

Contents

Overviewing Important Aspects of the Last Twenty Years of Research in Comparable Corpora	1
Serge Sharoff, Reinhard Rapp and Pierre Zweigenbaum	
 Part I Compiling and Measuring Comparable Corpora	
Mining Parallel Documents Using Low Bandwidth and High Precision CLIR from the Heterogeneous Web	21
Simon Shi and Pascale Fung	
Automatic Comparable Web Corpora Collection and Bilingual Terminology Extraction for Specialized Dictionary Making	51
Antton Gurrutxaga, Igor Leturia, Xabier Saralegi and Iñaki San Vicente	
Statistical Comparability: Methodological Caveats	77
Reinhard Köhler	
Methods for Collection and Evaluation of Comparable Documents . . .	93
Monica Lestari Paramita, David Guthrie, Evangelos Kanoulas, Rob Gaizauskas, Paul Clough and Mark Sanderson	
Measuring the Distance Between Comparable Corpora Between Languages	113
Serge Sharoff	
Exploiting Comparable Corpora for Lexicon Extraction: Measuring and Improving Corpus Quality	131
Bo Li and Eric Gaussier	
Statistical Corpus and Language Comparison on Comparable Corpora	151
Thomas Eckart and Uwe Quasthoff	

Comparable Multilingual Patents as Large-Scale Parallel Corpora . . .	167
Bin Lu, Ka Po Chow and Benjamin K. Tsou	
 Part II Using Comparable Corpora	
Extracting Parallel Phrases from Comparable Data	191
Sanjika Hewavitharana and Stephan Vogel	
Exploiting Comparable Corpora	205
Dragos Stefan Munteanu and Daniel Marcu	
Paraphrase Detection in Monolingual Specialized/Lay Comparable Corpora	223
Louise Deléger, Bruno Cartoni and Pierre Zweigenbaum	
Information Network Construction and Alignment from Automatically Acquired Comparable Corpora	243
Heng Ji, Adam Lee and Wen-Pin Lin	
Bilingual Terminology Mining from Language for Special Purposes Comparable Corpora	265
Emmanuel Morin, Béatrice Daille and Emmanuel Prochasson	
The Place of Comparable Corpora in Providing Terminological Reference Information to Online Translators: A Strategic Framework	285
Kyo Kageura and Takeshi Abekawa	
Old Needs, New Solutions: Comparable Corpora for Language Professionals	303
Silvia Bernardini and Adriano Ferraresi	
Exploiting the Incomparability of Comparable Corpora for Contrastive Linguistics and Translation Studies	321
Stella Neumann and Silvia Hansen-Schirra	