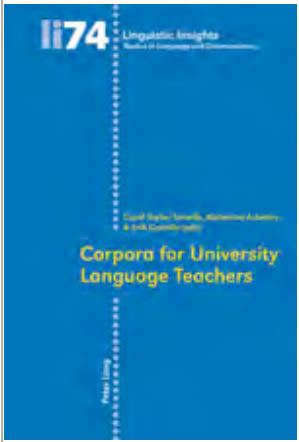


Corpora for University Language Teachers		
Author:	Carol Taylor Torsello, Katherine Ackerley & Erik Castello, Eds. (2008)	
Publisher:	Bern: Peter Lang	
Pages	ISBN	Price
Pp. 309	ISBN 978-3-03911-639-3 (paper)	\$80.95 U.S.



This collection of articles by researchers from nine Italian universities provides a useful overview of current approaches to using English corpora: searchable electronic collections of prose. The volume ended up, in effect, as a festschrift for Birmingham University linguist John Sinclair, who, before his death in 2007, was to have been a keynote speaker at the conference in Padua from which these papers were drawn. An introductory piece by Guy Aston not only recalls Sinclair and his influence—“John changed our view of the lexical item” (p. 17)—but also provides a good brief history of British English corpora, detailing in particular the relationship between the COBUILD reference books, the corpus they were based on, and the subsequent evolution of that corpus into the Bank of English; this introductory chapter also traces the later emergence of a competitor, the British National Corpus (BNC).

Although several recent collections have informatively discussed corpus work and language teaching (e.g., Sinclair, 2004), the current volume stands apart from other conference paper collections that simply report on a themed set of individual research projects. Since several of these papers were based on workshops, the book contains chapters that offer readers instructions on how to apply existing tools to their own corpus projects and language lessons. A later Aston essay, for example, reviews the new edition of the BNC, comparing its current texts to earlier versions and introducing the reader to XAIRA software for searching the XML tags used to code prose, thus allowing users to sort material by text variables such as genre, author, and date of composition. The first half of this article is an accessible introduction to the components of the BNC, whereas the second half assumes some experience with different

query formulas. “The BNC,” Aston observes, “is a prolific resource... learners [and, I would add, teachers] need to be trained to use it—to recognize and formulate problems, pose queries and interpret solutions” (p. 235).

While several chapters rely on results found in large general-language corpora like the BNC and the Bank of English, it is smaller, custom-made corpora that are discussed here most often. With much current ESL writing and vocabulary instruction emphasizing exposing students to specialized text types to help them gain mastery of the genres of their discipline, creating these Language for Special Purposes (LSP) corpora is well motivated. Some of the specialized corpora discussed in the book include the Padova Learner Debate Corpus (PLDC), which comprises computer forum posts by language learners engaged in debates (Dalziel & Helm). A set of four other corpora (Ulrych & Murphy) was gathered following the framework of mediated discourse analysis, (Scollon, 2001) to emphasize how monolingual texts as well as translated texts reveal editorial and social influences: (1) EuroParl, formal oral discourse from European parliamentary debates; (2) AbCoR, annual reports from multinational companies; (3) AMC, American movie transcripts and their dubbed Italian versions; and (4) EuroCom, essays, half of which were written by non-native English speakers working at the European Commission, the other half being versions of the same texts edited by native English speakers working as translators.

Focusing on another LSP corpus, Tognini Bonelli analyzes terms specific to economics writing in a dataset from *The Economist*. And Taylor compares speech features of the artificial exchanges found in the genre film and television transcripts to the use and distribution of the same speech features in exchanges within the Bank of English. Pushing the definition of textual corpora beyond written and spoken forms, Baldry explores how concordancing can make use of multi-modal material, which can be indexed in ways that help students reinforce their text-based language learning. For example, such corpora can be sorted by images or themes, aligning film clips and the metatext that explicates them, or linking web videos with thematically connected vocabulary items.

Focusing on the writing of language learners themselves, Castello created a corpus of 25 essays from both American and British ESL proficiency exams. These learner essays were gathered to measure features of textual complexity. In other work examining writing in a non-native language, D’Angelo created CADIS, the Corpus of Academic Discourse, to capture and compare the English of academic journal articles. That corpus allows the works to be sorted by both discipline as well as the native languages of the authors (English, Italian, or other first languages). Other chapters discuss not just the compiling of texts into a corpus, but using tags to annotate more specialized corpora: Prat Zagrebelsky discusses projects using tags to code common errors in language learners’ college essays. In another tagging endeavor, not student-based, Brunetti discusses creating XML tags to show the inflectional and syntactic relations of each lexical item in a corpus of Old English poems, as well as in its Italian gloss.

As with Brunetti’s chapter, some of the essays cover projects relevant for language-related curriculums for native-speaker students as well as for English language learners, though most

papers specifically focus on foreign language teaching and learning. For teachers planning to mine the results of this volume to model or help their students acquire individual English lexical items—to see, for example, how learners’ choices of modals compare to the edited usage of native speakers; which verbs most typically appear adjacent to the noun *survey*; or the different distributions of *fork out* vs. *pay*—it is important to keep in mind that the book’s contributors work mainly with British rather than North American varieties of English. American language practitioners who create or have created their own specialized corpus but seek a larger reference corpus of American phraseology should see Davies (2008), the Corpus of Contemporary American English (COCA), accessible on the web. However, as models of techniques for compiling a corpus based on specialized texts, and of tagging, concordancing, and searching for words that typically appear together in particular genres, these papers provide helpful guidelines for language teachers in any locale. These corpus creators successfully show how to bring to students’ attention patterns of usage found in disciplines ranging from movie transcripts and criticism to economics and news reporting, as well as in more traditional classroom text types such as poetry and academic essays. While several pieces are geared towards the comparison tasks of translators, all the chapters should prove especially relevant for those L2 classroom projects and assignments that value capturing real life constructions over grammar book examples.

References

- Davies, M. (2008). *The corpus of contemporary American English (COCA): 385 million words, 1990-present*. Available online at <http://www.americancorpus.org>.
- Sinclair, J. M. (Ed.). (2004). *How to use corpora in language teaching*. Amsterdam: John Benjamins.
- Scollon, R. (2001). *Mediated discourse: The nexus of practice*. London: Routledge.

Laurel Smith Stvan
The University of Texas at Arlington
<stvan@uta.edu>

© Copyright rests with authors. Please cite TESL-EJ appropriately.