

Table of Contents

List of figures and tables	x
Acknowledgements	xiii
1. Introduction	1
1.1 Book outline	2
1.2 How to use the DVD	6
2. Corpus linguistics and translation studies	7
2.1 A typology of translation-driven corpora	10
2.2 Corpus-based translation research	11
2.2.1 Regularities of translations	12
2.2.1.1 Simplification	14
2.2.1.2 Explicitation	16
2.2.1.3 Standardization	19
2.2.1.4 Translation of unique items	20
2.2.1.5 Untypical collocations	20
2.2.1.6 Interference	21
2.2.2 Regularities of translators	24
2.2.3 Regularities of languages	25
2.2.4 Learner translation corpora	28
2.2.5 Interpreting and multimodal corpora	30
2.3 Corpus-based translation teaching and learning	31
2.4 Computer-assisted translation and computational linguistics	32
2.5 Tasks	34
2.5.1 Experimenting with the TEC	34
2.5.2 Experimenting with COMPARA	35
2.5.3 Experimenting with the LTC	37
2.6 Further reading	39
3. Corpus design and acquisition	40
3.1 Corpus design	41
3.1.1 Size	42
3.1.2 Composition	44
3.1.3 Representativeness and comparability	45
3.1.4 Case study: the CEXI corpus	49
3.2 Corpus acquisition and copyright	52
3.3 Web corpora	55
3.3.1 The Web as corpus	57
3.3.2 The Web as a source of corpora	60
3.3.2.1 General Web corpora	62
3.3.2.2 Specialized Web corpora	64

3.4	Conclusions	68
3.5	Tasks	68
3.5.1	Corpus building project outline	68
3.5.2	Manual creation of a DIY monolingual corpus	69
3.5.3	Automatic creation of a DIY bilingual comparable corpus	71
3.6	Further reading	72
4.	Corpus encoding and annotation	74
4.1	Corpus-based translation studies and corpus annotation	76
4.2	Annotation for descriptive translation studies	78
4.2.1	Documentary information	84
4.2.2	Structural information	91
4.2.3	Text-linguistic information	93
4.3	Stand-off annotation	97
4.4	Conclusions	101
4.5	Tasks	101
4.5.1	Creating an XML TEI document	101
4.5.2	Adding a simple header	103
4.5.3	Marking-up text structure	104
4.5.4	Adding linguistic annotation	106
4.5.5	Indexing the corpus	107
4.5.6	Searching the corpus	108
4.6	Further reading	109
5.	Corpus tools and corpus analysis	110
5.1	Corpus creation and analysis tools	111
5.1.1	Text acquisition	111
5.1.2	Annotation	112
5.1.3	Corpus management and query systems	114
5.1.4	Data retrieval and display	115
5.2	Analysis of corpus data	117
5.2.1	Wordlists and basic statistics	117
5.2.2	Concordances	124
5.2.3	Collocations, clusters and clouds	130
5.2.4	Colligations and word profiles	135
5.2.5	Semantic associations	138
5.3	Conclusions	140
5.4	Tasks	141
5.4.1	Wordlists	141
5.4.2	Lists of lemmas	141
5.4.3	Keywords	142
5.4.4	Concordances	143
5.4.5	Collocations and clusters	144
5.4.6	Word profiles	145
5.5	Further reading and software	146

6. Creating multilingual corpora	149
6.1 Corpus acquisition	150
6.1.1 Comparable corpora	150
6.1.2 Parallel corpora	152
6.2 Alignment	154
6.2.1 Paragraphs and sentences	155
6.2.2 Approaches and tools	158
6.3 Case study: the OPUS corpus	166
6.4 Parallel corpora and translation memories	169
6.5 Alignment below sentence level	172
6.5.1 Alignment of comparable corpora	172
6.5.2 Word alignment	173
6.6 Tasks	174
6.6.1 Aligning a text pair	174
6.6.2 A parallel corpus of literary texts	175
6.6.3 Corpus creation checklist	178
6.7 Further reading and software	179
7. Using multilingual corpora	181
7.1 Comparable and parallel corpora	181
7.2 Display and analysis of parallel corpora	182
7.3 Case study: The Rushdie English-Italian parallel corpus	190
7.4 Case study: the OPUS Word alignment database	196
7.5 Multilingual corpora in translator training and practice	200
7.6 Tasks	202
7.6.1 Searching a parallel corpus of literary texts	202
7.6.2 Exploring the Europarl multilingual corpus	205
7.7 Further reading	205
8. Conclusions	206
References	209
Index	234