## Contents

Acknowledgements			viii	
List of Abbreviations				
1	Earl	y Corpus Linguistics and the Chomskyan Revolution	1	
_		Is Corpus Linguistics a Branch of Linguistics?	2	
		Early Corpus Linguistics	2	
		What Chomsky Said	5	
		Why is this Book being Written?	13	
		Corpus Linguistics from the 1950s to the early 1980s	20	
		Conclusion	24	
		Study Questions	25	
		Further Reading	25	
		Notes	26	
2	W/b	at is a Corpus and What is in It?	29	
		Corpora vs. Machine-readable Texts	29	
		Text Encoding and Annotation	32	
		Multilingual Corpora	69	
		Chapter Summary	73	
		Study Questions	73	
		Further Reading	74	
		Notes	74	
3	3 Quantitative Data		75	
	-	Introduction	75	
		Qualitative vs. Quantitative Analysis	76	
		Corpus Representativeness	77	
		Approaching Quantitative Data	81	
		Chapter Summary	98	

	3.6	Study Questions	99
	3.7	•	100
		Notes	100
4		Use of Corpora in Language Studies	103
	4.1	Corpora as Sources of Empirical Data	103
	4.2	Corpora in Speech Research	104
	4.3	Corpora in Lexical Studies	106
	4.4	Corpora and Grammar	109
	4.5	Corpora and Semantics	112
	4.6	Corpora in Pragmatics and Discourse Analysis	114
	4.7	Corpora and Sociolinguistics	115
	4.8	Corpora in Stylistics and Text Linguistics	117
	4.9	Corpora in the Teaching of Languages and Linguistics	119
	4.10	Corpora in Historical Linguistics	123
	4.11	Corpora in Dialectology and Variation Studies	125
	4.12	Corpora in Psycholinguistics	127
	4.13	Corpora and Cultural Studies	128
	4.14	Corpora and Social Psychology	129
	4.15	Chapter Summary	130
	4.16	Study Questions	131
	4.17	Further Reading	131
		Notes	132
_	Comm	one and Language Engineering	133
3	5.1	oora and Language Engineering Introduction	133
	5.2		133
		What have Corpora got to Offer?	
	5.3	Part-of-speech Analysis	135
	5.4	Automated Lexicography	142
	5.5	Corpora and Lexicography	144
	5.6 5.7	Parsing Multilianual Commun Familiation	145
		Multilingual Corpus Exploitation	151
	5.8	Conclusion	158
	5.9	Study Questions  Final on P. and in a	158
	5.10	Further Reading	163
		Notes	163
6	A Ca	ise Study: Sublanguages	165
	6.1	Formulation of our Hypothesis: Sublanguages	166
	6.2	Choosing the Corpora: The IBM, APHB and Hansard Corpora	167
	6.3	Processing the Corpora: Corpus Annotation and Manipulation	170
	6.4	Limiting Claims: What are we Looking At?	171
	6.5	Beginning Work: The Study	173
		,	

6.6	The First Study of Closure: Lexical Closure	173			
6.7	The Second Study of Closure: Part-of-Speech Closure	176			
6.8	The Third Study of Closure: Parsing Closure	180			
6.9	Drawing Conclusions: Findings	184			
6.10	Study Questions	184			
6.11	Further Reading	185			
	Notes	185			
7 Loo	7 Looking Backwards, Looking Forwards				
7.1	The Future Past	187			
7.2	The Future to Come	190			
7.3	Conclusion	195			
	Notes	195			
Glossary					
Appendix A: Corpora Mentioned in the Text					
Appendix B: Some Software for Corpus Research Appendix C: Suggested Solutions to Exercises Bibliography					
			Index		