link ADJ to-inf</b>	148
7.15	Parsing <b>v-link ADJ to-inf</b>	148
7.16	Parsing <b>v-link ADJ to-inf</b>	149
7.17	Parsing <b>v-link ADJ to-inf</b>	149
8.1	Verbs and Nouns with <i>months of</i>	163
8.2	Instances of Intensifying Phrases	164

# 1 Evaluative Language, Phraseology and Corpus Linguistics

This book examines the role of corpus linguistics in the study of evaluative language, prioritising that approach to corpus linguistics which focuses on phraseology.

There are clearly three terms in this opening sentence—evaluative language, corpus linguistics and phraseology—that require explanation. This is the purpose of this chapter.

## 1.1 EVALUATIVE LANGUAGE

Evaluative language is that language which indexes the act of evaluation or the act of stance-taking (Du Bois 2007). It expresses an attitude towards a person, situation or other entity and is both subjective and located within a societal value-system (Hunston 1994: 210). As a brief illustration of these points, here is an example of an evaluative act, taken from a newspaper review of a book which represents a new genre, ‘monster mash-ups’. These are books which combine characters and language from classical literature with plots taken from modern horror literature, for comic effect. The reviewer writes:

### Example 1.1

But the other obvious problem with monster mash-ups is that the joke very quickly grows old. *Pride and Prejudice and Zombies* is often very funny, but by the third or fourth chapter you’ve well and truly got the idea; by the time you come to *Sense and Sensibility and Sea Monsters*, the novelty has thoroughly faded. (Merritt 2009)

This paragraph performs an act of evaluation, fulfilling the purpose of the book review. It uses particular resources of English to indicate this function: the individual words *problem* and *funny*, and the word combinations *joke grows old* and *novelty has faded*. The evaluation is strengthened by the intensifiers *very*, *well and truly* and *thoroughly*.

A particularly interesting phrase is *you’ve well and truly got the idea*. This might be interpreted in a positive light—true understanding has been

## 2 *Corpus Approaches to Evaluation*

achieved. However, following as it does the comment *the joke very quickly grows old*, it implies a negative evaluation: the reader achieves an understanding that is too complete too early to convey maximum enjoyment of the book. In this way the paragraph appeals to a shared value between writer and reader: that jokes are funny when they are novel and that once a joke is familiar it is no longer as funny.

The paragraph gives information to the reader, but this is information that is entirely subjective. Another reviewer might offer a completely different assessment of the books concerned. Nonetheless, there is in this particular review no indication that the opinion held is purely personal, no overt concession to other points of view. A dialogue is enacted, between *Pride and Prejudice* and *Zombies* is often very funny and by the third or fourth chapter you've well and truly got the idea, but that dialogue does not admit of alternative viewpoints. To interpret the paragraph as 'this is what this writer thinks' rather than as 'this is what is objectively verifiable' depends upon understanding the nature of evaluation as a subjective act.

The paragraph does, however, engage the reader in addressing him or her as *you*. This underlines the paragraph as an act of persuasion; the expectation is that the reader will agree with the opinions expressed.

There are clearly a number of ways in which the brief explication of this paragraph could be extended, each of which corresponds to a tradition in the study of evaluative language. Perhaps the most obvious would be to study in more detail the particular language resources that are used to convey evaluation: the words, collocations and phrases. This is an endeavour that was recommended by Stubbs (1986) and that has been carried out in numerous contexts, for example, the extensive literature on stance, engagement and metadiscourse (e.g. Biber 2006a; Hyland 2009; Hyland and Tse 2004). Once such features have been identified, one set of texts can be measured against another to compare the amount and type of evaluative language in each (Biber 2006a; Hyland 2009; Charles 2006a, 2006b). A second approach might be to use identified linguistic resources to assess how a book such as *Pride and Prejudice* and *Zombies* (in Example 1.1) has been greeted by a large number of critics. This is the approach taken by Sentiment Analysis (e.g. Taboada and Grieve 2004), where the aim is to identify positive and negative evaluation in a very large collection of texts. The means to achieving that aim involves the automatic identification of evaluative language.

Most researchers find it advantageous and indeed instructive to break down the mass of language often referred to as 'evaluative' into different types. For example, G. Thompson (2001) distinguishes between 'interactive' and 'interactional', a distinction taken up by Hyland and Tse (2004). In other publications Hyland distinguishes between 'stance' and 'engagement'. Hunston (1994) identifies three functions of evaluation and suggests that they identify the epistemic object being evaluated (status), the value given to the object (value) and the relevance of parts of the text (relevance). Martin and White (2005) offer what is probably the most theory-grounded study of the functions and

forms of evaluative meaning in English, based on systemic-functional linguistics. They would point out that Example 1.1 can be discussed in a number of ways: as an example of a unilateral statement as opposed to a debate ('engagement'), as an intensified act of evaluation ('intensification') and as evaluation construed by assessing both the quality of the book—*funny*—and the reader's reaction to it—*the joke very quickly grows old* ('appraisal').

The study of evaluative language is important for a number of reasons. Indicating an attitude towards something is important in socially significant speech acts such as persuasion and argumentation. Taking a stance towards something and negotiating alignment or non-alignment is a crucial aspect of interaction between individuals. As linguistic study moves away from truth-value and towards a focus on the interactive, the importance of this aspect has become more apparent.

In spite of its importance, evaluative language can pose difficulties for linguistic theories. In some theories, language is classified into broad functions (or metafunctions). These may include an interpersonal function: building and expressing relationships. Evaluative language most obviously belongs to this function. On the other hand, Halliday and Hasan (1985) and Sinclair (1987) suggest that only some evaluation has an interpersonal, or interactive, function, while other acts of evaluation are located within the ideational function.

Evaluative language presents difficulties in analysis because there is no set of language forms, either grammatical or lexical, that encompass the range of expressions of evaluation. It is true that adjectives and adverbs frequently express evaluative meaning (e.g. Turney 2002; Conrad and Biber 2000), and, as shall be argued in Chapter 7, some patterns of use are associated with such meaning, but this does not mean that every adjective and adverb marks evaluation or that all evaluation can be identified in this way.

In fact, evaluation is frequently expressed cumulatively and implicitly. Example 1.2 is an example of cumulative evaluation (in which the names of the professor and the university have been replaced by pseudonyms):

#### Example 1.2

As I write this, Professor Smith, now a distinguished scholar, has her job under threat from the ghastly, grey accountants who run the University of Biggin-on-Sea. We are now in an epoch of production-line universities with celebrities paid fortunes to teach eight hours a week and genuine scholars dumped in the bin. (Ali 2008)

Evaluation of the University of Biggin-on-Sea, and of British universities in general, is cumulative: *ghastly, grey accountants, production-line universities, genuine scholars dumped in the bin*. This accumulation of evidence allows us to interpret *celebrities paid fortunes to teach eight hours a week* in a negative light.

Example 1.3 is an example of implicit evaluation:

#### 4 *Corpus Approaches to Evaluation*

##### Example 1.3

To determine whether any branched species [of alga] was differentially important to the establishment of surfgrass, I followed seedlings for 7 months. During this time 90% of the seedlings died . . . but the seeds attached to each algal species died in about the same proportion . . . These results suggest that all the branched species facilitate surfgrass and that the magnitude of the facilitation is proportional to the number of seeds originally attached to the alga. (Turner 1983)

In this paragraph, a question is asked—‘Does the species of alga affect how well surfgrass becomes established?’—and answered in the negative—‘Any species of alga helps surfgrass become established; it is the number of seeds that makes the difference’. Hunston (1993a) argues that evaluation here consists of setting a goal (‘To determine whether . . .’) and then noting that the goal has been achieved (‘The results suggest that . . .’), thereby evaluating the experiment as successful. Arguably, then, an act of evaluation has taken place that is marked by lexical repetition (*any branched species— all the branched species; important to—facilitate; differentially—proportional to*) but not by any recognisable instance of evaluative language.

These features of cumulation and implicitness make it appear that evaluative language is more suited to text-based than to corpus-based enquiry. The detail involved in explicating the evaluative language in a given text suggests that very close reading is required. It would seem, then, that the broader sweep adopted by corpus linguistic methods would be suitable only for adding quantitative detail to the more interesting small-scale studies. It is one of the tasks of this book to argue that this is not the case.

## 1.2 CORPUS LINGUISTICS

Corpus linguistics is, by now, a term covering a wide range of activities and approaches (e.g. Baker 2009; Sampson and McCarthy 2005; Teubert 2005). At its most basic, shared concerns include collecting quantities of text in electronic form so that they are open to data-manipulation techniques. Such techniques range from finding a search term and observing its immediate environments (key-word-in-context or concordance lines); to calculations of relative frequency (as in, for example, collocation studies); to annotation for such categories as word class, grammatical function or semantic class; and frequency calculations based on such categories. Frequencies of various kinds can be compared in different corpora, leading to observations about different registers or different languages and about the development of a language over time.

Corpus linguistics is more than a simple set of techniques, but it is a field where technological advancement and theoretical development go hand in hand. For example, one of the most basic corpus linguistics

practices—isolating a node word and its immediate co-text in each instance of occurrence, and manipulating that output to highlight similarity in co-text (i.e. sorting concordance lines)—was developed because words were considered to be more important than grammatical categories, and the immediate co-text of a word, including its significant collocations, was considered to offer the most important information about it. In turn, the prevalence of concordancing software has facilitated the observation of language patterning and so influenced the development of theories based on that phenomenon.

Corpus linguistics, then, is more than a way of investigating existing models of language. The methodologies it uses can be used both to change and to complement our understanding of those models. For example, from different perspectives, each of Biber et al. (1999), Carter and McCarthy (2006), Holmes and Nesi (2009) and Matthiessen (2006) focus on differentials between language in different registers or modes. The concept of register pre-dates large-scale corpus research of the kind used today. Corpus research has assisted, however, in quantifying the occurrence of various features in different registers (see Biber et al. 1999; Matthiessen 2006, though each is working with a very different concept of ‘register’), and in identifying features which are unique to a specific mode (Carter and McCarthy 2006).

On the other hand, research at the lexis-grammar interface (e.g. Sinclair 1991) offers a revision to previous models of language by suggesting that a sharp divide between syntax and semantics may not be advisable. More specifically, corpus linguistics has identified what in this book I shall call ‘phraseology’ as a key concept in the way that language works.

### 1.3 PHRASEOLOGY

‘Phraseology’ is a very general term used to describe the tendency of words, and groups of words, to occur more frequently in some environments than in others. As used in this book, the term takes its inspiration from Sinclair’s (1991) summary of a number of corpus studies that together indicate that (a) more language occurs in ‘fixed phrases’ than might otherwise be thought and, furthermore, that (b) ‘fixed phrases’ are more varied than might otherwise be thought. This observation leads on the one hand to a concern for multi-word units (MWUs), that is, stretches of language consisting of two, three or four (or more) words which occur frequently in a given corpus and which can often be seen to play a particular role in a given register. A large number of studies today focus on MWUs. On the other hand, it also leads to the identification of what Sinclair called ‘units of meaning’, that is, flexible word sequences which display some consistency in aspects of form but more so in aspects of meaning. Added to this is Hoey’s work on ‘lexical priming’ (2005). Hoey argues for the priming connected with individual words as a key psycholinguistic mechanism by which language is acquired,

## 6 *Corpus Approaches to Evaluation*

used and interpreted. Alongside the psycholinguistic arguments lie a set of observations which indicate that differential frequency relating to individual words is much more patterned than had previously been thought. Not only do words generally co-occur with specific others (collocation) and more frequently in one set of grammatical environments than others (colligation), but many words occur differentially in different parts of a text, such as in paragraph or text initial position. Some of these differential frequencies are register-independent and others are register-specific.

Under the very broad heading of phraseology, therefore, comes a range of diverse phenomena of which the following are examples:

- The noun *cold* co-occurs with the adjectives *bitter*, *icy* and *extreme*. The noun *heat* collocates with *intense*, *sweltering* and *extreme*.
- The verb *afford* mainly occurs with modal verbs such as *can* and *could*. On average, over half of occurrences have negative forms such as *cannot* or phrases such as *can ill afford*. This is remarkable when it is borne in mind that the ratio of positive and negative clauses in English is 9:1 (Halliday 1993).
- The noun *consequence* is much more likely to occur as the subject of a verb than as its object (Hoey 2005: 46–47).
- In US university discourse there are a number of MWUs or ‘lexical bundles’ that are frequently used by teachers to focus attention on a new topic (Biber 2006b: 142–143). These include *want to talk about* and *what I want to do is*.
- The two-word phrase *true feelings* is used in fairly restricted contexts, expressing difficulty or reluctance to express genuine emotion. The contexts are varied in form but all convey, often implicitly, this sense. Explicit examples include: *she hid her true feelings*; *when I’m able to reveal my true feelings*; *we lose the ability to express our true feelings*; *only her close female friends . . . had any idea of her true feelings* (Sinclair 2003).
- An MWU may be infrequent overall but may still have a quality of ‘phrase-ness’ because each element is frequent relative to the other elements. An example is *after a few moments of*, which occurs relatively rarely but in which *few* is a significant collocate of *moments of* and *after* is a significant collocate of *a few moments of* (Hunston 2002).

Although phraseology as a topic of research has gained considerably from the input of corpus linguistics, Gries (2008a) notes that it is also a central tenet in cognitive linguistics and construction grammar, and that it is not ignored either by generative grammarians. Commenting on the lack of agreement, or even explicitness, as to what might constitute the object of attention in phraseological studies (or ‘phraseologism’), Gries (ibid.: 4) identifies six features which provide a useful yardstick, not in the sense that all phraseological studies adopt the same attitude towards them, but in the

sense that such studies might legitimately be compared in terms of their stance towards them. The six features (quoted verbatim from Gries with comments from myself in parentheses) are:

- i. the *nature* of the elements involved in a phraseologism (studies may be divided into those that restrict themselves to wordforms or lemmas and those which adopt a wider brief, with elements that may be words, or grammatical forms, or broadly-defined elements of meaning);
- ii. the *number* of elements involved in a phraseologism (studies of collocation typically restrict themselves to two elements, not necessarily contiguous, while units of meaning as identified by Sinclair are not restricted in the number of words they may contain);
- iii. the *number of times* an expression may be observed before it counts as a phraseologism (Gries adopts the criterion that to be counted as a phraseologism, an expression must be observed more frequently than expected);
- iv. the permissible *distance* between the elements involved in a phraseologism (studies of contiguous collocations allow no distance whereas many studies specify discontinuous elements; Sinclair and Renouf (1991) is an example of the latter; Gries might have added ‘whether or not flexibility in the order of the elements is taken into account’);
- v. the degree of *lexical and syntactic flexibility* of the elements involved (as noted earlier, evidence suggests that most phrases admit considerable variation, and indeed much linguistic creativity depends on taking an apparently fixed phrase and treating it as flexible);
- vi. the role that *semantic unity* and *semantic con-compositionality/non-predictability* play in the definition (idioms are items that are both unified semantically—an idiom ‘means something’—and non-compositional—their meaning cannot be predicted from the meaning of each individual word; most studies of phraseology, however, apply the first but not the second of these restrictions, and work such as Hoey’s applies neither).

Looking at evaluative language from the perspective of phraseology is not new. Biber et al. (1999), for example, identify lexical bundles that express stance, such as *no doubt*, *in fact* and *according to*. (See also Biber 2006b for a more detailed discussion.) In addition, one of the insights owing to Sinclair’s ‘unit of meaning’ theory is that units of meaning express attitude, often implicitly. In this book it will be argued that phraseology as broadly defined plays a number of roles in the study of evaluative language.

#### 1.4 THE ORGANISATION OF THIS BOOK

The second and third chapters of this book focus on evaluative language as a discursal, rather than as a corpus, phenomenon. Chapter 2 expands the

## 8 *Corpus Approaches to Evaluation*

introduction to the ways that evaluative language has been studied without corpora. Chapter 3 provides an account of the concept of epistemic status, as an introduction to the work in Chapter 6.

The rest of the book focuses on corpus studies of evaluative language. Chapter 4 reviews the literature in this area. In Chapter 5, the role of phraseology in the expression of modal meaning is investigated. Evidence is put forward to support the argument that some verbs co-occur with modal meaning more frequently than others do, and that modal meaning can be expressed by a large number of grammatically varied phrases as well as by traditional modal verbs. In Chapter 6, it is argued that status nouns, which evaluate the epistemic status of a proposition, can themselves be observed to co-occur differentially with phrases that evaluate them in one respect or another. Chapter 7 considers the possibility of devising a local grammar of evaluation, based on the co-occurrence of words and grammatical units. Finally, in Chapter 8, the role of particular phrases in supporting and intensifying evaluative meaning is studied. Chapter 9 provides a conclusion to the book.

### 1.5 A NOTE ON CORPORA AND TERMINOLOGY

For the most part, the corpus information in this book is drawn from the Bank of English corpus held at the University of Birmingham. The corpus itself is jointly owned by the University of Birmingham and HarperCollins publishers. The version used in this book comprises 450 million tokens and consists of a number of sub-corpora consisting of, among others, news journalism from Britain, US and Australia, fiction and non-fiction books published in Britain and the US, spontaneous spoken English from Britain and the US and issues of a large number of special interest magazines. The study in Chapter 6 is based on the sub-corpus containing issues of *New Scientist* magazine. Otherwise, unless otherwise stated, information about collocation, frequency and usage is based on the whole Bank of English corpus, and examples and concordance lines are taken from there.

The software used to search the Bank of English determines some definitions of the terminology used in this book. A **random** sample of concordance lines takes each *n*th instance of the node word from across the corpus or the selected sub-corpora. For example, if there are 1,000 instances of a given word, and the search request specifies 100 examples, the software will take display every tenth occurrence. For calculations of **collocation**, the software specifies a span of four words each side of the node. For the most part, and unless otherwise indicated, I use raw frequency to place collocates in order, rather than measures of significance such as z-score, t-score or Mutual Information. In some discussions, reference is made to collocates that are in the L1 or R1 positions relative to the node (that is,

immediately preceding or following the node word). Where these are placed in order, again it is raw frequency that is used to establish the order.

Throughout this book, italics are used to indicate wordforms and capitals are used to indicate lemmas. A lemma is defined as the abstraction from a set of wordforms belonging to the same word class. Thus, *WALK* indicates *walk*, *walks*, *walking* and *walked* (but not the noun *walk*).