

CORPUS STYLISTICS  
THEORY AND PRACTICE

Dan McIntyre and Brian Walker

EDINBURGH  
University Press

# Contents

List of Figures	x
List of Tables	xii
List of Boxes	xv
Acknowledgements	xvii
<b>1 Combining corpus linguistics and stylistics</b>	<b>1</b>
1.1 Introduction	1
1.2 What is a corpus?	3
1.3 A brief history of corpus linguistics	6
1.4 The emergence of corpus stylistics and problems with its definition	11
1.5 Style and stylistics	16
1.6 Corpus stylistics versus cognitive stylistics	17
1.7 Aims of the book	20
1.8 Structure of the book	21
<b>2 Using corpora to support qualitative stylistic analysis</b>	<b>25</b>
2.1 Introduction	25
2.2 Strolling with intent: an exercise in corpus-informed stylistics	27
2.3 Collocation and semantic prosody	38
2.3.1 Semantic prosody as one aspect of extended units of meaning	40
2.3.2 Semantic prosody versus semantic preference	44
2.4 Case study: irony in ‘Aftermyth of war’	49
2.5 Conclusion	61

<b>3</b>	<b>Constructing and annotating corpora</b>	<b>64</b>
3.1	Introduction	64
3.2	Designing a corpus: representativeness, sampling and balance	65
3.2.1	Representativeness	66
3.2.2	Sampling and balance	67
3.2.3	Sample size and more about balance	72
3.3	Building a corpus: collecting and handling data	76
3.3.1	Converting and cleaning data files	76
3.3.2	Splitting and merging corpus files	78
3.3.3	Copyright	82
3.3.4	Ethics	83
3.4	Adding metadata to a corpus: mark-up and annotation	86
3.4.1	The format of mark-up and annotation	89
3.4.2	Descriptive metadata: headers	92
3.4.3	Automatic annotation	93
3.4.4	Manual annotation	95
3.4.5	Mark-up and annotation summary	96
3.5	Summary: suggested steps to building a corpus	97
3.6	Case study: investigating the <i>2000AD</i> comic strip 'Judge Dredd'	98
3.6.1	Scope of the study	99
3.6.2	Building the corpora	100
3.6.3	Results of the preliminary analysis	104
3.6.4	Further analysis	105
3.6.5	Case study summary	107
3.7	Conclusion	108
<b>4</b>	<b>Testing hypotheses and answering research questions</b>	<b>111</b>
4.1	Introduction	111
4.2	Corpora in stylistics	112
4.3	The concept of corpus-based and corpus-driven linguistics	114
4.4	The value of intuition	119
4.5	Hypotheses	121
4.6	Objectives and research questions	123
4.7	How big was Dickens's vocabulary?	124
4.8	Objectivity in stylistics	126
4.9	Case study: Hemingway's authorial style	129
4.10	Conclusion	141

<b>5</b>	<b>Making sense of significance</b>	143
5.1	Introduction	143
5.2	Vocabulary size	146
5.3	Keyness	154
5.3.1	Log-likelihood	154
5.3.2	Log ratio	158
5.3.3	Log ratio as a mechanism for reducing keyness lists	162
5.4	Collocation	164
5.4.1	Mutual Information	165
5.4.2	T-score	167
5.5	Case study: salience and foregrounding in David Peace's <i>1974</i>	168
5.5.1	<i>1974</i>	169
5.5.2	Keyness for text selection	169
5.5.3	Foregrounding in <i>1974</i>	173
5.6	Conclusion	178
<b>6</b>	<b>Cognition, character and corpus stylistics</b>	181
6.1	Introduction	181
6.2	A cognitive stylistic approach to characterisation	182
6.3	Constraints on subtitles	186
6.4	Case study: subtitling and characterisation in <i>Tinker Tailor Soldier Spy</i>	189
6.4.1	Plot, character and critical plaudits	190
6.4.2	Data preparation	191
6.4.3	Quantitative and qualitative corpus stylistic analysis	194
6.5	Conclusion	205
<b>7</b>	<b>Pedagogical corpus stylistics</b>	208
7.1	Introduction	208
7.2	Stylistics in the L2/FL classroom	209
7.3	Using corpora in the classroom	211
7.4	Corpus activities in the classroom	214
7.5	Case study: from internal deviation to external norms via Adrian Henri's 'Tonight at Noon'	217
7.5.1	Stylistics inside out	217
7.5.2	'Tonight at Noon' by Adrian Henri	219
7.5.3	A stylistic analysis of 'Tonight at Noon'	220
7.5.4	Teaching the poem to learners of English as a foreign language	221

7.5.5	Guided analysis	222
7.5.6	Round-up of the analysis so far	232
7.5.7	Analysis: case study summary and conclusions	233
7.6	Conclusion	235
<b>8</b>	<b>Historical corpus stylistics</b>	<b>237</b>
8.1	Introduction	237
8.2	Studying style diachronically and synchronically	238
8.3	Historical corpora and tools	239
8.4	Case study: discourse presentation in Early Modern English	242
8.4.1	Discourse presentation	243
8.4.2	Building the corpus	245
8.4.3	Annotating the corpus	247
8.4.4	Quantitative results	249
8.4.5	Discussion	255
8.4.6	Summary remarks	257
8.5	Case study: Frequencies and functions of modal auxiliary verbs in Early Modern English journalism	258
8.5.1	Building a corpus of Early Modern English news texts	259
8.5.2	Annotating the corpus	260
8.5.3	Quantification and results	262
8.5.4	Statistical analysis	265
8.5.5	Further analysis and discussion	267
8.6	Conclusion	273
<b>9</b>	<b>Real-world applications for corpus stylistics</b>	<b>275</b>
9.1	Introduction	275
9.2	The impact of applied stylistics	276
9.3	Language Unlocked	277
9.4	Wmatrix	277
9.4.1	Checking key concepts and key POS categories	280
9.5	Case study: the self-presentational style of the UK Green Party	281
9.5.1	Establishing research questions	281
9.5.2	Methods and data	282
9.5.3	Positive key semantic domains in GP-POL-2010	283
9.5.4	Negative key semantic domains in GP-POL-2010	289
9.5.5	Keywords in GP-POL-2010	297

9.6	Implications of the research	306
9.7	Conclusion	307
<b>10</b>	<b>The scope of corpus stylistics</b>	<b>309</b>
10.1	Introduction	309
10.2	The problem of definition	311
10.3	Corpus linguistics and stylistics: a symbiotic relationship	313
10.4	Conclusion	315
	Appendix 1: Corpora for corpus stylistics	317
	Appendix 2: Software for corpus stylistics	322
	References	324
	Index	353