Randi Reppen, Susan M. Fitzmaurice, and **Douglas Biber** (eds.). *Using corpora to explore linguistic variation*. Studies in Corpus Linguistics. Amsterdam and Philadelphia: John Benjamins Publishing Company, 2002. xii +274 pp. ISBN 90-272-22279-7. (Eur.) / 1-58811-283-7 (US). Reviewed by **Nelleke Oostdijk**, University of Nijmegen.

Using corpora to explore linguistic variation opens with an introductory chapter by the editors. They describe the linguistic scene, providing the background that ties together the papers presented in the present volume. They also explain the organizational principles adopted. In all, the introductory chapter is very informative and highly illuminating, as it clarifies what the book is about, how it is organized, and why it comprises the papers it does. The editors characterize the book as "a collection of papers that illustrate ways in which linguistic variation can be explored through corpus-based investigation" (p. viii). The organization of the book has been guided by the primary research questions addressed in the respective papers. Thus each of the papers in part I "focuses on the use of a particular linguistic feature (a single word, a set of related words, a grammatical construction, or the interaction between particular words and grammatical structures", while the papers in part II typically focus "on the overall characteristics of language varieties, either a single dialect or register, or the similarities and differences among a range of dialects/registers." (p. viii). In the third and final part, the same perspectives are applied in a historical context.

Below I briefly describe the contents of each of the papers in the three parts, before I go on to discuss the book along more general lines.

Part I: Exploring variation in the use of linguistic features

1. Deanna Poos and Rita Simpson, 'Cross-disciplinary comparison of hedging. Some findings from the Michigan Corpus of Academic English'
While Lakoff (1975) claims that hedging is one of the qualities of feminine speech, others have failed to find evidence for this claim. In their paper, Poos and Simpson investigate to what extent hedging is related to gender differences. More specifically, their research focuses on the use of *kind of* and *sort of* as prototypical examples of hedging devices in academic spoken English. Their analysis shows that academic discipline is a stronger predictor for the occurrence than gender, while the functions of these devices are rather diverse. Thus, apart from

expressing inexactitude, *kind of* and *sort of* may be used, for example, to soften the force of a stance or opinion, or to mitigate a criticism or request.

2. Fiona Farr and Anne O'Keefe, '**Would** *as hedging device in an Irish context: An intra-varietal comparison of institutionalised spoken interaction'*

Like Poos and Simpson, Farr and O'Keefe are also concerned with hedging. Their perspective, however, is rather different, as they look at the socio-cultural context as a factor in explaining why speakers hedge in discourse. Following an analysis of the hedging involving the use of *would* that occurs in two institutional face-to-face interactions in an Irish setting, they arrive at a tree-tiered model for the analysis of spoken interaction.

3. Michael McCarthy, 'Good listenership made plain: British and American non-minimal response tokens in everyday conversation'

McCarthy examines listeners' responses in exchanges between speakers in everyday conversations. In his research, he focuses on the role of adjectives and adverbs "which typically occur at points of speaker change in every talk, and which either account for the whole of the listener response or are the first item in the listener response" (p. 49). An examination of two varieties, viz. British and American spoken English, shows that 'good listenership' involves that the listener takes on an active role not only in acknowledging what the speaker says, but also in investing in what McCarthy describes as the relational aspects of discourse, creating and maintaining sociability and affective well-being in their responses.

4. Graeme Kennedy, 'Variation in the distribution of modal verbs in the British National Corpus'

Kennedy's large scale study of the distribution in the BNC of modal verbs and the verb phrase structures they occur in, confirms the findings of earlier studies which were based on smaller and/or less representative corpora. The analysis of some 1.45 million occurrences of modals shows that there is great deal of variation in their distribution in different genres and media. The use of different modals varies, depending on the meaning the modal carries, the texts and the genre it occurs in (spoken or written), the structure of the verb phrase, and whether the verb phrase is affirmative or negative. At the same time, however, the use of modals in complex verb phrase structures is found to be quite stable.

5. Ferdinand De Haan, 'Strong modality and negation in Russian'

In his study of modality and negation in Russian, De Haan examines the relation between the scope of negation and modality on the one hand and syntactic position on the other hand. The modal system in Russian is not as grammaticalized as in English and also the sentence structure is different. Modality in Russian is defined by its meaning, rather than the syntactic characteristics. On the basis of the results obtained De Haan reaches the conclusion that "it would appear that languages go from a syntactic approach (where placement of the negation in the sentence determines its scope) to a semantic approach (where the scope of the negation is determined by the modal verb)" (p. 108).

6. David Okey, 'Formulaic language in English academic writing: A corpus-based study of the formal and functional variation of a lexical phrase in different academic disciplines'

Over the years, the existence of ready-to-use strings (referred to as prefabricated strings, lexical phrases, etc.) has been acknowledged in many studies. Okey in his paper undertakes to "provide a clearer, less intuitive insight to these units" (p. 111). He uses a subset of the BNC to investigate the use of the lexical phrase *it is/has been (often) asserted/believed/noted that X* as it occurs in academic writing in the fields of social science, medicine and engineering. Apart from the topic priming function, four other discourse functions are identified that are associated with this lexical phrase.

7. Viviana Cortes, 'Lexical bundles in Freshman composition'

Lexical bundles as defined in Biber et al. (1999) are extended collocations, i.e. sequences of three, four, five or six words that statistically co-occur in a register. Cortes investigates the occurrence of four-word lexical bundles in the writing of freshman university students. Her findings do not confirm her working hypothesis, which predicts that the bundles used by the students probably resemble more closely the bundles found in conversation than those found in academic prose. Instead, students seem to "closely imitate" the most frequent bundles encountered in academic prose. However, a careful analysis of the findings reveals that there are pervasive differences in the way that freshman students use these bundles.

8. Charles Meyer, 'Pseudo-titles in the press genre of various components of the International Corpus of English'

In his paper, Meyer presents an analysis of the occurrence of pseudo-titles across seven different regional varieties of English. Finding its origin in Ameri-

can English press reportage, the use of pseudo-titles has spread to other varieties of English, including British English and New Zealand English. While the use of pseudo-titles in American English is considered unmarked, in British English it is stigmatized (pseudo-titles are found to occur mainly in tabloids; in more formal newspapers they are generally prohibited). Meyer's findings lead him to observe that "the spread of pseudo-titles in press writing not only shows that a grammatical construction can be borrowed from one variety to another but that once the construction is borrowed, the constraints on its usage can change, leading to new forms." (p. 148).

9. Susan Hunston, 'Pattern grammar, language teaching, and linguistic variation: Applications of a corpus-driven grammar'

Following a concise introduction to the principles of pattern grammar, Hunston presents an interesting discussion on the merits of this type of grammar and its application to the study of language variation on the one hand, and language teaching on the other. It is claimed that pattern grammar is "an approach to language which maintains the generalising characteristics of grammatical descriptions while prioritising the behaviour of individual lexical items" (p. 167). The discovery of patterns – a pattern is defined as "a sequence of grammar words, word types or clause types which co-occur with a given lexical item" (p. 169) – benefits from the availability of large corpora such as COBUILD, although, as Hunston is careful to point out, intuition also comes into play in this, as the co-occurrence of lexis and pattern is not random but is associated particularly with meaning, while this association is not predictive. Moreover, there is evidence that patterns change over time.

Part II: Exploring dialect or register variation

10. Chandrika Rogers, 'Syntactic features of Indian English: An examination of written Indian English'

Rogers investigates three syntactic features that have previously been identified as characteristic features of Indian English. They are: use of the progressive with stative verbs, use of the present tense and the past perfect, and use of prepositional verbs. Her present study, which is based on the use of the stative verbs *have*, *know*, *want*, *like*, *hear* and *look* in an 800,000 word corpus of written Indian English, does not confirm earlier findings. In comparison with British and American English, in the Indian English data the progressive is more frequent in general, i.e. not specifically with stative verbs. The corpus comprises

insufficient data to draw conclusions on the use of the present and past perfect. Rogers suggests that an investigation of spoken data might yield rather different results. With respect to the use of prepositional verbs and patterns of preposition use, the data show Indian English to be markedly different from British and American English.

11. Eniko Csomay, 'Variation in academic lectures: Interactivity and level of instruction'

Csomay sets out to investigate the linguistic characteristics of academic lectures as they actually occur in real settings (as opposed to experimental settings which have been used in earlier studies). The present study involves 23 features that have been identified in Biber (1988) as characteristic of academic prose and conversation. An analysis of data from 176 lectures taken from the T2K-SWAL Corpus brings to light two situational parameters that have an effect on the linguistic features present in the lectures, viz. the degree of interactivity and the level of instruction.

Part III: 'Historical variation'

12. Susan Fitzmaurice, 'The textual resolution of structural ambiguity in eighteenth-century English: A corpus linguistic study of patterns in negation'

Within a context in which two grammatical systems for the formation of negative clauses co-exist, Fitzmaurice investigates whether this co-existence potentially gives rise to ambiguity and, if so, how speakers deal with this. The older of the two systems is the *do*-less one in which the main verb is followed by *not*, while the newer system is the one that uses *do*-support. The older system is understood to be recessive. An in-depth study of the different patterns in which the negative can occur reveals that the two systems occur side by side without the older system getting in the way of the newer one.

13. Christer Geisler, 'Investigating register variation in nineteenth-century English: A multi-dimensional comparison'

Geisler follows in Biber's footsteps in his multi-dimensional analysis of the development of English registers through the nineteenth century, a period which in other studies so far has largely been neglected. Adopting the sets of co-occurring grammatical features identified in Biber (1988) and the four dimensions associated with these, Geisler investigates the development of seven registers over three time periods: 1800-1830, 1850-1870, and 1870-1900. His findings

show that some of the registers are rather heterogeneous. The results obtained in this study only in part confirm the findings for other time periods.

Using corpora to explore linguistic variation is a book that clearly belongs in the tradition of what can be characterized as 'the Biber school', although in some contributions also the influence of Sinclair's work is apparent. With one or two exceptions, all papers build upon and extend previous research carried out by Biber and others (especially his work published in Biber 1988 and 1995, but also the joint publications with Finegan on historical English, incl. Biber and Finegan 1989, 1992 and 1997), while frequent reference is also made to the *Longman grammar of spoken and written English* (LGSWE 1999). Biber's research on the dimensions of linguistic variation and the linguistic features that characterize these, together with the findings published in the LGSWE with respect to the frequency and distribution of linguistic structures, lexical bundles, etc. is highly influential and pervasive in the research presented in the papers in this volume, as much in the research questions that are being investigated as in the (methodological) approach adopted.

Central to all papers is their use of corpus data. Some of the research reported on in this volume clearly benefits from the availability of (relatively) novel resources, such as the Michigan Corpus of Academic Writing (MICASE), the T2K-SWAL Corpus and CONCE (Corpus of Nineteenth-Century English). Although some researchers explore well-known corpora such as the BNC, the International Corpus of English (ICE) or COBUILD, others for their specific research find a need to compile special data collections. This seems to support the claim that, although there are so many corpora available already, still more corpora are needed.

While the editors in their introduction describe the methodological challenges that researchers encounter in analysing the influence of contextual factors on linguistic variation, the authors of the individual papers should be complimented on their work: without exception, they appear to have a strong awareness of the methodological sanity of what they are doing. They are quite ready to point out any limitations of the data they have used, or to identify possible flaws or defects of their investigative approach. All the papers in this volume report on substantial work. There is ample reference to the linguistic literature, and much care is taken to relate the present findings to results obtained in earlier studies.

Most of the authors of the papers included were present at the Second North American Conference on Corpus Linguistics and Language Teaching held at Northern Arizona University in Flagstaff, Arizona in the spring of 2000. This

might explain why many papers also pay (some) attention to the implications their results (may) have for pedagogical applications and teaching methods. For the time being, I think, the papers contribute to raising an awareness of different aspects of linguistic variation. Before the results presented here can be put to any practical use, however, much more work will yet have to be done.

References

Biber, Douglas. 1988. *Variation across speech and writing*. Cambridge: Cambridge University Press.

Biber, Douglas. 1995. *Dimensions of register variation*. New York: Cambridge University Press.

Biber, Douglas and Edward Finegan. 1989. Drift and evolution of English style: A history of three genres. *Language* 65: 487-517.

Biber, Douglas and Edward Finegan. 1992. The linguistic evolution of five written and speech-based English genres from the 17th to the 20th centuries. In M. Rissanen, O. Ihalainen, T. Nevalainen, and I. Taavitsainen (eds.). *History of Englishes: New methods and interpretations in historical linguistics*, 688-704. Berlin: Mouton de Gruyter.

Biber, Douglas and Edward Finegan. 1997. Diachronic relations among speech-based and written registers in English. In T. Nevalainen and L. Kahlas-Tarkka (eds.). *To explain the present: Studies in the changing English language in honour of Matti Rissanen*, 253-275. Helsinki: Societé Néophilologique de Helsinki.

Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad, and Edward Finegan. 1999. *Longman grammar of spoken and written English*. London: Longman.

Biber, Douglas, Randi Reppen, V. Clark, J. Walter. 2001. Representing spoken language in university settings. The design and construction of the spoken component of the T2K-SWAL Corpus. In R. Simpson and J. Swales (eds.). *Corpus linguistics in North America*, 48-57. Ann Arbor MI: University of Michigan Press.

Lakoff, Robin. 1975. *Language and woman's place*. New York: Harper and Row.