Contents

List of figures *page* x
List of tables xi
Acknowledgements xii
Preface xiii

1 What is corpus linguistics? 1
1.1 Introduction 1
1.2 Mode of communication 3
1.3 Corpus-based versus corpus-driven linguistics 5
1.4 Data collection regimes 6
1.5 Annotated versus unannotated corpora 13
1.6 Total accountability versus data selection 14
1.7 Monolingual versus multilingual corpora 18
1.8 Summary 21
Further reading 21
Practical activities 22
Questions for discussion 23

2 Accessing and analysing corpus data 25
2.1 Introduction 25
2.2 Are corpora the answer to all research questions in linguistics? 27
2.3 Corpus annotation 29
2.4 Introducing concordances 35
2.5 A historical overview of corpus analysis tools 37
2.6 Statistics in corpus linguistics 48
2.7 Summary 53
Further reading 54
Practical activities 55
Questions for discussion 55

3 The web, laws and ethics 57
3.1 Introduction 57
3.2 The web and legal issues 57
3.3 Ethical issues 60
3.4 Summary 69
Further reading 69
Practical activity 70
Questions for discussion 70

4 English Corpus Linguistics 71
4.1 Introduction 71
4.2 University College London (UCL) 74
4.3 Lancaster University 76
4.4 University of Birmingham 79
4.5 Université Catholique de Louvain 81
4.6 University of Nottingham 84
4.7 Northern Arizona University and the USA 88
4.8 Summary 91
Further reading 91
Practical activities 92
Questions for discussion 92

5 Corpus-based studies of synchronic and diachronic variation 94
5.1 Introduction 94
5.2 Diachronic change from Old English to Modern English 94
5.3 Diachronic variation in contemporary Modern English 96
5.4 The multi-dimensional approach to variation 104
5.5 Corpora and variationist sociolinguistics 115
5.6 Summary 118
Further reading 119
Practical activities 119
Questions for discussion 120

6 Neo-Firthian corpus linguistics 122
6.1 Introduction 122
6.2 Collocation 122
6.3 Discourse 133
6.4 Semantic prosody and semantic preference 135
6.5 Lexis and grammar 142
6.6 Corpus-as-theory versus corpus-as-method 147
6.7 Summary: Sinclair's contribution to corpus linguistics 162
Further reading 164
Practical activities 164
Questions for discussion 165

7 Corpus methods and functionalist linguistics 167
7.1 Introduction 167
7.2 Functionalism in linguistics: a brief overview 168
7.3 Corpus-based research from a functionalist perspective 171
7.4 Corpora and typology 176
7.5 Corpora and cognitive approaches to linguistics 179
7.6 Corpora in the analysis of metaphor 185
7.7 Summary 188
Further reading 189
Practical activities 189
Questions for discussion 191

8 The convergence of corpus linguistics, psycholinguistics and functionalist linguistics 192
8.1 Introduction 192
8.2 Corpus methods and psycholinguistics 193
8.3 The convergence of neo-Firthian corpus linguistics and functionalist linguistics 210
8.4 Summary 221
Further reading 222
Practical activities 223
Questions for discussion 224

9 Conclusion 225
9.1 Introduction 225
9.2 The story of corpus linguistics, from past to future 225
9.3 Revisiting old friends: computational linguistics 227
9.4 The textually mediated world: the humanities and social sciences 230
9.5 The challenge ahead: integrating corpora with new methods in linguistics 233
9.6 The final word 236

Glossary 238
Notes 254
References 259
Index 292