BOOK REVIEWS

Aijmer, K. (Ed.) 2009. *Corpora and Language Teaching*. Amsterdam/Philadelphia: John Benjamins. (viii + 232 pp.)

1. Introduction

The recent surge in attention for the link between corpora and language teaching with the publication of volumes such as Aston et al. (2004), Sinclair (2004), Campoy & Luzón (2007) and Lombardo (2009) is long overdue. Corpus research has established itself as a mature discipline over the past few decades, and since the late nineties learner corpus research has, slowly but steadily, gained ground. Still, the gap between (corpus) linguists and language teaching professionals has shown no noteworthy signs of narrowing. At the same time the fences between corpus linguistics and second language acquisition (SLA) research have remained upright, although the need for an interdisciplinary approach has been advanced repeatedly in the past (see e.g. Leech 1998, Myles 2005, Hugon 2008). Aijmer's (ed.) *Corpora and Language Teaching*, which draws on contributions to a symposium at the University of Gothenburg in 2005, addresses these issues from different perspectives, corresponding to its four parts: corpora and second-language acquisition, the direct corpus approach, the indirect corpus approach and new types of corpora. First, I shall briefly discuss each contribution, after which I shall evaluate the volume as a whole, which will picture it as an excellent reflection of and introduction to a fascinating yet still somewhat imperfect field of research.

2. Synopsis of the volume

In the first contribution on **corpora and second-language acquisition** Granger aptly observes that learner corpus research is inherently at the crossroads of four disciplines: corpus linguistics, linguistic theory, foreign language teaching and second language acquisition. She argues a convincing case for researchers in second language acquisition to renounce their objections to learner corpus research as "a bona fide data type in SLA research" (p. 17), although she notes that learner corpus studies have to make an effort to remain attractive by adopting strict design criteria for corpora and by working towards some form of harmonisation in the field. Granger also considers the relationship between learner corpora and language teaching. In determining the extent to which a learner corpus can be

used in a pedagogic context, learning needs, teaching objectives and “teachability” should always be borne in mind. Learner corpora can contribute at least to three dimensions of learner proficiency: accuracy (with a revamped form of error analysis), (lexical and grammatical) complexity and fluency (e.g. through phraseology). Finally, Granger sets three priorities for future research. These include a need for more learner corpora (especially longitudinal ones), a need for some form of standardisation (e.g. of mark-up and annotation systems) and a need for thorough analyses based on sound theoretical frameworks as well as pedagogic tools that meet actual needs in classrooms.

Johansson discusses the affinities between the idea, stemming from error analysis, that language learning is a process of hypothesis testing and more recent approaches that see students as researchers who can analyse corpora in a language learning setting. He too advocates the inclusion of corpus research in SLA, because corpora and their usage-based approach “strengthen natural processes of language acquisition in that they make it easier for learners to notice and experience repeated instances of use” (p.39). In Johansson’s view, corpora are capable of dissolving the classic dichotomy between implicit and explicit teaching methods, as formal instruction and discovery of language use and language principles through corpora can easily be reconciled. He does issue a fair word of caution, though: the use of corpora in the classroom “is vindicated to the extent that it agrees with what we know about language and language acquisition” (p.42), and there is urgent need for studies that test the benefits and drawbacks of the corpus approach in language teaching.

The second part of the present volume consists of three contributions that focus on how corpora can be used in the language classroom (= the **direct corpus approach**). Granath reports on how she uses corpora in an English course for first-term university students. Her pedagogic corpus method is threefold. First, the students have to master specified basic requirements (e.g. identify word classes and identify which hits are relevant in a corpus search). Second, a corpus can be used to answer students’ questions, which “makes the language come alive in the classroom in a way that is not possible by just relying on reference books and made-up examples” (p.55). Finally, students get hands-on training in the language lab to enable them to analyse concordances independently. The main advantages of this approach are that the students’ insight into the value of corpora is fostered, and that the exposure to new, real-life examples brings more variation to the language class. However, evaluations of this approach by students yield a picture of mixed feelings: most of them find corpus work difficult, and few are convinced of its added value. Granath believes that especially advanced students, current and future language teachers will benefit from taking this approach.

Ebeling presents Oslo Interactive English (OIE), an interactive learning environment for undergraduate students of English, which is meant to provide supplementary exercise material to on-campus teaching. The exercises (multiple choice, gapfilling, error correction and open exercises) are based on a corpus of about 7 million words of 20th century English fiction and non-fiction. Film scripts are included to represent “spoken” material. Students who submit their answers get a score and a key, which always includes links to online theory as well as to corpus searches. Usage statistics show that relatively few students make use of OIE, which could be overcome by integrating it into on-campus teaching. Ebeling contends that only the best students may benefit from this approach due to the demanding nature of corpus analysis. However, the range of exercises in OIE is sufficiently broad, and although some exercises require students to perform a corpus search, for others the corpus merely serves as an optional resource.

In order to gauge what ELT teachers need and how corpora can help satisfy those needs, Römer conducts a survey among teachers in German secondary schools. They express a need for (native-speaker) consultation in marking exams, better resources to determine the acceptability of a structure in English, and textbooks that reflect actual language use. In Römer’s view corpus linguistics can offer around-the-clock native-speaker advice through online corpora as well as reliable reference tools. Corpora can also contribute to the development of better teaching materials, and support teachers adequately in developing materials of their own. The main problem that remains, however, is that few language teachers are aware of these resources and of how they can be exploited through corpus analysis.

In the third part of this volume, the role of corpora as a resource for generating insight into learner language is investigated (= the **indirect corpus approach**). Herriman & Boström Aronsson have studied the Swedish sub-corpus of the International Corpus of Learner English (SWICLE) and its native speaker reference corpus (LOCNESS) to compare thematic organisation in the writing of Swedish advanced learners of English to native speaker practice. After scrutinising the way in which the language learners select clause themes, employ subjective and objective interpersonal metaphors (e.g. *I think, it is important that*) and cleft constructions, they distil some features of non-native argumentative writing. These are characterised as typical of spoken language, reflecting an interpersonal, speaker-involved style, which is often more emphatic and persuasive than in native writing. On the basis of their research, Herriman & Boström Aronsson suggest that teaching could be improved if students were offered the opportunity to learn about differences in cultural norms of writing between their mother tongue and the target language as well as about register differences in general.

Hasselgård selects a sub-corpus of the Norwegian part of ICLE (NICLE) to conduct an exploratory study of thematic choice in English argumentative texts

written by Norwegian learners. She uses the Integrated Contrastive Model (Granger 1996), which involves comparing the learners' mother tongue (Norwegian) with the target language (English) on the one hand, and the learners' interlanguage with that of learners with a different mother tongue (in this case Swedish), on the other. As her approach is corpus-driven, Hasselgård decides to include expressions of stance in her analysis as well in order to test hypotheses arising from her investigation of thematic choice. The results show that the language learners in this study can build grammatical structures of sentence openings, but have not quite mastered the discourse conventions of argumentative writing, which, Hasselgård argues, should become the focus of instruction at this advanced level.

Hunston's chapter discusses the usefulness of certain corpus-based descriptions for language learners, in particular when these involve mention of relative frequencies. She exemplifies this issue with reference to phraseology: a phrase may not be frequent in itself, yet it may be significant as a unit to each of its components. A further point that Hunston addresses is that of semantic sequences. These "express 'what is often said'" (p. 148), and are frequent patterns that contain word forms that may not be frequent in themselves. Although the abstract description of such sequences (e.g. 'modal meaning + *decide* + wh + potential decision' as in *It's up to you to decide what to do*) has little informative value for languages learners, it may have its merits for those who devise new teaching materials.

In the fourth and final part of the volume three **new types of corpora** are introduced. Cheng zooms in on small, specialised corpora in an endeavour to determine the "aboutness" of particular texts in two English economic and financial services corpora. After dismissing n-grams and skipgrams as appropriate tools to draw up the lexical profile of specialised texts, she posits the concept of the con-gram, which is able to account for both constituency variation and positional variation, and can be generated automatically with specific software. Cheng argues that this not only yields information which is useful for ESP writers, but can also find a place in data-driven learning.

Meunier & Gouverneur present a new corpus of textbook material, based on 32 volumes of best-selling titles on the international ELT market: the TeMa Corpus. It includes over 700,000 words of texts, spoken data, vocabulary exercises, and even the guidelines to the exercises. The vocabulary sub-corpus is pedagogically tagged. This corpus unveils only part of the input that language learners are exposed to, but it allows scrutiny of at least this vital element in the language-learning process.

Finally, Mukherjee highlights three areas in which spoken learner corpora can forge a place of their own in language pedagogy, which he claims has long favoured written over spoken media. His starting point is the German component of the Louvain International Database of Spoken English Interlanguage (LINDSEI-

GER). He compares the use of verb-noun collocations in learner speech and writing, examines the discourse marker *you know* in LINDSEI-GER as a whole and in individual performances, and sets off performance phenomena such as repetitions in the learner corpus against the native speaker reference corpus of LINDSEI (LOCNEC). For each of these Mukherjee discusses a number of interesting possibilities for pedagogic applications.

3. Evaluation

Aijmer has succeeded in bringing together ten noteworthy contributions that allow the volume (and its readers) to take stock of the state of the current relationship between corpora and language teaching. Such a broad field mandates some heterogeneity in the articles that are meant to represent it. *Corpora and Language Teaching* is well-written, and is cleverly divided into four parts, which are aptly balanced. Granger's and Johansson's contributions provide an excellent reflexive background against which the more practically oriented articles can be read more insightfully. It is hard not to agree with Granger's view of learner corpus research as "inherently interdisciplinary" (p. 15), and it is telling for the slow evolution in the field that her plea for an integration of learner corpora in SLA strongly echoes the argument she put forward over a decade ago (Granger 1998). The discussion of crossovers between corpus linguistics, SLA and FLT is long overdue. Perhaps it is unlikely, at least in the short term, for current scholars to specialise in more than one of these, but the time seems right to set up extensive interdisciplinary projects that bridge this gap, which appears to have been oddly insurmountable so far. Interestingly, Granger does not only rebut some SLA criticisms of corpus linguistic methodology, but equally calls for self-reflection within her own discipline to improve some of its techniques.

The arguments about the potential of using corpora in language teaching are well balanced, and the contributors all appear to take a realistic stance towards this issue. Johansson sets the tone by making the case for systematic studies that test the extent to which a corpus approach in language teaching can be beneficial. Contrary to Granath, who proposes an ambitious project in which students conduct fairly complex corpus analyses, Ebeling cautions that corpus-driven learning only works if it is integrated into regular classes and if the sights are not set too highly, in that secondary-school pupils cannot be expected to carry out full-fledged corpus analyses. In other words, if it is employed in language teaching, a corpus approach has to match the teaching and learning objectives specific to the target group. In this context, Cheng's suggestion to ask learners to draw up congrams of and on their own may not be readily achievable and might generate

varying levels of success depending on the level of the students. Such tasks, and more importantly the subsequent analyses of the results, would require a sound knowledge of corpus linguistic methodology as well as a solid linguistic insight.

The studies reported in the volume are all corpus-based or corpus-driven, and they deal with a wide variety of linguistic phenomena, including syntactic structures, phraseological patterns, text structures, discourse strategies and various genres. The range of areas that are covered in the individual chapters highlights the fact that corpus linguistics is capable of investigating a large number of linguistic phenomena, all of which bear relevance to language learning in one way or another.

The present volume attests to the maturing nature of learner corpus research in that it demonstrates that a large-scale project like ICLE, which started when the field was still in its infancy, is now beginning to bear fruit. The international dimension of corpus research and its application to the language and teaching context is best reflected in the studies by Herriman & Boström Aronsson and Hasselgård. The latter partly replicates a study by Boström Aronsson (2005) carried out for a different sub-corpus of ICLE, but also adds a new dimension to the investigation. The identical compilation procedure and mark-up of the components of the corpus should yield an unrivalled insight into learner language and second language acquisition processes, which are arguably relevant for contrastive linguistics as well. Future projects following the same design should add further to this emerging field of research.

In most of the studies that deal with the direct corpus approach to language teaching the focus is exclusively on native speaker corpora. However, Ebeling and Mukherjee rightly point out that learner corpora can serve a purpose in the language classroom as well. Whereas the former limits the use of such corpora to error correction exercises, Mukherjee argues in favour of using learner corpora as a source of correct language use as well, and argues that “learners react much more positively to the use of positive evidence obtained from learner data because they do not get the impression of learner output being treated exclusively as a hotch-potch of mistakes and errors” (p.213).

Anyone interested in exploring new avenues of corpus research or corpus exploitation in an educational setting will find this book utterly inspiring. It contains countless practical suggestions for corpus research that are readily usable in a range of different contexts. Römer’s contribution most notably supplies some excellent food for thought towards developing pedagogic applications of corpora.

Spoken corpora are under-represented in this volume. Unfortunately, this may well be merely an accurate reflection of their share in the field of corpus linguistics generally. Compiling a spoken corpus is evidently much more time-consuming than a written corpus, and the transcription process makes it even more laborious.

Automated (POS-)tagging for spoken corpora is largely non-existent at this moment, and there are many more barriers to the development and analysis of spoken corpora.

If the dearth of spoken (learner) corpora persists, the consequences for language teaching may be detrimental. Due to an acute shortage of authentic spoken language material, learners' exposure to real-life spoken language is bound to remain limited. Moreover, spoken learner corpora are desperately needed to determine how learners use a foreign language in a range of spoken contexts, and thus how their awareness of different pragmatic phenomena can be raised. One can only hope that spoken corpora like LINDSEI will grow to be as influential as its written counterpart ICLE.

Strikingly, with the exception of Römer, the type of 'learner' discussed in this volume seems to be restricted to that of an advanced learner of English, often represented by undergraduate students. However, for want of (unfortunately rather utopian) large-scale longitudinal studies, inclusion of and comparison with learner performance in earlier stages of the acquisition process is vital if the aim is to gain a better insight into the acquisition process and possible improvements to it. The contributors to this volume have argued for initiatives that make language teachers (Römer, Herriman & Boström Aronsson, Cheng, Mukherjee), future language teachers (Granath, Römer) and materials designers (Hunston, Cheng, Meunier & Gouverneur, Mukherjee) more sensitive to the benefits of corpora in pedagogic practice, but also to the specific nature of learner language as attested by corpus research. For this effort to stand a chance of success it has to be based on data that are relevant to a specified target audience, but the shortage of data that cater for the diverse range of learners might turn out to be a real barrier to the spread of corpus approaches in language teaching.

This volume is concerned mainly with foreign language teaching, although this is not necessarily reflected in the title. It would be interesting to see how some of the techniques and strategies laid out in *Corpora and Language Teaching* can be transferred and applied to L1 language teaching. It may be that an inherently different type of language learner calls for a different approach, if only because they are likely to encounter problems of another nature.

There is thus clearly room for extending the scope of this volume, although the focus of the studies presents an accurate reflection of the current state of the research field in question, which has only recently begun to shrug off its fledgling status. As such it is an important read for anyone genuinely interested in corpus linguistics and its pedagogic implications, even without the requirement of extensive prior knowledge of corpus linguistics, applied linguistics or second language acquisition. Chances are that a similar book published in a decade or so will demonstrate that the discipline has coped with some of the current shortcomings. For

now, this volume presents a collection of cutting edge research which is likely to be highly influential in shaping future directions in the area of corpus linguistics and language teaching.

References

- Aston, G., Bernardini, S. & Stewart, D. (Eds.) 2004. *Corpora and Language Learners*. Amsterdam/Philadelphia: John Benjamins.
- Boström Aronsson, M. 2005. *Themes in Swedish Advanced Learners' Written English*. PhD Dissertation. Göteborg: Göteborg University.
- Campoy, M. C. & Luzón, M. J. (Eds.) 2007. *Spoken Corpora in Applied Linguistics*. Bern: Peter Lang.
- Granger, S. 1996. "From CA to CIA and back: An integrated approach to computerized bilingual and learner corpora". In K. Aijmer, B. Altenberg & M. Johansson (Eds.), *Languages in Contrast. Papers from a Symposium on Text-based Cross-linguistic Studies. Lund 4–5 March 1994*. Lund: Lund University Press, 37–51.
- Granger, S. 1998. "The computer learner corpus: A versatile new source of data for SLA research". In S. Granger (Ed.), *Learner English on Computer*. Harlow: Addison Wesley Longman, 3–18.
- Hugon, C. 2008. "High-frequency verbs: Starting block or stumbling block for advanced L2 communication?". In G. Rawoens (Ed.), *Taal aan den Lijve. Het Gebruik van Corpora in Taalkundig Onderzoek en Taalonderwijs*. Gent: Academia Press, 69–98.
- Leech, H. 1998. "Preface". In S. Granger (Ed.), *Learner English on Computer*. Harlow: Addison Wesley Longman, xiv–xx.
- Lombardo, L. 2009. *Using Corpora to Learn about Language and Discourse*. Bern: Peter Lang.
- Myles, F. 2005. "Interlanguage corpora and second language acquisition research". *Second Language Research*, 21 (4), 373–391.
- Sinclair, J. McH. (Ed.) 2004. *How to Use Corpora in Language Teaching*. Amsterdam/Philadelphia: John Benjamins.

Reviewed by Lieven Buysse, Hogeschool-Universiteit Brussel