

**Joybrato Mukherjee.** *Korpuslinguistik und Englischunterricht: Eine Einführung* (Sprache im Kontext series 14). Frankfurt: Peter Lang, 2002. 214 pp. ISBN 3-631-39346-6. Reviewed by **Hilde Hasselgård**, University of Oslo.

*Korpuslinguistik und Englischunterricht* is, as the title says, an introduction to corpus linguistics aimed at teachers of English, particularly in German upper secondary schools. At the same time, the author takes the opportunity to advocate the use of corpus-based materials and corpus methods in English language teaching.

The author sets up four aims in the preface to the book: (i) to give a presentation of the development of modern corpus linguistics, with particular regard to implications for language teaching; (ii) to discuss the potential of corpus linguistics in relation to didactic concepts and models; (iii) to describe concrete uses of corpus data and corpus methods in English language teaching; and (iv) to make

some suggestions for future English language teaching and the education of English teachers in the light of the corpus revolution.

The stated aims are reflected in the organization of the book into five chapters (of which two are devoted to the third aim). Chapter 1 gives an introduction to some fundamentals of corpus linguistics. The presentation includes a survey of various English language corpora, from the beginnings of corpus linguistics with the Survey of English Usage and the Brown corpus to present-day projects such as the International Corpus of English. Furthermore, some of the principles of corpus-based language description are outlined, and there are examples and case studies. These illustrations come either from the author himself or from the work of others. Particularly Kennedy (1998) is frequently cited in this part of the book. The topics as well as the examples are selected with regard to a readership of present and prospective teachers.

Chapter 2, "From corpus to classroom", discusses some central concepts of language teaching, such as the use of authentic material, communicative competence, learner autonomy, and the choice of the native or the intercultural speaker as a model of proficiency. Some of the assumptions underlying corpus linguistics are also dealt with, such as the emphasis on empirically grounded language description, collocations and other kinds of lexicogrammatical patterns, and genre differences (cf. Biber et al. 1999).

Chapters 3 and 4 show how corpus methods can be applied in the classroom by the teacher (3) or by the students (4). The author suggests ways of exploiting corpus data indirectly, through use of corpus-based teaching materials, or directly in that the teacher and/or the students use corpora in their work with the English language. Many of the ideas for corpus use are explicitly linked to stated aims in the curriculum for the upper secondary school in one of the German Bundesländer (Nordrhein-Westfalen). There are concrete suggestions as to how corpus work can be justified within the existing plan, for example to encourage students to write a corpus-based "Facharbeit", a long essay that is usually written within literature or culture studies, but which might be linguistically oriented. Each chapter is concluded with a section suggesting further reading, which will be useful to all those wanting to take up corpus linguistics after reading this book.

The main impression of *Korpuslinguistik und Englischunterricht* is that of a carefully thought out volume with a firm basis both in the author's own work and in related, reputable course books and reference works (e.g. Kennedy 1998, Biber et al. 1999, Tognini-Bonelli 2001). In spite of its slender appearance, it is also a rich volume. It is ambitious, in that the author wants to convince his read-

ers and cover a lot of ground, and it is comprehensive in its survey of both corpus linguistics and English language teaching.

In the following few paragraphs I will, however, present some critical points. The first of these has to do with the coverage of various topics. It is clearly impossible to do everything within the confines of one book, and priorities have to be made. Although, by and large, I agree with Mukherjee's selection of topics to be presented, I have certain reservations as to the relative weighting of them. The reason is that the target group consists of English teachers (and students), who presumably have little prior experience with corpora.

Chapters 1 and 2 are survey chapters, of corpus linguistics and concepts of foreign language teaching, respectively. They provide good overviews and interesting discussions, at the same time as arguing forcefully for the relevance of corpus data and corpus methods in the teaching of English. However, it takes rather a while to reach the more practical 'applications' section. Furthermore, it will perhaps be a disappointment to some readers that this section takes up slightly less than a third of the book, even though ideas for corpus work in the classroom can be found elsewhere as well. Thus, the first two chapters might have been reduced in favour of a more comprehensive 'how-to-do-it' section, and some of the theoretical considerations taken care of by means of reference to other works particularly within the didactics of language teaching, but also to some extent course books in corpus linguistics as a research field.

An example of a section that would benefit from more extensive practical advice is found on p. 159, where the author lists some grammatical topics that may usefully be investigated in corpora by students. One of these is the variation between the simple present and the present progressive – a phenomenon that is notoriously difficult to search for, because it goes beyond a mere lexical search. If given this task, students (and possibly their teacher as well) would need concrete tips on how to find useful material in the corpus.

As indicated by the author in his brief discussion of "corpus literacy" towards the end of the book, it is not a trivial matter to teach students how to perform corpus searches to get at the material they want/need or how to judge their results in relation to the type of material they have, to mention but a few areas. So if the book is going to be a "manual" for teachers wanting to use corpora in their teaching, I think an extended practical section might have been welcome.

Another point of criticism concerns the treatment of the issue of copyright, with the consequences it has for the limitations on the use and distribution of corpora. In an innocent-looking footnote (p. 155) the author remarks that the ICAME corpora, for reasons of copyright, can only be used by licence holders

and only for research. As many of the preceding case studies are based on material from these corpora, this footnote may seem to undermine much of the argumentation by severely restricting the possibilities of using real corpora in class. It would have been a good idea to make this reservation earlier on, and emphasize other corpora/text archives that actually *can* be used by teachers and students in a secondary school setting. To be fair, newspaper archives are given some attention (p. 36), as well as the freely available “Simple search of BNC World” and the Cobuild corpus concordance sampler. Besides, it seems that the BNC as well as the ICE corpora can be used in teaching. These points should, however, have been stated more clearly and at an earlier stage of the book. It is important that readers should be confident that there are corpora that they can use for their own purposes, and also where to find them, if corpus linguistics is going to find its way into the classrooms.

The copyright issue might have been given some attention also in another connection, namely where the author suggests that the teachers may collect corpus materials themselves in connection with English for special purposes or the study of particular genres (p. 134). Although a lot of material may be easily downloaded from the Internet or scanned from books, it is not necessarily unproblematic to store and distribute it. Teachers who want to do this should at least be made aware of the most common limitations imposed by copyright restrictions.

In teaching English as a foreign language, *parallel corpora* have obvious advantages in that they can illustrate similarities and differences between English and the learners' L1. Parallel corpora are briefly presented on p. 60. I was surprised to see the English-Swedish Parallel Corpus (ESPC) cited as the only example, as the ESPC relies heavily on its sister project in Norway (the English-Norwegian Parallel Corpus), where much of the methodology and all the software were developed (cf. Johansson et al. 1999/2001). Furthermore, the Chemnitz Internet Grammar and Translation Corpus might have been given more attention. As mentioned briefly by the author (p. 151), the Internet Grammar is based on a parallel corpus of German and English. From its website, the *Chemnitz Internet Grammar* does not seem to be limited to academic users, and should thus be relevant to German learners of English. The learner will get access to corpus examples, although the whole corpus is open only to linguists.

Mukherjee makes a good case for using corpus methods on “non-corpus” texts in a combination of content analysis and linguistic investigation. One example is a comparison between a speech by George W. Bush and an essay by Arundhati Roy (p. 101 ff.). With word frequency lists for the two texts as a starting point, the author proceeds to identifying patterns which some of the most

frequent content words enter into. One may wonder, however, if all the conclusions drawn by the author are based on the concordance work alone, or whether they presuppose knowledge of the whole text. The suspicion that the latter may be the case is even stronger in the case of the analysis of *The Cure's* texts (p. 137), where the concordance in fact tells the uninitiated reader very little about content.

The idea of using corpora in cultural studies is not new (cf. e.g. Stubbs 1996: 157), but certainly interesting in a TEFL setting. Thus, the case study of the collocational patterns of *Kashmir* in ICE-GB and ICE-IND is a good example of a corpus-linguistic basis for comparing British and Indian politics. It is, however, more doubtful if differences in patterns of preposition and article use (p. 99) can be classified as intercultural phenomena; rather, I would say they belong to the field of regional variation (indeed another fruitful area of corpus investigation suggested by the author).

In recommending corpus-based teaching materials (Chapter 3), Mukherjee devotes a lot of well-deserved space to the Cobuild range. As his bibliography, like his examples, is generally impressively up-to-date, I was surprised to see that the 3rd edition of the Cobuild dictionary (2001) is not included, while there is frequent reference to the second edition (1995). The author's point about continually updating corpus-based dictionaries so as to include a new word such as *road rage* (absent from the 1995 edition) could have been better made by reference to the 2001 edition, where the word indeed has an entry.

In spite of the above reservations, I think *Korpuslinguistik und Englischunterricht* gives a nice introduction to the field of corpus linguistics and pays due attention to its stated readership (teachers in secondary schools). It should also be able to inspire reflection on what kind of English students should learn and how they should learn it. One can only hope that it will reach its target group and that corpora will find their natural place in the EFL classroom.

## References

- Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad and Edward Finegan. 1999. *Longman grammar of spoken and written English*. London: Longman.
- Chemnitz Internet Grammar and Translation Corpus:  
<http://www.tu-chemnitz.de/phil/InternetGrammar/> [accessed 21.11.02]
- Johansson, Stig, Jarle Ebeling and Signe Oksefjell. 1999/2001. *English-Norwegian Parallel Corpus manual*.  
<http://www.hf.uio.no/iba/prosjekt/ENPCmanual.html> [accessed 20.11.02]

- Kennedy, Graeme. 1998. *An introduction to corpus linguistics*. London: Longman.
- Schmied, Joseph et al.: *The Chemnitz Internet Grammar*.  
<http://www.tu-chemnitz.de/phil/InternetGrammar/publications/index.html>  
[accessed 21.11.02]
- Sinclair, John (ed.). 2001. *Collins English dictionary for advanced learners* (3rd edition).
- Stubbs, Michael. 1996. *Text and corpus analysis*. Oxford: Blackwell.
- Tognini-Bonelli, Elena. 2001. *Corpus linguistics at work*. Amsterdam: John Benjamins Publishing Company.