

# Table of contents

Acknowledgements	XIII
Introduction	1
0.1 Discourse and discourse analysis	2
0.2 Corpus linguistics	5
0.2.1 What it is and what it does	5
0.2.2 Quantity, frequency, comparison and recurrence (or patterning)	8
0.2.3 Serendipity	9
0.3 Corpus-assisted discourse studies or CADS	10
0.3.1 Definition and aims	10
0.3.2 A comparison between traditional corpus linguistics and CADS	11
0.4 The corpora and tools for analysing corpora	14
0.4.1 The corpora	14
0.4.2 Corpus annotation	16
0.4.3 Tools for analysing corpora	17
0.5 Guide to the contents of this book	20
CHAPTER 1	
The two principles of discourse organisation: Chunk recall and inductive reasoning	25
1.1 Introduction	25
1.2 Grammatical organisation	26
1.2.1 Open choice	26
1.2.2 The idiom principle and coselection	26
1.2.3 Issues with idioms	29
1.3 Script theory	30
1.3.1 Learning and memory	30
1.3.2 Understanding discourse	32
1.4 Inductive knowledge-driven reasoning	34
1.4.1 Needs, goals and plans	34
1.5 Parallels	35
1.5.1 Open choice and logical induction: <i>Rule-driven</i> behaviour	35
1.5.2 The idiom and the script principles: Lexical priming	36
1.6 Conclusion	40

## CHAPTER 2

## Evaluation in discourse communication

43

- 2.1 For good and for bad, for better and for worse 43
- 2.2 Point of view 46
- 2.3 Evaluation working in discourse 47
- 2.4 Categories of evaluative lexis 52
- 2.5 Note: The evaluator and evaluative voices 54
- 2.6 Evaluation and cohesion; evaluative consistency or *harmony* 55
- 2.7 Evaluative prosody 58
- 2.8 Embedding and nesting 61
- 2.9 Conclusion 63
  - Further Research 64

## CHAPTER 3

## Evaluation and control

67

- 3.1 Control: The linguistic unit 67
- 3.2 Control and power relations 71
- 3.3 The control feature and evaluative prosody: Examples 73
  - 3.3.1 *Set in* 73
  - 3.3.2 *Sit through* 75
  - 3.3.3 *Undergo* 77
  - 3.3.4 *Budge* 78
  - 3.3.5 *Persistence/Persistent* 80
  - 3.3.6 *Break out* 81
  - 3.3.7 *Outbreak* 82
  - 3.3.8 *End up* 84
  - 3.3.9 *Cause* 84
  - 3.3.10 *Fuel* 86
  - 3.3.11 *Fickle and flexible* 88
  - 3.3.12 *Orchestrate* 89
  - 3.3.13 *True feelings* 90
- 3.4 Conclusions 94

## CHAPTER 4

## Investigating rhetoric in discourse I: Irony

97

- 4.1 Irony explicit and implicit 97
- 4.2 Suitability of data 98
- 4.3 Case study I: Explicit irony 99
  - 4.3.1 What is irony? Ask the people 99
  - 4.3.2 The evaluator 102
  - 4.3.3 Reversal of evaluation 103

4.4	Case study 2: Implicit irony	105
4.4.1	Using corpus techniques to find episodes of implicit irony	105
4.4.2	Reversal of evaluation in implicit irony	106
4.4.3	Verisimilar ironies: Litotes	107
4.4.4	Irony in questions	109
4.5	Conclusions on explicit and implicit irony	110
4.6	Case study 3: The form, function and exploitation of phrasal irony	112
4.6.1	Evaluative clash with the phrase	112
4.6.2	Evaluative oxymoron	113
4.6.3	Substitution by evaluative opposite in well-known phrases	115
4.6.4	The “popularisation” of the ironic usage of a phrase	116
4.6.5	Replacing an expected negative element of the template with something positive	116
4.6.6	Replacing an expected positive element of the template with something negative	119
4.6.7	How such ironic uses become popular	122
4.6.8	A final twist: When is evaluative reversal ironic clash and when simply a counter-instance?	123
4.6.9	Ratio, inherent hyperbole, critical intent	126
4.6.10	Conclusions on phrasal irony	127
	Suggestions for further research	128

## CHAPTER 5

	Investigating rhetoric in discourse 2: Metaphor	131
5.1	Corpus linguistics and metaphor: Methodologies	131
5.2	Corpus linguistics and metaphor: Challenges and potential pitfalls	132
5.3	Case study 1: Metaphors of <i>anti-Americanism</i>	133
5.3.1	Why analyse metaphor in this context?	133
5.3.2	Corpora and methodology	134
5.3.3	Results	135
5.3.4	Conclusions to case-study 1	145
5.4	Case study 2: Metaphor and humour in review articles	146
5.4.1	Metaphors and meaning potential	146
5.4.2	Humour and metaphor resources	152
5.4.3	Figurative language as part of the humorous style	154
5.4.4	Incongruous comparison	155
5.4.5	Intensification and hyperbole	159
5.5	Conclusions	162
	Suggestions for further reading	163
	Suggestions for further research	163

## CHAPTER 6

<b>Corpus-assisted stylistics: Investigating author style</b>	<b>165</b>
6.1 The comic prose of P.G. Wodehouse	168
6.2 The corpora	169
6.3 Formality – informality	170
6.4 Hyperbole and litotes	175
6.5 Playing with degrees of precision	176
6.6 Colourful imagery	177
6.7 Playing with co-occurrence	180
6.8 Conclusion	182
Appendix: The contents of the three literary corpora	185

## CHAPTER 7

<b>Cross-linguistic discourse analysis: Investigating the representation of migrants in the UK and Italian press</b>	<b>187</b>
7.1 Cross-cultural/cross-linguistic CADS	187
7.1.1 Previous research	187
7.1.2 What methodological challenges might the researcher face?	189
7.2 Representation of migrants in the Italian and UK press	191
7.3 Corpora	192
7.4 <i>Racism and xenophobia</i>	194
7.4.1 Racism	194
7.4.2 Xenophobia	198
7.5 RASIM and ICES geographical identities: Frequencies	200
7.5.1 UK data	201
7.5.2 Italian data	202
7.6 Conclusions	206
Suggestions for further research	206

## CHAPTER 8

<b>Interactive spoken discourse 1: Managing the message</b>	<b>209</b>
8.1 Introduction and review	209
8.2 The grammar of spoken discourse: Is it distinct from most forms of writing?	213
8.3 Studying institutional adversarial talk	216
8.4 White House press briefings	216
8.5 The Hutton Inquiry	217
8.6 Similarities and differences between the two discourse types	218
8.7 Asserting the administration's message, imposing primings in briefings	220

- 8.8 Repeated messages and forced primings in Hutton respondents' discourse 227
  - 8.8.1 Impersonal constructions 228
  - 8.8.2 Key nouns: Vagueness in reference 229
  - Suggestions for further research 232

## CHAPTER 9

## Interactive spoken discourse 2: CADS &amp; (im)politeness 239

- 9.1 Overview of corpus linguistics and (im)politeness 239
- 9.2 A case study: When "politeness" is not being polite 241
  - 9.2.1 Introduction 242
  - 9.2.2 Negative politeness 243
  - 9.2.3 The discourse context 245
  - 9.2.4 The corpora and corpus interrogation tools 246
  - 9.2.5 Identifying impoliteness 248
  - 9.2.6 Looking for meta-pragmatic comment 249
  - 9.2.7 Looking for shifts from transactional to interactional mode 252
  - 9.2.8 Two illustrative markers of negative politeness 254
    - 9.2.8.1 with (\*) respect 254
    - 9.2.8.2 Vocatives 257
  - 9.2.9 Mock politeness 259
- 9.3 Conclusions 261
  - Suggestions for further research 262

## CHAPTER 10

## Modern diachronic corpus-assisted discourse studies (MD-CADS) 1: Comparisons over time in lexical grammar and discourse practices 265

- 10.1 Comparing the 1993, 2005 and 2010 corpora: Corpus wordlists and keywords 267
- 10.2 The methodology of set identification: Evaluative lexical keywords 269
- 10.3 Language in the press: Patterns in the keywords list 270
- 10.4 Informalisation 271
- 10.5 Language in the press: Hyperbolic evaluation 272
- 10.6 The keyword sets 273
  - 10.6.1 Hyperbole and extremes in evaluation in the keywords 273
    - 10.6.1.1 Size and ranking and relative importance 274
    - 10.6.1.2 Positive and amplified evaluation 274
    - 10.6.1.3 Intensification and emphasis 275

- 10.6.2 Vagueness 276
- 10.6.3 Vague and informal evaluative lexis 277
- 10.7 Evaluative meanings in the keywords and diachronic conclusions 279

CHAPTER 11

**Modern diachronic corpus-assisted discourse studies (MD-CADS) 2:  
Comparisons over time of social, political and cultural issues 283**

- 11.1 Antisemitism: The longest hatred 283
  - 11.1.1 The statistical consistency of discourses around antisemitism 283
  - 11.1.2 A working definition of *antisemitism* 284
  - 11.1.3 Aversion and attribution 285
  - 11.1.4 Methodology 285
  - 11.1.5 Blending stretches containing duplicated text 286
  - 11.1.6 When not to remove duplicated text 288
  - 11.1.7 The procedure we adopted 288
  - 11.1.8 Looking for similarities across the datasets 289
  - 11.1.9 Discourses on *antisemitism* in AS93 290
  - 11.1.10 Discourses on *antisemitism* in 2005, 2009 and 2010 291
  - 11.1.11 The perpetrators 293
  - 11.1.12 "The global Jewish conspiracy" 297
  - 11.1.13 Differences in focus of the three newspapers 298
  - 11.1.14 Discussion and conclusions on *antisemitism* 300
- 11.2 Case study 2: Girls and boys in the UK press 301
  - 11.2.1 Why search for similarity? 301
  - 11.2.2 Classic ways of searching for difference in corpus linguistics 303
  - 11.2.3 Ways of searching for similarity 304
    - 11.2.3.1 Word comparison 305
    - 11.2.3.2 Consistency analysis 306
    - 11.2.3.3 Key keywords 307
    - 11.2.3.4 Consistent (or wide-distribution) collocates 308
    - 11.2.3.5 Lockwords 309
    - 11.2.3.6 Alternative keyword calculations 309
  - 11.2.4 Previous research into the use of gender terms 310
  - 11.2.5 Data analysis 310
    - 11.2.5.1 C-collocates 312
    - 11.2.5.2 Word Sketches 315
    - 11.2.5.3 Thesaurus 316
    - 11.2.5.4 C-clusters/c-ngrams 318
  - 11.2.6 Discussion and conclusions on *girl/s* and *boy/s* 319

- 11.3 Conclusion 320
- Suggestions for further research 321

## CHAPTER 12

- Conclusion 323
- 12.1 CADS and discourse theories 323
  - 12.1.1 Discourse organisation and the idiom/open choice principles 323
  - 12.1.2 Lexical priming 325
  - 12.1.3 Evaluation 326
- 12.2 The eclecticism of CADS research 328
- 12.3 Corpus-assisted discourse studies: More than the sum of discourse analysis + computing 331
  - 12.3.1 Adding value to discourse analysis: Keeping us honest and the “culture of the counterexample” 331
  - 12.3.2 Actively looking for counterexamples 335
- 12.4 CADS: From research to the teaching of discourse analysis 339

## Appendix: Resources 343

- 1. Corpus compilation 343
  - 1.1 BootCaT 343
- 2. Annotation and interrogation 343
  - 2.1 UAM Corpus Tool 343
- 3. Mark up 343
- 4. Corpus interrogation 344
  - 4.1 AntConc 344
  - 4.2 Concgram 344
  - 4.3 Sketch Engine 344
  - 4.4 Wordsmith Tools 344
  - 4.5 Wmatrix facilities 344
- 5. Publicly available corpora which are referred to in the book (alphabetical order) 345
  - 5.1 BASE 345
  - 5.2 BNC 345
  - 5.3 Brown Family Corpora 345
  - 5.4 Corpus of Contemporary American English (CoCA) 345
  - 5.5 CORD historical Corpus of English Dialogues 346
  - 5.6 Hong Kong Corpus of Conversational English (HKCCE) 346
  - 5.7 London-Lund 346

5.8	MICASE	346	
5.9	SiBol/Port	346	
6.	Selected resources for downloading data used in corpora employed in this book	346	
7.	Other resources	347	
	Appendix references	347	
	<b>References</b>		<b>349</b>
	<b>Author index</b>		<b>365</b>
	<b>Subject index</b>		<b>369</b>