

# Contents

Acknowledgements	viii
List of Abbreviations	ix
1 Early Corpus Linguistics and the Chomskyan Revolution	1
1 Is Corpus Linguistics a Branch of Linguistics?	2
2 Early Corpus Linguistics	2
3 What Chomsky Said	4
4 Why is this Book being Written?	11
5 Through the Interregnum to the Revival of the Early Eighties	16
6 Conclusion	17
7 Study Questions	18
8 Further Reading	19
Notes	19
2 What is a Corpus and What is in It?	21
1 Corpora vs. Machine-readable Texts	21
2 Text Encoding and Annotation	24
3 Multilingual Corpora	57
4 Chapter Summary	59
5 Study Questions	60
6 Further Reading	60
Notes	60
3 Quantitative Data	61
1 Introduction	61
2 Qualitative vs. Quantitative Analysis	62
3 Corpus Representativeness	63
4 Approaching Quantitative Data	66
5 Chapter Summary	84

6 Study Questions	84
7 Further Reading	85
Notes	86
4 The Use of Corpora in Language Studies	87
1 Corpora as Sources of Empirical Data	87
2 Corpora in Speech Research	88
3 Corpora in Lexical Studies	90
4 Corpora and Grammar	93
5 Corpora and Semantics	96
6 Corpora in Pragmatics and Discourse Analysis	98
7 Corpora and Sociolinguistics	99
8 Corpora and Stylistics	101
9 Corpora in the Teaching of Languages and Linguistics	103
10 Corpora in Historical Linguistics	106
11 Corpora in Dialectology and Variation Studies	109
12 Corpora in Psycholinguistics	111
13 Corpora and Cultural Studies	112
14 Corpora and Social Psychology	113
15 Chapter Summary	114
16 Study Questions	115
17 Further Reading	115
Notes	116
5 Corpora and Computational Linguistics	117
1 Introduction	117
2 What have Corpora got to Offer?	118
3 Part-of-speech Analysis	119
4 Automated Lexicography	127
5 Corpora and Lexicography	129
6 Parsing	129
7 Machine Translation	136
8 Conclusion	139
9 Study Questions	140
10 Further Reading	144
Notes	144
6 A Case Study: Sublanguages	147
1 Formulation of our Hypothesis: Sublanguages	148
2 Choosing the Corpora: The IBM, APHB and Hansard Corpora	149
3 Processing the Corpora: Corpus Annotation and Manipulation	152
4 Limiting Claims: What are we Looking At?	153
5 Beginning Work: The Study	155

6 The First Study of Closure: Lexical Closure	155
7 The Second Study of Closure: Part-of-Speech Closure	158
8 The Third Study of Closure: Parsing Closure	162
9 Drawing Conclusions: Findings	165
10 Study Questions	166
11 Further Reading	166
Notes	167
7 Where to Now?	169
1 The Past and the Present	169
2 The Future	170
3 Conclusion	175
Notes	175
Glossary	177
Appendix A: Corpora Mentioned in the Text	181
Appendix B: Some Software for Corpus Research	189
Appendix C: Suggested Solutions to Exercises	193
Bibliography	197
Index	207