

Reviews

Annelie Ädel and **Randi Reppen** (eds.). *Corpora and discourse. The challenges of different settings*. Amsterdam and Philadelphia: John Benjamins. 2008. 295 pp. ISBN 978-90-272-2305-0. Reviewed by **Mikko Laitinen**, Universities of Helsinki and Jyväskylä.

The past few years have seen the publication of a number of studies combining corpus linguistics with discourse analysis (Baker 2006; Biber, Connor and Upton 2007; Hoey *et al.* 2007). The time certainly seems to be ripe for such a methodological combination (Aijmer 2007). The editors of *Corpora and discourse*, Annelie Ädel and Randi Reppen, begin by acknowledging this recent change and cross-fertilization between the two fields. They point out that corpus linguistics seems to be a likely partner for co-operation since representative data samples, automated data retrieval and analysis offer a range of solutions for various subfields of linguistics.

Corpora and discourse brings together scholars exploring discourse at various levels, ranging between the prosodic and textual, and targeting a range of language varieties in different settings. The common methodological factor is that each contributor offers empirical examinations of actual patterns of language use, taking into account higher-level discourse structures or discourse organization and investigates linguistic forms quantitatively and qualitatively using systematically compiled collections of data.

The volume is composed of a brief introductory chapter by the editors and eleven contributions divided into four sections, followed by three types of index. Section 1 concentrates on explorations of academic discourse, and consists of three articles. In the first one, Steve Walsh, Anne O’Keeffe and Michael McCarthy explore markers of vague language in a range of speech events in academic settings, comparing them with data from everyday spoken language. The focus is on the non-lexicalized forms of vague category markers (such as *and so on*, *et cetera*, and *and things like that*), and they investigate these markers in a one-million word corpus of spoken academic English, the Limerick-Belfast Corpus

of Academic Spoken Language (LIBEL CASE). They first identify forms and functions of vague language in academic discourse and then proceed to compare these frequency-based findings with corpus results from everyday spoken language in the Limerick Corpus of Irish English (LCIE) and the Cambridge and Nottingham Corpus of Discourse in English (CANCODE). The data analysis sets out from the quantitative identification of vague markers in academic discourse and rough comparisons with spoken language. The results show that conversations contain a wider range of vague markers and in higher frequencies than academic discourse, which restricts itself to the use of certain forms only. They then move on to a close analysis of these forms using the framework of Self-Evaluation of Teacher Talk, developed by Walsh (2006), which separates four types of classroom mode, each with different forms and functions of vague words that can be linked with the various pedagogical micro contexts used in class.

Marina Bondi's contribution looks into the role of emphatics in academic discourse. She targets the uses of adverbs (focusing on *invariably*, *significantly*, and *undoubtedly*), investigating these in her two specialized corpora, which comprise all the English-language articles of ten journals in two disciplines (history and economics) published in 1999–2000 (pp. 35–36). The corpora are each 2.5 million words in size. In its methodology, Bondi's contribution seeks to complement comparative analyses of variation in academic texts by integrating corpus-linguistic methods with some of the tools used in discourse analysis. The actual analysis consists of two phases. In the first one, she uses quantitative methods to search for unusually frequent or infrequent uses of emphatic adverbs in one corpus when compared with the other one (i.e. a keyword analysis defined in Scott 2008). After identifying two sets of adverbs as keywords in her two corpora (p. 38), she continues to the second stage of her approach. In this phase, she makes use of a concordance-based analysis of adverbs and their semantic roles, their lexico-semantic patterns of collocation and semantic preference, and examination of their text-level pragmatic functions and positions in linear texts. Her results show considerable variation between the two disciplines not only in the stock of adverbs and their frequencies but also on the functional level, so that the textual data from economics, for instance, tend to be more frequently self-referential than those from history.

Tamsin Sanderson's long article, "Interaction, identity and culture in academic writing", also combines corpus-linguistic methods and the analysis of special corpora with research on academic discourse. Her analysis is based on SCEGAD (Synchronic Corpus of English and German Academic Discourse), a corpus of journal articles in which the texts have been drawn from five academic

disciplines in the humanities, written in American and British English, and in German. In addition to the language of the article, this corpus contains background parameters on the authors' gender, age and academic status. Sanderson's analysis deals with first and second person pronouns, and she begins with a quantitative analysis of her findings, making extensive use of the background parameters in the corpus and carrying out statistical significance testing and factor analysis. She then proceeds to qualitative analyses of the results, dividing all the instances of person reference in her corpus into eleven heuristically constructed discourse categories. The main finding of this thorough methodological scheme is that the 'I-taboo' tendency is present only in her corpus of German academic texts in the humanities, but not in the English and American counterparts (pp. 88–89).

The second section of the volume, "Exploring discourse in workplace settings", contains three contributions. Elaine Vaughan's data-driven contribution, "Got a date or something?", highlights the uses of humor and laughter at work. Her analysis of the lighter side of life is based on authentic institutional interaction by English language teachers in two contexts. Her corpus consists in part of recordings made among the English department staff of a public university in Mexico and in part of comparable data from a private language school in Ireland (p. 96). All the informants are native speakers of English representing speakers of different geographical and national varieties (American, English, Scottish, Irish, Ugandan, Jamaican, etc.). The recordings were carried out during meetings, each with roughly ten participants, the entire corpus consisting of around 40,000 words. After first discussing the complexities of identifying humour in spoken language, Vaughan examines the total of 73 humorous episodes located in the material (p. 103). Her results show that in material of this kind the use of various types of humour depends on the institutional roles of the discourse participants.

Lynne Flowerdew's article, "Determining discourse-based moves in professional reports", presents her top-down approach, combining a genre analysis of entire texts with corpus-based methods. She makes use of her 225,000-word corpus of sixty environmental recommendation-based reports that have been commissioned from various consultancy companies by the public administration in Hong Kong. Flowerdew's argument is that the reports follow the rhetorical pattern of Problem–Solution, evident through her keyword analysis (p. 118). The article presents her coding scheme accounting for the three layers in texts (Macrostructure, Genre structure and Textual patterning) (p. 119), which is then used to extract possible keywords reflecting the larger structural units in the texts. The data analysis is based on the textual distribution of two sets of word

(*problem/problems* and *impact/impacts*) that she relates to the coding scheme used to understand the textual macrostructure of her data.

Winnie Cheng and Martin Warren's contribution in combining corpora and discourse is an examination of the interface between the phraseological characteristics of language and intonational patterns of communication, using data consisting of a quarter of a million-word subcorpus of the prosodically transcribed Hong Kong Corpus of Spoken English (HKCSE). They combine discourse intonational patterns with word associations. The underlying idea is to test Brazil's (1997) claim that lexical bundles, in addition to carrying patterns of choice at the lexical, grammatical and semantic levels, have a context-dependent intonational value. Cheng and Warren test this notion by looking at the ten most frequently occurring phrases (both lexically and grammatically rich ones) in the corpus, and investigating their realization in tone units and their prominence (pp. 142, 144). Their analysis shows the emergence of discernible discourse intonational patterns, but more importantly, the data show that such patterns may change over time, as the status of phrases changes. The case in point is their illustration of the phrase *Asia's world city* (p. 143).

The third section of the volume consists of three contributions. In "Who's speaking?", Gregory Garretson and Annelie Ädel tackle the issue of hearsay evidentiality and the controversial question of bias and balanced reporting in newspaper articles reporting on politics during the 2004 presidential campaign in the U.S. Their thorough-going study investigates a corpus of newspaper articles collected in the U.S. and related to the 2004 campaign. This synchronic corpus contains nearly 1,800 articles published in the eleven highest-circulation daily papers, collected during the 30-day span leading up to election day. It consists of over 1.7m words, and includes over 40,000 tokens of evidentiality in the form of reporting words (pp. 162–163). The data were analyzed manually and automatically, making use of a coding taxonomy created for the various sources of evidentiality. Their results indicate a minor tipping of the balance favouring or disfavouring one candidate or the other in some papers, but overall, no statistically significant differences suggestive of bias in reporting.

Testing the usability of television as a source of spoken data for pedagogical ESL/EFL purposes, Paulo Quaglio's contribution, "Television dialogue and natural conversation", compares linguistic similarities and functional differences in dialogues from an American television comedy with a large corpus of spoken American English. The data sources are (a) transcripts of a situational comedy "Friends" (c. 600,000 words), made available for entertainment purposes by online fan clubs, and (b) the spoken component of the Longman Grammar Corpus (c. 4m words) (pp. 190–191). The results of his multidimensional analysis

show that the transcripts of *Friends* offer a “fairly accurate” representation of spoken conversation for language teaching purposes (p. 198). A detailed functional analysis of vague language on the one hand and expletives/taboo words on the other indicate that there are differences between the two data sources, which Quaglio mainly contributes to the nature of TV as a medium.

In her contribution, “A corpus approach to discursive constructions of a hip-hop identity”, Kristy Beers Fägersten introduces the reader to a corpus of message-board postings of hip-hop websites. Her study is an analysis of the various strategies used by writers posting messages through which they construct their identities as members of the hip-hop community. Her method is a corpus-based discourse analysis, and the data consist of the message boards of five websites devoted to hip-hop culture (p. 216). Fägersten’s analysis, based on data exceeding 100,000 words and making use of frequencies, keywords and collocations, identifies three strategies in identifying with this sub-culture: message openings and closings, the uses of expletives and context-specific slang, and a range of manifestations of creative verbal art.

The fourth section shifts the focus to exploring discourse through specific linguistic features. Christine Johansson takes a diachronic look at the uses of *it*-cleft constructions in 19th-century English. She investigates such clefts in two historical corpora, A Corpus of Nineteenth-Century English (CONCE) and the Helsinki Corpus of English Texts. Her results show that *it*-clefts became more frequent during the 19th century, particularly in such speech-related genres as trials, debates and drama, but surprisingly not in letters (pp. 252–253). Instead of highlighting speech style, she proceeds to a detailed investigation of information focus as the underlying reason for the frequency of *it*-clefts in the 19th century trials in CONCE.

William J. Crawford’s contribution, “Place and time adverbials in native and non-native English students writing”, investigates claims about the spoken nature of learner writing. After one-to-one data retrieval, he explores two English place adverbs (*here/there*) and two time adverbs (*now/then*), comparing them quantitatively and qualitatively in two corpora: the International Corpus of Learner English (ICLE) and the Louvain Corpus of Native English Essays (LOCNESS). He then compares the corpus findings with the frequencies of these sets of adverbs in published academic writing and face-to-face conversations as reported in the *Longman grammar of spoken and written English* (Biber *et al.* 1999). His results show that non-native writers tend to use these adverbs in frequencies which are closer to academic writing than conversation, whereas differences between native and non-native writers are not visible in relative fre-

quencies but instead in functional differences in making use of the range of functions of these place and time adverbs.

The value of this volume lies in the various ways in which the contributors make use of corpora in discourse analysis. Whereas some contributors actually seem to be doing corpus linguistics *per se*, describing language in use and linguistic forms in context (cf. the various ways of defining ‘discourse’ in Schiffrin, Tannen and Hamilton 2001: 1), many writers take a more ambitious approach and aim at benefiting from corpus-linguistic methods by investigating larger textual units beyond phrases and sentences. It is obvious that combining two methodological approaches that are inherently as distinct as quantitative corpus linguistics and discourse analysis of larger textual units in use is not always easy. At the same time, as many of the contributions in this volume show, such incorporation of the two methodologies can provide a number of insights into discourse phenomena that could be generalized across texts and other contexts. Such insights include, in particular, Sanderson’s quantitative and qualitative analyses of the various discourse patterns in academic journals (Chapter 3) and Quaglio’s comparisons between scripted television dialogue and natural conversation (Chapter 9).

In their brief introduction of this volume, Ädel and Reppen (p. 1) point out that discourse phenomena necessarily depend on both co-text and context and are highly sensitive to the researcher’s interpretations, requiring therefore plenty of interaction with the data, starting from the selection of material that actually represents the particular domain being studied (cf. Biber, Connor and Upton 2007: 17–18). Similarly, if one thinks of corpus linguistics and corpus compilation, which require that the researcher knows the language variety/varieties sampled and is aware of the limitations in terms of the kinds of generalizations that can be made, it is equally obvious that there are a number of issues in common between the two approaches. At the core of this enterprise is the question of retrieving discourse-relevant data, a question discussed by Ädel and Reppen (pp. 2–4) in their introduction, where they list possible methods of locating linguistic forms linked with a particular function in a corpus database. The ones included in their list are one-to-one searching of forms with clear functions, sampling examples of the linguistic phenomenon under scrutiny, sifting (manually pruning initial search results), and making use of frequency lists as the basis for contextual analysis, all of which must be familiar to anyone working with corpora in language description. Some of the contributors, Garretson and Ädel in particular (Chapter 8), actually tackle this question of how to decide and identify the units for analysis through automated coding.

The articles in this volume can offer a corpus linguist a number of insights, one of the most important of which is highlighting the analysis of specialized registers and contexts. As is apparent in these studies, the real value of combining corpora and discourse, at this stage at least, seems to be the focus on and investigation of relatively narrowly-defined and specialized genres. Discourse analysis and explorations of discourse organization are typically based on a detailed analysis of individual texts, so the research starts out by identifying an interesting discourse phenomenon in a more or less isolated data source. The research process can then lead to the collection of a corpus database of a number of such data sources, eventually ending up with a versatile range of texts and data, which in turn can offer a fresh perspective on the general textual population sampled and included in corpus databases. So instead of (occasionally) relying too much on the canonized text types in the large well-known corpora, the contributions in this volume have selected and used data from highly interesting sources. A case in point is for instance Kristy Beers Fägersten's (Chapter 10) data of message-board postings by the subculture of hip-hoppers. Alternatively, in Quaglio's article (Chapter 9), the primary material consists of transcripts of popular sitcom shows, made publicly available by devoted fans all over the world, which are then used for research purposes. Both are prime examples of innovative solutions that can enrich corpus linguistics: if databases are sizable enough and if quantitative methods are used to analyze such representative data, the results should be generalizable across texts.

References

- Aijmer, Karin. 2007. Review of *Using corpora in discourse analysis* by Paul Baker. *ICAME Journal* 31: 170–173.
- Baker, Paul. 2006. *Using corpora in discourse analysis*. London and New York: Continuum.
- Biber, Douglas, Ulla Connor and Thomas A. Upton. 2007. *Discourse on the move: Using corpus analysis to describe discourse structure*. Amsterdam and Philadelphia: John Benjamins.
- Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad and Edward Finegan. 1999. *Longman grammar of spoken and written English*. Harlow: Pearson Education.
- Brazil, David. 1997. *The communicative value of intonation in English*. Cambridge: Cambridge University Press.

- Hoey, Michael, Michaela Mahlberg, Michael Stubbs and Wolfgang Teubert. 2007. *Text, discourse and corpora: Theory and analysis*. London: Continuum.
- Schiffrin, Deborah, Deborah Tannen and Heidi E. Hamilton (eds.). 2001. *The handbook of discourse analysis*. Oxford: Blackwell.
- Scott, Mike. 2008. *WordSmith Tools help*. Liverpool: Lexical Analysis Software. Available at [http://www.lexically.net/downloads/version5/HTML/proc_tag_handling.htm] (last accessed on 15 Jan. 2009).
- Walsh, Steve. 2006. *Investigating classroom discourse*. London: Routledge.

Sonja Adolphs. *Corpus and context. Investigating pragmatic functions in spoken discourse* (Studies in Corpus Linguistics 30). Amsterdam and Philadelphia: Benjamins, 2008. x + 151 pp. ISBN 978-90-272-2304-3. Reviewed by **Anna-Brita Stenström**, Bergen University.

This book highlights the crucial role of spoken corpora for analysing and understanding the relationship between lexico-grammar, utterance function and discourse context. A new approach to the analysis of utterance function is developed, based on orthographic transcripts from the five-million-word Cambridge and Nottingham Corpus of Discourse in English (CANCODE). The focus of the book is on lexico-grammatical strings that have traditionally been connected with a particular speech act, here strings that realize suggests, such as *Why don't you* and *Why not*. It challenges the distinction between direct and indirect speech acts, arguing for a reconsideration of indirectness based on analyses within a larger discourse-based framework. Moreover, it queries whether pragmatic theory is really necessary and useful for understanding utterance function and could be replaced by a framework based on the extraction of text-internal patterns. Since orthographically transcribed data alone is insufficient for an adequate description of utterance function and perception, the last chapter is devoted to multi-modal, i.e. video-recorded, corpus data. All the chapters pay ample attention to previous work related to the areas dealt with.

Chapter 2 starts by an overview of the traditional approach to the interpretation of language function and the difference between pragmatics and corpus linguistics. The main difference between the two is said to lie in the way they interpret meaning. Whereas pragmatics assumes that function and meaning can

be separated, corpus linguistics assumes that the two are inseparable, which leads to different methods of describing language. This chapter maintains that the distinction between semantic and pragmatic meaning is unnecessary, if patterns of collocation, colligation and discursal placement are taken into consideration. Special attention is paid to the concept of ‘routinisation’, defined as “the recurrent use of a particular speech act expression to realise a particular function” (p. 28). Routinisation can give rise to conventional indirect speech acts, which also rely on the context for their interpretation, however. This puts special demands on the particular corpus used. CANCODE, which has been classified in terms of *context-type*, based on the degree of intimacy or distance between the speakers, and *interaction type*, is said to fulfil such demands.

Chapter 3, which deals with the act of suggesting, argues that, although certain forms can realize this particular speech act function, this is not always the case, judging by the surrounding discourse. But by studying corpus data, it is possible to discover whether certain forms are related to certain functions. It is emphasized that a broad functional category has to be defined as an organising principle before considering the lexico-grammatical strings that typically realize a particular speech act. The parameters said to form the functional profile of a speech act expression consist of collocation, functional distribution and contextual distribution. Influenced by the concept of ‘semantic prosody’ introduced by Sinclair (1996) for describing speech act expressions, this chapter proposes the notion of *functional prosody* to enable a description of speech act expressions judging by their patterns and distribution. It is pointed out that, since the profile of a speech act expression often depends on the preceding discourse, a concordance line is not enough for identification. Collocations, i.e. regularly cooccurring lexical items, are said to be useful for establishing the prototypical function of a particular speech act expression. The study of collocations showed for instance that the main function of *Why don’t you* is suggesting and not questioning. Two ways of disambiguation are proposed: substitution by *for what reason*, which makes it into a question, and by the type of response, which is more detailed if a question precedes. In addition, prosody can have a decisive effect as well as the type of verb that follows. The study also showed that speech act expressions are often accompanied by ‘internal modifiers’, such as the down-toner *just*, or vague language (e.g. *kind of, or something like that*), which modify the propositional content of the utterance, or by various modal expressions with a modifying effect (e.g. the verb *think* or the adverb *perhaps*).

The relationship between speech act expressions and the situational context is examined in more detail in Chapter 4, which first outlines how contextual categories might influence the functional profile of a speech act expression and

then goes on to discussing this relationship with reference to previous approaches. Special emphasis is put on the importance of corpus design for judging the robustness and validity of the functional profile of a particular speech act expression. A comparison of the frequency of the speech act expressions investigated in relation to the different speaker relationship configurations in the corpus, in terms of intimate, socio-cultural, professional, transactional and pedagogic, shows that *Why don't you* is roughly twice as common as *Why not* and *How about* in the intimate category (where the speakers know each other well), while the speech act verb *suggest* is typically used in the pedagogic category (e.g. tutorial, lecture) and least often in the intimate category. The sample analysis in this chapter is represented by the two speech act expressions *why not + Verb* and *how about + Verb*, which have traditionally been regarded as suggestions. The study shows, however, that the main function of both expressions is that of a question, judging by their use in casual conversations between close friends. All in all, the speech act episode is said to be a suitable starting-point for discussing the relationship between lexico-grammar, discourse and genre.

Chapter 5 considers the effect of a corpus-based description of pragmatic functions and outlines a preliminary framework. The starting-point is the traditional view of the relationship between particular lexico-grammatical strings and specific speech act functions, which proves to be inconsistent with corpus data. It is pointed out that a systematic account requires a dynamic model that describes processes rather than products and reveals the interactive nature of the exchange. The unit of analysis discussed is the suggestion episode, with the main focus on 'agency' (expressed through the relationship between the speaker who elicits a suggestion, the participant who makes the suggestion and the person who is named as the agent) and the 'time frame' of the suggested action. Special attention is paid to (1) goal-types and related text-types, (2) agency and (3) the sequential organisation, which can be broken down into episodes. Goal-type and text-type are said to be strongly related, although different aspects of generic activity are addressed. The role of agency is to mark the relationship not only between context-type and goal-type but also between context-type and text-type.

Replies to suggestions are given special attention. All in all, responses to speech acts are said to have received little attention by conversational analysts. Burton's (1980) 'supporting' and 'challenging' moves are said to suit everyday conversation better than the responses in Sinclair and Coulthard's (1975) framework, developed for classroom interaction. A supporting move supports either the topic (the ideational aspect) or the speaker (the interpersonal aspect), while the challenging move holds up the progress of the episode. Other response types

are clarifying moves, which ask for clarification, and detaching moves, which seek to terminate the interaction. Here, the status of backchannels is problematic. The question whether they should count as supporting moves or not leads to the conclusion that minimal responses that only reflect attention and acceptance of information, which is the role of backchannels, should not be regarded as supporting moves, only those that explicitly support the topic or the speaker.

As was stated in the introduction to the book, reliance on orthographic transcriptions is not enough to capture what is going on in spoken interaction. An accurate analysis of the functions of speech act expressions and episodes relies heavily on intonation and body language, a type of analysis that is largely missing to date. Not only is there a need for an 'integrated' approach to the analysis of conversation, which would be made possible with access to a video-recorded corpus, but also a need for mark-up tools for coding both verbal and non-verbal elements. These aspects are considered in Chapter 6, which discusses the difficulties involved in collecting a multi-modal corpus and the advantages of using multi-modal conversational data. The advantages of using a multi-modal corpus are illustrated by an analysis of backchannels. A framework based on Irish and British English multi-modal data is provided for classifying the various back-channel functions, which reflect different levels of speaker engagement, from continuer (*mm*) to information receipt (*okay*). The role of accompanying head movements and intonation patterns is emphasized: they can not only change the function of the backchannels but also affect the surrounding discourse. It is pointed out that fully searchable multi-modal spoken corpora are gradually becoming available thanks to today's advanced technology.

Chapter 7, finally, emphasizes again the crucial importance of corpora for the description of discourse patterns and speech act expressions and also "for testing and evaluating existing theories and claims made about the way in which they are used." (p. 131). But certain limitations are also pointed to. One is that only form but not function can be searched for in a corpus, a second that enough evidence for a robust description of lexico-grammatical strings requires a much larger corpus. The fact that spoken corpora are often analysed in the same way as written corpora, which of course excludes intonation and gestures, is put forward as yet another limiting factor. The chapter ends with a section proposing future challenges. One is to define an adequate unit of analysis for pragmatic functions, and the analysis of speech act expressions is suggested as a good starting-point. Another challenge is to collect and process representative multi-medial spoken corpora that are large enough for the description of pragmatic functions.

The book was a pleasure to read, and I have very few critical comments. My only ‘serious’ objection refers to the lack of a diagram that gives a clear and concise summary of the preliminary model described in Chapter 5 for analysing suggestion episodes. I found the model adequately described in the text and aptly illustrated by examples. Still, a concise summary in the form of a figure or diagram would have been a helpful supplement. In this chapter it is pointed out that responses to speech acts have been paid very little attention. Exceptions are for instance Stenström (1984, 1994), which also show that a modified version of Sinclair and Coulthard’s model for analysing classroom interaction works very well for casual conversation. One thing that is not entirely clear is the relation between frequency and corpus size. CANCODE is said to consist of five million words, but on page 57 there is a reference to a subcorpus consisting of 2.5 million words. It is not clear whether this is the only place where a subcorpus is used. If it is, this would mean, for instance, that *Why don’t you* occurred only 182 times (p. 61) in the entire five-million-word corpus, a figure that seems very low. Finally, a minor point concerning the layout. Figure 1 (p. 75), representing the frequency of different speech act expressions and their function, is slightly bewildering in that the speech act term ‘suggest’ itself is quite unexpectedly represented by a bar in the bar chart.

Irrespective of the above comments, this book, which is a comprehensive account of the development of pragmatics and its relation to corpus-based research, is an excellent demonstration of the importance of spoken corpora for interpreting utterance functions in relation to lexico-grammatical strings, and not least of the importance of multi-media spoken corpora. The framework developed here for analysing suggestions is likely to be applicable to the analysis of a number of other speech act expressions. All in all, the book can serve as input to more research aiming at answering the question put forward in the introduction, whether pragmatic theory is necessary, or useful, for understanding utterance functions, or whether corpus-based analysis can provide a framework for describing such functions.

References

- Burton, Deirdre. 1980. *Dialogue and discourse: A sociolinguistic approach to modern drama dialogue and naturally occurring conversation*. London: Routledge and Kegan Paul.
- Sinclair, John. 1996. The search for the units of meaning. *Textus* 9 (1): 75–106.
- Sinclair, John and Malcolm Coulthard. 1975. *Towards an analysis of discourse*. London: Oxford University Press.

Stenström, Anna-Brita. 1984. *Questions and responses in English conversation*. Malmö: Liber.

Stenström, Anna-Brita. 1994. *An introduction to spoken interaction*. London: Longman.

Stefan Dollinger. *New dialect formation in Canada. Evidence from the English modal auxiliaries*. Amsterdam and Philadelphia: Benjamins, 2008. xxii + 355 pp. ISBN 978-90-272-3108-6. Reviewed by **Arja Nurmi**, University of Helsinki.

This volume sets out to do three things. First, it provides corpus-based evidence of the development of early Canadian English, which has so far been sorely lacking. Second, it tests Trudgill's (2004) theory on new-dialect formation. Third, more marginally, it examines Abraham's (2001) theory of root loss in modals in a new context.

The book starts with extensive background information on Canadian English. This overview shows clearly how patchy and lacking earlier research is, particularly with the historical stages of the variety. The majority of previous studies are focussed on phonology and lexis, and many are anecdotal, impressionistic descriptions of the 'peculiarities' noted by the authors. Because of the vast amount and varied nature of research summarised, covering the different levels of language from phonology to language attitudes, the chapter is somewhat disjointed. The introduction of the Chinook Jargon, for example, in the section discussing regional surveys of phonology, seems misplaced. But this is a minor concern, since this chapter is certainly a comprehensive overview of earlier research on Canadian English and, as such, a good starting point for anyone interested in the topic.

The next section of the book discusses the external language history of Ontario. Early Ontario English is, quite reasonably considering Canadian settlement history, the main focus of the study. The different waves of immigration from England, Ireland and Scotland, the role of migrants from America and the influence of other settlers are all described and discussed. Once again, presenting this much information in such a concise manner provides its problems, but the thoroughness has great value for the reader.

In Chapter 4 the Corpus of Early Ontario English is introduced. The corpus introduction is thorough in that it discusses the selected texts' Canadianness, choice of genres and sampling principles. The availability of any other genres as evidence of early Ontario English is not mentioned. On the whole, the corpus is well balanced and there is some attempt to take into account the dialect background of informants. Since the ARCHER corpus and the Corpus of Late Eighteenth-century Prose are used as comparison material, it would have made sense to introduce them in this context as well, rather than 50 pages later in the chapter discussing Late Modern English modals. In fact, for a reader not familiar with historical corpora of English, a slightly more thorough introduction of the corpora used might have been useful.

The following chapter describes the main theoretical underpinning of the volume, Trudgill's new-dialect formation. There are two main points where this study diverges from the scenario described by Trudgill. Firstly, Canadian English was not formed in a *tabula rasa* situation as many southern-hemisphere Englishes were, because, while there were no prior speakers of English in the territory studied, there was an existing English-speaking population south of the border, with its own variety and with constant communication and migration between the Ontarian and American settlements. Dollinger proposes that the situation in Ontario may be described as a *semi-tabula rasa*. Secondly, most of the studies on dialect formation focus on phonetic forms. This study attempts to apply the processes of new-dialect formation described by Trudgill to grammatical forms and their functions (i.e. modal auxiliaries).

The final chapter before the empirical part is an introduction to the modal auxiliaries in Late Modern English. This includes a brief discussion of types of modality and the general trends of the history of modals observed in earlier research. Here also Abraham's root loss theory is introduced.

The following four chapters present one selected group of modals each. These groups are in some ways treated as one variable, but this is, of course, difficult in the case of such a complex linguistic phenomenon as modality. The problems related to different types of modality and different linguistic conditions attached to each verb are discussed in detail. Each chapter starts with a description of the earlier history of the modals in question, starting from Old English and making frequent reference to Present-Day English as well. The range of earlier research is well covered, and includes all major corpus-based works. Each variable also has specific questions related to it, which are well accounted for. So, for example, in the case of *must/have to*, the fact that *have to* does not have epistemic uses in this time period means that the author discusses other possible expressions of strong epistemic modality as competitors of *must*

in this field. Similarly, in the case of *will* and *shall*, the contemporary normative grammars and attitudes are taken into account. Here the thorough background work comes well into use, since the grammars used in early Ontario are known, and their influence can be gauged.

The discussion in the empirical chapters constantly places the results gained from the Canadian material and the comparison corpora into the wider contexts of the volume: the development of Canadian English, the influence of American and British English, the different stages of new-dialect formation and Abraham's theory. It is unfortunate that the American data does not cover the whole time period studied, and that the amount of text written by British immigrants from the particularly interesting areas (Scotland, Ireland, Northern England) is so limited, but this is simply a matter of available data which all historical corpus linguists must learn to cope with. Dollinger has certainly made the most of the corpora used in his study.

The final chapter sums up the findings of the empirical part and presents the overall picture emerging from the study. It is clear that there is much new information on the development of English modal auxiliaries in the Late Modern period in all three varieties covered by the corpora. The theory of root loss also finds some support from this study, although it must be remembered that the expansion of epistemic uses has been a centuries-long process, which is only partially covered by this data. The only caveat to these results is that many of the trends and differences observed are not statistically significant, and in some cases there was too little data to draw any conclusions. On the whole, the trends observed seem to point in the same direction, however, which seems to lend even the non-significant results some credence.

The conclusion also weighs different hypotheses presented in earlier research concerning the focussing and development of Canadian English. Dollinger confirms that Trudgill's notion of colonial lag in the first generation of settlers seems to fit the data. This does not make early Canadian English conservative on the whole, however; rather it seems more progressive than British English. Most of the changes covered in the study seem to be cases of drift or parallel development in the three varieties compared, and this, in fact, seems to be the most important factor affecting the modal system. Apart from the shared developments, the influence of the Loyalist post-1776 settlers, who spoke American English, seems to have been clearly more important than British input in the early stages of Canadian English. Even independent Canadian developments outrank the influence of the mother country. This multi-layered picture of the various contributing trends is also a realistic description of the complexities of language variation and change.

As to the theoretically most wide-reaching question, the applicability and validity of Trudgill's new-dialect formation model, the results are promising. It seems that while some adaptations need to be made, not only because of the different linguistic environment from the one the model is based on, but also because of the very different level, the model itself has wider applications than the situation it was originally developed for.

The volume, despite its obvious merits, contains also many little things that make reading harder. Each little mistake in itself would be negligible, but the constant re-reading because of them is distracting. There are constant typographical errors and even several cases of missing words, obvious results of incomplete editing. One typical example of this is on page 130: "one cannot start with the concept that are varieties are ultimately descendants of one", where the first *are* is presumably meant to be *all*. Equally annoying is the information structure of the book, where many facts (those relating to Canadian history and geography, for example) are first mentioned as if they were already familiar to the reader, and explained only later. There are also cases where the reader is left guessing. With Figures 5.1 and 5.2, it is never made clear whether these figures, presenting spelling variation in American and British English, are based on the ARCHER corpus (which is only introduced much later), and if so, which parts of it. Since spelling in the Late Modern period varied considerably between published and private writing, this is a relevant concern. It is unfortunate that mistakes like these, easily corrected by a thorough editor, are left to make the reader's task that much harder.

While there are problems with the volume, its merits far outweigh them. We have gained new insight into the history of Canadian English and the process of new-dialect formation. Many questions are outlined for further research, both for the applications and refinement of Trudgill's model and the development of Canadian English, not to mention the paths of the various modal auxiliaries.

References

- Abraham, Werner. 2001. Modals: Toward explaining the 'epistemic non-finiteness gap'. In R. Müller and M. Reis (eds.), *Modalität und Modalverben im Deutschen*, 7–36. Hamburg: Buske.
- Trudgill, Peter. 2004. *New dialect formation. The inevitability of colonial Englishes*. Edinburgh: Edinburgh University Press.

Thomas Egan. *Non-finite complementation. A usage-based study of infinitive and -ing clauses in English* (Language and Computers 65). Amsterdam and New York: Rodopi, 2008. xii + 432 pp. ISBN 978-90-420-2359-8. Reviewed by **Nelleke Oostdijk**, Radboud University Nijmegen.

The book presents an account of a synchronic usage-based study of non-finite complement clauses in which the methods of corpus linguistics and the theoretical stance of cognitive linguistics are combined. The study investigates the distribution and meaning of infinitive and *-ing* participle clauses in Present-Day English that function as clausal objects or non-optional adverbials and seeks to explain their use in the contexts in which they occur. The study is based on data from the British National Corpus (BNC).

Before this study there have been many other studies directed at investigating verb complementation and/or the form and function of non-finite clauses. None of these studies, however, have offered the degree of comprehensiveness aimed for in the present study, nor can they compete when it comes to the scale of this study. Egan focuses on constructions containing active voice matrix verbs and infinitive and *-ing* complements. The two main types of construction containing active voice matrix verbs in English are same-subject (equi-subject, 'SS') constructions, in which the subject of the complement clause is identical to the subject of the matrix verb, and the different-subject ('DS') constructions, in which the subject of the complement predicate is profiled as distinct from the subject of the matrix verb. The study encompasses as many as 310-odd matrix verbs and is based on the evidence found for the various types of construction in the 100-million-word BNC.

The book comprises eight chapters, three appendices, two of which are concerned with data and another one which is a list of technical terms that are used, a list of references and an index. Once you get into the book, and particularly after having read the introductory chapter ("The purpose and the scope of the study") and Chapter 2 ("Classification of the constructions"), it is clear that the structure and organization of the book could not have been any different. After a discussion of the literature in Chapter 3, in Chapter 4 the various complement types and complementisers are described. In Chapters 5 through 7 the focus is on different forms of non-finite constructions containing either identical or near-synonymous matrix verbs. Chapter 8 summarizes the main objectives of the study and the conclusions arrived at. The order of the chapters is perfectly logical and the decision to have a selection of data as examples in the body of the text and further information in the appendices makes sense.

Right from the start the author takes the reader by the hand and embarks with him on an adventure, exploring the realm of non-finite complement clauses. The author acts as a guide and at the same time as a companion, presenting his ideas and findings, arguing his case and illustrating it with a great many (868!) examples, while occasionally sharing his hesitations as to the maintainability of particular analyses that have been offered previously in the literature. The study and the way in which it is presented is an invitation to the reader to join the author in his analysis. It is not until the concluding chapter that Egan reveals his intentions in this respect:

the reader has been presented in the course of the book with large amounts of real language data. He or she may well disagree with some of the author's interpretations of these data. One of my aims, indeed, has been to provide the reader with sufficient data to do just that. (p. 305)

The material presented, which comprises theoretical insights, data and analyses, provides ample food for thought. I mean this in a strictly positive sense. In the same way as when I say that I do not find it an easy book to read. This is not in any way a shortcoming of the author. It is simply because the subject matter is complex and requires your full attention, and the amount of data and the information presented is rather overwhelming. The author actually has done a commendable job in presenting insights from previous research and novel findings in a highly structured manner. Upon arriving at the summary and conclusions, the reader is likely to subscribe to Egan's observation that he has "been struck by how neat the distribution actually is, how form and function seem to go hand in hand, in an area sometimes said to be characterized by no little degree of chaos and arbitrariness" (p. 308).

The study stands out by its very sound methodological underpinning. Throughout the book the author makes an effort to clarify various methodological issues and the approach he has adopted. Key issues here are the nature of the study (synchronic, usage-based), the methods employed (for example in defining and delimiting the scope of the study and the selection of the data), and the theoretical stance (cognitive linguistics). In the next few paragraphs I will try to summarize the relevant points that are brought to bear.

The first issue is related to the nature of the present study and concerns the inherent limitations of a synchronic study. As the author points out, such a study considers only the present state of affairs, while not putting the findings in the broader perspective of on-going change in language. Previous work by, for example, Mair (2002) and Fanego (2004) has shown that the *-ing* form is still

advancing in certain contexts at the expense of the *to*-infinitive form. This begs the question what insights may be gained from a synchronic study investigating non-finite complement clauses. Egan puts it as follows:

So, given that the English non-finite system of complementation is still evolving and that this evolution is reflected in synchronic variation; and given the fact that the meaning of the complement forms will not necessarily be equally apparent in all instantiations; can we make any worthwhile generalisations about the meaning of the complement types and complementisers? (p. 90)

And his answer to this question:

I think that there is enough regularity in the usage in the BNC to answer this question in the affirmative. Provided we remain aware of the possible shortcomings inherent in the viewing things from a purely synchronic perspective, we may establish fairly robust characterisations of each of the complement types. (p. 90)

Related to this point is the question how to deal with instances that do not appear to fit in with the classification that is arrived at. The author has opted for including such instances rather than sweeping them under the carpet. Thus readers can judge for themselves whether they think they are sufficient evidence for overturning the proposed classification. Evidently Egan is not at all convinced that this is the case and is quite firm in stating that “exceptions like these are clearly just that – exceptions. They are so rare as to render it unnecessary to revise the classification proposed” (pp. 196–197).

The present study is a usage-based study in that it “aims to describe the structures of non-finite complement clauses in English as they are manifested in the actual experience of speakers and writers of present-day British English” (p. 3). It is also usage-based in the sense that the data upon which it is based derive from a corpus of contemporary British English which includes data from a wide variety of written and spoken genres and registers. The advantages are obvious: the corpus instances are guaranteed to be genuine and contemporary, while all utterances are contextualized. Moreover, all findings presented in this study may be verified by going back to data in the original data source. In this respect, the choice of the BNC is well-motivated as it widely available and widely used.

In defining and delimiting the scope of the present study, Egan has decided to focus on post-verbal infinitive and *-ing* participle complement clauses. *Wh*-clauses were excluded from the study on the grounds that as a class they are well-defined and can easily be recognized, while the alternation they display

between different complement types is minimal. The starting point for extracting relevant instances from the corpus was a list of matrix verbs compiled by Rudanko (1989) on the basis of information found in Visser (1963) and supplemented by further verbs taken from Quirk *et al.* (1985). For reasons of economy, the verb *seem* was excluded from the final list. The next step then was to extract for each of the verbs in the list instances where the verb occurs with a non-finite complement clause. Because of the size of the BNC, it was practically impossible to extract all instances for all verbs. Therefore, Egan decided to restrict his data set to a random selection of maximally 1,000 instances per verb. The data set includes both occurrences where the verb is actually followed by a non-finite clause and occurrences where this is not the case. From the frequencies and distributions of non-finite complement clause constructions that are observed in this data set, Egan arrives at what he refers to as projected totals for the BNC as a whole. Thus, very cleverly, without having actually extracted all instances for each construction type in this study, we gain insight as to their frequency and distribution. For example, just to give you an idea, for different-subject *-ing* clauses the projected total is some 35,000; for equi-subject *to*-infinitives the projected total is somewhere around 300,000.

The author has adopted a cognitive approach. This is clearly reflected in his interpretation of the data which is led by the cognitive-functional factors that come into play while considering different types of non-finite complement clauses. As Egan is careful to point out, it has not been his aim to determine “whether taking these features into account can help us provide a better account of the system of non-finite complementation in English than alternative approaches. The strengths and weaknesses of various other approaches of non-finite complementation were illustrated in Chapter 3. To the extent that the classifications and explanations account satisfactorily for the data, this may be taken as evidence of the efficacy of the cognitive approach.” (p. 307).

While usually I am not too keen on very bulky appendices, in this case the inclusion of Appendix 1 (“The Matrix Verbs”) which comprises as many as 87 pages is fully justified. The appendix lists all the matrix verbs that featured in the present study. For each verb we find information about the number of tokens of the verbs in the BNC, the various syntactic types of non-finite clauses they occur with, an indication of the frequency of each, their semantic classification in the present study, page references, and an example of each type.

There can be no doubt that the book has a great deal to offer for anyone interested in non-finite complement clauses. Personally, I am very much impressed with what Thomas Egan has achieved. The study is of exemplary

quality and Egan's attention to his data, his interpretations and discussion are a real treat.

References

- British National Corpus*. 2001. Oxford: Oxford University Computing Services.
- Fanego, Teresa. 2004. Is Cognitive Grammar a usage-based model? Towards a realistic account of English sentential complements. *Miscelánea. A Journal of English and American Studies* 29: 23–58.
- Mair, Christian. 2002. The changing forms of complementation in late Modern English: A real time study based on matching text corpora. *English Language and Linguistics* 6 (1): 105–131.
- Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech and Jan Svartvik. 1985. *A comprehensive grammar of the English language*. London: Longman.
- Rudanko, Juhani. 1989. *Complementation and case grammar: A syntactic and a semantic study of selected patterns of complementation in Present-Day English*. Albany: State University of New York Press.
- Visser, Fredericus. 1963. *An historical syntax of the English language*. Leiden: E.J. Brill.

Gaëtanelle Gilquin, Szilvia Papp and María Belén Díez-Bedmar (eds.). *Linking up contrastive and learner corpus research* (Language and Computers 66). Amsterdam and New York: Rodopi, 2008. ISBN 978-90-420-2446-5. xi + 282 pp. Reviewed by **Hilde Hasselgård**, University of Oslo.

This volume contains ten corpus-based studies of learner language, most of which were presented at a workshop with the same title as the volume, held in September 2005 at the University of Santiago de Compostela. The volume has four parts: Methodology, Learner lexis, Learner syntax, and Learner discourse, with most papers relating to English L2 writing by advanced learners from a variety of language backgrounds. The International Corpus of Learner English (ICLE) is central to most of the studies, but is not the only corpus used.

A clear merit of the volume is its explicit emphasis on methodology in the contrastive study of learner language. Already in the early days of contrastive linguistics, contrastive studies were linked to language teaching (e.g. Lado

1957). With the study of error analysis popular in the 1970s, the focus shifted from language teaching to language learning, and more recently the advent of corpora of learner language has greatly facilitated the study of interlanguage. Since the study of interlanguage often involves considerations of the learner's first language (e.g. in the study of transfer), contrastive studies have remained a pillar in learner language research. Studies of learner language can be contrastive in a number of ways, all of which are represented in the present volume:

- a. The learner's first language in contrast to the target language (= the language to be learnt) with the aim of predicting or explaining learning difficulties.
- b. The learner's interlanguage in contrast to similar output by native speakers, with the aim of identifying overuse, underuse and misuse.
- c. The interlanguage of one learner group in contrast to that of one or more other groups (Contrastive Interlanguage Analysis; cf. Granger 1996), with the aim of distinguishing L1-specific from more universal features of interlanguage.
- d. The integrated contrastive model (ICM), as outlined by Granger (1996) and developed by Gilquin (2000/2001), combining the study of bilingual corpora with that of learner corpora, thus combining comparisons (a) and (b) above.

The latter approach presupposes the existence of a parallel corpus containing the learner's first language with the language to be learnt, and it is presumably this approach that gave rise to the title of the volume. In the opening paper of the volume, Gaëtanelle Gilquin takes the ICM as her starting point and combines it with Jarvis's (2000) model for studying transfer. The resulting Detection-Explanation-Evaluation (DEE) model implies that the ICM should be supplemented with comparison (c) above. The usefulness of the model is demonstrated through a case study of the overuse of the conjunction *even if* by French learners of English, compared to native speakers and compared to other groups of learners. The frequency of *even if* in the English interlanguage of French learners is similar to that of *même si* in a comparable corpus of native French. Cross-linguistically, the correspondence between *even if* and *même si* is only partial, as the latter also corresponds to *even though*. These combinations thus bring out the overuse of *even if* by French learners as an example of negative transfer. In contrast, the overuse of the expression *more and more*, which might be regarded as transfer if only comparisons (a) and (b) are carried out, turns out to be a more general characteristic of learner language according to comparison (c), and is

thus a developmental feature. The paper illustrates clearly how different corpora and methods can usefully complement each other in the study of learner language. Moreover, its close and critical attention to methodology and the interpretation of results is instructive.

The second study, by Annelie Ädel, also has a strong focus on corpus methodology and the combination of approaches. As the title suggests (“Involvement features in learner writing: Do time and interaction trump register awareness?”), the author explores the use of interaction signals identified in several studies based on ICLE material. She compares the Swedish subcorpus of ICLE with a similar corpus of English argumentative writing by Swedish students (the Uppsala corpus of learner writing – USE). The two corpora differ in two important respects: in contrast to SWICLE, the essays in USE are untimed and the students were given topical texts as a starting point for argumentative writing; thus the differences concern task-setting as well as intertextuality. It turns out that the frequency of involvement features is much lower in the USE essays. An important conclusion is that “the overuse of involvement features cannot just be attributed to lack of register awareness” (p. 46), but also to task-setting (timing) and intertextuality. The paper ends with two pedagogical recommendations, namely that learners will achieve a more native-like style of formal writing if they are not timed and if they are given “other texts as input and to serve as point of departure for their writing” (p. 46). The paper is also a useful reminder that a corpus, and thereby the results gained from investigating it, should always be evaluated critically, and that studies may usefully be replicated using a different corpus. This last point is also voiced at the end of the book, by Demol and Haderman, who speculate that the disagreement between their own results and those of previous research concerning clause linking and clause integration among native speakers and learners of Dutch and French may be due to the proficiency levels of the learners represented in the corpora used.

Space does not permit a detailed review of each of the eight studies of lexis, syntax and discourse. John Cross and Szilvia Papp study verb + noun combinations as evidenced by Chinese learners of English compared to learners whose first language is German or Greek. They find that the Chinese learners produce more combinations that diverge from a native speaker norm, and further that erroneous combinations produced by German and Chinese learners differ in that the Germans may use combinations creatively to communicate meanings, while the Chinese may be trying to reproduce memorised chunks to show off their knowledge. Cristóbal Lozano and Amaya Mendikoetxea compare Spanish and Italian learners of English as to their production of postverbal subjects (including notional subjects in sentences with an anticipatory subject), finding that both

learner groups produce such constructions only with unaccusative verbs, in contrast to unergative ones.

John Osborne reports on a rather comprehensive study of adverb placement in the English of French learners as compared to that of native speakers of English. Like Ädel, he uses two different corpora from each language variety, thus checking the validity of the findings from each corpus. The findings for the French learners are also compared to the ICLE subcorpora for learners with a variety of other L1 backgrounds, thus approaching Gilquin's DEE model for investigating transfer. Interestingly, adverb placement seems to group the learners according to typological distinctions in their L1; thus speakers of Germanic languages resemble each other, as do speakers of Romance languages, while a third group is made up by speakers of Slavic languages plus Finnish. Even if transfer is evident in the adverb placement of the French learners, one of Osborne's findings is that learners may produce patterns that are ungrammatical in their L1 as well as in the target language. Although the study is placed in the syntax section it also takes into account a number of discourse features (such as weight in the case of heavy NP shift).

María Belén Díez-Bedmar and Szilvia Papp investigate the degree to which Chinese and Spanish learners master the use of the English article system. The two learner groups have different starting points, as Chinese lacks an article system almost entirely while the Spanish article system differs in some respects from that of English. As expected, the Spanish group performs better than the Chinese group. Mention should be made of the methodology applied in this study, whereby all contexts for the use of the definite, indefinite or zero articles were identified (i.e. for each noun phrase in all the texts). The authors were then able to quantify the learner's degree of success, in terms of a proportion of appropriate uses, and also to identify the problem areas of both groups; the most problematic contexts for Chinese learners are particularly those that require the indefinite article with generic reference but also those that require the indefinite article for specific reference or the zero article with generic reference. The Spanish learners also make mistakes in generic contexts that require zero or the indefinite article, but to a lesser degree than the Chinese.

The integrated contrastive model is demonstrated by Christelle Cosme, who looks at the role of transfer as regards participle clauses in learner English. Her starting point is a study by Granger (1997) of participle clauses among learners of English, and Cosme provides a contrastive analysis involving two trilingual corpora of English, French and Dutch to try to explain the learner data. It is found that present participle clauses are extremely rare in the Dutch corpora, while the French present participle is at least less frequent than the English one,

particularly with adnominal function. Adnominal past participle clauses, on the other hand, are almost three times as frequent in French as in English, but in Dutch they are only half as frequent as in English. Adverbial past participles are rare in all three languages. The underuse of participle clause observed by Granger can thus be nuanced: French learners can be expected not to underuse adnominal past participles, only present participles, while Dutch learners will be expected to underuse both types.

Marcus Callies deals with the use of raising constructions in the English of German and Polish L1 learners as compared to native speaker data from the LOCNESS; more specifically 'subject-to-object' raising (*We believe them to retire next week*) and 'tough movement' (*This problem is difficult to see*). The constructions are identified according to 'believe'-type verbs and the adjectives characteristically used in tough movement constructions. Particularly the tough movement construction is underused by the learners, apparently with the exception of those connected with the adjective *difficult*. 'Subject-to-object' raising is overused by Polish learners in the case of *find* and 'subject-to-subject' in the case of *consider*. These oddities are ascribed to 'transfer of training', as the constructions tend to be explicitly taught to Polish learners.

Mike Hannay and Elena Martínez Caro explore thematic choice in the written English of advanced Spanish and Dutch learners as compared to native speakers of English. 'Theme' is defined as the initial part of the sentence extending up to, but not including, the finite verb. This is done by identifying word order patterns in the theme zone. It is found that the three groups vary mostly in that the learners seem to overuse the so-called 'level 2' pattern with one constituent preceding the subject, while all three groups have a similar use of the 'level 3' pattern, with two constituents preceding the subject.

The only paper to deal with target languages other than English is Annemie Demol and Pascale Hadermann's exploratory study of discourse organisation in French L1, Dutch L1, French L2 and Dutch L2 written narratives, more specifically the use of parataxis, hypotaxis and condensation in sentence construction. They find less difference than hypothesised (on the basis of previous studies) between French and Dutch L1 as well as between French L1 and L2 and Dutch L1 and L2. They suggest that the surprising results may be at least partly due to the different corpora used for the studies, including writer maturity (student vs. professional for the L1 data) and learner proficiency for the L2 data.

As indicated at the outset of this review, I find the volume interesting and rewarding. A few problems nevertheless need pointing out. One concerns the naming of corpora in several of the studies, where the authors refer somewhat misleadingly to learner corpora (of English) as, for example, the German or the

Bulgarian corpus; in Díez-Bedmar and Papp's paper two corpora are even referred to as Chinese NNS and Spanish NNS, which would seem to suggest non-native varieties of these languages rather than English. As one of the aims of the present volume seems to be establishing a common framework for studying learner language, one should perhaps also decide on a way of talking about learner corpora which does not presuppose that all language learning has English as its target. (As long as a subcorpus of ICLE is used, it would be convenient to refer to it as e.g. the Polish ICLE or its acronym PICLE). Since the studies all use corpora, they also involve some quantitative data, whose presentation is almost inherently difficult. Many of the studies supply measures such as chi square to validate their statistics. However, it is sometimes difficult to interpret the quantitative findings. This may have to do with the use of comparison of differently-sized samples, in which percentages may neutralize the differences, but sometimes at the same time unduly enlarging very small numbers (e.g. in Lozano and Mendikoetxea's study where percentages are calculated from samples of 52 and 15 unaccusatives) or the use of percentages rather than normalized figures (as evidenced e.g. in Cross and Papp's study), or indeed the use of percentages without giving the raw figures, as in Demol and Hadermann's study. Finally, although the division of the chapters into lexis, syntax and discourse looks neat, the papers demonstrate that such divisions are somehow inadequate. The only paper in the lexis section (Cross and Papp) deals with a phraseological matter which is also partly syntactic. On the other hand, Callies's paper, in the discourse section, takes a lexical starting point in order to identify syntactic constructions with different discourse implications. This last point is, however, not so much a criticism as a recognition that the study of language often defies narrow categorisation.

All in all, the volume is well worth reading, and the editors – who were also the organisers of the workshop preceding the publication – should be commended for their efforts not only in bringing together researchers in the field representing a variety of L1 backgrounds as well as theoretical approaches, but above all for developing the discipline by taking a critical and constructive view of its foundations – its research material and its methodology.

References

- Gilquin, Gaëtanelle. 2003. The Integrated Contrastive Model: Spicing up your data. *Languages in Contrast* 3: 95–123.
- Granger, Sylviane. 1996. From CA to CIA and back: An integrated approach to computerized bilingual and learner corpora. In K. Aijmer, B. Altenberg and

- M. Johansson (eds.). *Languages in contrast* (Lund Studies in English 88), 37–51. Lund: Lund University Press.
- Granger, Sylviane. 1997. On identifying the syntactic and discourse features of participle clauses in academic English: Native and non-native writers compared. In J. Aarts, I. de Mönnink and H. Wekker (eds.). *Studies in English language and teaching. In honour of Flor Aarts*, 185–198. Amsterdam and Atlanta: Rodopi.
- Jarvis, Scott. 2000. Methodological rigor in the study of transfer: Identifying L1 influence in the interlanguage lexicon. *Language Learning* 50: 245–309.
- Lado, Robert. 1957. *Linguistics across cultures: Applied linguistics for language teachers*. Ann Arbor, MI: University of Michigan Press.

Sebastian Hoffmann, Stefan Evert, Nicholas Smith, David Lee and Ylva Berglund Prytz. *Corpus linguistics with BNCweb – a practical guide* (English Corpus Linguistics 6). Frankfurt am Main: Peter Lang GmbH, 2008. xviii + 290 pp. ISBN 978-3-361-56315-1. Reviewed by **Hans van Halteren**, Radboud University Nijmegen.

On the basis of the title of the book, I expected a manual for the BNCweb system, obviously for doing research with the BNC. I assumed that the wording of the title was just an attempt to avoid a deadly title like *The BNCweb user manual*. However, I was soon proved wrong in this. The authors do actually have the intention to introduce corpus linguistics in general, using BNCweb as a tool, so that of course all functionality of the tool has to be explained. They want to entice the readers to try corpus linguistics, to convince them sufficiently so that they will keep using it and in the end (fortunately) also to warn them to “be careful out there”. Below I will try to answer the question if both goals, teaching BNCweb and teaching corpus linguistics, are adequately reached. Before that, I will take you on a short tour along the chapters in the book.

Chapter 1 introduces not only the book but also the very use of corpora itself, by way of a few example questions which are best answered on the basis of a corpus, such as how the word *shall* is used in Present-Day English. Chapter 2 goes on to describe corpus linguistics as a whole, addressing subjects like representativeness and the relation to theoretical linguistics. Chapter 3 refocuses on the BNC, with an extensive description of how it was built and what we can find

in it. Here, it also becomes clear that the authors do not stop at telling us what commands to use, but spend ample attention on potential pitfalls in corpus use, such as drawing premature conclusions about gender-specific language and at the same time ignoring the fact that most text spoken by men in the BNC is in the more formal context-governed component, while most text spoken by women is in the more informal and spontaneous demographically sampled component.

After these more introductory chapters the book turns towards teaching us the use of BNCweb. Chapter 4 presents the simplest of explorations, showing how to make a query and how to see the results. Chapter 5 seems to move away again from BNCweb, but at the right point and for an excellent reason. Right at the beginning, before we dare skip the rest of the book to just start exploring, the authors explain some basic statistical aspects, such as the need to use normalized frequencies, the investigation of the precision and recall of our queries and the concept of statistical significance. Only after we have been warned to take practically nothing at face value, can the instruction proper of BNCweb continue. Chapter 6 presents the construction of queries in the so-called *Simple Query Syntax*. Although called simple, this already allows wildcards in tokens, access to parts-of-speech and lemmas, and even regular expressions on the token sequence level. Chapter 7 explains that we do not necessarily always have to read the full concordance, but that BNCweb can bring order to our exploration with sorting options and automatic distribution analyses. Again we are warned, though, that we have to keep thinking critically too and that, in the end, some manual inspection is always advisable. Chapter 8 goes on to the automatic analysis of collocation patterns. It also presents, at a very basic level, the various statistical formulae used for this analysis, giving us at least some insight into which one(s) to use and when.

The next three chapters are concerned with exploration beyond the beaten path. Chapter 9 shows how we can add our own annotation to concordance lines, in case we want to investigate something not addressed by the BNC markup. This can be done inside BNCweb, but we are also allowed to export data to, for instance, Excel, annotate and then reimport them. Unfortunately, both methods are admittedly imperfect. Chapter 10 shows how we can create and use a subcorpus, i.e. our own selection of samples from the BNC corpus. The system provides many ways to define subcorpora and make working with them more efficient. Chapter 11 shows how we can compare word use in two parts of the corpus. We can look at frequency lists for the parts, but BNCweb can also provide so-called *Keywords*, the words which occur more frequently in either of the selected parts.

The last two chapters are presented to achieve completeness, but their nature seems to be more of the need-to-know-basis type. Chapter 12 presents the more complex *CQP Query Syntax*, provided by the Corpus Query Processor of the IMS (Institut für Maschinelle Sprachverarbeitung) Open Corpus Workbench. This syntax enables us to access many more types of BNC annotation and to use many more involved interrelations. However, its use comes dangerously close to programming. Reading the chapter, I got the impression that the authors, possibly subconsciously, also have the feeling that this functionality should not be used by the novice. Although outwardly it seems to claim anyone can do this, the text is written much more densely, and I doubt that anyone other than more widely experienced users will be able to use CQP Query Syntax on the basis of what they find here. Chapter 13, finally, mainly explains to system administrators what their possibilities are, and what their duties are. Although useful for administrators, the chapter feels a bit misplaced in this book.

After these chapters we find a short bibliography, a glossary and several appendices, describing David Lee's genre classification scheme, the part-of-speech tag list, a four-page quick reference for the Simple Query Syntax and a list of HTML-entities for less common characters.

For the BNCweb user, the book is an invaluable resource. It provides a complete description of the system's functionality, and it does so in an extremely useful way. All aspects are presented with working examples, instructive tasks and chapter-final exercises. Wherever appropriate, it is explained why things are the way they are. Finally, and this is where this book outshines many others, the user is warned where the mistaken conclusions lurk and often how they can be avoided.

I am less ecstatic about the book's ability to serve as an introduction to corpus linguistics. True, it does a nice job of introducing a particular kind of corpus linguistics: exactly that kind which you can do with the BNC and for which the BNC was created. But some topics in corpus linguistics are not mentioned at all. Let me just mention aligned multilingual corpora and annotation addressing linguistic constructs beyond the word level. Seeing that the book is also an introduction to corpus linguistics, a bit more information in Chapter 2 would have been nice, that is, more than the seven pointers to the bibliography (which, by the way, are marred by some errors, listing two books with different years of publication than in the bibliography and two which are missing there).

This leads me to the final question: for which users and in which contexts is this a good book to buy? As already implied, if you are using or going to use the BNC, buy this book and cherish it! It will help you do your research in a proper and efficient way. I would hope that the average reader of the *ICAME Journal*

should already be familiar with the pitfalls, but you never know if something is mentioned you missed so far. I must warn you, though, that the tone is very much that of an undergraduate text book – and probably aimed at insecure undergraduates at that. If there is one thing which annoyed me in the book, it is the need apparently felt to emphasize the user-friendliness of BNCweb; if I ever get my hands on a machine-readable version of the book, I am going to measure how much *easy, easily, simply, straightforward*, etc. are being overused here. Still, the content should make the tone acceptable. Then, if the tone is that of an undergraduate textbook, should this be the book you make your students buy for the corpus linguistics course? That mostly depends on what you want to do in the course. An ideal situation would be one where you have an introductory and an advanced course. In that case, yes, use this book for the former, extended with some literature about the ‘other’ corpus linguistic activities. If you have just one course, though, I think you do not want to spend all the students’ time on just this. You should not have them buy the book, but you yourself should buy it and base some course material on it. Finally, the book might also be useful for linguists (or other researchers) who need to lose their corpus virginity. Obviously, readers of the *ICAME Journal* will not be of this type. But if you know someone else who is, you can safely advise this book.

Fanny Meunier and Sylviane Granger (eds.). *Phraseology in foreign language learning and teaching*. Amsterdam and Philadelphia: John Benjamins, 2007. x + 259 pp. ISBN 978-90-272-3244-1. Reviewed by **Göran Kjellmer**, University of Gothenburg.

The present volume is the first of three to emanate from a conference on phraseology in Louvain-la-Neuve in 2005. A second volume is entitled *Phraseology: An interdisciplinary perspective*, and a third is *La phraseologie dans tous ses états*. The focus of all three volumes is thus on phraseology; as the title of the current volume indicates, it concerns itself chiefly with pedagogical aspects of phraseology.

The book, which is fittingly dedicated to the memory of John Sinclair, is a collection of articles by contributors to the conference, coming from Europe, Asia, the U.S. and New Zealand. It is organised into four sections. After the Preface and Introduction follow Section I “Extracting and describing phraseo-

logical units”, II “Learning phraseological units”, III “Recording and exploiting phraseological units” and IV “Concluding remarks”.

To lead into the discussions of various aspects of phraseology and its place in language pedagogy, Nick Ellis gives a fact-filled and very useful overview of the history of phraseology in linguistics, from structuralist insistence on language patterns over generative views of constructions and formulas as being peripheral in a system governed by Universal Grammar, to the present emphasis on phraseological issues in foreign-language learning and teaching. His survey thus places the study of phraseology in its historical (and epistemological) context.

Graeme Kennedy gives a down-to-earth demonstration of the benefits of a phraseological approach by extracting the collocations of a few common verbs (*enjoy, give, receive, start, begin, stop, end* and *finish*), which show that “high frequency lexical verbs tend to be associated with other words having particular grammatical features or belonging to particular semantic domains” (p. 33). The question why phraseology is or should be considered an important element in language-teaching is not a major concern in the book but is taken for granted by most writers. The main themes of the book are the definition, description and extraction of phraseological units on the one hand, and their implementation in the teaching and learning process on the other. It could be said that the extraction part is on safer ground, theoretically speaking, than the application of the results to the classroom. Finding suitable phrases in text material and processing them for use in a pedagogical context is naturally an important aspect of phraseological pedagogy. Even if some writers allow for teacher intuition as a contributing mechanism in the selection of phrases, statistical procedures are mostly used to extract them. Thus for example, Susanne Handl presents a largely statistical method for determining the ‘collocational factor’ of word combinations.

It is a common experience among scholars that learners’ use of phrases is not sufficiently nativelike. Magali Paquot finds, for example, that “there are striking differences in use between native and learner writers in English” (p. 108), and Graeme Kennedy (p. 37), Céline Gouverneur (p. 224) and Dirk Siepmann (p. 194) concur. Siepmann sums up the inadequacies of learners’ phraseology found in studies by other scholars and produces a list of such shortcomings.

In “Phraseology effects as a trigger for errors in L2 English: The case of more advanced learners”, John Osborne investigates samples of university-level L2 English, looking for four types of errors or deviations from standard English (here represented by the LOCNESS corpus): nominalised adjectives (*vasts amounts*), pluralisation of uncountables (*clever advices*), omission of 3rd person

singular –s (*technology prevent humans' ability*), and placement of adverbs between the verb and its object (*they know perfectly the TV programmes*). His goal is to find if such errors are due to chance or if it is possible to identify specific contexts in which they occur. His analysis of the relevant cases leads him to conclude that there are three types of phraseological effects: *blending*, when elements forming a larger unit share or transfer grammatical features among themselves (*the values or advices*), *bonding*, when lexical or grammatical elements that have formed associations in the learner's lexicon or grammar appear in adjacent positions (*follow blindly everything*), and *burying*, when elements which are embedded inside larger units become less salient and so lose grammatical features (... *loves when a tender and careful woman waits for him at home and after a busy day meet him* ...). He concludes: "Taken collectively, [phraseological error-inducing effects] may not be as random as they seem." (p. 82).

When it comes to assessing the nativelike or nonnativelike character of learners' English, Alison Wray and Tess Fitzpatrick have devised a method called "Deviation profiling", where Chinese and Japanese learners are required to memorise a number of genuine English phrases and render them as exactly as possible on a later occasion. Deviations from the original are then recorded and classified in three dimensions (morphological, lexical and phrasal); they form the basis of a final assessment of the learners' proficiency. JoAnne Neff van Aertselaer and Magali Paquot, whose interest lies in studying the influence of L1 on learners' English, adopt a different approach. They both use a three-ways comparison of text produced by groups of different nationalities. Paquot sums up (p. 117): "The study shows that L1-related effects contribute significantly to learners' use of multi-word units."

If now there is a general consensus that learners' proficiency in the phraseological domain falls short of a native-English standard, a natural consequence is then to ask what could be done about it, as Averil Coxhead does when she refers to recent studies, each with part of the answer. One of them is Michael Hoey's work on priming, which, however, is criticised (p. 155) by Coxhead for being too vague: "If direct learning is needed what kind is best? How many encounters and what kind of exposure exactly and when?" (Hoey might answer: "As many encounters, as many kinds of exposure and as often as possible"!) Kennedy also advocates implicit autonomous language learning, especially through reading. "It may be that reading of all kinds, including literary works, may make an overdue return for greater attention in language pedagogy, for reading does provide the kind of exposure which facilitates implicit learning" (p. 40).

Two of the aids to such exposure are textbooks and dictionaries. Céline Gouverneur takes on the question of the presentation and use of collocations in three

sets of textbooks. She creates a database in which all the relevant elements are tagged, which allows for detailed analysis of the frequencies and treatment at different levels of collocations involving the verbs *make* and *take*. The pedagogical implications of the treatment are made apparent by Gouverneur's skilful use of the categories resulting from the tagging system. She finds that textbook designers play down the importance of such collocations in exercises intended for advanced learners.

As for dictionaries, Dirk Siepmann and Mojca Pecman are highly critical of the present state of affairs. Siepmann makes spot checks of learners' dictionaries to see to what extent they cover fully transparent collocations and finds them inadequate. He also finds that foreign learners' use of such collocations is inadequate. As he thinks that the poor quality in this regard of learners' dictionaries is at least partly to blame for the students' performance, he suggests that new dictionaries should be created, based on vast materials from the Internet. The dictionaries should be onomasiological rather than semasiological to enable students to find groups of more or less synonymous phrases gathered under a common head. (It could be mentioned that the *Longman activator*, 1993, Harlow, Essex: Longman, seems to be one such dictionary.) Mojca Pecman records a great number of obstacles to and difficulties in processing bilingual phraseology, i.e. points to be remembered in the process of making a bilingual phraseological dictionary. She then lists the advantages in having bilingual phraseological dictionaries of sublanguages, such as the specific combinatorial properties of sublanguages and highly developed styles of conventions. In her project, aimed at producing a tool for French learners of English, she makes use of a parallel corpus, extracts multi-word units and classifies them semantically. Each such unit is linked to its hyperonym, which then allows the production of both a semasiological approach and an onomasiological approach to the material. The result is a doubly useful tool for the learner.

David Wible, too, is critical of existing dictionaries, which he finds are static, centralised and passive, particularly with regard to multi-word expressions. He proposes the creation of digital dictionaries, whose multi-functionality is immediately helpful to the learner. Wible demonstrates how a device called Collocator can identify, by means of a large database, phrasal sequences in a text and present them to the learner for consideration if he so wishes. The database can even be the Internet itself. Learners, Wible argues, are often not aware of the existence of common collocations and therefore do not look them up. Collocator will present them to the reader.

There is a general consensus among the contributors that foreign-language pedagogy should aim at improving phraseological proficiency among students

(though John Osborne finds that phraseological patterns can trigger errors), but the issue is discussed in a more general way by the editors. Unlike many advocates of phraseology as the be-all and end-all of language pedagogy, they recommend principled eclecticism: “We thus call for a harmonious combination of technology (corpora, statistical measures, etc.), common sense and teachers’ experience in selecting relevant units for teaching” (p. 250). In particular they see the danger of an over-enthusiastic emphasis on things phraseological crowding out the teaching of grammar from the curriculum and recommend a considered balance between lexis and grammar in the classroom.

Some important points stand out in *Phraseology in foreign language learning and teaching*:

- Mastery of multi-word expressions is a crucial part of language proficiency.
- The phraseological proficiency of ESL/EFL learners is generally inadequate.
- Teaching methods therefore need to be developed to give more prominence to phraseology.
- Textbooks and dictionaries also need to be developed in that direction.
- Such developments are under way.
- A balance should be sought between lexis and grammar in the curriculum.

It will have appeared that the book contains a stimulating collection of articles, several of which give impetus to further research in corpus-based phraseology and in pedagogical applications. It presents an impressive picture of the great volume of work going on in the area and is practically free of misprints and other blemishes. In the articles it is a recurring theme that we don’t know enough, that there are many gaps in our knowledge of the field waiting to be filled. *Phraseology in foreign language learning and teaching* triumphantly fills a great many such gaps.

Christoph Rühlemann. *Conversation in context. A corpus-driven approach.* London and New York: Continuum, 2007. 272 pp. ISBN 978-08-264-9713-0. Reviewed by **Karin Aijmer**, University of Gothenburg.

Grammar has traditionally dealt with written language only, although lip service has sometimes been given to spoken language when there are differences. As a result the grammar of conversation has been poorly described. In the ‘core grammar’ approach represented for example in Quirk *et al.*’s *Comprehensive grammar of the English language* (1985) the focus is on the commonalities of different varieties of English and describing this common core by grammatical rules. The situation is now changing. In the corpus-based *Longman grammar of spoken and written English* (Biber *et al.* 1999) and the more recent *Cambridge grammar of English* (Carter and McCarthy 2006) the uniqueness of conversational data is prominent. Rühlemann’s study owes a great deal to the Nottingham corpus team around Ron Carter and Mike McCarthy and to the authors of the *Longman grammar*. The approach is strictly functional. Conversational grammar is separate from written grammar and is described on the basis of the constraints imposed by the interaction itself and the way spoken language is processed.

Quoting Rühlemann (p. 1), “In spite of the present surge of interest in spoken language sparked off by the availability of spoken corpora no comprehensive account of conversational grammar has however been formulated to date.”. His aim is “to contribute to the linguistic discussion of British conversation by exploring the conversational subcorpus of the BNC” (p. 1). The aim includes:

- to provide an overview of characteristic features of British conversation
- to explore some neglected features of British conversation
- to explore social differentiation in the use of some selected features
- to show how conversational language is adapted to needs arising from specific types of constraints on speakers in conversational situations (the adaptedness hypothesis).

The volume consists of nine chapters and an introduction by Mike McCarthy. The first two chapters provide the theoretical background for the case studies in Chapters 4–8. Chapter 9 is a summary and conclusion.

Chapter 1 introduces an approach to the grammatical description, which is dependent on register and situation. Neither speech nor writing are monolithic concepts but there is a wide range of overlapping registers, each with their own grammatical and lexical characteristics. Following the Hallidayan framework, the variation between registers is associated with the variation in the types of context and situation. Another characteristic of registers is that they are marked by a set of linguistic features which can vary depending on the situation. The

main thrust of the study is thus “to define the factors that constitute the situation type that gives rise to conversation and to construct a situational framework based on which the distinctive linguistic features of conversation can be adequately understood” (p. 8). Conversation is regarded as the archetype of spoken language; it is “a feature of everyday life used virtually by everybody both productively and receptively” (p. 8). It is also the point of departure for more specialized varieties such as spoken media interactions. Moreover, conversation is fuzzily delimited from adjacent registers, and it typically integrates ‘subregisters’ such as telephone conversation, narratives, language in action. Central to the study is the use of corpora which enable a focus on variation and language varieties (a variety grammar). The corpus perspective is also visible in the close attention to lexis and its association with grammar (a lexical grammar).

Chapter 2 describes data and methods. Rühlemann has mainly used the (demographically sampled part of) the large BNC-corpus. This makes it possible to make generalizations about conversational phenomena on the basis of a number of coded factors in the corpus (including sociolinguistic factors such as the age, sex, region and social class of the speakers). The corpus-specific methods include the study of ‘association patterns’ (collocations) and discourse factors (factors involving the larger context). The empirical case studies were chosen because they illustrate the ‘adaptedness hypothesis’ (i.e. how lexis and grammar are modified or constrained by situational factors or the interpersonal goals of conversation). The author provides a list of features *not* studied in the present work with reference to previous work. However, there is little discussion of how and why the current topics and conversational features were chosen.

Chapter 3 lists the factors making up the conversational situation type. There are five factors governing conversation: shared context (speakers are face-to-face), co-construction (e.g. turn-taking and role rotation), real-time processing, discourse management and relation management. The shared context includes prosody, kinesics (gaze, gestures, facial expressions) and vocal features (laughter, sighing). However, many shared non-verbal phenomena are difficult to discuss on the basis of the available spoken corpora (such as the BNC). Work aiming at the compilation of multimodal corpora may provide an opening for an analysis of conversation integrating verbal and non-verbal communication. Discourse management (a factor less basic than the others since it arises from co-construction and real-time processing) has to do with the need to create discourse coherence and is manifested in the use of discourse markers. Relation management is another complex factor, since it includes politeness, emotive communication, and involvement. Involvement and emotive communication could perhaps have been considered as a separate factor. Conversational features

or patterns are not the result of abstract grammatical rules, but they can be explained as adaptations to the constraints set by these factors. Several constraints can be present simultaneously. It is therefore not self-evident where a particular linguistic phenomenon should be dealt with. Laughter (treated in Chapter 4 as an example of adaptation to the shared context) is shown to have both a discourse management and a relation management function. The situational factors are well motivated both theoretically and empirically (which does not mean that they are necessarily exhaustive). Other linguists may propose different factors (and either fewer or more factors). The factors are not unique to or distinctive of conversation but salient in the description of conversation as shown by the comparison with other registers such as lectures. There is no discussion of the question whether the same factors can be used to study other registers of spoken communication.

Chapters 4 to 8 demonstrate how a variety of linguistic features can be related to one or more of the situational factors. Chapter 4 (adaptation to the shared context) explores the role of deixis in conversation, in particular the high frequency of personal pronouns (*I, you*) in the conversational context. Several reasons can be given for this, for example that *I* and *you* are integrated parts of discourse markers (*I see, you know*), and that they are frequent because they occur in direct speech presentation. Context can also manifest itself in non-verbal features such as laughter. Both ‘between-speech laughter’ (BSL) and ‘within-speech laughter’ (WSL) (using a laughing tone) are frequent in the BNC and perform discourse-management functions there. The types of laughter were also found to be differently distributed with regard to the sociolinguistic variables sex, age and class.

The co-construction phenomena discussed in Chapter 5 include question tags (inviting co-construction), backchannels (listening to speakers) and ‘ways of co-constructing utterances syntactically’. An example of ‘sharing syntax’ is the use of sentence relatives to expand the same speaker’s or the previous speaker’s utterance. The two types of sentence relatives are shown to fulfil different roles in terms of co-construction and relation management and occur with distinct association patterns (collocations).

Chapter 6 is a necessarily incomplete description of some discourse markers against the background of their function to “indicate what course upcoming discourse is intended to take in relation to prior discourse” (p. 116). The term discourse marker is poorly defined and has been used or misused about many different pragmatic phenomena (see e.g. the discussion on terminology in the contributions to the anthology edited by Kerstin Fischer 2006). According to Rühlemann, discourse markers are realisations of discourse deixis and used to

establish discourse coherence by signposting how utterances or discourses are related to each other. Rühlemann argues convincingly that phrases such as *I/He/She/said/says* can be regarded as discourse markers since they “perform vital discourse-deictic functions by indexing that upcoming discourse needs to be interpreted ... as a presentation of anterior speech, thought or emotion” (p. 117). *Like* is picked out for closer investigation in the BNC because it has been shown to have a high frequency in speech compared with writing, which suggests that it is an ‘emergent’ discourse marker. However, *like* is notoriously multifunctional as a discourse marker. It does not only have loose or approximative meaning but is a co-constructive device keeping the conversation flowing ‘in a reciprocal manner’. The marker has also acquired a quotative function (‘presented discourse function’ in the author’s terminology). The quotative *like* appears in two different contexts: actual speech and to signal a potentially recurring episode in narratives. The new discourse marker functions are also shown to be interesting in a sociological and grammaticalization perspective.

Chapter 7 discusses a fairly neglected topic in conversational studies, namely the study of pauses and other disfluency phenomena which can be subsumed under the pragmatic concept of speech management. Speech management phenomena are reflections of the processing taking place in the planning (managing online pressures prospectively) and of failures to execute the messages reflected in restarts and repair (managing online phenomena retrospectively). Pauses are filled or unfilled. Filled pauses are *er* and *erm*. Speech management phenomena should not be interpreted too narrowly. Also discourse markers such as *well*, *I don’t know*, *I mean* function as signals about the planning and production of sentences. The chapter also discusses phonetic and grammatical reduction. The grammatical phenomena include substitution, omission (ellipsis) and generalisation. Generalization is said to be at work in the use of the discourse marker *I says* (the third person is generalized to first person).

The concluding chapter 9 discusses some themes in more depth and raises questions for future research. To begin with, the sociolinguistic analysis has shown that many conversational features are differentiated socially. The features investigated from this perspective were for instance more used by females than by males, which suggests that the grammar of British conversation is a ‘female grammar’. However, other features such as social class probably need to be looked at more. Secondly, deixis (along with the adaptedness hypothesis) can explain a number of phenomena in conversational grammar. Deixis in spoken language is a flexible notion which can be used in many different ways. Speakers use deixis to ‘slip out of their own view-point’ and assume someone else’s point of view (what Brown and Levinson 1987 refer to as point-of-view opera-

tions or point-of-view ‘flips’). A number of phenomena in conversation can be explained in this perspective (and probably many more than are discussed in this study). The frequency of direct speech presentation and the use of the historical present are examples of point-of-view switches or ‘deictic centre oscillation’ resulting in greater involvement or vividness (cf. Brown and Levinson’s positive politeness). Thirdly, conversational grammar also provides a challenge for language teaching. Although spoken language has become a major new field of research, the findings are only slowly trickling into language teaching. Written grammar (the grammar of the standard language) is still the norm. Many features of spoken language are stigmatized and there is a thin line between what is accepted as conversational and non-standard grammar. The variation between ‘standard grammar’ and conversational grammar also raises a number of questions about grammaticality and the model or norms in EFL teaching.

To sum up, *Conversation in context* is not only another book about British conversation. Rühlemann provides a coherent theoretical framework for analysing grammatical or discoursal phenomena in conversation as shown by the variety of case studies. The empirical case studies do not only fit and strengthen the model but they represent a new way of looking at grammar and offer a new and deeper understanding of many grammatical phenomena. It remains to be seen what other features (if any) are needed to account for spoken or conversational grammar and to test these factors on the basis of more data. The book can be recommended to anyone interested in pragmatics (and sociopragmatics), discourse analysis and corpora. The book is well-written with useful and frequent to-the-point summaries. It opens up new vistas for research on grammar in conversation and other registers and is likely to have an impact on many areas in applied linguistics.

References

- Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad and Edward Finegan. 1999. *The Longman grammar of spoken and written English*. Harlow: Pearson Education.
- Brown, Penelope and Stephen C. Levinson. 1987. *Politeness. Some universals in language use*. Cambridge: Cambridge University Press.
- Carter, Ron and Mike McCarthy. 2006. *The Cambridge grammar of English*. Cambridge: Cambridge University Press.
- Fischer, Kerstin (ed.). 2006. *Approaches to discourse particles*. Amsterdam: Elsevier.

Hoey, Michael. 2005. *Lexical priming*. London: Routledge.

Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech and Jan Svartvik. 1985. *A comprehensive grammar of the English language*. London: Longman.