Paul Baker, *Sociolinguistics and corpus linguistics*. Edinburgh: Edinburgh University Press, 2010. Pp. ix, 189. Pb. $24.50.

Reviewed by Mercedes Durham
English Linguistics, University of Aberdeen
Aberdeen, AB24 3UB, Scotland
mercedes.durham@abdn.ac.uk

Although sociolinguistics and corpus linguistics "overlap in terms of their epistemology, focus and scope" (9), as Paul Baker points out, the practitioners of each tend to view one another as having different research aims and are often not completely cognizant with the methods and practices of the other. This is unfortunate, as language in society cannot be examined without collecting data (a corpus) and conversely a corpus cannot be fully studied without considering the people who are the source of it (society). Thus, the two fields study different parts of the same thing and, as this book makes clear, there is much potential for fruitful exchange between the two. *Sociolinguistics and corpus linguistics* aims to provide students and scholars of both fields with "an additional tool to their work belt" (12). It is successful in this respect, and the detail with which various methodological processes are described is exemplary. The volume abounds in examples, detailed explanations, and basic introductions to concepts that are potentially new for students, as well as honest discussions about the pitfalls and issues that arise when using corpora. The book is written in a clear and lucid style, and the use of the author's own research—some of it conducted expressly for the book—gives freshness to his expositions and provides accessible demonstrations of how the concepts and methods may be applied.

The seven chapters cover a wide range of topics and could certainly provide ideas for teaching students with broad sociolinguistic and corpus-based interests. Moreover, students could easily replicate many of the case studies presented.

The first chapter provides a cursory introduction to sociolinguistics before turning to the terminology and methodology of corpus linguistics and providing a detailed presentation of some of the basic statistics used in both fields. The section on corpus myths will be particularly useful to students new to the field, as will the detailed explanations of the distinction between corpus-based and corpus-driven research. The presentations of the statistical procedures are clear and generally provide enough detail for students to replicate them later.

The three following chapters treat core concepts in sociolinguistics and corpus linguistics studies. Ch. 2 presents the British National Corpus (BNC), a number of studies conducted using it, as well as frequency analysis. It also introduces students to some of the more detailed methodology and concepts that are necessary for corpus linguistics. Ch. 3 discusses diachronic variation and provides a very good description of the Brown family of corpora; both their contents and the justifications

for their construction are covered in detail. This will be very useful for students hoping for some inspiration about where to obtain data, and overall the volume provides a good discussion of corpora available for which researchers might have missed otherwise. One aspect, however, which is overlooked in the discussion of corpora, is the wealth of sociolinguistic corpora available (see Beal, Corrigan, & Moisl 2007a,b; Anderson 2008). Although these are not as freely available as the ones Baker mentions, it would have been useful to give students of sociolinguistics and corpus linguistics at least an inkling that such material exists. Ch. 4 deals with synchronic variation in terms primarily of comparisons between different varieties of English (American and British especially) and how differences in the corpora can (tentatively) be extrapolated to cultural differences in some cases. It also presents ways of comparing spoken and computer-mediated varieties of language.

Chs. 5 and 6 turn from language variation and change to what Baker terms interpersonal communication, looking more closely at what discourse studies can gain from corpus analysis. This offers new avenues both for corpus linguists and the more traditional type of language variation and change researchers. The discussion of words that collocate with *man* and *woman* in the BNC makes for particularly fascinating reading and provides a good balance of information and analysis. However, the conflation of micro- and macrosociolinguistics masks very considerable differences in their research methods and aims.

The concluding chapter summarizes the main points of the previous chapters, but also provides a practical application of the methodologies of corpus linguistics, by presenting a corpus-based analysis of the first six chapters: that is, an analysis of the most frequent words and collocations used in them. This is an ingenious and entertaining way to make students immediately see the possible applications of the methodology by using a corpus that they have just become acquainted with.

While the detail with which the studies and methodologies are presented is commendable, a caveat must be added. By attempting to cater for a range of audiences—"undergraduate and postgraduate students of sociolinguistics, or corpus linguists who wish to use corpora to study social phenomena" (back cover)—the volume runs the risk of leaving some of them dissatisfied. At times it is difficult to discern which of the potential audiences is being addressed. That Baker is writing primarily as a corpus linguist and not a sociolinguist is often an advantage: the range of studies within corpus linguistics is impressive and insightful. But this bias becomes problematic when it comes to some of the book's potential uses for undergraduate students of sociolinguistics. The discussion of sociolinguistic research and methodologies is sometimes perfunctory and incomplete. While this is not an issue for seasoned sociolinguistic researchers, it might prove frustrating for students who do not already have a thorough grounding in sociolinguistic methodology. It also means that corpus linguists hoping to conduct sociolinguistic research will gain only a partial view of the field and of the investigative procedures available to them.

This shortcoming is most clear in literature cited as being representative of the field of sociolinguistics. For example, although the authors Baker mentions

(Cheshire, Labov, the Milroys, Trudgill); p. 31) as using the "variationist approach" are certainly core to the field, the research cited is over twenty years old; more recent sociolinguistic research, which often is more closely tied to more corpus-based research, is not mentioned. Similarly, Bloome & Green (2002) deals particularly with the sociolinguistics of reading, and thus has a rather different focus than the spoken language that tends to be examined in mainstream sociolinguistics, and as such would not necessarily be a sociolinguist's first choice to define the field. It would have been helpful, too, alongside undergraduate textbooks like Wardhaugh's (2010) *Introduction to sociolinguistics*, to cite more advanced guides. Where else will undergraduates discover sources of more in-depth research, if not in the textbooks we give them?

Another example of the mismatch between the book and its supposed audience of undergraduate sociolinguists is the fact that Labov's principle of accountability is not mentioned until p. 97 and then only in passing. This principle dictates that we should not merely examine the instances of a variant that interest us or that are easy to extract, but that the entire variable context must be considered. This is a central tenet of sociolinguistic research and it should have been dealt with more fully, even if some of the corpus linguistic research cited found ways of circumventing this issue. The discussion of the zero form as an object clause link (i.e. variation between *that* and zero complementizer; p. 76) notes that many corpus studies considered only the most frequent verbs due to difficulties extracting the zero forms. It could, however, have added that some sociolinguistic researchers (Tagliamonte & Smith 2005) did in fact examine every single instance across a range of corpora and then compared the merits and advantages of both methods.

This no doubt reflects some general differences of opinion and methodology between sociolinguistics and corpus linguists. At the heart of sociolinguistics is a desire to understand how (and why) people make use of the linguistic options available to them; witness, for example, recent work on stance and indexicality (Eckert 2008; Jaffe 2009), and this social side is precisely what is often lacking in corpus-based research. There are clear advantages to being able to sample a large section of texts, speakers, and/or periods, but this methodology at the same time restricts the possibility of fully understanding the social forces underlying the patterns observed in that sample. Yet it is only such an analysis that can yield a complete social understanding of the data. Although much can, and is, gained from linguistic analysis of a wide section of data, a true understanding of the social dimension of language requires an understanding of the people using the language and not just the language they use. This is what sociolinguistics aims to provide, a methodological point not sufficiently emphasized by Baker. An introductory volume may not be the place for this, but one would have expected a volume that claims to bridge both fields to give a clearer indication of what sociolinguistics has to offer corpus linguists.

Despite these lacunae, the volume does have much to offer to sociolinguistics and corpus linguists alike, and opens avenues for further discussion and synergy between the two fields. If it is not fully suitable as the sole textbook for a course

on sociolinguistic corpus analysis, it will certainly provide an additional source of ideas and inspiration for teachers and students when used in combination with additional sociolinguistic material.

REFERENCES

Anderson, Wendy (2008). Corpus linguistics in the UK: Resources for sociolinguistic research. *Language and Linguistics Compass* 2(2):352–71.

Beal, Joan; Corrigan Karen; & Hermann Moisl (eds.) (2007a). *Creating and digitizing language corpora. Vol. 1: Synchronic databases*. Basingstoke: Palgrave Macmillan.

———; ———; & ——— (eds.) (2007b). *Creating and digitizing language corpora. Vol. 2: Diachronic databases*. Basingstoke: Palgrave Macmillan.

Bloome, David, & Judith Green (2002). Directions in the sociolinguistic study of reading. In P. D. Peason, R. Barr, M. L. Kamil, & P. Mosenthal (eds.), *Handbook of reading research*, 395–421. Mahwah, NJ: Lawrence Erlbaum.

Eckert, Penny (2008). Variation in the indexical field. *Journal of Sociolinguistics* 12(4):453–76.

Jaffe, Alexandra (ed.) (2009). *Stance: Sociolinguistic perspectives*. Oxford: Oxford University Press.

Tagliamonte, Sali, & Jennifer Smith (2005). No momentary fancy! The zero 'complementizer' in English dialects. *English Language and Linguistics* 9(2):289–309.

Wardhaugh, Ronald (2010). *An introduction to sociolinguistics*. 6th edn. Oxford: Blackwell.

(Received 11 April 2011)

ADAM JAWORSKI & CRISPIN THURLOW (eds.), *Semiotic landscapes: Language, image, space*. London: Continuum, 2010. Pp. ix, 314. Hb. $160.

Reviewed by DURK GORTER
University of the Basque Country – Ikerbasque
Donostia-San Sebastián, Gipuzkoa, Spain
d.gorter@ikerbasque.org

The field of linguistic landscape studies is rapidly expanding. The texts we see around us on signs in the public sphere of cities all over the world are no longer a neglected source of sociolinguistic data. Investigations of how languages function in signage are based on diverse theoretical and methodological approaches. Diversity can be seen as a built-in characteristic of the field of linguistic landscape studies. Its results offer fresh sociolinguistic perspectives on issues such as urban multilingualism, globalization, minority languages, and language policy. The current book is a valuable addition to a growing number of publications about linguistic landscapes. The editors refrain from the term "linguistic landscape" for the title of their collection, selecting "semiotic landscape" as an alternative, although they also state that "all landscape is semiotic," which would make the term pleonastic. It is interesting to note that the editors themselves as well as most contributors to the book continue to use the term "linguistic landscape" repeatedly. The subtitle