HANS LINDQUIST, *CORPUS LINGUISTICS AND THE DESCRIPTION OF ENGLISH*
(EDINBURGH UNIVERSITY PRESS, 2009)
A DETAILED EVALUATION

By Marlies Gabriele Prinzl

Summary

In *Corpus Linguistics and the Description of English* Hans Lindquist offers another introductory book to corpus linguistics, but aims it specifically at 'university students of English at intermediate to advanced levels who have a certain background in grammar and linguistics, but who have not had the opportunity to use computer corpora to any great extent' (Lindquist 2009, xvi). He proposes that the book, especially certain sections of it, may also be of interest to students of literature. *Corpus Linguistics and the Description of English* comprises ten chapters. Chapters 1-5 cover the basics, introducing corpus linguistics as a discipline, discussing its methods and explaining key terms, while Chapter 6-10 delve into more specific and different subject matters ranging from corpus-based metaphor studies to the applications of corpora in sociolinguistics. Readers new to corpus linguistics would therefore benefit from reading the first section of the book, but might opt to peruse only chapters relevant to their studies from the second part. That said, Chapters 6-10 provide a valuable overview of the different possibilities within corpus linguistics for anyone new to the field. All chapters are set up in an identical fashion and include, in addition to a discussion of the topic covered, a chapter summary, study questions, suggestions for further reading as well as online corpus exercises on the books supplementary webpage.

Evaluation
The first chapter introduces corpus linguistics as a field. It is established from the beginning that the name does not so much indicate what is being studied, but the methodology that is being used. However, Lindquist also notes that 'it cannot be denied that corpus linguistics is also frequently associated with a certain outlook in language' (Lindquist 2009, 1). Furthermore, the author emphasises that the book's focus is on transmitting the 'joy and fascination that lie in the description of the English language' (Lindquist 2009, 1). As other introductory books – e.g. Kennedy's *An Introduction to Corpus Linguistics* (1998) or Meyer's *English Corpus Linguistics: An Introduction* (2002) – Lindquist commences with a historical overview of the field and lists the first corpora. The overview is refreshing to read, not only because of personal anecdotes, but because some updates are provided. When introducing the historically important Brown Corpus and the LOB corpus that was modelled on it, it is pointed out that new corpora - Lancaster 1931 and Lancaster 1901 – are now being compiled based on historical ones. The chapter covers all the essential basics, such as concordances, frequency, the distinction between corpus-based and corpus-driven, different corpora types (spoken, general, specialised, historical, parallel) and is comprehensive to the extent that the use of dictionaries, text archives and the web as corpora are also mentioned. Criticism of corpus linguistics is acknowledged, although Lindquist's choice of phrasing ('armchair linguists', 8) leaves no doubt about what side he is on. Although links are provided online for the resources used as part of the exercises section, Lindquist unfortunately does not include all the websites of the various corpora he mentions in the chapter. Admittedly, some of these - such as the Bank of English - are not publicly available; however, their websites still contain further information and, in some cases, demo versions or details on how to gain access.

In the second preparatory chapter, 'Counting, calculating and annotatin', Lindquist provides further bases for corpus research, delineating quantitative and qualitative methods. Word frequency is discussed with an example application, surely appreciated by the book's student readers, of how such information is used by mobile phones 'guessing' words when texting. The

indispensable question of 'What makes a word?' is also raised. A significant part of the chapter is devoted to managing and comparing frequency data by using statistical methods such as significance testing and the measurement of strength of lexical association. It is noted that the extent of use of such methods in the field varies greatly as language scholars have not traditionally received training in statistics and that some researchers (Gries 2006) have called for 'greater sophistication' (Lindquist 2009, 37) being needed in this area. Lindquist's own introduction into statistics covers only the basics and, again, resources provided can be incomplete, e.g. the availability of online tutorials is mentioned, but no further details are given. Furthermore, no specific sources are recommended for critical values of significance for the chi square tests, something that individuals with little statistical background might have appreciated. That said, Lindquist's emphasis on good statistical practices is important and all significant, related aspects (distribution, representativity, use of percentages, normalising practices) are considered. Finally, the chapter also introduces students to corpus annotation.

Chapter 3, "Looking for Lexis" discusses the uses of corpora for lexicographers. It explores the different meanings of words through the example of "squeeze", concluding that 'the meaning of a word can only be ascertained by looking at the contexts in which it occurs' (Lindquist 2009, 57). This observation subtly hints at that 'certain outlook in language' associated with corpus linguistics, which Lindquist already alluded to in Chapter 1. As part of this discussion several more important terms are introduced ('collocation', 'colligation, 'semantic preference' and 'semantic prosody'), all of which are carefully defined and illustrated by examples. Lindquist moreover does not omit to mention that there is controversy about some of the terms even if not much elaboration is given on this point. The chapter also considers lexical changes over time, a topic that is picked up again in Chapter 9. Finally, an account of how corpus techniques can be used not only to study language as a whole, but to examine how it is used by a specific writer or within a single work should be of particular interest to students with a background in literary studies.

As the title 'Checking collocations and colligations' already indicates, Chapter 4 more thoroughly explores two terms introduced in the previous section. Much of the focus is in fact on collocations, as only two pages at the end are dedicated to colligations. Lindquist commences with a discussion on native-like fluency, stating that '[t]he ability to combine words in the right way is the key to native-like fluency; (Lindquist 2009, 71). With this, collocations and their challenge to Chomskyan (generative) linguistics is put forward. Lindquist's treatment on 'collocation' goes further back in time than that of other introductory books as he attributes the first usage to the educationalist H.E. Palmer in 1933, who defined the term as 'a succession of two or more words that must be learnt as an integral whole and not pieced together from its component parts' (title page). The author then proceeds, more familiarly, with Firth's usage of the term collocation that emphasises how the meaning of individual words is influenced by other words frequently occurring with it, offering a second definition: 'The more-frequent-than-average co-occurrence of two lexical items within five words of the texts' (Krishnamurthy in Sinclair et. al, 2004, xiii). Lindquist observes thus that Palmer's and Firth's definitions point to different concepts, but that linguists often use 'collocation' without making a distinction.

The final general chapter, "Finding Phrases", continues with language patterns by focusing on phrases, that is, 'more or less fixed strings which are used over and over again' (Lindquist 2009, 91). After naming some of the many terms used to refer to the phenomenon of phrases, Lindquist briefly explains John Sinclair's open choice principle and idiom principle (Lindquist 1991, 100), emphasising the view of language that has emerged through corpora – that there is a significant amount of linguistic repetitiveness and that language users will frequently rely on conventionalised utterances even when other possibilities exist. Most of the chapter is devoted to examining examples of idioms and recurrent phrases, allowing readers to gain insight into how and what kind of corpus research can be done. Lindquist uses both more established methods

(querying different types of corpora for complete as well as incomplete phrase units and n-grams) as well as an emerging one (using Google to search for country-specific variants of 'storm in a teacup'), with the latter serving as an introduction to a method explored more thoroughly in Chapter 10. Students from a literature-based background will also appreciate that Lindquist's final example draws on prose fiction, discussing Mahlberg's 2007 study of Charles Dickens's recurrent long phrases in twenty-three of his novels.

As the first of the more specialised chapters, Chapter 6, 'Metaphor and Metonymy', commences with a mention of Lakoff and Johnson's influential *Metaphors We Live By*, which has motivated increased interest in metaphor since its 1980 publication. Lindquist provides fine definitions of metaphor and the related concepts of simile and metonymy, being particularly detailed in his explanation of the first term. The rest of the chapter is then devoted to using corpora for researching metaphors. Three different procedures are presented (starting with the source domain, starting with the target domain, starting from a manual analysis) and give students a good understanding of some of the options in corpus-based metaphor studies. What disappoints however, is that the chapter - entitled 'Metaphor and Metonymy' after all - does not serve readers interested in the second literary term mentioned. Not even the Further Reading section includes any recommendations for any students wanting to find out more about corpus-based studies of metonymy.

Chapter 7 illustrates the possibilities that corpora offer for studying grammar. Lindquist states that, unlike areas explored up to this point, which involved searching for individual words or strings of words, investigating morphology and particularly syntax in a corpus can be more complicated. This observation is somewhat of a simplification, as research done in other areas (e.g. metaphors) is not always word-based. Lindquist includes different sample studies on pronouns (who/whom), get-passives, adjective complementation, et cetera, most inquiries being diachronic in nature and some comparing American and British usages. Although Lindquist primarily presents studies done by other researchers, he also replicates them - providing helpful step-by-step instructions – and critically discusses differences in results, setting students up for good practice. However, all examples involve querying for individual items and a grammatically specified search (Nelson et al.'s 2002 study using a parsed corpus) is only briefly referenced at the end of the chapter. This means that readers unfortunately do not get any practice in non-word based searching, though the omission may be the result of parsed corpora being few in number and often not publicly accessible.

The next chapter, 'Male and Female', investigates the application of corpora in sociolinguistics, specifically in relation to gender-specific differences in language. The usefulness of corpus metadata such as a speaker's social class, educational level and age are highlighted. Lindquist immediately also notes that the availability of such data is lacking in most corpora and, therefore, that possibilities for sociolinguistic research are still quite limited. The chapter then looks at a number of different studies investigating gender both in terms of how men and women talk and are talked about, focusing, as in the previous section, on diachrony. As in other parts of the book, plenty of data tables are included, and inconclusive results are inspected. The chapter also offers helpful suggestions for studies that are beyond the scope of the book but may be of interest to readers seeking ideas. The "Further Reading" section points out that though numerous resources on language and gender exist, few are corpus-based.

After many examples exploring language change throughout *Corpus Linguistics and the Description of English*, Chapter 9 is exclusively dedicated to the topic. It commences with the essential explanation of 'synchronic' and 'diachronic' perspectives. The focus is, of course, on the latter, and a distinction is made between the two major ways to study change in language through corpora: the study of 'change in real time' and 'change in apparent time'. The difficulty of identifying causes of language - whether internal or external - is considered, and many sample

studies are discussed. The most notable of these is perhaps the study presented in section 9.4, which, instead of relying on modern and historic corpora, uses the online Oxford English Dictionary (OED) as a source for data. With its very particular design and an uneven coverage, the OED is not a source without problems, but with entries going as far back as one thousand years, it may be an interesting option for some students. Citing this study further underlines Lindquist's aim to be as exhaustive as possible, something also evinced in Chapter 10.

The final chapter ventures into territory not treated in most older, introductory books on the subject: the world wide web. It is an area that has only recently started to gain traction, but one, as Lindquist notes, that fills certain gaps as for some linguistic research 'standard corpora, even if they contain 100 million words or more, do not provide enough data' (187). A useful distinction is made between 'web as corpus' (using searching engines to trawl the web as a corpus) and 'web for corpus' (using the web as a resource to create a corpus). The chapter covers ways for using the web in corpus linguistics and includes sample studies such as Mair's 2007 research on preposition use in different regional varieties of English. While the advantages of the web for corpus linguistics (text types not found elsewhere, quantity of results, et cetera) are touted, Lindquist also considers drawbacks and issues (replication, biased random sampling, lack of linguistic annotation), concluding that 'an important part of corpus linguistics in the future will be web-based in one way or another'" (205). Corpus research using the web has its doubters, but Lindquist's conclusion is realistic, although the intended audience for *Corpus Linguistics and the Description of English* probably least needs to be convinced of this.

Verdict
Corpus Linguists and the Description of English provides an introduction to corpus linguistics that is highly accessible for university students of English at different levels. The book meets the goals that it set for itself and is very much a hands-on guide with a multitude of sample studies and clear step-by-step instructions. Exercises in every chapter allow readers to check their understanding of concepts introduced and provide them with the opportunity to actually query corpora themselves. In terms of content, the book – a slim volume – manages to be surprisingly comprehensive, presenting a wide range of topics, including some options (OED as corpus, the web as corpus) that make it an updated introduction to a still evolving field. Commendably, linguistic as well as literary applications of specific methods are discussed. A more thorough exploration of topics would have been useful on occasion, as in Chapter 6 for corpus-based approaches of metonymy. However**,** there is little else to criticise, and the following suggestions are more of a wish list. A glossary would be a welcome addition for future editions as students new to the field would surely find a checklist for all the specialist terms very helpful – many terms are introduced in *Corpus Linguistics and the Description of English* and this can quickly feel overwhelming. A key to exercises should also be included, even when the tasks set are as straightforward as in this volume. The website of Corpus Linguistics and the Description of English (http://www.euppublishing.com/series/ETOTELAdvanced/Lindquist) is already a wonderful supplement to the book, but more thorough use could be made of it: all web-based resources mentioned by Lindquist should be listed there, including the online tutorials for statistics, which are currently only vaguely referenced. With print editions listings of online resources can be problematic as links quickly become outdated or inactive; however, the supplementary website for this book means that it should be fairly easy to keep recommendations current.

All things considered, Lindquist's *Corpus Linguistics and the Description of English* is an excellent book for anyone wishing to become acquainted with corpus linguistics and its wide range of applications. As it can be read either from cover to cover or perused selectively, it is suitable for many types of readers. The book will doubtlessly be appreciated by individuals with little or no background in the subject and, because it succeeds at transmitting that 'joy and fascination' (1) and provides plenty of ideas for new projects, it is bound to inspire students to explore the field

further, in exactly the way that suits them best.

A shorter version of this review originally appeared in the LINGUIST List at http://linguistlist.org/issues/21/21-3562.html

Bibliography

Gries, Stefan Th. "Some proposals towards a more rigorous corpus linguistics," *Zeitschrift für Anglistik und Amerikanistik*. 54.2. 2006. 191-202.

Lakoff, George and Mark Johnson. *Metaphors We Live By*. Chicago, London: University of Chicago Press, 1980.

Mahlberg, Michaela. "Corpus stylistics: Bridging the gap between linguistics and literary studies," *Text, Discourse and Corpora*. Eds. Michael Hoey, Michaela Mahlberg, Michael Stubbs and Wolfgang Teubert. London: Continuum, 2007. 219-246.

Mair, Christian. "Change and variation in present-day English: Integrating the analysis of closed corpora and web-based monitoring," *Corpus Linguistics and the Web*. Eds. Marianne Hundt, Nadja Nesselhauf and Carolin Biever. Amsterdam: Rodopi, 2007. 233-247.

Nelson, Gerald, Sean Wallis and Bas Aarts. *Exploring Natural Language Working with the British Component of the International Corpus of English*. Amsterdam: Benjamins, 2002.

Palmer, Harold E. *Second Interim Report on English Collocations*. Tokyo: Kaitakusha, 1966 (1933).

Sinclair, John. *Corpus Concordance Collocation*. Oxford: Oxford University Press, 1991.

Sinclair, John, Susan Jones and Robert Daley. *English Collocation Studies: The OSTI Report*. Ed. Ramesh Krishnamurthy. London: Continuum, 2004.