Williams, C. (2002). *Non-progressive and Progressive Aspect in English*. Fasano di Puglia: Schena Editore.

Wright, S. (1994). The mystery of the modal progressive. In D. Kastovsky (Ed.), *Studies in Early Modern English* (Topics in English Linguistics 13) (pp. 467–485). Berlin/New York: Mouton de Gruyter.

CHRISTOPHER WILLIAMS, *University of Bari (Italy)*

~

McEnery, T., Xiao, R. & Tono, Y. (2006). *Corpus-based Language Studies: An Advanced Resource Book*. London/New York: Routledge.

The book is one of the Routledge *Applied Linguistics Series* which are designed to guide a number of key areas (e.g. translation, second language acquisition and intercultural communication) in the field of applied linguistics. In keeping with other books in this series, this book follows the 'introduction-extension-exploration' template, which explains key terms and concepts ('introduction'), introduces and comments on selected core readings ('extension'), and puts theory into practice in student-oriented case studies ('exploration').

The first chapter offers an insightful overview of corpus linguistics. This begins with a useful clarification of the term *corpus* but moves on to focus on several heated debates over (i) the intuition-based and corpus-based approaches to language studies, (ii) whether corpus linguistics should be regarded as a methodology or a theory, and (iii) the corpus-based vs. corpus-driven linguistics. While those well-established independent branches of linguistics such as phonetics, syntax, semantics or pragmatics study a certain aspect of language use, corpus linguistics does not; it affords a wide range of applications across all branches of linguistics. It is in this sense that McEnery, Xiao and Tono convince us that corpus linguistics is a methodology rather than a theory, though, as they admit, this view is not shared by all scholars (e.g. Mahlberg 2005).

The remaining chapters in Section A are an exposition of key concepts in corpus linguistics — representativeness, balance, sampling, corpus mark-up and annotation, to name but a few. Sticking with the theme of 'a resource book', the authors clarify confusing terminology: *parallel corpora* (L1 texts and their translations), *comparable corpora* (L1 texts collected from different languages using the same sampling frame), *comparative corpora* (varieties of the same language), *development corpora* (L1 learner data) and *learner corpora* (L2 learner data), and guide readers through a range of pragmatic considerations

and dominant practices (e.g. the availability of machine-readable texts and data for a particular text type, copyright clearance, the use of web-based corpus-processing tools for downloading data from the Internet, etc.) involved in the creation of a DIY ('do-it-yourself') corpus. Yet the authors' advice on how to deal with the notoriously vexatious copyright issues in corpus construction does not seem to be particularly appealing. This is perhaps inevitable given no advances in existing fair-use provisions, which hardly allow any reproduction of published works except for short extracts not more than 400 words. As McEnery, Xiao and Tono admit, it is only a matter of time before the right balance between copyright and fair use for corpus-building could be reached. Also of interest is an update of earlier works (see, among others, Kennedy (1998) and Meyer (2002)) on available corpus resources around the world and across languages. Another important theme of the book, as its title suggests, 'corpus-based language studies', is evidenced in a whole chapter in Section A (Unit A10) as well as Part 2 of Section B (selected, original, influential journal articles) on the applications of corpora in some areas of language studies.

Section B, as noted above, is basically composed of excerpts from published material, seeking to provide a more thorough grounding in key corpus-linguistic concepts. Nonetheless, this section is by no means 'descriptive' and 'uncritical', as it appears to be. Having demonstrated their views on a particular issue in Section A, McEnery, Xiao and Tono, in their careful selection of published work, give readers an opportunity to understand other viewpoints and form their own views. Part 1 'Important and controversial issues' discusses that external (or situational, social or extra-linguistic) criteria rather than internal (or linguistic) criteria should be used in initial corpus design, by drawing upon two highly relevant works, namely Biber's (1993) 'Representativness in corpus design' and Atkins et al. (1992) 'Corpus design criteria'. Yet another long-standing controversy — the role of corpora in linguistic analysis, language teaching and learning — is also discussed by including extracts from the debates between Henry Widdowson, Michael Stubbs and John Sinclair. While the debate is still on-going, it is difficult to deny, as McEnery et al. rightly note, that the value of corpus data for linguistic investigations depends ultimately on the research question researchers intend to address, as is the case for every methodology.

Part 2 'Corpus linguistics in action' presents published language studies using corpora and corpus analysis techniques. Like Part 1, this part of Section B serves as a bridge between Sections A and C. In lexical studies, excerpts taken from Krishnamurthy (2000) and Partington (2004) are used to illustrate the terms *collocation* ('the relationship between a node and individual words') and *semantic prosody* ('semantic sets of collocates') respectively, which are taken

up earlier in Section A (p. 84). These two studies are selected because they provide background knowledge for Case Study 1 in Section C. Likewise, in grammatical studies, Carter and McCarthy's (1999) account of the English *get*-passives in spoken discourse (based on the CANCODE spoken English corpus) and Kreyer's (2003) study of genitive and *of*-construction in written English pave the way for Case Study 2 in Section C, an exploration of the syntactic conditions which influence the choice between a *to*-infinitive and a bare infinitive following HELP. The Case Study 2, exploring also language change and regional differences of the alternations between HELP and HELP *to*, has a bearing on three other excerpts: Kilpiö (1997) traces the developments in the functions of the verb BE from Old English to Early Modern English; Mair, Hundt, Leech and Smith (2002) report on shifts in part-of-speech frequencies; in particular, frequency changes among nouns and verbs; Lehmann (2002) is a corpus-based (the spoken BNC and the *Longman Spoken American Corpus*) analysis of subject relatives with a zero relativiser in American and British English. Yet another published research on language variation is Biber's (1995a) framework of multifeature/multidimensional (MF/MD) analysis of register variation, the results of which are compared to those using WordSmith Tools in Case Study 5 of Section C. Case Study 6 is directly related to the excerpt of McEnery, Xiao and Mo (2003), which explains how comparable corpora are used to investigate aspect markers cross-linguistically. The three excerpts from published material about a corpus-based approach to language teaching and learning (i.e. Gavioli & Aston (2001), Thurstun & Candlin (1998) and Conrad (1999)) demonstrate that corpora are valuable resources for both teachers (who decide what to teach based on corpus evidence) and learners (who learn from concordance-based materials), in anticipation of Case Study 3, which engages readers' interest in a learner language analysis on the basis of the *Longman Learners' Corpus*.

As you may have noticed, Case Study 4 in Section C 'Swearing in modern British English' seems to have no place at all in these four major areas of linguistics aforementioned: lexical (Case Study 1 'Collocation and pedagogical lexicography') and grammatical studies (Case Study 2 'HELP or HELP *to*: what do corpora have to say?'), language variation (Case Study 5 'Conversation and speech in American English'), contrastive and diachronic studies (Case Study 6 'Domains, text types, aspect marking and English-Chinese translation'), and language teaching and learning (Case Study 3 'L2 acquisition of grammatical morphemes'), which have benefited most from corpus data. Even McEnery, Xiao and Tono did not have any mention of its connection with Section B. However, given that this case study explores swearing in two registers, both spoken and written British English, it clearly plays a part in language variation,

as does Case Study 5. Most importantly, it proves the usefulness of the corpus-based approach in sociolinguistics, where the operationalisation of sociolinguistic theory into measurable categories suitable for corpus research appears to be problem-prone; this case study examines the patterns of swearing in modern British English with respect to such variables as user age, gender, social class of speakers and writers as encoded in the BNC.

In the last section of the book, Section C 'Exploration', McEnery et al. offer, in each case study, a detailed, step-by-step practical guide to explore a corpus research question using a corpus exploration tool and/or a statistics package. This is particularly helpful to a reader who has just started to learn the worth of the corpus linguistic methodology. The Concord and Keyword functions of the corpus-processing tool *WordSmith* are explained in both Case Study 3 and 5, whereas how to use the *BNCWeb* (the World Edition of the BNC corpus) to make complex queries are demonstrated in Case Study 1 and 4. Case Study 2 introduces another corpus-analysis tool, *MonoConc Pro* and a commonly used statistics package SPSS. To handle parallel texts, *ParaConc* is recommended to use as in Case Study 6. Also worthy of note is that at the end of each case study, readers are given some tasks to gain first-hand experience of using the tools and techniques just learned to solve language problems.

This book is arguably the first of its kind, combining theoretical issues with a practical guide of corpus-analysis software packages currently available. The marriage of 'technical knowledge' with 'practice' makes it stand out from other existing introductory books in corpus linguistics. If the titular promise of 'a resource book' is not entirely fulfilled, this is testament solely to the constraints of book length and to the authors' goal to keep their discussion as clear and concise as possible.

References

Mahlberg, M. (2005). *English General Nouns: A Corpus Theoretical Approach*. Amsterdam/Philadelphia: John Benjamins.

Meyer, C. (2002). *English Corpus Linguistics: An Introduction*. Cambridge: Cambridge University Press.

Kennedy, G. (1998). *An Introduction to Corpus Linguistics*. London/New York: Longman.

May L-Y Wong, *City University of Hong Kong*