Contents

List of figures ix
List of tables xi
To readers xvi
List of abbreviations xviii

1 Corpus linguistics 1
1.1 Introducing corpus linguistics 1
1.2 The professor's shoeboxes 2
1.3 The tape-recorder and the computer 3
1.4 What can we get out of corpora? 5
1.5 Criticism from armchair linguists 8
1.6 Types of corpora 10
1.7 Very large and quite small corpora 22
1.8 Summary 22
Study questions 23
Corpus exercises 23
Further reading 23

2 Counting, calculating and annotating 25
2.1 Qualitative and quantitative method 25
2.2 Frequency 26
2.3 Comparing frequencies 37
2.4 Distribution in the corpus 40
2.5 Using percentages and normalising 41
2.6 Representativity 43
2.7 Corpus annotation 44
2.8 Summary 50
Study questions 50
Corpus exercises 50
Further reading 51

3 Looking for lexis 52
3.1 The role of the lexicon in language 52
3.2 How lexicographers use corpora 52
3.3 The meaning of words 53
3.4 Semantic preference, semantic prosody and evaluation 58
3.5 How words change in frequency over time 60
3.6 How words spread between varieties of English 63
3.7 How authors use words 66
3.8 Summary 69
Study questions 70
Corpus exercises 70
Further reading 71

4 Checking collocations 72
4.1 Two types of collocations 72
4.2 Collocations in a window 74
4.3 Adjacent collocations 79
4.4 Summary 88
Study questions 88
Corpus exercises 89
Further reading 89

5 Finding phrases 90
5.1 Phraseology 90
5.2 Idioms 92
5.3 Recurrent phrases 96
5.4 A literary application: Dickens's recurrent long phrases 107
5.5 Summary 108
Study questions 109
Corpus exercises 109
Further reading 109

6 Metaphor and metonymy 111
6.1 Introduction 111
6.2 Using corpora in the study of metaphor 119
6.3 Using corpora in the study of metonymy 129
6.4 Summary 134
Study questions 135
Corpus exercises 135
Further reading 135

7 Grammar 137
7.1 Introduction 137
7.2 *Who* and *whom* 137
7.3 *Get*-passives 140
7.4 Adjective complementation 143
7.5 Prepositional gerund or directly linked gerund 145
7.6 Using a parsed corpus: passives revisited 153
7.7 Summary 154
Study questions 154
Corpus exercises 155
Further reading 155

8 Male and female 156
8.1 Introduction 156
8.2 Referring to men and women 157
8.3 The way men and women use language 165
8.4 Summary 171
Study questions 172
Corpus exercises 172
Further reading 172

9 Language change 174
9.1 Introduction 174
9.2 What is likely happening with *likely* 176
9.3 Grammaticalisation: The history of *beside*(*s*) 180
9.4 The *OED* as corpus: Starting to say *start to* and *start* V-*ing* 188
9.5 Sociolinguistic explanations of language change: The rise of third person singular -*s* 191
9.6 Summary 193
Study questions 193
Corpus exercises 194
Further reading 194

10 Corpus linguistics in cyberspace 195
10.1 Introduction 195
10.2 The web as corpus 196
10.3 Using commercial search engines for linguistic research 199
10.4 Regional variation: Agreement with collective nouns 201
10.5 Grammar: Adjective comparison 203
10.6 Dialect and non-standard language 206

10.7 Web genres and compiling corpora from the web 208
10.8 Summary 216
Study questions 217
Corpus exercises 217
Further reading 217

References 219
Index 232