Paul BAKER, Andrew HARDIE & Tony MCENERY, A
Glossary of Corpus Linguistics

Edinburgh University Press Ltd, 2006, 192 pages

François Maniez

Electronic version

URL: <http://journals.openedition.org/lexis/1319>

DOI: 10.4000/lexis.1319

ISSN: 1951-6215

Publisher

Université Jean Moulin - Lyon 3

Electronic reference

François Maniez, « Paul BAKER, Andrew HARDIE & Tony MCENERY, *A Glossary of Corpus Linguistics* », *Lexis* [Online], Book reviews, Online since 26 February 2007, connection on 24 September 2020. URL : <http://journals.openedition.org/lexis/1319> ; DOI : <https://doi.org/10.4000/lexis.1319>

This text was automatically generated on 24 September 2020.

Lexis is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

Paul BAKER, Andrew HARDIE & Tony MCENERY, *A Glossary of Corpus Linguistics*

Edinburgh University Press Ltd, 2006, 192 pages

François Maniez

REFERENCES

Paul Baker, Andrew Hardie & Tony McEnery

. Edinburgh University Press Ltd, 2006. ISBN : 978-0748620180, Prix :??€,192 pages

- 1 This paperback is the tenth in a series that includes glossaries of various subspecialties in the field of linguistics (Peter Truggill's *Glossary of Sociolinguistics* and Laurie Bauer's *Glossary of Morphology*, to name but two). The authors include Tony McEnery, who collaborated with Andrew Wilson on *Corpus Linguistics*, a reference work in the field, published in 1996 by Edinburgh University Press. The entries focus on the following broad groupings: important corpora, key technical terms, key linguistic terms, key statistical measures, key computer programs/retrieval systems and standards applied within the field of corpus linguistics.
- 2 The number of entries is not specified, but it is probably close to 500, with about three entries per page on average. No definition is longer than two pages, which is the size of the entries devoted to such core concepts as *collocation* and *concordance*. The glossary is preceded by a very useful list of more than 100 acronyms and followed by a bibliography containing more than 300 references.
- 3 The contents are quite heavily weighted on the side of online resources, whether they are corpora or retrieval tools (the book is quite comprehensive on the subject of part-of-speech tagging, for instance), and a bit lighter on the side of theoretical linguistics. It is however worthy of note (especially in light of his oft-quoted derogatory comments about corpus linguistics, which have been the subject of many threads on the Corpora

list in the past) that three of Noam Chomsky's books are included in the references, and that he is mentioned in the entries devoted to *competence-performance*, *corpus linguistics*, *empiricism* and *grammar*.

- 4 The book's forte is quite clearly the high number of useful online corpus treatment tools, many of which were unknown to this reviewer. However, some of the numerous online corpora that are mentioned (the *Guangzhou Petroleum English Corpus* or the *Thai English Learner Corpus* come to mind) actually refer to either comparatively small corpora or relatively specialized uses of language and seem to be rather obscure contributions to the field, at least judging from the number of online quotations retrieved for them on the Web's search engines.
- 5 Considering that the authors are all UK-based, it will come as no surprise that this glossary is generally British-centered: most of the corpora and bibliographical references that are mentioned (with the notable exception of Douglas Biber) concern work that was conducted within British universities or research centers. This bias probably explains the absence of such items as Smadja's Xtract collocation retrieval tool or Tact, one of the very first concordance software packages that was freely available on the Web. Another item that certainly deserved to be included is William Fletcher's *Phrases in English* (<http://pie.usna.edu/>), an online resource that allows the user to search the British National Corpus for various word and/or part-of-speech patterns (an entry is, however, devoted to Fletcher's other main contribution to the field of corpus linguistics: the online concordancer KwiCFinder). Another surprising omission is *stop words*, which at least deserved a mention as a synonym in the entry devoted to *function words*. Finally, since the LOB and Brown word frequency lists are mentioned in the entry devoted to the ICAME word lists, it would have been logical to mention their BNC equivalent on Adam Kilgarriff's home page at <http://www.kilgarriff.co.uk/bnc-readme.html>.
- 6 Apart from the UK, there are a few references to work conducted in the United States and Scandinavian countries. Researchers in corpus linguistics who specialize in the study of the French language will be pleased to see that some of their work is mentioned in such entries as *European Language Activity Network*, *French Interlanguage Database*, *GlossaNet*, *MULTEXT* tools or *Trésor de la Langue Française Informatisé*. A number of European projects are also described, notably in such entries as those devoted to the *European Corpus Initiative*, *European Network of Excellence in Human Language Technologies* and *European Science Foundation Second Language Databank*.
- 7 One of the subfields that seems to have been left relatively unexplored in the glossary is that of terminology extraction, which is defined in six lines, without a single example of related software. Likewise, the entry devoted to *term bank* only mentions Eurodicautom, but not such major realizations as Termium or the Canadian *Grand Dictionnaire* (another Canadian resource that perhaps deserved to be included is the Hansard English-French bilingual corpus, all the more so as there is an entry for the *Canadian Hansard Treebank*).
- 8 Since the book is a glossary, its entries are quite logically sorted in alphabetical order, but after reading it from cover to cover, this reviewer still wonders if that was indeed the best possible way to present the wealth of information that it contains. The dictionary format is obviously useful for anybody who wants a quick definition of a linguistic concept like *lexical density* or of more technical terms like *granularity*, *mutual information* or *robust*, but on the other hand, if one is interested in a particular corpus,

looking it up on line will clearly provide a lot more information, and the abundance of entries referring to online resources may discourage the reader who is interested in concepts or notions from perusing the book sequentially. Bearing this in mind, there might have been a case for separating the book into a theoretical linguistics glossary (possibly subdivided into general notions and statistical measures) and a list of online resources, which could itself be divided into corpora and corpus treatment tools (a further division might even have been that between free and paying access, a useful distinction for researchers who work in under-subsidized university systems).

- 9 In spite of the various criticisms made above, this glossary will definitely be useful to intermediate or advanced students in corpus linguistics, as it provides them with brief, clear definitions for most of the commonly used terms in the field. As for other researchers, they may be pleased to discover a fair amount of online resources of which they were unaware.

AUTHORS

FRANÇOIS MANIEZ

François Maniez, Department of Applied Languages, University Louis Lumière-Lyon 2, France. Pr. François MANIEZ has been teaching translation, terminology and corpus linguistics at the University Louis Lumière-Lyon 2 since 1990. He is the head of the Master's Degree Applied Languages Program and director of the *International Communication in the Health Sciences* Master's Degree program. He is also a member of the CRTT (Terminology and Translation Research Center) and of ATALA (Natural Language Processing Association). His research interests include lexicology, terminology, translation studies and corpus linguistics applied to English for Specific Purposes, especially in the medical field.