

Contents

Author's acknowledgements	ix
Publisher's acknowledgements	x
CHAPTER ONE: <i>Introduction</i>	1
1.1 Corpora	3
1.2 The role of computers in corpus linguistics	5
1.3 The scope of corpus linguistics	7
CHAPTER TWO: <i>The design and development of corpora</i>	13
2.1 Pre-electronic corpora	13
2.1.1 Biblical and literary	13
2.1.2 Lexicographical	14
2.1.3 Dialect	15
2.1.4 Language education	16
2.1.5 Grammatical	17
2.2 Types of electronic corpora	19
2.3 Major electronic corpora for linguistic research	23
2.3.1 First generation corpora	23
2.3.1.1 The <i>Brown Corpus</i>	23
2.3.1.2 The <i>Lancaster–Oslo/Bergen (LOB) Corpus</i>	27
2.3.1.3 Other first generation corpora modelled on the <i>Brown Corpus</i>	29
2.3.1.4 The <i>London–Lund Corpus (LLC)</i>	31
2.3.2 Corpora of English compiled for specialized purposes	33
2.3.2.1 Corpora for lexicography	33

2.3.2.2	Dictionaries as corpora	35
2.3.2.3	Corpora for studying spoken English	36
2.3.2.4	Diachronic corpora	38
2.3.2.5	Corpora for research on language acquisition	40
2.3.2.6	Other corpora for special purposes	43
2.3.3	Second generation mega-corpora	45
2.3.3.1	The <i>Cobuild</i> project	46
2.3.3.2	The <i>Longman Corpus Network</i>	48
2.3.3.3	The <i>British National Corpus (BNC)</i>	50
2.3.3.4	The <i>International Corpus of English (ICE)</i>	54
2.4	Electronic text databases	57
2.5	Issues in corpus design and compilation	60
2.5.1	Static or dynamic	60
2.5.2	Representativeness and balance	62
2.5.3	Size	66
2.6	Compiling a corpus	70
2.6.1	Corpus design	70
2.6.2	Planning a storage system and keeping records	75
2.6.3	Getting permission	76
2.6.4	Text capture	78
2.6.4.1	Written texts	78
2.6.4.2	Spoken texts	80
2.6.5	Markup	82
2.7	Organizations and professional associations concerned with corpus design, development and research	85
CHAPTER THREE: <i>Corpus-based descriptions of English</i>		88
3.1	Lexical description	91
3.1.1	Pre-electronic lexical description for pedagogical purposes	93
3.1.2	Computer corpus-based studies of the lexicon	97
3.1.3	Collocation	108
3.2	Grammatical studies centred on morphemes or words	121
3.2.1	Verb-form use for tense and aspect	122
3.2.2	Modals	130
3.2.3	Voice	134

3.2.4	Verb and particle use	137
3.2.5	Subjunctive	137
3.2.6	Prepositions: <i>of, at, from, between, through, by</i>	139
3.2.7	Conjunctions: <i>since, when, once</i>	144
3.2.8	<i>More and less</i>	152
3.3	Grammatical studies centred on the sentence	154
3.3.1	Sentence length	157
3.3.2	Syntactic processes	159
3.3.2.1	Clause patterning	159
3.3.2.2	Noun modification	162
3.3.2.3	Conditionality	165
3.3.2.4	Causation	167
3.3.2.5	Negation	171
3.3.2.6	Clefting	173
3.4	Pragmatics and spoken discourse	174
3.5	Corpus-based studies of variation in the use of English	180
3.5.1	Comparisons of spoken and written English	182
3.5.2	Comparisons of regional varieties of English	190
3.5.3	Variation in registers and genres	197
3.5.4	Studies of language change	199
CHAPTER FOUR: <i>Corpus analysis</i>		204
4.1	Corpus annotation and processing	206
4.1.1	Lemmatization	206
4.1.2	Word-class tagging	209
4.1.3	Semantic aspects of tagging	225
4.1.4	Parsing	231
4.2	Procedures used in corpus analysis	244
4.2.1	Word lists	245
4.2.2	Concordances	247
4.2.3	Statistics in corpus analysis	258
4.3	Corpus search and retrieval software	259
4.3.1	The <i>Oxford Concordance Program (OCP)</i>	260
4.3.2	<i>WordCruncher</i>	261
4.3.3	<i>TACT</i>	262
4.3.4	Other widely used software for special purposes	263
4.3.5	New generation software	265

CONTENTS

CHAPTER FIVE: <i>Implications and applications of corpus-based analysis</i>	268
5.1 Goals of linguistic description and the effect of corpora on methodology	270
5.1.1 Language as possibility and language as probability	270
5.1.2 The description of English	273
5.2 Corpus linguistics and computational linguistics	276
5.3 Corpus-based approaches to language teaching	280
5.3.1 The content of language teaching	282
5.3.2 Language teaching methodology	288
References	295
Index	310