Book and Software Reviews / Critiques de livres et de logiciels

McEnery, Tony, Xiao, Richard, & Tonio, Yukio. (2006). *Corpus-Based Language Studies: An Advanced Resource Book*. New York: Routledge. Pp. xx, 386, US$33.95 (paper).

The corpus-driven revolution in applied linguistics continues apace, and along with it the paradox that while corpora are changing the face of applied linguistics (most dictionaries, grammars, and course books now claim to be corpus based), this is occurring largely without the participation of practitioners. Only a few teachers or researchers have ever built a corpus or delved through concordance lines. Possibly in a bid to remedy this hands-off problem, the book display at the recent conference of the American Association of Applied Linguists (AAAL) in California in spring 2007 offered roughly a dozen books promising to enlist language teachers and graduate students in the revolution at a more basic level. The two I walked off with were the volume under review here and O'Keeffe, McCarthy, and Carter's (2007) *From Corpus to Classroom*.

It quickly became apparent that the latter volume was basically a set of conclusions about how corpus evidence can inform teaching practice, based on insights drawn from the impressive Cambridge and Nottingham CANCODE corpus of written and spoken British English, but that there would be no hands-on opportunities to construct further insights of one's own from this corpus, owing to its total inaccessibility to the broader professional community. So it was a relief to find that access and hands-on were main themes in the other volume.

McEnery, Xiao, and Tonio's *Corpus-Based Language Studies* is one in a new series by Routledge/Taylor & Francis dealing with themes in applied linguistics (intercultural communication, translation, grammar and context, and SLA) and bearing the common subtitle 'an advanced resource book.' Access and support for hands-on projects are main themes of the series and are arguably key to a

new-approach show of force from a publisher (the new owners of Lawrence Erlbaum) bidding to enter the language education field as competition for Cambridge University Press. So how accessible and hands-on does the McEnery et al. volume get?

The series format involves presenting first concepts and procedures, then related excerpts from actual research articles, and finally fully worked example research projects that can be replicated. Topics include corpus types, corpus building, corpus tagging (adding part-of-speech, or POS, markers), corpus statistics, corpus controversies, and corpus analysis for various purposes. Research article excerpts include a thoughtful and relevant collection of true area classics by Widdowson, Biber, Stubbs, Carter, McCarthy, and others. Research projects unpacked in detail include a methodology for corpus-based lexicography; a study in the sociolinguistics of British swearing, made possible by the huge and finely subdivided British National Corpus (BNC); and a replication using the Longman Learner's Corpus of the empirical morpheme acquisition sequences studies of Krashen and others in the 1970s—to mention fewer than half of those available. No other volume that I know of has gone so far to bring corpus research into focus for practitioners, or to make at least the beginning of a serious research project actually possible.

But there are probably limits to how accessible such an involved methodology can actually be made. For one thing, many of the book's analyses involve the use of Mike Smith's text-analysis program Wordsmith, which must be purchased, to get beyond the demo limitations, for about CAD$100. There are, of course, Web concordancers, and, indeed, the swearing study mentioned above is based on the University of Zurich's Web concordancer, which apparently runs the full BNC. Getting onto this site requires a password, however, which the book does not actually mention; when I e-mailed the site administrator, I received a irritated reply complaining about the number of users this book was sending them, as well as a reluctantly given username and password, which did not work).

The biggest access problem in this area, of course, is not software (several excellent free concordancers are available for download, such as Lawrence Anthony's *AntConc*) but, rather, corpora, such as the Japanese EFL division of the Longman Learner Corpus that was used in the morpheme study already mentioned. That study begins with a casual mention that this now rather dated corpus is publicly available and launches straight into a description of some complex operations that must be performed to tag and otherwise bring it up to date for a

modern treatment. But where to get the corpus is not mentioned. There are two companion Web sites for the book (the publisher's and the authors'), so perhaps it can be downloaded from one of these? These sites do offer corpus materials specifically linked to some of the projects in the book, but, in the case of the Longman Learner Corpus, after 20 minutes of scrolling and clicking I finally managed to get merely a link to the Longman site (which was 'under construction'). Readily available, however, is a tool for processing this corpus in some way that the analysis requires, should the reader find a way to obtain the corpus. So no one should buy this book expecting to start a complex research project tomorrow morning. Some preparatory digging will still be required.

Nor should anyone buy this book expecting to find awareness-raising or other hands-on activities for their language learners. The book is about doing research, not about developing teaching materials. There is, admittedly, a smallish section on pedagogical uses of corpora, which raises an interesting quibble with the authors, and with which I will end. The book throughout promotes the benefits of using tagged corpora for every purpose; for instance, the difference between *bank_VERB* and *bank_NOUN* is coded directly in the corpus so that an alphabetical sorting of these homonyms will keep them distinct. It is probably important for research concerns that the corpus be pre-parsed in this and many other ways, but the point of corpus work for learners is that they deal with the raw word form (as they must in real life) and use the context to decide for themselves which *bank* is involved in a particular utterance. I have always believed that applied linguists in the tradition of this volume do not give adequate place to the uses of an untagged or flat corpus as a constructivist learning tool. Thus, it is satisfying that none of the pedagogical studies cited appears to have involved learners' working with a tagged corpus!

Bravo on a fine book, one that gives more access than ever before to a fascinating and important set of research methods and questions.

Reference

O'Keeffe, A., McCarthy, M., & Carter, R. (2007). *From corpus to classroom: Language use and language teaching*. Cambridge: Cambridge University Press.

Tom Cobb, *Université du Québec à Montréal*