Contents

1 Introduction — 1

2 Linguistic fundamentals of anaphors and anaphora — 3
2.1 Basic definitions — 3
2.2 The relationship between anaphor and antecedent — 6
2.2.1 Coreference — 6
2.2.2 Substitution — 7
2.2.3 Comparison of coreference and substitution — 8
2.2.4 Sense relations — 10
2.3 Related concepts: cataphora, endophora, exophora and deixis — 11
2.3.1 Cataphora — 11
2.3.2 Endophora and exophora — 12
2.3.3 Deixis — 13
2.4 Anaphors as cohesive devices in texts — 16
2.4.1 Texts and their features — 16
2.4.2 Cohesion — 18
2.4.3 Cohesive devices as a form of reduction — 20
2.5 Anaphors in the present book — 22

3 Types of anaphors — 23
3.1 Central pronouns — 24
3.1.1 Personal pronouns — 25
3.1.1.1 Subjective and objective forms — 25
3.1.1.2 Person, number and gender — 25
3.1.1.3 Anaphoric and non-anaphoric use — 28
3.1.1.4 Cataphoric use — 32
3.1.1.5 Relationship between anaphor and antecedent — 34
3.1.1.6 Summary — 35
3.1.2 Possessive pronouns — 35
3.1.2.1 Determinative and independent possessive pronouns — 35
3.1.2.2 Person, number and gender — 37
3.1.2.3 Anaphoric/cataphoric and non-anaphoric use — 38
3.1.2.4 Relationship between anaphor and antecedent — 39
3.1.2.5 Summary — 40
3.1.3 Reflexive pronouns — 40
3.1.3.1 Basic and emphatic use — 41
3.1.3.2 Person, number and gender — 42
3.1.3.3 Anaphoric/cataphoric and non-anaphoric use — 43
3.1.3.4 Summary — 44
3.1.4 Summary of personal, possessive and reflexive pronouns — 44
3.2 Reciprocal pronouns — 52
3.2.1 Compound and split construction — 53
3.2.2 Anaphoric/cataphoric and non-anaphoric use — 53
3.2.3 Summary — 54
3.3 Demonstrative pronouns — 55
3.3.1 Dependent and independent demonstrative pronouns — 55
3.3.2 Grammatical and referential aspects — 56
3.3.3 Anaphoric/cataphoric and non-anaphoric use — 57
3.3.4 Relationship between anaphor and antecedent — 59
3.3.5 Summary — 60
3.4 Relative pronouns — 62
3.4.1 Form and function — 62
3.4.2 Types of clauses and their anaphoric and non-anaphoric use — 63
3.4.3 Further non-anaphoric uses — 64
3.4.4 Gender and case — 65
3.4.5 Summary — 67
3.5 Adverbs — 68
3.5.1 Here and there — 68
3.5.2 Now and then — 70
3.5.3 Where, when, while and why — 71
3.5.4 Summary — 73
3.6 Noun phrases with a definite article — 75
3.6.1 Syntactic characteristics — 76
3.6.2 Anaphoric/cataphoric and non-anaphoric use — 76
3.6.3 The relationship between anaphor and antecedent — 78
3.6.4 Summary — 79
3.7 Proper names — 80
3.7.1 Proper names and proper nouns — 80
3.7.2 Syntactic features — 81
3.7.3 Anaphoric and non-anaphoric use — 81
3.7.4 Summary — 83
3.8 Indefinite pronouns — 83
3.8.1 One and ones — 84
3.8.2 Other, others and another — 85
3.8.3 Both, all and each — 86
3.12.3.1 Anaphoric/cataphoric and non-anaphoric use with regard to clause and phrase functions — 122
3.12.3.2 Further non-anaphoric uses — 123
3.12.3.3 Summary — 123
3.12.4 Summary of to-infinitive, -ing-participle and -ed-participle clauses — 124
3.13 Conclusion — 126

4 Anaphors in hypertexts — 131
4.1 Previous research — 131
4.2 Hypertexts — 133
4.2.1 What are hypertexts? — 133
4.2.2 Cohesion and coherence in hypertexts — 136
4.2.3 Linguistic characteristics of hypertexts — 137
4.2.4 Classifications of hypertexts — 139
4.3 Corpus of hypertexts — 142
4.3.1 Corpus design — 142
4.3.2 “Wikipedia” — 143
4.3.3 “Blogs” — 147
4.3.4 “(Traditional) Websites” — 151
4.4 Results of the corpus analysis — 154
4.4.1 Frequency of anaphors in the corpus — 154
4.4.1.1 Distribution of the twelve anaphor types — 154
4.4.1.2 Central pronouns — 161
4.4.1.3 Reciprocal pronouns — 164
4.4.1.4 Demonstrative pronouns — 165
4.4.1.5 Relative pronouns — 167
4.4.1.6 Adverbs — 168
4.4.1.7 Proper names — 170
4.4.1.8 Indefinite pronouns — 170
4.4.1.9 Other forms of coreference and substitution — 172
4.4.1.10 Verb phrases with do and combinations — 172
4.4.1.11 Ellipses — 174
4.4.1.12 Non-finite clauses — 175
4.4.2 Ratio of items occurring anaphorically and non-anaphorically — 178
4.5 Conclusion — 183
5 Text retrieval and its handling of anaphors — 185
5.1 What are text retrieval systems? — 185
5.2 Text retrieval models — 186
5.3 Evaluation of text retrieval results — 190
5.4 Natural language processing methods in text retrieval — 193
5.4.1 Sentence delimitation and tokenisation — 194
5.4.2 Stop word detection — 195
5.4.3 Stemming and lemmatisation — 197
5.4.4 Part-of-speech tagging — 200
5.4.5 Parsing — 201
5.4.6 Natural language processing on the Web — 202
5.5 Text retrieval and the Internet — 203
5.5.1 Crawling — 203
5.5.2 Indexing — 204
5.5.2.1 Processes of feature extraction — 205
5.5.2.2 Example of a feature extraction — 207
5.5.2.3 Index — 209
5.5.3 Queries and search results — 211
5.6 Conclusion — 214

6 Approaches to and uses of anaphora resolution — 217
6.1 Uses of anaphora resolution systems — 217
6.1.1 Machine translation — 218
6.1.2 Information extraction — 220
6.1.3 Question answering — 221
6.1.4 Text summarisation — 223
6.1.5 Text retrieval — 224
6.2 Structure of anaphora resolution systems — 226
6.3 Anaphora resolution approaches — 228
6.3.1 Rule-based approaches — 229
6.3.2 Data-based approaches — 233
6.3.3 Comparison of anaphora resolution approaches — 235
6.4 Anaphora resolution in text retrieval — 237
6.5 Evaluation of anaphora resolution systems — 238
6.5.1 Measures for evaluation — 238
6.5.2 Comparing different algorithms — 240
6.5.3 Annotated corpora for evaluation — 242
6.5.3.1 Current annotated corpora — 242
6.5.3.2 An annotated corpus for all anaphor types — 244
6.6 Conclusion — 245
Development of extensive linguistic rules for anaphora resolution: the example of non-finite clause anaphors — 247

7.1 Identifying anaphoric items — 247
7.1.1 Rules for -ing-items and their evaluation — 250
7.1.2 Rules for -ed-items and their evaluation — 253
7.1.3 Rules for to-items and their evaluation — 254

7.2 Identifying antecedents — 256
7.2.1 Rules for -ing and their evaluation — 257
7.2.2 Rules for -ed and their evaluation — 266
7.2.3 Rules for to and their evaluation — 268

7.3 Conclusion — 272

8 Conclusion — 275

Bibliography — 281

Index — 297