

Working with Specialized Language

A practical guide to using corpora

Lynne Bowker and Jennifer Pearson

Contents

<i>List of figures</i>	ix
<i>Acknowledgements</i>	xiv
Introduction	1
<i>How to use this book</i>	2
<i>Outline</i>	2
PART I	
Setting the scene	7
1 Introducing corpora and corpus analysis tools	9
<i>What is a corpus?</i>	9
<i>Who uses corpora?</i>	11
<i>Are there different types of corpora?</i>	11
<i>Are there tools for investigating corpora?</i>	13
<i>Why use corpora to investigate language?</i>	13
<i>Key points</i>	20
<i>Further reading</i>	21
<i>Exercises</i>	21
2 Introducing LSP	25
<i>What is LSP?</i>	25
<i>Who uses LSP?</i>	27
<i>Why learn an LSP?</i>	28
<i>How can a corpus help with LSP acquisition?</i>	31
<i>Key points</i>	39
<i>Further reading</i>	40
<i>Exercise</i>	40

PART II

Corpus design, compilation and processing	43
3 Designing a special purpose corpus	45
<i>Corpus size</i>	45
<i>Text extracts vs full texts</i>	49
<i>Number of texts</i>	49
<i>Medium</i>	49
<i>Subject</i>	50
<i>Text type</i>	50
<i>Authorship</i>	51
<i>Language</i>	51
<i>Publication date</i>	52
<i>Corpus design: an example</i>	52
<i>Key points</i>	53
<i>Further reading</i>	55
<i>Exercises</i>	55
4 Compiling a special purpose corpus	58
<i>Copyright and permissions</i>	59
<i>Electronic resources</i>	61
<i>Corpus compilation: an example</i>	68
<i>Key points</i>	73
<i>Further reading</i>	73
<i>Search engines and meta-search engines</i>	73
<i>Tutorials for effective web searching</i>	73
<i>Potential LSP resources for corpus building</i>	74
<i>Exercise</i>	74
5 Markup and annotation	75
<i>What is markup?</i>	75
<i>What is annotation?</i>	83
<i>Key points</i>	89
<i>Further reading</i>	90
<i>Addresses of tagsets discussed in this chapter</i>	90
<i>Exercises</i>	91
6 Bilingual and multilingual corpora: pre-processing, alignment and exploitation	92
<i>Bilingual and multilingual corpora</i>	92
<i>Why use a parallel corpus?</i>	93
<i>Who uses parallel corpora?</i>	94
<i>Creating your own parallel corpus</i>	95

<i>How does alignment work?</i>	100
<i>Using your parallel corpus</i>	103
<i>Key points</i>	107
<i>Further reading</i>	107
<i>Where to get information about corpus resources</i>	108
<i>Where to find out more about bilingual concordancers</i>	108
<i>Exercises</i>	108

7 Introduction to basic corpus processing tools 109

<i>Producing word lists</i>	109
<i>Concordances</i>	120
<i>Key points</i>	129
<i>Further reading</i>	130
<i>Concordancing software</i>	130
<i>Exercises</i>	130

PART III

Corpus-based applications in LSP 135

8 Building useful glossaries 137

<i>What is a glossary?</i>	137
<i>Why do you need to produce a glossary?</i>	138
<i>What about dictionaries and comparable documents?</i>	139
<i>Traditional methods of producing a glossary</i>	140
<i>Practical terminology work</i>	143
<i>Storing your data</i>	161
<i>Summary of glossary compilation</i>	162
<i>Key points</i>	164
<i>Further reading</i>	164
<i>Exercise</i>	164

9 Term extraction 165

<i>Monolingual term extraction</i>	166
<i>Bilingual term extraction</i>	171
<i>Using term extraction techniques with basic corpus analysis tools</i>	172
<i>Key points</i>	173
<i>Further reading</i>	174
<i>Some products containing term extraction tools</i>	174
<i>Exercises</i>	175

10 Using LSP corpora as a writing guide	177
<i>Creating a small corpus</i>	177
<i>Writing science</i>	178
<i>Writing a product review</i>	185
<i>Key points</i>	191
<i>Further reading</i>	192
<i>Exercise</i>	192
11 Using LSP corpora as a translation resource	193
<i>Using a parallel corpus as a translation resource</i>	193
<i>Using a monolingual corpus as a translation resource</i>	198
<i>Key points</i>	209
<i>Further reading</i>	209
<i>Exercise</i>	210
12 Other applications and future directions	211
<i>Creating LSP learner corpora</i>	211
<i>Using LSP corpora to identify and study neologisms</i>	213
<i>Using corpora to investigate metaphor in LSP</i>	215
<i>Customizing tools to meet LSP needs</i>	217
<i>Key points</i>	220
<i>Further reading</i>	221
<i>Exercises</i>	222
Appendix	225
<i>Tools</i>	225
<i>Resources</i>	226
 <i>Glossary</i>	 229
<i>Bibliography</i>	234
<i>Index</i>	239