

Reviews

Karin Aijmer. *Corpora and language teaching* (Studies in Corpus Linguistics 33). Amsterdam and Philadelphia: John Benjamins, 2009. 232 pp. ISBN978-90-272-2307-4. Reviewed by **Gudrun Rawoens**, University of Ghent.

The volume *Corpora and language teaching*, edited by Karin Aijmer, in the series *Studies in Corpus Linguistics*, brings together a number of papers, selected from presentations made at a symposium held at Gothenburg University in 2005, as well as a number of commissioned papers.

The starting point for the volume is that, in spite of the great wealth of corpora and the advanced technological possibilities for using them for various purposes, the knowledge and expertise related to the actual use of corpora, notably in a variety of pedagogical settings, is still fairly limited. Given this starting point, the aim of the volume has been to collect a number of contributions that report on the state of the art of the use of corpora for pedagogical purposes, either in applied settings (classroom teaching) or for pedagogical research purposes, and to suggest a number of possible improvements for the future. Common to all papers is that they deal with English in general and with EFL (English as a Foreign Language) or ESL (English as a Second Language) in particular – although they are no doubt familiar to the intended readership of scholars and teachers working in the field, it might have been helpful to see a number of these common abbreviations written out in the introduction. Even though the intended readership quickly becomes clear from the content, it could have been an advantage to see this reflected in a subtitle of the book or have it mentioned more explicitly in the introduction.

The volume contains eleven contributions in total preceded by an introduction written by the editor and followed by an index. The papers are organized into four parts, each containing two or three papers sharing a common theme. It can be noticed that the subtitles labeling these four parts as used in the table of contents are not the same as the ones used in the introduction. What is more, in the introduction the contribution of Winnie Cheng is presented together with the

papers from Part III, whereas it is classified in Part IV in the book itself (as well as in the table of contents). This does not alter the consistency of the volume itself, but it might be slightly misleading to the reader.

Part I, “Corpora and second-language acquisition”, contains two contributions that discuss the implication of corpora in language teaching from a number of theoretical angles and applications. The focus of the first paper by Sylviane Granger is on learner corpora research in general and its pedagogical applications in particular. More specifically, the author argues in favour of narrowing the gap between learner corpus research and what it has to offer in pedagogical contexts, such as second language acquisition and foreign language teaching. The other paper in this section is by Stig Johansson, who elaborates on the value which corpora can have in second-language acquisition and devotes a section to different types of corpora. As for the integration of corpora in language teaching, the author discusses two general approaches – the inductive and the deductive approach – and argues for an approach that combines elements of these.

The papers in Part II, “The direct corpus approach”, address issues associated with corpora in the classroom. In the first paper of this section Solveig Granath takes the reader stepwise through a series of corpus-related exercises used in EFL teaching at university level. In doing so, the author provides an account of concrete examples from the classroom which can contribute to raising the awareness of the possibilities of learning by means of corpora. Signe Oksefjell Ebeling gives an account of Oslo Interactive English (OIE), an interactive learning environment which can be used either in on-campus teaching or outside the classroom. OIE is remarkable in that the exercises are data-driven and cannot be done without consulting a corpus, and that the corpus is immediately accessible via OIE. Through using this application the students not only become acquainted with corpus data, but also learn how to interpret them. The final paper of this part by Ute Römer is an investigation of the results of a survey conducted with English language teachers at a secondary school in Germany. The results make clear that the teachers feel a need for better teaching materials and reference tools as well as greater possibilities for native speaker advice. Römer suggests that this need can be met by introducing corpora in the classroom. Corpora cannot only serve as a basis for creating suitable data-driven learning materials, but they can also be relied on as a reference tool when a native speaker is not at hand.

The third part of the book is titled “The indirect corpus approach” and is concerned with the use of corpora for applied linguistics research. It opens with a paper by Jennifer Herriman and Mia Boström Aronsson who examine thematic choice and thematic variation in argumentative writing by Swedish

advanced learners of English in comparison with the same type of writing as produced by native English speakers. This study is based on the corpora ICLE and LOCNESS and is conducted within the framework of Systemic Functional Grammar (SFG), which is used to define different types of theme. The analysis reveals for instance that the non-native speakers use a greater variety of themes and themes that contain interpersonal information. An overuse of cleft constructions with the learners is also attested. In an attempt to explain these findings, the authors rely on factors such as influence from the mother tongue and learner strategies. The paper by Hilde Hasselgård continues in the same vein, with an examination of thematic choice in student writing described within the theoretical framework of SFG. In this paper, however, the focus is on Norwegian learners of English and the study described makes use of the ICM (Integrated Contrastive Model). The study reveals that Norwegian learners tend to use Norwegian patterns in their English writing, as witnessed by an overuse of initial adverbials. Besides, there is an overuse of first person pronouns. The paper by Susan Hunston explores how data derived from corpora is presented to learners and teachers, with particular attention given to corpus frequency and its link to phraseology. The author focuses on multi-word units and claims that much information can be derived from the internal collocation between the words in a cluster instead of adopting a limited view of the overall frequency of the cluster.

Finally, Part IV, labeled “New types of corpora”, includes three papers. In the first paper Winnie Cheng describes how a corpus-driven approach can be applied to the learning and teaching of phraseology in English for Specific Purposes (ESP). In particular, the paper deals with how the ‘aboutness’ of a text can be determined by studying and calculating the phraseological patterns. The analysis is conducted on the basis of two specialized corpora from the fields of economics and financial services. A particular type of corpus and its possible pedagogical applications is presented by Fanny Meunier and Céline Gouverneur, to wit the TeMa corpus which is a computer-readable corpus containing textbook material. The corpus is remarkable in that it is the largest and most diversified pedagogic corpus available at present. What is more, part of it is pedagogically tagged, facilitating research for various vocabulary related purposes, as illustrated by means of a few examples. The authors also express their hope that this kind of research in its turn can contribute to enhancing the quality of the textbooks. The paper by Joybrato Mukherjee on the grammar of conversation brings the volume to a close. Starting from three case studies based on the German component of the LINDSEI-corpus containing spoken English, the author shows that there are differences indeed between spoken English used by German learners as compared to English native speakers. Therefore, the

author argues in favour of increasing the awareness of these features typical of spoken English in classroom teaching.

The biggest asset of this book is that it brings together a variety of topics, both related to corpus research and to concrete language teaching situations. This variety is reflected in that some of the papers take a theoretical stance towards the issues raised, whereas others describe more practical applications. Another valuable point is that most papers are characterized by ample exemplification of research findings and possible applications in language teaching. What is more, they offer a good ground for discussions, since most authors have considered both advantages and weaknesses. Also, the strong pedagogical implication is surely something which language teachers and scholars can benefit from. Even though it is not the first volume dealing with this topic, it is a vital link in the discussions and findings in this field, which is changing rapidly. Also, whereas previous studies have mainly dealt with beginners, this volume focuses on advanced learners.

In summary, this work provides a valuable contribution in filling the still existing gap between the wealth of theoretical knowledge on the one hand and practical applications on the other. The essays in the volume have been written by established scholars with great expertise in this area of research, and who else than Karin Aijmer, with her expertise and broad range of publications in the field of corpus research and language teaching, could have been better placed to bring these together in a volume.

Winnie Cheng, Chris Greaves and Martin Warren. *A corpus-driven study of discourse intonation. The Hong Kong Corpus of Spoken English (Prosodic)*. Amsterdam and Philadelphia: John Benjamins, 2008. 325 pp. ISBN 978-90-272-2306-7. Reviewed by **Anne Wichmann**, University of Central Lancashire.

The availability of prosodically transcribed spoken corpora is limited. These days, most transcription is left by corpus compilers to the consumer, partly because of the inordinate time and expertise needed to annotate accurately, and partly because there is no universal agreement on which prosodic system to use. The publication of such a large corpus of prosodically transcribed English is therefore something to be welcomed. This book describes the results, and some analyses, of a very ambitious project – the Hong Kong Corpus of Spoken English (prosodic), consisting of 0.9 million words of prosodically transcribed

speech. The corpus and the software with which to analyse it are provided on a CD which accompanies the book, and will be of great value to very many researchers into discourse prosody.

Unfortunately, the same cannot be said of the sample analyses offered in several chapters of this book. These display little understanding of intonation and almost no awareness of the fact that the variety of English being described differs prosodically in important ways from the Southern British English (SBE) for which the transcription system was designed. As the authors explain, the corpus is intercultural, containing conversation between speakers of different varieties of English, mainly Hong Kong Chinese (HKC) and Native Speakers of English (NES) from a number of countries, but with an overall preponderance of HKC speakers. The authors claim in their conclusion that there are only very slight differences in the distribution of intonation features, but they are clearly unaware of the striking differences, such as accent placement in relation to information status, and the treatment of adverbials, that leap from the page in many of the sample illustrations.

Of the nine chapters, four are devoted to describing the corpus, its transcription and the software designed to search it. The remaining five are devoted to quantitative and qualitative analyses of the main features of intonation: tone units, accent placement (prominences), tone choice (e.g. between rise and fall), key and termination (relative pitch height) and a brief conclusion. Of the descriptive chapters, I shall focus on the account of the transcription process (Chapter 3) and of the remaining I shall focus on the analyses of tone units, prominence, tones and key/termination (Chapters 5, 6, 7 and 8).

Chapter 3 deals with the transcribing process. As every transcriber knows, the clearest of systems is often hard to apply to real speech. Even un-messy data, such as read aloud texts, proves challenging, and it is not surprising that in practice all aspects of the transcription proved difficult. David Brazil's intonation system (e.g. 1997) is an approach similar to that of Halliday (e.g. 1967), in that it is explicitly discourse oriented, and formally similar to other British approaches, such as currently used by, for example, Wells (2006). The only features unique to Brazil are the concepts of 'key' and 'termination' (relating to pitch height of onset and nucleus respectively), which signal relationships between successive utterances and some potential attitudinal consequences.

The detailed account of the transcribing process suggests that the transcription is internally consistent. There are, however, some problematic decisions that have consequences for subsequent analysis. This includes the definition of the tone unit, which involves assigning a boundary to every pause. The authors claim, rightly, that "there are no properties of the sound substance which can

uniquely identify a tone unit boundary” (p. 38), but there are many phonetic phenomena that can contribute to its identification, including pitch discontinuity, syllable lengthening and absence of connected-speech phenomena. There is also much in the literature to indicate that pauses are not a reliable indicator of a tone-unit boundary. The location of a boundary has, of course, implications for the identification of the ‘first’ and ‘last’ accented syllable (onset and nucleus), and hence what constitutes key and termination, so the identification of a tone unit impacts on all other analyses.

In the identification of tones, some of the problems described here are familiar. For example, it is well known that some post-nuclear syllables (part of the nuclear ‘tail’) appear prominent due to final lengthening, as in | I’ll be home \late this ·evening |, where the nuclear fall is on ‘late’ and the first syllable of ‘evening’ is slightly lengthened. These can be marked as ‘stressed but not accented’, as in the Spoken English Corpus (SEC) or with a non-starred pitch accent, as in the autosegmental system. It appears that the decision here is to ascribe a level, and therefore nuclear, tone to such syllables. This changes the location of the nucleus, and could explain why the authors observe so many level nuclei in the corpus. It also means that future analysts should be cautious when dealing with quantitative information derived from the transcription. Most importantly, it has implications for the analysis of the composition of tone units and any conclusions drawn.

Despite these caveats, my impression is that the transcription is consistent, and that an intonation specialist could derive very useful information, especially about the location of stress and the choice of tone. I will now turn to the chapters that present analyses of parts of the data.

In Chapter 5 the overview of tone-unit lengths indicates that most tone units contain between one and five words. The high number of one-word units is mainly ascribed to the use of interjections, which occur more in spontaneous speech than in more formal discourse. However, the high number of conjunctions (*and, but, because, or*) that appear in separate tone groups, and particularly personal pronouns (*I, you, we, they*) is likely to be a function of the decision to mark a boundary wherever there is a pause. In other systems (e.g. SEC), this would be marked as an internal pause, ascribed to mental planning or an interactional strategy, such as floor holding. The observations are therefore more likely to reflect pause distribution than tone unit boundaries, and are thus to some extent a question of performance rather than of phonology.

The distinction between phonology and realisation is blurred when the authors consider the examples of ‘or something’ as a tag (e.g. ‘maybe some earrings or something’). They observe that the tag is sometimes in a separate tone

group and sometimes part of the post-nuclear unstressed material, and claim that this choice reflects how far the referent is limited. My hunch is that a number of these ‘separate tone groups’ are in fact simply due to the presence of a pause, and therefore still melodically part of post-nuclear unstressed material. This kind of anomaly is well known in relation to reporting clauses (e.g. ‘... , he said’) – a functional view would simply posit a tone-group internal ‘rhetorical’ pause, while the autosegmental solution is to posit a separate tone group with ‘copied tones’ (see Gussenhoven 2004). Both solutions are, of course, anomalous – one results in a pause inside a tone group, while the other results in a tone group without a nucleus. This would not matter so much if pauses had been identified as such in the transcription, since this would allow users to make their own decisions as to what constituted a tone group in some cases.

The discussion of accent placement in Chapter 6 is both interesting and tantalising. Throughout the book, the authors provide samples of the transcription to illustrate their analyses, and in a number of cases they provide seemingly unwitting evidence for some interesting differences between Hong Kong English and Southern British English. Such differences appear to lie both in accent placement and in tone choice. The following extract is a good example of the latter. The utterance ‘er at the end of the term or something’ (line 71 Fig 5.1), is transcribed as

{ \ [ER] the END of the < TERM > } { \ [OR] < SOMETHing > }¹

This can be roughly simplified as ‘er the end of the \TERM | or \SOMETHing’, indicating that the ‘tag’ imitates the nucleus in the previous tone group. Assuming that the transcription is accurate, this indicates that this variety of English has a very different pattern of accent choice and placement from SBE, which would most probably de-accent the tag. Future researchers may well find this a very fruitful line of investigation. The same is true of Example 6.6 , which reveals a pattern of tone sequence that would not be acceptable in SBE: (simplified) ‘This course is the third \ CHOICE | \ ACTually’. An independent ‘actually’ with a falling tone is not possible in SBE – such adverbials are either de-accented altogether or, in final position, have a (falling-)rising tone.

The difference in accent placement is most obvious in relation to information structure. The de-accenting of given information, familiar in SBE, does not seem to operate in HKE. For example, 6.2 (simplified): ‘I’m sure you will go through a lot of HAPpy times, but I will guarantee you that you will go through a lot of miserable TIMES’. This pattern of accentuation, assuming it is correctly transcribed, does not conform to the SBE pattern of de-accenting given information, which would produce ‘... a lot of MISerable times’.

Some egregious errors undermine whole sections, such as the potentially interesting section on prosodic ‘convergence’. The discussion focuses on the changing stress pattern of ‘front OFFICE’ and ‘FRONT office’ in the course of a conversation, and ascribes it to convergence, a kind of ‘accommodation’ between speakers. Unfortunately, the analysis ignores the very normal stress shift well known to occur in certain rhythmic environments. This is the ‘intermediate accent rule’ (see e.g. Knowles 1987: 124) which accounts for the change from ‘indePENDent’ to ‘INdependent WITness’, and applies equally to the example reported here of a conversation about working in a hotel: ‘front OFFice’ changes to ‘FRONT office co-ORDinator’, ‘FRONT office dePARTment’, ‘FRONT office opeRATION’ in a similar metrical environment. Errors such as this severely undermine the validity of the analyses contained in the book.

The first part of Chapter 7 offers quantitative overviews of tone frequency in the corpus. I have already indicated that the frequency of level tones should be treated with caution, and future analysts should be aware of the transcription conventions used. The second part of the chapter offers some analyses including a comparison of tone groups with units in Linear Unit Grammar (Brazil 1995). Given my uncertainty concerning the transcription of tone groups it is difficult to judge how far the mapping is useful.

The analyses are sometimes elaborate speculation about things that have simpler explanations. There is a lengthy section discussing the use of rise and rise-fall tones, both of which are said by Brazil to be choices (over fall-rise and fall respectively) related to dominance and control. Since the rise-fall is rare in the corpus, the focus is mainly on rise tones. Three examples (7.25, 7.26, 7.27) illustrate the use of the supposedly ‘dominant’ rise tone in lists (e.g. ‘first we have the refu/GEES, the /RIots, the /WATER shortage, the \BANK runs’). It is true that listing is a way of holding the floor, but I am not convinced that the choice of a rise instead of a fall-rise makes the floor-holding strategy any more effective. Two further examples (7.29, 7.30) illustrate the use of a rise with questions in service encounters (Can I /HELP you sir) glossed as “exerting pressure on hearer to speak” (p. 152) and “probably perceived as warmer” (p. 153) than a fall-rise. On the whole, it is in the nature of questions to require an answer, and so the choice of tone must have a different function. The attitudinal nuances generated by the choice of ‘Can I help you’ with a rise, with a fall-rise, or even with a fall, need better explanation than we have here. The quantitative analysis (Fig 7.1) showing that the rise tone is used more frequently by dominant partners in unequal encounters (e.g. academic supervisions) needs to be related to the kind of speech acts performed in these encounters. Dominance, in

Brazil's sense, can only be expressed through tone when the speaker has a choice. Taking short passages where speakers appear to do normal things, like listing or asking questions, is not sufficient evidence. Brazil's insights are not counter-intuitive, but these analyses do not serve him well.

A particularly interesting feature of Brazil's system is that of 'key' and 'termination', discussed in Chapter 8. The categorisation of three different pitch heights on certain accented syllables (high, medium and low) is something generally assumed to be a realisational phenomenon (and gradient) rather than part of the phonology. Brazil's analysis does, however, constitute an early indication – missing from many accounts – of the pitch convergence that often occurs between speakers. In Conversation Analytic terms, 'pitch concord' would be seen as 'affiliative' behaviour, and recent applications of Social Accommodation Theory to prosody show that the underlying observations made by Brazil were extremely important. Frustratingly, much of the chapter consists of reiterating what Brazil said and then by verbalising tables of distribution figures. There seems to be little point in quantifying how much any speaker or group of speakers uses a certain pitch height on onset or nucleus, since this is clearly a syntagmatic issue – what matters is what follows what. The tables do show that 'mid termination' constitutes over 80 per cent of cases, which suggests that low and high are marked, but this is not discussed.

To summarise, this book contains a very rich source of information for anyone interested in intonational variation across different Englishes. It provides a large amount of prosodically transcribed data, an excellent software package with which to search it, and an appendix with over one hundred pages of preliminary quantitative information. However, readers would do well to focus on the corpus itself rather than the analyses that accompany it.

Note

1. Capitals indicate the prominent syllable.

References

- Brazil, David. 1995. *A grammar of speech*. Oxford: Oxford University Press.
- Brazil, David. 1997. *The communicative role of intonation in English*. Cambridge: Cambridge University Press.
- Gussenhoven, Carlos. 2004. *The phonology of tone and intonation*. Cambridge: Cambridge University Press.
- Halliday, M.A.K. 1967. *Intonation and grammar in British English*. The Hague: Mouton.

- Knowles, Gerry. 1987. *Patterns of spoken English: An introduction to English phonetics*. London: Longman.
- Knowles, Gerry, Briony Williams and Lita Taylor. 1996. *A corpus of formal British English speech: The Lancaster/IBM Spoken English Corpus*. London: Longman.
- Wells, John C. 2006. *English intonation*. Cambridge: Cambridge University Press.

Peter Collins. *Modals and quasi-modals in English* (Language and Computers 67). Amsterdam and New York: Rodopi, 2009. ix + 193 pp. ISBN: 978-90-420-2532-5. Reviewed by **Udo Fries**, University of Zurich.

This is an important book for students of modality in present-day English. As a corpus-linguistic study it has all the advantages one can draw from an electronic corpus, but it also shows the limitations of large corpora. The size of the three corpora Collins uses is certainly impressive. One million words each, taken from the British and the Australian components of the *International Corpus of English* (ICE), and another (almost) 200,000 words of texts from a corpus of American English, put together for the purpose of this study – from the *Santa Barbara Corpus* and the *Freiburg-Brown Corpus of Written American English* (Frown) – form the basis of this investigation. These 2,200,000 words yield altogether 46,121 modals and quasi-modals. It must have taken Collins a long time to study each and every one of these instances in their contexts. Unfortunately, he does not tell us how he managed to do this, and the reader has to rely on his judgement. What we do get is a thorough analysis of altogether 548 examples, in which the author shows over and over again that his individual interpretations are sound, that he has no scruples in taking over the argumentation of his predecessors in the field, or that sometimes there may be alternative interpretations. But Collins also offers new interpretations and he can be quite outspoken when he believes other views are wrong. All this makes the book worth reading.

Collins sees his study as “an exercise in corpus linguistics” (p. 5), as against corpus-driven studies. He makes use of the divisions of the ICE corpora into spoken and written texts, dialogues vs. monologues, and non-printed vs. printed texts. Further distinctions made available by the ICE corpora, though he lists

them early on (p. 2), hardly ever reappear in the study: for example the distinctions between academic, popular, instructional, persuasive, or creative writing.

In Chapter 2, “Theoretical preliminaries”, Collins discusses, albeit briefly, his choice of items, defines modal auxiliaries, and explains the term and the range of quasi-modals, which, according to him, consist of semi-modals and lexico-modals (related to Quirk *et al.*'s (1985: 141–146) distinction of modal idioms and semi-auxiliaries). The modals under investigation are *can*, *could*, *may*, *might*, *need*, *ought to*, *shall*, *should*, *will* and *would*, and the quasi-modals are *be able to*, *be about to*, *be bound to*, *be going to*, *be supposed to*, *be to*, *had better*, *have got to*, *have to*, *need to*, and *want to*. They range from the most frequent item, *would*, with 7,775 instances, followed by *can* with 7,663 instances, to the smallest, *be bound to*, with just 27 instances.

Collins discusses modal meanings, modal strength, degrees of modality, and the distinction between subjectivity and objectivity. He frankly admits that it is not always possible to distinguish between epistemic, deontic and dynamic meanings (his major categories), which probably explains the large number of indeterminate meanings: I counted, in the various tables scattered throughout the book, altogether 3,126 of them, corresponding to 6.7 per cent of all instances. It may be well worth studying these in another book, for which I would suggest to have the individual figures in the tables right-aligned, which would make the figures appear much more reader-friendly, especially when there are two sets of figures in one column.

The main part of the study is divided into chapters on “Necessity and obligation” (Chapter 3), “Possibility, permission and ability” (Chapter 4), and “Prediction and volition” (Chapter 5). In each chapter the relevant modals and quasi-modals are discussed one after the other: their deontic, epistemic and dynamic uses (where applicable), additional features (where necessary), and regional and stylistic variation. Regional variation refers to British, American and Australian English, which are sometimes referred to as *varieties* of English and sometimes – less fortuitously – as *dialects*.

Collins works with figures representing raw frequencies, frequencies normalized to tokens per one million words, and percentages. Percentages make sense when we have large groups of examples; with smaller ones they may be misleading, when, for example, 8.7 per cent of one usage is opposed to 28.6 per cent of another, and the former consists of 489 instances out of a total of 5,603, and the latter only six out of 21 (p. 136). Figures are important, especially in a corpus-linguistic study, but I find it very reader-friendly that not everything that could be counted has been counted. For large groups of phenomena we find phrases like *the majority of...cases* (p. 39), *normally* (p. 43, p. 48), *there is no*

shortage of examples (p. 101), *there is no shortage in the data* (p. 112), *are common* (p. 102), and – perhaps less satisfactory, for smaller groups, *occasional examples* (p. 46, p. 71), *the corpora yielded a small number of tokens* (p. 153), or *rare in the corpus* (p. 72, p. 74). Traditional grammarians may like this type of vagueness. And note that even in large corpora not everything that is theoretically possible is also attested, e.g. *should* in a *lest* clause (p. 49), *had better* with preceding negative (p. 79), or various uses of *shall* with a 2nd or 3rd person subject (p. 137).

A few formal things got wrong. There is something strange in the phrase *if it view turns out to be wrong* (p. 8), in *AmE display a marked dispreference for...in comparison to BrE and AmE* (p. 68), the singular *displays* would be appropriate and instead of the second *AmE* read *AusE*. There are some misprints, many of them typical of a manuscript proof-read by a computer programme instead of a human being: *million-word corpora* instead of *one million word corpora* (p. 1), *each of 2,000- words* (delete hyphen) (p. 2), *how they are they manifested* (delete second *they*) (p. 8), *I wish to thank to Edgar Schneider* (delete second *to*) (p. 9), *the issue is not one that that has ramifications* (delete second *that*) (p. 31), *that that the correlation* (delete second *that*) (p. 36), *does not not rival* (delete second *not*) (p. 45), *that we dealing with tendencies* (insert *are* after *we*) (p. 59), hyphen missing in ICEGB in example (227) (p. 84), several times italics missing for *be to*, (p. 86, 6th line from bottom), for *can* contrasts with *may* (p. 97, 8th line from bottom), for *can* (three times) in example (30) (p. 99), for *might* in example (83) (p. 110).

Some of the literature quotes in the text do not find their counterparts in the list of references at the end of the book: Römer 2006 (p. 5/6) may be Römer (2005), Berglund 1999 (p. 6) is probably Berglund (2000) in the list of references, Close 1981 (p. 55) and Aarts and Aarts 1995 (p. 153) are not listed at all in the reference section.

Examples (30) and (48) in Chapter 3 are identical, but their source is not: (ICE-GB 89 and 108), *British and American corpora* should read *British and Australian corpora* (p. 72), the frequency figures at the bottom of p. 91 got curiously wrong: for 7763 read 7663, for 3357 read 3557, and for 1499 read 1599. There is a bracket missing on p. 128 (line 13) after the colon, and the second example in this line should read *she is seventeen* instead of *she'll be seventeen*. For *Volitional* read *Volitional* p. 137 (line 11), for (86) read (85) p. 142 (line 5), for *moreso* read *more so*, p. 143 (line 6), the reference to *Sections 5.4.2. and 5.4.3 below*, p. 150, should be to *Sections 5.5.3 and 5.5.4*.

The main strength of this book lies in its many individual observations. The best thing for a reader to do is to begin with the “Conclusion” (Chapter 6),

which gives the major results and relates the many individual observations to major trends, the relationship between the three varieties of English and their diachronic development in the second half of the 20th century. From there one should turn to the individual sections and subsections of Chapters 3, 4, and 5 for a sound interpretation of a fascinating section of English grammar.

Reference

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

Chris Greaves. *ConcGram 1.0: A phraseological search engine* (Studies in Corpus Linguistic Software 1). Amsterdam and Philadelphia: John Benjamins, 2009. CD-ROM. ISBN 978-90-272-4027-9. Reviewed by **Andrew Kehoe**, Birmingham City University.

1 Background

Over the past decade there have been many technological advances which have impacted on corpus linguistics, in terms of the size of corpora processable on a desktop PC and the range of analytical functions available to the researcher. Despite these advances, there remain relatively few software packages designed specifically for the automatic extraction of phrases from corpora. One reason for this, and one to which this review will return, is that the extraction of certain kinds of phrasal structure requires resources still beyond the capacity of current PCs. Another reason is that the concept of ‘phrase’ remains somewhat loose, defined in different ways by different scholars. The situation has not moved on significantly from that described by Altenberg in 1998, who began his study of the phraseology of spoken English in the following way, before limiting it to the computationally-tractable analysis of continuous strings of words recurring in identical form:

Phraseology is a fuzzy part of language. Although most of us would agree that it embraces the conventional rather than the productive or rule-governed side of language, involving various kinds of composite units and ‘pre-patterned’ expressions such as idioms, fixed phrases,

and collocations, we find it difficult to delimit the area and classify the different types involved. (1998: 101)

Gries (2008) provides a useful overview of phraseological developments over the past decade in various fields of linguistic study. He proposes a more rigorous definition of phraseology based on six parameters, the most significant of which for the present review is “the permissible distance between the elements involved in a phraseologism” (p. 4). Gries notes that *n*-gram-based studies, including work on word clusters and lexical bundles (Biber *et al.* 2004), concern themselves only with immediately adjacent words. The upper limit of *n* is usually five. Work on collocation widens the span of analysis but, as Gries comments, it is common for researchers working on collocation to restrict themselves to phrases of a particular length or detectable within a chosen span. With this in mind, Gries concludes that

a top-down, or *a priori*, approach may not always be the most useful strategy. Sometimes it may be more revealing to let the data – rather than the preconceptions of any particular researcher – decide what the potentially most revealing pattern is. (2008: 21)

2 *ConcGram: Introduction*

ConcGram is a software package which attempts to facilitate the kind of corpus-driven analysis of phraseology proposed by Gries. The concgramming approach first came to the attention of corpus linguists in the 2006 paper “From *n*-gram to skipgram to concgram” (Cheng *et al.* 2006). The intermediate step in this sequence, the skipgram, is one which Gries does not mention in his 2008 survey. A skipgram, or gapped *n*-gram, is a refinement of the *n*-gram approach designed to detect non-contiguous word associations. The free kfNgram program (Fletcher 2007) is a skipgram implementation available for corpus linguists to run on a desktop PC. This works by first generating *n*-grams from a corpus and then matching these against each other to detect ‘phrase frames’. For example, 4-grams like ‘as well as the’, ‘as far as the’ and ‘as long as the’ are grouped by kfNgram under the phrase frame ‘as * as the’. Whilst this is an advance on the basic *n*-gram approach, the limitations, as Cheng *et al.* point out, are that kfNgram allows only one word position to be skipped and it cannot handle positional variation (i.e. it can detect ‘A * B’ sequences but not ‘B * A’).

ConcGram is designed to overcome these limitations by detecting “all of the permutations of constituency variation and positional variation generated by the association of two or more words” (Cheng *et al.* 2006: 414). The software devel-

oper describes ConcGram as “the first program able to identify up to five co-occurring words, irrespective of either constituency or positional variation, in a text or corpus fully automatically” (p. v).

It must be said at this stage that the sub-title of the program, “a phraseological search engine”, is somewhat misleading. The term ‘search engine’ does describe the operation of the program in one sense but it is now so closely associated with the web that the sub-title may lead some potential users to conclude that ConcGram searches the web for phrases. It does not and, whilst it would be possible to run ConcGram on texts downloaded from the web, there is no feature in the program for downloading web texts (like the ‘WebGetter’ utility in WordSmith Tools, for instance).

The ConcGram software is supplied on CD-ROM, an increasing rarity in an age of digital downloads and web-based applications. The setup program on the CD installs the program itself on the user’s PC, along with tutorial files and a comprehensive 90-page user manual in PDF format (the latter also freely available online: <http://www.benjamins.com/jbp/series/CLS/1/manual.pdf>). I have been running ConcGram for several months and have worked through the manual, trying the various features of the program on a variety of corpora. Space does not permit a full discussion of every feature, so this review focuses on the main strengths and weaknesses of the program as I perceive them, beginning with a look at its basic operation.

3 Basic operation

Upon running ConcGram, the user is presented with a standard Windows interface (Figure 1). The program has the appearance of a text editor, with font and formatting options on the toolbars, and it does, in fact, handle its input (corpus) and output (wordlist, concgram) files in text format. It is in the menu of commands across the top where ConcGram’s main features can be found, particularly under the ‘Concordance’, ‘Concgrams’ and ‘Statistics’ headings. The ‘Help’ menu brings up a useful help file explaining the commands, which contains some of the same material as the separate PDF manual.

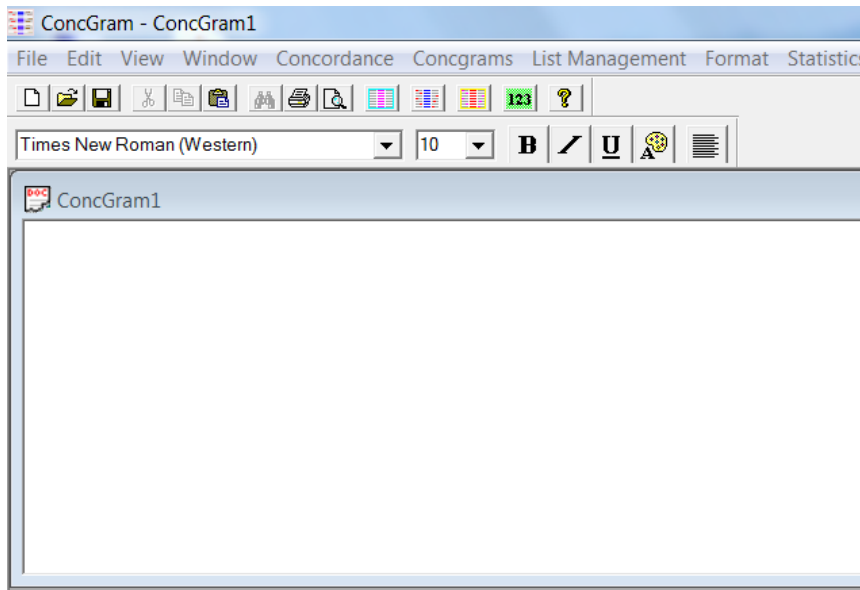


Figure 1: Basic ConcGram interface

The manual states that “ConcGram has all of the functions usually associated with traditional corpus linguistics software, such as the generation of word frequency lists, the determination of the specificity (i.e. ‘keyness’) of single words (plus two-word concgrams), the generation of single origin concordances, mutual information values, t-scores, and so on” (Warren’s introduction to Greaves *op. cit.*, p. 5). The most obvious referent here is WordSmith Tools (Scott 2009), the software which popularised the concept of ‘keyness’. A full feature-by-feature comparison of ConcGram and WordSmith is beyond the scope of this review, which instead focuses on the ‘unique feature’ (Greaves *op. cit.*, p. 4) of the former: the analysis of concgrams.

The generation of concgrams is an iterative process, beginning with 2-word concgrams and building to 5-word concgrams. 2-word concgrams are created by identifying all unique words (types) in the corpus and then generating the full set of concordance lines for each of these node words, using a default span of 50 characters on either side of the node (Cheng *et al.* 2006: 415–416). For each node, those words appearing within the span (an average of 9 or 10 words each side according to Cheng *et al.*) are paired in a 2-word concgram with the node.

Tutorial 1 in the manual guides the user through this process, using as its corpus one of two text files packaged with the program containing speeches by the Chief Executive of Hong Kong. Figure 2 shows a frequency-sorted 2-word conigram list for this file:

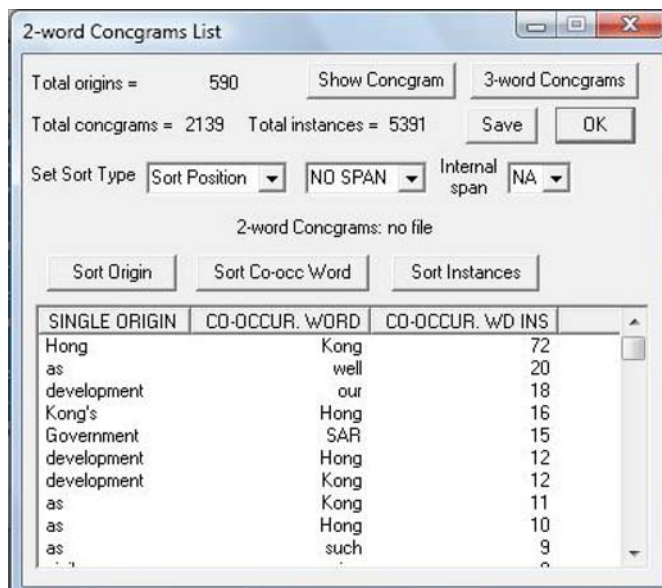


Figure 2: Sorted 2-word conigram list box (p. 72)

This output shows that 2,139 2-word conigrams have been generated from 590 origins (types), the most frequent of which is *Hong/Kong*. It is possible to view a conigram concordance display by selecting a pair and pressing the 'Show Conigram' button. The example given in the tutorial is *development/our* (Figure 3):

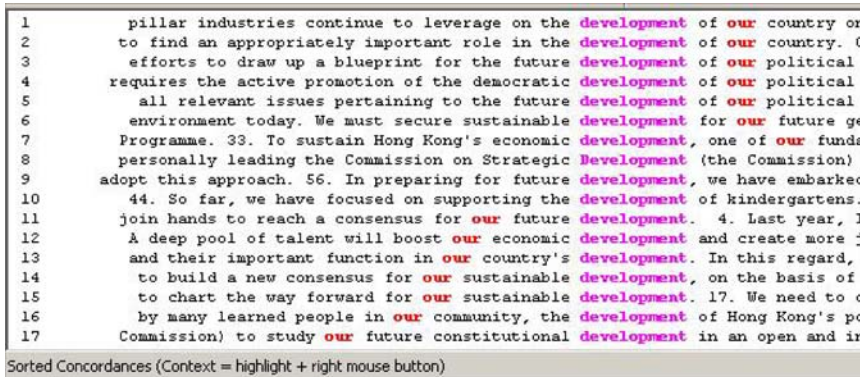


Figure 3: 2-word concgram for 'development/our' (p. 74)

This illustrates how ConcGram captures positional and constitutional variation, but potential users may be forgiven, at 2-word concgram level, for asking what this approach offers over standard collocational analysis, albeit collocational analysis with a wider span than usual. One may also question how concgram output differs from the familiar KWIC output produced by several other software packages. The response of the authors is that

concgrams represent a serious challenge to the current view about word co-occurrences that underpins the KWIC display. Studying KWIC displays, which only highlight the node (i.e. the centred word), has unintentionally created, in the minds of some users, a hierarchical approach which puts the node as the centre of attention and the words associated with the node as subordinate to it (Warren in Greaves 2009: 4)

Many would agree with this argument and welcome the reassessment of concordance display offered by the program. ConcGram provides a useful feature which enables the user to re-centre the output on any word in the concgram at the click of a button. The problem is that concgram output *looks* so much like standard KWIC output. The novelty of the concgram approach would perhaps have been more apparent had the software presented output in a radically different way. The situation is not helped by the ambiguous terminology used by the program, with both the word associations (e.g. the 2-word association *development/our*) and the concordance output (as in Figure 3) referred to as 'concgrams'.

The main strengths of ConcGram come to light when the analysis is extended to 3-, 4- and 5-word associations. In generating these, the program introduces the notion of ‘origins’. For the initial 2-word concgram search, each type in the corpus was used as a search origin. The 3-word concgram search takes each 2-word concgram as an origin and repeats the process, generating all concordance lines for the double-origin search and counting occurrences of other words within the concordance span. Similarly, 4-word concgram search uses all triple origins from the 3-word output, and 5-word concgram search uses quadruple origins from the 4-word output. Figures 4 and 5 show examples of 4- and 5-word concgrams respectively, taken from the manual:

4-Word Concgrams List

Unique origins = 7307 Show Concgram All 5-word Concgrams

Unique concgrams = 24998 Total instances = 81737 Specific 5-word CG

Set Sort Type: Sort Position USE SPAN Set span: 4 Save

4-word Concgrams: no file

Sort Triple Origin Sort Co-occ Word Sort Instances OK

TRIPLE ORIGIN	CO-OCC WORD	CO-OC WD INS
as / such / and	in	26
carried / out / to	the	26
engineering / structural / and	of	26
Hong / Kong / and	between	26
Hong / Kong / in	used	26
Hong / Kong / The	Kong's	26
international / Proceedings / on	Symposium	26
material / structural / micro	the	26
as / such / a	of	25
building / model / of	the	25

Preview of concordance lines on the right: ...s of the Second, ...s of the Fourth, ...s of the Second, ...gs of the Sixth, ...gs of the Ninth, ...gs of the Sixth, ...ngs of the 10th, ...ngs of the 10th, ...ngs of the 10th, ...ngs of the 10th, ...ngs of the 10th, ...ngs of the 8th, ...ngs of the 5th, ...ngs of the 8th, ...ngs of the 4th, ...ngs of the 4th, ...ngs of the 9th, ...ceedings of the, ...ceedings of the, ...roceedings, 2nd, ...Proceedings of

Figure 4: 4-word concgram for ‘international/Proceedings/on’ with ‘Symposium’ (p. 32)

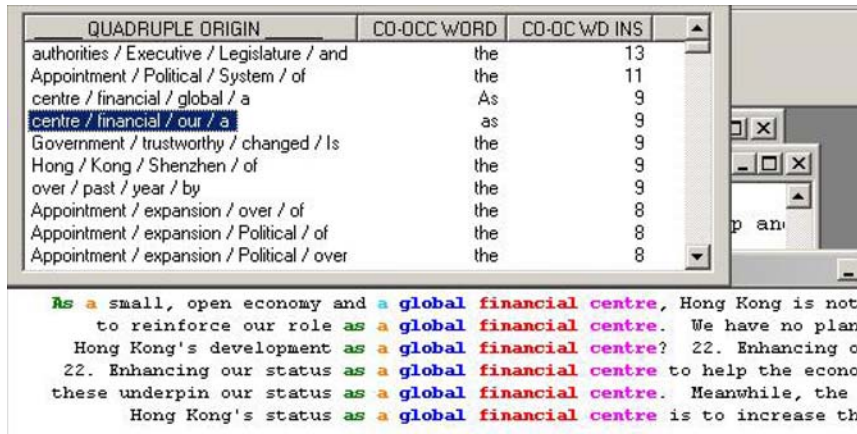


Figure 5: 5-word concgram for 'centre/financial/our/a' with 'as' (p. 34).

Figure 5 was perhaps not the best example to illustrate ConcGram's capabilities in the manual, as all but one of the cases shown are contiguous examples of 'as a global financial centre'. Also note that the wrong line is highlighted in the list: 'centre/financial/our/a' instead of 'centre/financial/global/a'.

This is a recurrent problem in both the ConcGram manual and help file, which contain numerous examples of incorrect Figure numbers, frequencies of occurrence in screenshots which differ from those mentioned in the text, and missing steps in tutorials. In some cases, the instructions are correct in the manual but not in the help file, or vice-versa, but neither is consistently correct. For example, Tutorial 1 in the manual instructs the user to create the initial 2-word concgram output by selecting 'CONCGRAMS > CREATE NEW CONCGRAM LIST (AUTOMATIC) > USING ALL THE WORDS IN TEXT > WITH NO INITIAL UW LIST'. However, 'WITH NO INITIAL UW LIST' is not available as an option at this stage. The version of this tutorial in the help file contains the correct instructions, but omits important steps in other tutorials. This is unfortunate as some of the concepts being discussed are new and quite difficult to grasp on first reading; yet the user is often hindered by unclear or incorrect instructions.

4 Kinds of research possible

The manual describes the contribution ConcGram can make to our understanding of phraseological variation, with reference to Sinclair’s idiom principle (1987) and three categories of phraseology: meaning shift units (Sinclair 2007), collocational frameworks (Renouf and Sinclair 1991), and organisational frameworks (Greaves and Warren 2008).

The analysis of constituency and positional variation facilitated by ConcGram is of tremendous benefit in the study of meaning shift units. The example given in the manual is *play/role*, for which the software generates variants including ‘play a role’, ‘play a key role’, ‘play a much more significant role’, ‘an important role to play’ and ‘the role major companies should play’.

Turning to collocational frameworks, the manual lists the five most frequent cases in a 5 million word sample of the BNC (‘the...of’, ‘of...the’, ‘in...the’, ‘a/an...of’, ‘the...of...the’). It then presents “examples of the top two” in the form of the 2-word concgram *of/the* (Figure 6):

```

1         a Japanese motor manufacturer by a member of the public (ASA Ca
2         <u who=PS1HH> Ah! <u who=PS000> First of all the [unclear]
3         she had a house to start with. When you think of all the families o
4         that’s just enforcing just merely a question of enforcing the law.
5         serious implications as to what we do in terms of improving the high
6         to get there) thoroughly. There is no other way of anticipating the p
7         careful planning, especially in the allocation of committee rooms f
7         law. Sections “B” and “C” outline the provisions of the civil and cri
9         gradually increased, including the resumption of coffee cultivation
10        of advice in recent years about the importance of a healthy diet, a
11        secondary, day or boarding. The vast majority of schools are “gover
12        shareholders as dividends then the capital base of the business has b

```

Figure 6: 2-word concgram ‘of/the’, illustrating collocational frameworks (p. 7)

Whilst constituency variation, or variable distance between the fixed slots, is undoubtedly useful in the analysis of collocational frameworks, the importance of positional variation is debatable. The above output blurs the distinction between the two separate frameworks ‘the...of’ and ‘of...the’ by merging them into a single concgram *of/the* (the constituents written in alphabetical order in the software’s notation). In such cases, it could be argued that ConcGram makes phraseology more fuzzy rather than less so, in Altenberg’s terms.

The manual also describes how the software can be used to examine the aboutness of a text. Several recent studies have used concgrams in this way,

including work by Greaves and colleagues and by other researchers (e.g. O'Donnell *et al.* 2008). However, the manual does not give a convincing example of 'aboutgrams' (a term suggested to Greaves and colleagues by John Sinclair). Greaves selects twelve concgrams manually from the frequency-sorted 2-word output produced in Tutorial 1 (see Figure 2) and suggests that these "show the 'aboutness' of the speech" (p. 73): *Hong/Kong, development/our, Government/SAR, development/support, support/families, air/quality, Chief/Executive, Council/Legislative, Executive/Chief, Government/Development, community/support, development/economic*. Whilst some of these concgrams are indicators of aboutness, on the whole, they offer little more than a frequency sorted wordlist.

A more methodologically sound approach would be to compare the test corpus against a larger reference corpus to generate 'key' concgrams, akin to the key words and key clusters produced by WordSmith Tools. There is a 'keyness' facility in ConcGram, referred to as 'specificity', but this works with single words only rather than with concgrams. This is contrary to Warren's comment in his introduction to the ConcGram manual, cited in Section 3 of this review. The calculation of specificity is based on the 'weirdness' formula (Ahmad 2005) and not the more familiar log-likelihood. This simple calculation divides the percentage occurrence of a word in the test corpus by its percentage occurrence in the reference corpus. Thus, a score greater than 1 means that the word is specific to the test corpus, whereas a score less than 1 means it is not. In the example given in Tutorial 4 of the ConcGram manual, there are many words with 'infinite specificity', meaning that a word from the test corpus does not appear in the reference corpus. Rather confusingly for novice users, this is displayed by the program as '1.#INF', which looks like a calculation error. An unfortunate effect of this is that, when Greaves presents his specificity output and states that "the word 'technology' has the highest specificity" (p. 83), the user must look beyond the many cases of infinite specificity to the 820th item in the list.

Another statistical tool is provided by ConcGram for testing the significance of co-occurring words, using *t*-score and Mutual Information (MI) calculations. These are available only for 2-word concgrams as "the formulas for calculating both...only provide values for the co-occurrence of 2 words" (p. 53). Thus, for 3-word concgrams and above, frequency is the only measure available.

5 User interface

As stated, ConcGram has a standard Windows interface and, as a result, learning how to use the software is a relatively straightforward process. There are, how-

ever, a few quirks which may hamper the user. One minor annoyance is that the main menu options are visible only if a text file has been opened. In Figure 1, an 'empty' text file is open (called 'ConcGram1'), but if the user closes this file, all menu options except 'File', 'View' and 'Help' disappear. This would make sense if the missing menu items were only usable with a corpus file open, but this is not the case. There are many useful features in the 'Tools' and 'List Management' menus, and several features in the other menus, which do not require a file to be open but which are not available if a file is not open. Compounding this problem is the fact that there are several points in the accompanying tutorials where the user is instructed to 'close all open files', thus causing the menu options to disappear.

The fact that the program handles all input and output as plain text files is a potential strength and should mean, for instance, that the user is able to import wordlists generated by other programs. In reality, the situation is not quite as straightforward since ConcGram uses a specific format for its wordlist files different from that used by other programs. A more serious problem with the text file approach is that ConcGram does not discriminate between corpora, wordlists and concgram output. If the user has a corpus file open and saves concgram output, the resulting text file opens in the main window. If the user then runs a concordance search, the program searches the corpus file but also searches the concgram output file, producing meaningless concordances. This pitfall is noted in the manual but the point is that the design of a program should prevent the user from making such obvious errors.

Any software program, especially one as complex as ConcGram, will contain some bugs, and there will be some aspects of its user interface which can be refined by the developer as common user errors emerge. For example, Mike Scott, developer of WordSmith tools, releases new 'builds' of the software on a regular basis, which fix bugs and add new features, and are free of charge to registered users. There does not appear to be a similar mechanism for ConcGram at present, but it would be beneficial if one were established.

6 Technical requirements

I installed ConcGram on PCs of varying specification running Windows XP, Vista, and 7 and the process was a straightforward one, the program requiring a modest 18.5MB of hard disk space. The CD case does not list any minimum system requirements for running the program but the manual states that it will take one whole day to create the initial 2-word concgram list for a corpus of 1 million tokens (18,000 types) on a Windows XP PC with a Pentium 4 3GHz processor

and 2GB of RAM. This is an average specification for a modern home PC, although it should be noted that many university PCs are of a lower specification and may struggle with the memory-intensive ConcGram software.

The developer seems to be aware of this and has included a whole section in the manual on “monitoring virtual memory”. He suggests that, “to avoid getting an ‘Out of virtual memory’ error”, the user should open the ‘Task Manager’ in Windows and manually inspect the ‘Physical Memory Usage History’ chart. “If the blue line reaches the top line”, Greaves advises, “the search results should be stopped and saved before resuming later” (p. 43). This is not good software development practice. A Windows program should generate an ‘Out of virtual memory’ error only in exceptional circumstances, and a PC user should not be expected to monitor memory usage in the manner suggested.

Even running ConcGram on a top-of-the-range PC is no insurance against the kind of virtual memory error discussed above, given that the program is expected to run for a whole day (or perhaps longer) as part of its normal operation. Desktop PCs are simply not designed to run large processes for such a long period of time and anyone running ConcGram on a PC of average specification would be unable to use their computer for other tasks during that time (checking email, browsing the web, writing Word documents, etc).

No specific technical details are given in the documentation but the main reason for the memory-intensive nature of ConcGram is apparently that it does not perform any indexing of corpora or caching of results. The corpus is read in as a raw text file and the program searches this from the beginning on each iteration, from 2-word concgram to 5-word concgram. An example of apparent duplication of effort appears in the description of the ‘Configurations List’ for 2-word concgrams. One of the more useful features of the programs, this displays frequencies for positional and constitutional variants of each pair in the 2-word concgram list (i.e. ‘A B’, ‘A * B’, ‘A * * B’, ‘B A’, ‘B * A’, ‘B * * A’). However, to calculate these frequencies, ConcGram performs a full text search of the corpus for each concgram, despite the fact that the full text has already been searched to generate the concgrams in the first place. Indexing or caching may have increased speed considerably in such cases.

The ConcGram manual does suggest several methods for speeding up search. These include the specification of an exclusion list to stop the program generating concgrams containing high frequency words (although this is not possible when looking at collocational frameworks of course) and the running of the program on smaller batches of words (perhaps concgrams beginning with a single letter). The main way the user can speed up search is by specifying the initial input, either as a single word or as a list of words, instead of generating

congrams for all types in the corpus. In this ‘user-nominated’ mode, the user has more control over the congram options and can search using single character wildcards and ‘prefixes’/‘suffixes’ (words beginning or ending with a particular string rather than affixes in a morphological sense). Fully automated exhaustive search is the default mode in ConcGram but perhaps the program would have been more user-friendly had the less memory-intensive user-nominated mode been the default.

7 Summary

The developer of ConcGram claims that “the automatic identification of congrams remains its main distinguishing feature” (p. v) and this is certainly presented as the ‘unique selling point’ of the program. In fact, the most recent version of WordSmith Tools (version 5) can also generate congram output, through the ‘WSConcGram’ utility, and it was this tool that was used to produce congrams in the research on aboutness mentioned above (O’Donnell *et al.* 2008). The WordSmith congram implementation uses a corpus index to increase processing speed and can handle corpora of up to 10 million words (Mike Scott, personal communication). When tested on the 1 million word Brown corpus, it took just over six minutes to produce 823,489 congram partnerships. WordSmith displays congrams in a different way than the ConcGram program, listing individual words and allowing the user to select a word or group of words before displaying congrams of all lengths containing the word(s). The WordSmith version does not provide the same range of configuration options as ConcGram and the concordance output format is not as flexible. Neither program provides a perfect solution but the point is that WordSmith Tools is a package which many corpus linguists will already own. ConcGram cannot compete with WordSmith on WordSmith’s terms as an all-round corpus analysis tool but, due to the severe performance limitations, nor can it compete with WordSmith on its own terms. It is not clear that ConcGram offers enough significant advantages to be worthy of recommendation as a stand-alone product.

On paper, congrams are an exciting concept and one which has the potential to transform the study of phraseology. However, the implementation of this concept in the ConcGram software package is flawed and, in its current state, the program has a rather experimental feel. The examples in the manual hint at the potential of congrams as a new research paradigm and the tiny corpus (8,389 tokens) used in the tutorial serves to demonstrate basic program function-

ality. The problem is that, as a result of technical limitations, the user is unable to use ConcGram to its full effect on corpora of any meaningful size.

References

- Ahmad, Khurshid. 2005. Terminology in text. Paper presented at the Tuscan Word Centre International Workshop: *Dial a Corpus*. Certosa di Pontignano, Italy.
- Altenberg, Bengt. 1998. On the phraseology of spoken English: The evidence of recurrent word combinations. In A.P. Cowie (ed.). *Phraseology. Theory, analysis, and applications*, 101–122. Oxford: Clarendon Press.
- Biber, Douglas, Susan Conrad and Viviana Cortes. 2004. *If you look at...: Lexical bundles in university teaching and textbooks*. *Applied Linguistics* 25: 371–405.
- Cheng, Winnie, Chris Greaves and Martin Warren. 2006. From n-gram to skip-gram to concgram. *International Journal of Corpus Linguistics* 11 (4): 411–433.
- Fletcher, William H. 2007. *kfNgram software*: <http://www.kwicfinder.com/kfNgram>
- Greaves, Chris. 2009. *ConcGram 1.0: A phraseological search engine – user manual*: <http://www.benjamins.com/jbp/series/CLS/1/manual.pdf> (also on ConcGram CD).
- Greaves, Chris and Martin Warren. 2008. Beyond clusters: A new look at word associations. *IVACS 4, 4th International Conference: Applying Corpus Linguistics*. 13–14 June 2008, University of Limerick, Ireland.
- Gries, Stefan Th. 2008. Phraseology and linguistic theory: A brief survey. In S. Granger and F. Meunier (eds.). *Phraseology: An interdisciplinary perspective*, 3–25. Amsterdam and Philadelphia: Benjamins.
- O'Donnell, Matthew, Mike Scott and Michaela Mahlberg. 2008. Exploring text-initial concgrams in a newspaper corpus. *7th International Conference of the American Association of Corpus Linguistics*, Brigham Young University, Provo, Utah, USA, 12–15 March, 2008.
- Renouf, Antoinette and John McH. Sinclair. 1991. Collocational frameworks in English. In K. Ajimer and B. Altenberg (eds.). *English corpus linguistics. Studies in honour of Jan Svartvik*, 128–143. Harlow: Longman.
- Scott, Mike. 2009. *WordSmith Tools* version 5. <http://www.lexically.net/wordsmith>.

Sinclair, John McH. 1987. Collocation: A progress report. In R. Steele and T. Threadgold (eds.), *Language topics: Essays in honour of Michael Halliday*, 319–331. Amsterdam and Philadelphia: Benjamins.

Sinclair, John McH. 2007. Collocation reviewed. Ms, Tuscan Word Centre, Italy.

Andreas H. Jucker, Daniel Schreier and Marianne Hundt (eds.). *Corpora: Pragmatics and discourse*. Amsterdam and New York: Rodopi, 2009. 516 pp. ISBN: 978-90-420-2592-9. Reviewed by **Thomas Egan**, Hedmark University College.

This volume contains a selection of papers from the 29th ICAME conference, arranged by the University of Zurich at Ascona in May 2008. There were almost a hundred presentations at the conference, in the form of papers, work-in-progress reports and posters. A large number of papers were reworked for publication and, from these, the editors selected twenty-two for inclusion in this volume. The stated theme for ICAME 29 was *Corpora: Pragmatics and Discourse*, so it comes as no surprise that this is the title of the book, nor that papers dealing with aspects of pragmatics and discourse studies account for some half of the contents. The editors have, however, also found space for papers dealing with other aspects of corpus-based language studies, as will be apparent from the discussion below. In a review of this length it is, of course, impossible to do justice to any, let alone all, of the individual contributions. Let me therefore state at the outset that I found this a very stimulating volume, with a wide range of state-of-the-art papers on various aspects of English corpus linguistics.

The volume is divided into three sections. The first of these is entitled “Pragmatics and discourse”, mirroring both the title of the book and the theme of the conference, and contains ten papers. The second, containing nine papers, is called “Lexis, grammar and semantics”, while the third and shortest section contains just three papers, grouped together under the heading “Corpus compilation, fieldwork and parsing”. The editors have chosen to include in the first section only papers whose main focus is on either pragmatics or discourse. Thus, the fact that papers are placed in sections two or three does not mean that they do not touch on pragmatics. Several of the papers in the second section, among them Arja Nurmi’s on *may* and Daniël Van Olmen’s on imperatives, explicitly

do so. However, although one could argue that these two papers belong in the first section, the editors' more stringent criteria for inclusion therein do perhaps make for more homogeneous sections. In the following I will, however, not adhere to the order of presentation in the book in my discussion of the various papers.

Section One begins with an overview of recent work in the field of historical corpus pragmatics by Thomas Kohnen. As Kohnen points out, all work in this field must be termed 'recent', the disciplines of corpus linguistics and historical pragmatics only coming together some fifteen years ago. Given the relative newness of the field, its theoretical and methodological parameters have not yet been set in stone. Having presented approaches taken by various scholars, Kohnen presents some of the challenges posed by the task of studying speech acts from a historical perspective. He stresses the need for a 'coherent research agenda' to pursue the aim of establishing a history of English in terms of genres and speech acts and advocates adopting what he terms 'a genre-based empirical bottom-up methodology'.

One genre that has been the subject of growing attention in recent years is that of medical texts, following the compilation in Helsinki of corpora of such texts from the Middle English and Early Modern English periods. Irma Taavitsainen, joint-leader of this corpus project, discusses the weakening role played by classical authorities in the early modern period in a paper entitled "The pragmatics of knowledge and meaning: Corpus linguistic approaches to changing thought-styles in early modern medical discourse". She first employs corpus methodology to trace the incidence of authorities such as Galen and Hippocrates and more generalised sources of authority such as 'philosophers' and 'physicians' from 1500 to 1700. She notes a general decline in the former and an increase in the latter, but this development is by no means uniform across all six of the categories into which the corpus is divided. On the subject of categories one minor quibble is that the author refers to subcategories '2a and 2d' (on p. 47), without having explained to the reader the composition of these subcategories. A qualitative analysis of works from the various categories shows that the category of specialised treatises contains several works from the late sixteenth century with passages reminiscent of medieval commentaries. Taavitsainen examines one of these in detail, plotting the occurrences in the text of references both to classical authorities and contemporary authors. She points out that passages containing references to both of these are particularly revealing of changing thought styles.

The Corpus of Early Modern English Medical Texts also furnishes the data for Jukka Tyrkkö and Turo Hiltunen's paper on the frequency of nominalization

in medical writing. They concentrate on suffix-marked word-level nominalizations and show that there was a steady, statistically significant increase in their employment across the two-hundred year period from 1500 to 1700, with the increase starting as early as the sixteenth century. They also note that, contrary to what they had expected, there is no significant difference in the increased use of nominalizations between the three subcorpora of specialised texts, health guides and surgical texts. They do, however, point to significant differences between individual writers in the employment of nominalizations and make the interesting observation that “the writers who nominalize most flexibly tend to be some of the most noted writers of seventeenth-century medicine – a fact which may partly explain why the writers who came after them endeavoured to affect a similar style” (p. 311).

Another genre for which a family of corpora has been assembled at Helsinki, that of correspondence, is the topic of three papers, by Minna Nevala on the sociopragmatics of *friends* in the seventeenth and eighteenth centuries, by Minna Palander-Collin on self-reference in the sixteenth and eighteenth centuries and by Arja Nurmi on the development of *may* from the fifteenth to the eighteenth century. Nevala shows that, whereas friends were often acquired by birth or marriage in the seventeenth century, they were much more likely to be acquired for reasons of affection or even for strategic considerations in the eighteenth. She divides people referred to as ‘friend’ into four broad categories, ‘family’, ‘intimate friends’, ‘acquaintances’ and ‘more distant’, and shows that whereas the use of *friend* to refer to the first of these declined from the seventeenth to the eighteenth century, it rose for each of the other three groups. She states that the figures for the four groups are all statistically highly significant. The significance attached to the group as a whole is, however, entirely due to the categories of ‘family’ and ‘intimate friends’. The difference for the other two categories is not significant.

In her paper entitled “Self-reference and mental processes in early English personal correspondence: A corpus approach to changing patterns of interaction”, Palander-Collin compares the use by male writers of the pronoun ‘I’ in the sixteenth and eighteenth centuries. She demonstrates that it increased in frequency and that much of this increase is related to its use with mental process verbs such as *hope*, *think*, *believe*, *wish*, etc. Although it might appear that the increase in the use of ‘I hope’, for example, could be taken to signal a more self-centred attitude, Palander-Collin shows that this is not in fact the case, with the hopes expressed often being for the well-being of the addressee. Self-reference is thus shown to often serve interpersonal purposes. Interpersonal purposes, specifically purposes of indicating the possession of authority, were also served by

the auxiliary *may*, discussed by Arja Nurmi in her paper “*May*: The social history of an auxiliary”. Nurmi charts the use, and gradual decline, of *may* according to the gender, social status and educational background of the letter writers, as well as their relationship to the recipients of the letters. In a nutshell, she finds that the use of *may* declined across the board after 1500, with women and uneducated men from the lower ranks of society leading the way. Educated males of higher rank also use *may* less as time passes, but they continue to use it more often than other letter writers. It is no doubt this fact that made the form attractive to social climbers. As Nurmi points out: “steep aspirers increase their use of *may* longer than letter writers in general, with a peak in the seventeenth century” (p. 328).

In addition to medical texts and correspondence, a third genre that is explored in several papers is that of religious texts. Both Tanja Rütten and Ingvilt Marcoe base their papers on subsections of the Corpus of English Religious Prose, under compilation at the University of Cologne. Rütten traces the part played by exhortation in sermons and treatises from the fourteenth to the seventeenth century and Marcoe the use of subordinating conjunctions in treatises and prayers. Rütten shows that both the incidence and form of exhortation change with the reformation. The amount of exhortation declines and it becomes more indirect. With the change in the role of the priest from God’s mouthpiece in His dealings with the congregation, to the congregation’s mouthpiece in its dealings with God, it is not surprising to find priests adopting a more circumspect tone in their dispensation of moral instruction. Moreover, as Marcoe notes in regard to a rise in conditional and concessive clauses in religious tracts: “In religious texts of the EModE period, it is no longer sufficient to simply state the principles of the Christian doctrine to instruct the reader in the basic tenets of the Christian faith. With the diversification of beliefs, writers feel compelled to argue their points of view and to add justifications for the beliefs they wish to disseminate” (p. 389).

There is one more paper with a historical focus in the book. In “Digital editions for corpus linguistics: Representing manuscript reality in electronic corpora”, Alpo Honkapohja, Samuli Kaislaniemi and Ville Marttila introduce a new project, the aim of which is to produce online editions of historical manuscripts in a form that facilitates research not only by corpus linguists, but also by historians. The motivation for the project is what are perceived to be weaknesses in traditional historical corpora, such as their containing texts from critical editions incorporating features from various manuscript witnesses. As envisaged by the three authors, digital editions will be flexible and user-friendly, will conform to the guidelines of the Text Encoding Initiative and will only use open source soft-

ware. They will contain normalised textual content, thereby allowing for all the usual corpus searches, as well as faithful renditions of the manuscript texts. They will also contain links between the two levels and all the normal apparatus of edited manuscripts. The authors are at present working on three texts, and they envisage more texts being produced in this format in the future, thus leading to an ever-expanding pool of digital editions.

Thirteen of the papers in the book deal with various aspects of Present-Day English. Three of them are based on ICE corpora, Carolin Biewer's on ICE-Fiji, Dagmar Deuber's on two ICE corpora from the Caribbean, and Daniël Van Olmen's on data from the spoken part of ICE-GB, which he compares with data from the Spoken Dutch Corpus. Van Olmen's paper is called "A contrastive look at English and Dutch (negative) imperatives". He begins, however, by considering all forms of imperatives in the two languages and finds that positive imperatives in Dutch outnumber their English counterparts by over 50 per cent. However, as he points out, "The higher number in Dutch is the result of the relative proliferation of so-called imperative discourse markers in the language" (p. 411). Unlike positive imperatives, negative imperatives are more frequent in spoken English than in Dutch. To further explore these, Van Olmen looks at a translation corpus containing five English and five Dutch plays. He finds that Dutch negative imperatives are much more likely to be translated into English by negative imperatives than vice versa. One particular usage stands out as being much more common in English than in Dutch. This is the expression of support for the addressee in the form of a negative imperative, as in 'Don't worry about it'.

In her paper "Passive constructions in Fiji English: A corpus-based study", Carolin Biewer compares both *be* and *get* passives in Fiji English with similar constructions in British and New Zealand English. As the spoken part of ICE-Fiji is not yet ready to be used for research, she restricts her comparison to written texts. She shows that the number of *be*-passives is very similar in newspaper texts in the three varieties, although there are not many long passives in Fiji English. She relates this finding to substrate influence, Fijian lacking a passive construction with oblique encoding of the agent. As for *get*-passives, ICE-Fiji contains more of these than ICE-GB, but fewer than ICE-NZ. Dagmar Deuber's paper on ICE-Jamaica and ICE-T&T (Trinidad and Tobago) is concerned with the actual compilation of the corpus, and in particular with the question of how best to adhere to the demands of the ICE design that the corpus should contain conversations between educated speakers of English. The problem is that when educated speakers converse, they tend either to speak Creole or to code-switch between English and Creole. In order to ensure that the dialogues contained as much English as possible, the compilers used semi-formal interviews. Neverthe-

less, as Deuber shows in a richly illustrated discussion of intertextual variation in a small selection of interviews, interviewees occasionally switched to Creole, particularly when engaged in talk about more personal topics.

Spoken learner corpora are investigated in two papers, one by Karin Aijmer on *I don't know/dunno* in the spoken English of Swedish learners, the other by Naixing Wei on phraseology in Chinese learners' spoken English. Aijmer uses the recently compiled Swedish component of the LINDSEI corpus and compares the Swedish learners' usage of *I don't know/dunno* as a pragmatic marker to that of native speakers as registered in the LOCNEC corpus. She discusses the various pragmatic functions that may be served by the form and points out that, unlike native speakers, who often employ it for reasons of politeness (to avoid giving a direct answer), Swedish learners tend to use it more for speech management purposes, i.e. when they are not sure exactly how to say what they want to say. Naixing Wei looks at chunks of three words or more in COLSEC, a 700,000 word corpus of Chinese learners' spoken English based on recordings of 302 oral examinations in which a student is interviewed by a teacher. The author refers frequently to an article by Altenberg from 1998, comparing his results to those of Altenberg who looked into similar chunks in the London-Lund corpus. Indeed, so frequent are the references to Altenberg's study that the article would have benefited had his figures been included in the tables for purposes of comparison. Although there are many similarities between the Chinese learners and native speakers in their use of fixed phrases, there are some chunks that stand out as typical for the Chinese learners. One particularly interesting one is 'with the development of' which, as Wei points out, appears to have a very positive prosody for the learners. All progress, in their world view, appears to be good progress!

In her paper "*Go to V: Literal meaning and metaphorical extensions*", Sara Gesuato analyses some 1,400 tokens of the non-progressive 'go to V' construction in the Bank of English corpus. She finds that some 88 per cent of them encode the physical motion sense of *go*. The most common infinitival predicates are *see, live, show, get, visit* and *meet*, in that order. One of these verbs, *show*, differs from the others in occurring mainly with inanimate subjects, and the sense of 'providing evidence' encoded by *goes to show* is one of four minor senses of the construction. The other three are 'physical transfer', 'contribution towards' and 'proceeding to'. Gesuato is interested in whether the motion sense typical of the non-progressive construction may also be encoded by its progressive counterpart, the prototypical use of which is to encode futurity. To investigate this question she devised twenty sentences with 'be going to V', using as raw material tokens from her corpus, and asked twelve

native speaker informants whether they would find these sentences acceptable as answers to the question ‘Where are you going?’. The degree of judged acceptability ranged from unanimity in the case of ‘to look at the garden’ to seven positive responses out of twelve in the case of ‘to meet a senior official’. Allowing for the possibility that examples like the latter could have been rejected on the general grounds of their being insufficiently informative (why should the speaker conceal the identity of the senior official he or she is going to meet?), a possibility alluded to by the author in a footnote, we may agree with Gesuato that her acceptability test shows that the motion sense may on occasion be encoded by ‘be going to V’.

The evolution of British attitudes to the EU, as reflected in three broadsheet papers, is the topic of Anna Marchi and Charlotte Taylor’s paper “Establishing the EU: The representation of Europe in the press in 1993 and 2005”. Their data come from two subcorpora of the Siena-Bologna Modern Diachronic Corpus of British newspapers. The first is from 1993, the year of the Maastricht treaty, which contains almost 99 million words. The second is from 2005, the year of the French and Dutch referenda on the European constitution, containing over 146 million words. Their methodological approach is that of Corpus-Assisted Discourse Studies, which combines qualitative and quantitative methods, with a view to unearthing implicit attitudes to the topic under investigation. Their paper contains a wealth of quantitative data, concerning, among other things, the frequency of references to the two treaties, the frequency of various keywords related to Europe and the frequency and semantic prosody of *euro*-compounds. Their main finding is that, while there is not much change from 1993 to 2005 in the (generally negative) attitude towards the EU displayed by the Times and Telegraph, in particular, there does seem to be less interest in European affairs in 2005 than in 1993.

Political discourse also provides the data for Anita Fetzer’s study of *sort of* and *kind of*, in which she examines their occurrence in ten speeches by, and twenty-two interviews with, British politicians. Her data yield a total of sixty-four tokens of the two forms, making it much less frequent in political discourse than in everyday conversation. Fetzer interprets the forms as serving primarily as contextual cues, with the context influencing our perception of them as indicating a greater or lesser degree of fuzziness. She discusses a wide range of examples, concentrating on the ‘more fuzzy’ type, where *kind of/sort of* is taken to signal non-prototypicality of the referent. She mentions on page 131 that *sort of* can signal a metaphorical rather than a literal reading, and I wonder whether several of her examples would not be better interpreted with respect to degrees of metaphoricity than degrees of fuzziness. When, for example, a retired politi-

cian speaks of the possession of a ‘sort of return ticket’ to government (p. 144), the meaning, though figurative, is perfectly clear, and in the example ‘we’re more sort of spun against than spinning’ (p. 139), in which Fetzer takes *sort of* to function as a hedge, it could also be interpreted as a cue to an upcoming witticism.

There are five articles in this volume based on data from the BNC, either exclusively or in conjunction with other corpora. Magnus Levin and Hans Lindquist compare the three text-organising phrases *on the face of it*, *on its face* and *in (the) face of* in British and American English, with data from the BNC, COCA and the Time magazine corpus. Perhaps the most surprising result of their investigation for many readers will be that the phrase *on its face* actually exists. Although it is confined to legal texts in BE, it has spread to other genres in AE. Steven Coffey’s paper “*A nightmare of a trip, a gem of a hotel: The study of an evaluative and descriptive frame*” is concerned, as its title indicates, with the *N of a N* construction. Leaving aside tokens which he regards as fully lexicalised units, such as *a whale of a time*, tokens in which the first noun is quantitative rather than descriptive or evaluative, such as *a ghost of*, and tokens beginning *a hell of a*, he examines 380 expressions with an eye to their internal composition and function as well as the text types in which they occur. He finds that the construction is not as frequent as one might have expected, at least in the BNC, and that the most common first nouns are *gem*, *giant*, *devil*, *bitch* and *brute*, while the most common second nouns are *man*, *woman*, *girl* and *song*.

Karin Axelsson is working on a project based on real-life and fictional dialogue in the BNC. Her paper deals with the problems concerning the retrieval of examples of the latter. In the first place, one has to distinguish between direct speech and direct thought (there is no real-life thought in the BNC!). In addition, there are only a few corpora with a mark-up of direct speech and the BNC is not one of them. She points out problems this lack of mark-up poses for frequency calculations. Another paper that addresses problems related to frequency counts is Magali Paquot and Yves Bestgen’s “*Distinctive words in academic writing: A comparison of three statistical tests for keyword extraction*”. How often has a corpus linguist felt satisfaction at finding an unusual item in his or her corpus, only to discover on closer investigation that this particular item is actually idiolectal? Problems of interpretation caused by the uneven spread of words or constructions, in their case keywords in academic discourse, across a range of texts, prompted the authors to test out alternatives to the commonly-used log-likelihood statistical test. The two tests they consider are the t-test and the Wilcoxon-Mann-Whitney test. They conclude that, for their purposes of

identifying keywords in academic texts in the BNC, it is the t-test that proves most suitable.

The final paper in the volume, by Hans Martin Lehmann and Gerold Schneider, examines how a syntactic parser, originally developed by Schneider, can be applied to the BNC to produce a data-base of lexical dependencies, which can then be used to investigate syntax-lexis interactions. In this paper they concentrate on the relations between subject and verb in both active and passive constructions, the relationship between verb and object, as well as subject-verb-object combinations. Their investigations throw up a wealth of fascinating data, which they maintain can serve as raw material for linguistic analysis despite the fact that the parser output only contains between 80 per cent and 90 per cent of possible tokens. One particular advantage they point to is the suitability of the parser as a tool to investigate the semantic prosody of low-frequency items.

It has only been possible in the course of this review to scratch the surface of the very interesting papers in this volume. The editors are to be congratulated on their choice of such a thought-provoking collection and also on the very high standard of editing. As a final comment, I would point to the number of papers (nine of twenty-two) that contain statistical calculations. Recent years have seen an increase in pleas, not least by Stefan Gries, for more statistics in corpus linguistics. On the evidence of this volume, it would seem that these prayers are in the process of being answered.

Geoffrey Leech, Marianne Hundt, Christian Mair and Nicholas Smith. *Change in contemporary English. A grammatical study.* Cambridge: Cambridge University Press, 2009. ISBN: 978-0-521-86722-1. xviii + 341 pp. Reviewed by **Magnus Levin**, Linnaeus University, Sweden.

The present volume is one of a number of original and well-researched volumes on change and variation in World Englishes. While Leech *et al.*'s book focuses on recent changes in American (AmE) and British English (BrE), Peters *et al.* (2009) covers Australian and New Zealand English and their relations to American and British English from a synchronic perspective, and Rohdenburg and Schlüter (2009) take a synchronic view of AmE and BrE. The three volumes partly overlap in their choice of topics (e.g. subjunctives and progressives), while some areas are only covered in one of them. In the discussions below

some minor comparisons between the findings in the different volumes will be made.

Because of their focus on small corpora (by modern standards), Leech *et al.* mainly explore high-frequency core phenomena of syntax and morphology. However, in some instances the authors stray outside the subtitle's restriction to a 'grammatical study', as in the lexical investigation of expanded predicates (*take or have a look at something*) in Chapter 8, the decreasing use of the titles *Mr, Mrs* and *Miss* (pp. 259–261) and changes in punctuation (pp. 244–246). This is not a real problem, however, since these phenomena tie in with the other topics covered, and are linked to some of the general trends discussed such as colloquialization and democratization. Because of the focus on a few general trends connected to social and cultural changes (colloquialization, densification of content, democratization and 'Americanization') and some language-internal factors (grammaticalization), the book comes across as an integrated whole with a number of cross-references between the chapters and a lengthy overview chapter at the end revisiting these tendencies. Although all four authors are "jointly responsible for the whole work" (p. xxiv), the number of chapters each author took "particular responsibility" for is given as Geoffrey Leech five, Marianne Hundt three, Christian Mair two and Nicholas Smith one.

The main focus is on results from the 'Brown family', consisting of the one-million Brown and LOB corpora from 1961 and the 1991–92 Frown and F-LOB corpora, but one of the strengths of this volume is that a very large number of comparisons are made with other corpora, such as the spoken component of the British National Corpus (BNC), the Longman Corpus of Spoken American English (LCSAE), parts of the London-Lund Corpus (LLC) and the International Corpus of English – Great Britain (ICE-GB). The spoken material is organized to provide a comparable diachronic corpus of BrE speech. Throughout the volume the findings are presented either in figures or small tables, while more detailed tables have been wisely moved to a 30-page appendix.

Due to the availability of corpus data, BrE is covered in slightly more depth than AmE. For some features, the authors supplement their findings with the 1931 clone of the LOB corpus. The latter corpus is so recent that there is still some uncertainty about the naming of it. It is variously referred to as *Lanc-31* (p. 255 (Table 11.3)), *Lanc-31 (B-LOB)* (pp. 108 fn., 255) and *B-LOB* (pp. xxv, 10, 188, 312).

Another minor point regarding the presentation of data is that there are some inconsistencies in the order of the bars in the bar charts, where sometimes the 1961 data from Brown and LOB are presented to the left and the 1991 data from Frown and F-LOB to the right (pp. 257, 297 (Figures A6.3 and A6.4)), some-

times the AmE findings (Brown and Frown) are presented first followed by the BrE data (from LOB and F-LOB) (pp. 53, 84, 197, 223), and, finally, sometimes the BrE data are given first followed by the AmE findings (pp. 174, 297 (Figures A6.1 and A6.2)).

Chapters 1 and 2 provide the background for the volume, where the enterprise is labelled “short-term diachronic comparable corpus linguistics” (p. 24). The authors point out that most of the ongoing changes are reflected in statistical changes rather than in categorical innovations, and argue that “frequency evidence is far more important in tracing diachronic change than has generally been acknowledged in the past” (p. 50). Changes are typically observed, it is argued, as changing stylistic preferences in different genres, and, as expected, there are few categorical either-or contrasts between varieties.

The authors admit that there are some problematic features with the Brown family. First of all, representativeness of corpora is a rather elusive concept, and although the corpora are based on a wide sampling frame comprising a wide variety of published writing, the composition of the corpora is not ideal. They are small by modern standards, and in 1961 the knowledge of genre and register was much less well developed, as illustrated in the inclusion of some ill-defined categories such as ‘Popular Lore’. Furthermore, diachronic changes in genre conventions make comparisons with B-LOB / Lanc-31 and earlier time periods difficult (p. 30). As becomes clear in the course of the book, the large number of genres in the corpora, the time-depth and the tagging nevertheless allow unique insights into the changing and variable nature of AmE and BrE grammar during the 20th century.

The main body of the volume begins in Chapter 3, which concerns two instances of the subjunctive mood moving in opposite directions in modern English: the increasing mandative subjunctive (*he called again, demanding that she sing him his favourite song*) and the decreasing *were* subjunctive (*if I weren't such a workaholic*). As shown in previous studies, the mandative subjunctive has been revived in AmE and is also increasing, but still lagging behind, in BrE. In spoken and written BrE, there is variation between the subjunctive, *should* periphrasis and, less expectedly, the indicative (*I suggested that Jack goes in*). One major factor in this development, it is suggested, is that the mandative subjunctive is losing its formal connotations. In contrast to mandatives, *were* subjunctives decreased in BrE between 1961 and 1991 while remaining stable in AmE.

Chapters 4 and 5 concern two interconnected areas: modal auxiliaries, which are generally decreasing, and semi-modals, which are on the increase. The decline of one category cannot readily be explained by the growth of the other,

because, as becomes clear from comparisons within the Brown family, modals are considerably more frequent than semi-modals. However, comparisons with spoken data show that particularly in AmE the semi-modals have been catching up with the modals, and the authors argue tentatively that this will affect writing after some delay. For the core modals there is a strikingly consistent pattern of decrease (the only exception being *can*). *Shall*, *may* and *must* are examples of modals that are most rapidly decreasing. The evidence suggests that *may* may be heading towards monosemy as a marker of epistemic possibility (*we may be able to hide you*) and deontic *must* is also decreasing rapidly compared to epistemic *must*.

Semi-modals like *BE going to*, *HAVE to*, *NEED to* and *WANT to* present one of the most fruitful areas of on-going change in Present-Day English. AmE speech appears to be spearheading the increasing use of these. Typical features of grammaticalization processes are found with semi-modals. *Going to* and *want to* are becoming reduced to *gonna* and *wanna* in speech and there are also signs of generalization of meaning. One of the functional explanations given by Leech *et al.* for the increases of *NEED to* and *WANT to* is their use as weak deontic devices, as in *I'm not a feminist, but I do think you need to hear a balanced view of matters* where an imposed obligation is 'camouflaged' as being in the listeners best interest, and in *You want to watch out that you don't get burned to an ash, first sunny day* where *want to* could be replaced by *need to*. A major factor behind this change is that speakers are now choosing less face-threatening modes of expression, a fact which is related to the trend of democratization.

One of the most striking instances of ongoing change, the spread of the progressive, is covered in Chapter 6. Although a rapid change is taking place in this area, the differences between the varieties seem to be minuscule. While Leech *et al.* found slightly more progressives in BrE than in AmE (p. 122), Collins (2009: 117), who compared ICE-GB with the not-yet-completed ICE-US supplemented by additional spoken material, found the opposite. Overall, there has been a 10 per cent increase in progressives between 1961 and 1991 in the Brown family. As could be expected, informal and speech-like genres contain the highest frequencies of progressives. The clearest increase is seen in the present progressive active. Progressive passives increase in BrE, but not in AmE, where *be*-passives are decreasing. A further significant increase is seen with progressives preceded by modal auxiliaries.

Since the progressive is most typical in informal and speech-like genres, the spread of the progressive is a case of colloquialization of writing, and because the biggest increases are found in the core areas, most notably the present progressive active, it can be seen as a case of grammaticalization where a form is

becoming generalized. Special pragmatic uses of the progressive, such as attitudinal / expressive uses (*You're always whingeing*), interpretive uses (*When she said that, she was lying*), politeness (*I was just wondering if...*) and statives (*John's being silly*), are at the most considered as “minor players” (p. 142) in the increase, and they only grant interpretive uses any major importance. This appears to be in contrast to Collins (2009: 120), who argues that “as a set the special uses are likely to have impacted on the growth of the progressive”.

The increasing progressive and other equally puzzling instances of drift make the authors conclude in the final chapter that “[i]t is tempting to adopt the thesis that frequency changes take place under their own momentum: that increase begets increase and decrease begets decrease” (p. 269). Such hypotheses require further testing in large-scale corpora.

The investigation of passives in Chapter 7 reveals that there has been a sharp decrease in *be*-passives between 1961 and 1991 in written English. The decrease is most marked in academic AmE, a fact which can be attributed to prescriptive influence. *Get*-passives, in contrast, are increasing in use but can be said to be replacing just a tiny fraction of the lost *be*-passives. *Get*-passives are therefore deemed not to have become a stylistically and semantically equivalent alternative.

In Chapter 8 expanded predicates of the type *take/have a look* are explored. The results indicate that speakers of BrE prefer to *have* rather than *take a look* while Americans have opposite preferences, as also shown by Rohdenburg and Schlüter (2009: 399–400). While no significant diachronic trend emerges from the Brown family and there are relatively small differences between the varieties in writing, the data from the spoken corpora reveal an almost categorical difference between BrE (*have*) and AmE (*take*). Moreover, although such predicates are only slightly more frequent in written BrE than in written AmE, the occurrences in spoken BrE outnumber the ones in AmE five to one. Findings like this clearly illustrate the need to use corpora of both speech and writing when comparing varieties.

Chapter 9 is entitled “Non-finite clauses”, but since twenty pages are devoted to verb complementation patterns (e.g. *help* NP *to*-inf./bare inf.) and only two pages to *to*-infinitives and prepositional gerunds, the heading is something of a misnomer. Both parallel developments, as in the case of increasing bare infinitive after *help* in both AmE and BrE, and diverging trends, as with BrE developing *stop* NP *V-ing* (as opposed to *stop* NP *from V-ing*), are found in the material. The corpora furthermore reveal that catenative *start* and *stop* are increasing in frequency overall. Although the Brown family of corpora is too small to provide more than hints of change with some the less frequent phenom-

ena in this area, they allow researchers to “challenge widely held assumptions about the supposedly inevitable Americanization of British English” (p. 205).

Chapter 10, which is the most original of the ‘core’ chapters, covers a range of topics related to noun-phrase structure and changes in the distribution of word classes. Both the written and spoken corpora show that there have been highly significant increases in the frequency of nouns overall in both AmE and BrE writing and BrE speech. Interestingly, this tendency towards densification of content words goes against the general tendency towards colloquialization. These two apparently contradictory changes indicate a denser packaging of information and a tendency to “‘sell’ this fairly heavy intellectual diet in a somewhat more informal/colloquial style than used to be the case” (p. 252). There is nevertheless some evidence that the shift towards densification is reaching its end, since it appears to be slowing down. The trend of content densification is reflected specifically in the increasing use of acronyms, proper nouns, noun sequences (*real estate tax shelter sales people*) and increasing *s*-genitives at the expense of *of*-genitives. The authors suggest that, apart from compressing information, the increase in personal names may also be a reflection of the increased importance of the identity of the individual in Western society. At least one change in noun-phrase structure can be connected to both colloquialization and prescriptive influence. *Wh*-relatives are decreasing in use and *that*-relatives are increasing. This is a typical example of prescriptive influence affecting language use, since the substantial decrease in *which* in written AmE is due to American prescriptivists leading the fight against relative *which*.

The volume ends with Chapter 11 tying together findings from the preceding chapters with additional instances of change and the various explanatory processes postulated by the authors. This chapter is different from the previous ones since it uses a ‘top-down’ approach going from theory to the data rather than the ‘bottom-up’ procedure typically used in corpus studies. Thus various language-internal, discourse-based and socio-linguistic factors, such as grammaticalization, colloquialization, densification, dialect contact, prescription, Americanization and democratization, are used as starting points for a look at different phenomena. Apart from those features discussed above, the chapter in rapid succession covers additional topics including contraction, negation, questions, nominalization, split infinitives and adjective comparison.

Different trends sometimes act in the same direction, as is the case with colloquialization and Americanization, but, as the authors clearly demonstrate, some results do not easily lend themselves to such interpretations, and some trends, such as colloquialization and densification, appear to work in opposite directions. The overall findings that written English is colloquializing and that

AmE is generally ahead of BrE tally with Rohdenburg and Schlüter's (2009: 421) overview of almost 50 different features in AmE and BrE. As in the present volume, Rohdenburg and Schlüter found AmE to be in the lead in most cases, being the more progressive in 35 of 48 features, while BrE prefers the more formal alternatives in 32 of 43 cases. Leech *et al.*'s findings of AmE usually being ahead of BrE are thus supported by Rohdenburg and Schlüter, but, as Leech *et al.* quite rightly point out, the term 'Americanization' may still be misleading since the changes observed may well be instances of parallel developments without dialect contact affecting BrE.

Finally a few words about a classic example of 'drift' from synthetic to analytic language discussed in Chapter 11. The tagging of the corpus allows easy retrieval of all periphrastic and inflectional comparative adjectives, and the results indicate minor increases (AmE 3.9%, BrE 2.2%) in periphrastic forms (p. 265). However, as these numbers appear to include all instances of comparatives, i.e. not only variable (disyllabic) adjectives, but also invariable cases such as *bigger* and *more beautiful*, and because the variation with variable adjectives is so strongly dependent on different lexical items (e.g. *happy* being mainly inflectional and *likely* typically periphrastic), it is doubtful to what extent Tables 11.5 and 11.6 (p. 265) really reveal anything about this potential drift. The more relevant, though also inconclusive, findings are given in Table 11.7 (p. 267) where the number of adjectives occurring with both types of comparison is shown to be diminishing. Thus, there is some evidence of a decrease in variation in this area, but the small corpora used in the present volume are too limited to allow conclusions to be drawn with any certainty.

The minor points of criticism aside, Leech *et al.* is an excellent volume that will serve as an indispensable model for future studies of short-term changes. All chapters dealing with central phenomena of the grammar have a firm empirical basis, are methodologically sound and aim at explaining the changes in a wider context. Throughout the analyses, the authors weigh different explanations against each other and connect the areas covered to each other in a lucid manner. The inclusion of both written and spoken data from BrE and AmE makes it possible for the authors to make a very large number of interesting comparisons. There is nevertheless scope for further research into the topics covered, and the volume will thus serve as inspiration for further studies.

References

Collins, Peter. 2009. The progressive. In P. Peters *et al.* (eds.), 115–123.

- Peters, Pam, Peter Collins and Adam Smith (eds.). 2009. *Comparative studies in Australian and New Zealand English. Grammar and beyond*. Amsterdam and Philadelphia: Benjamins.
- Rohdenburg, Günter and Julia Schlüter. 2009. New departures. In G. Rohdenburg and J. Schlüter (eds.), 364–423.
- Rohdenburg, Günter and Julia Schlüter (eds.). 2009. *One language, two grammars? Differences between British and American English*. Cambridge: Cambridge University Press.

Ute Römer and Rainer Schulze (eds.). *Exploring the lexis-grammar interface* (Studies in Corpus Linguistics 35). Amsterdam and Philadelphia: John Benjamins, 2009. 321 pp. ISBN 978-90-272-2309-8. Reviewed by **Joe Trotta**, University of Gothenburg.

Exploring the lexis-grammar interface (henceforward referred to as ELGI) is a collection of articles that highlight the inseparability of grammar and lexis. The research articles in this volume, originally presented at a conference at Leibniz University of Hannover in 2006, are informed by the recent theories, assumptions and analytical tools that derive from scholars who prefer a ‘contextualized’ view of linguistic study and challenge the widely-accepted reduction and compartmentalization of language into artificial, idealized components such as ‘lexis’ and ‘grammar’. Various theoretical and methodological approaches, such as, for example, the *idiom principle* (cf. Sinclair 1991), Hunston’s and Francis’s *Pattern grammar* (2000), Michael Hoey’s *Lexical priming* (Hoey 2005), *Construction grammar* (Goldberg 1995/1996) and *Collostructional analysis* (Stefanowitsch and Gries 2003), are sometimes employed in isolation, sometimes in conjunction and sometimes side-by-side to discuss and interpret various issues in the lexis-grammar interface. According to the editors, the present volume is not an attempt to accentuate the value of one theoretical framework over another, but rather, “...this volume seeks to uncover those linguistic mechanisms which help to overcome the traditional conceptual separation of core areas of language” (p. 2).

Aside from the very informative introduction by the editors (‘Zooming in’), which gives a succinct, informative overview of the field and puts each of the articles into context, this volume consists of fifteen papers, organized into two

main sections. Part I, called “Setting the Scene”, contains five pieces that set out the main aims, assumptions and commitments of the lexis-grammar-interface enterprise, while Part II, “Considering the particulars”, comprises a collection of diverse case studies that exemplify representative research in this area. There is, however, some overlap in the ‘setting-the-scene’ articles and the ‘considering-the-particulars’ articles, i.e. some of the papers in Part I are underpinned by specific case studies, and conversely, most of the articles in Part II give quite a bit of background information on the study of the lexis-grammar interface.

The starting point for Part I is Michael Stubbs’s article “Technology and phraseology: With notes on the history of corpus linguistics”, which outlines some of the theoretical and methodological concerns lying at the core of all the work included in the book. It begins with a sketch of some of the main historical developments in corpus linguistics, in particular those which are lesser known or overlooked, and follows with some illustrations of how new corpus tools (such as William Fletcher’s Phrases in English (or PIE) data-base) can be used to provide data on form-function relations. Stubbs concludes with some interesting suggestions for future corpus linguistic research, such as a detailed comparison of *pattern grammar* (which has an inductive, corpus-driven approach) and *construction grammar* (which is deductive and largely based on intuition). He also calls for corpus-driven studies to explore some of the well-known philosophical theories of language as social action, such as those found in the scholars like Ludwig Wittgenstein, John Austin and John Searle.

Hoey’s contribution, “Corpus-driven approaches to grammar: A search for common ground”, follows well from Stubbs’s opening article – in it the author looks for commonalities in three different approaches to the lexis-grammar interface, namely Sinclair’s *idiom principle*, Hunston and Francis’s *pattern grammar* and Hoey’s own *lexical priming*. Using data from different corpus-based studies, he compares the approaches and illustrates that no important incompatibilities exist between the three models; and in fact, despite their differences, in many ways they support and complement each other.

In “Valency: Item specificity and idiom principle”, Thomas Herbst uses insights from his work on the *Valency dictionary of English* to discuss the role of idiomatic and idiosyncratic aspects of language. Though it may seem reasonable to question whether valency falls under the purview of lexis or grammar, Herbst implies that attempting to subsume valency under one or the other heading is not fruitful since, along the lines of Sinclair and Halliday, he views grammar and lexis as different ends of the same continuum. Thus valency patterns show a gradient from the easily generalizable (and thus more grammar-like) to the obviously collocational (making it more lexis-like). Also, while valency can be

viewed as the syntactic potentiality of a lexical unit, Herbst is eager to draw attention to the fact that the meaning of a lexical unit can depend on its occurrence in a particular valency configuration.

Schröder and Busse's contribution, "Fowler's *Modern English usage* at the interface of lexis and grammar", takes the view that usage guides are the natural meeting grounds for issues that straddle the lexis-grammar divide. The article, which compares usage books, dictionaries and grammars (more specifically, three editions of *Modern English usage* (1926, 1965 and 1996), three contemporaneous editions of the *Concise Oxford English dictionary* (1924, 1964 and 1995), and three well-respected reference grammars (Quirk *et al.* 1985; Biber *et al.* 1999; Huddleston and Pullum 2002)), shows that several issues of divided usage are treated more explicitly and in a more user-friendly way in usage guides than in grammars or dictionaries. The authors also note that the boundaries between grammars, dictionaries and usage guides are becoming blurred and that usage guides have become less prescriptive in nature.

Concluding the introductory papers in Part I of ELGI is the very impressive report by Ellis, Frey and Jalkanen called "The psycholinguistic reality of collocation and semantic prosody (1): Lexical access". This work uses a lexical decision task in two different experiments to assess whether word recognition and lexical access are sensitive to particular collocations and to the generalizations of semantic prosody. In experiment 1, volunteers were tested to see if word recognition/lexical access is sensitive to collocational frequency. The results demonstrated that language in this task is clearly sensitive to patterns of usage of particular collocations, i.e. speakers judged the status of letter strings as words (rather than non-words) in sequences of normal collocation usage faster than they did in non-typical collocational pairings. Using a similar procedure, experiment 2 was designed to test whether word recognition/lexical access is sensitive to semantic prosody. The results of this test, however, did not demonstrate that speakers process positive valence words faster after positive prosody verbs. The authors interpret this to mean that memory for particular lexical associations affords fluent lexical access (based on experiment 1) but that there are no top-down semantic generalizations upon this level of processing (experiment 2).

The first article in the case studies subsumed in Section II is Joybrato Mukherjee's investigation into some new norms that have emerged in the lexicon of present-day Indian English. Mukherjee examines incipient multi-word units in Indian English such as new collocation patterns (e.g. *educational officer* rather than *education officer*), new prepositional verbs (e.g. *to discuss about sth*; *to visit to sb*, etc.), new ditransitive verbs (e.g. *inform sb sth*, *provide sb sth*) and new verb complementation patterns, for example the growing use of

give as a monotransitive. The paper documents and discusses these new forms and shows that none of the changes can be attributed to superstrate retention from colonial (or earlier) forms of English, but rather that they are genuine structural nativizations in present-day Indian English.

The use of *as* as a complementizer (e.g. *When I started to repair them I thought, Well, to think as we drink water as is pumped through these pipes*) and a relativizer (e.g. *...to think as we drink water as is pumped through these pipes...*) is examined in Daniela Kolbe's paper entitled "The semantic and grammatical overlap of *as* and *that*: Evidence from non-standard English". She questions whether the development of the *as*-complementizer was caused by analogy with relativizer *as* (same form – same function) or by parallels in the semantics of *as* and *that*. Her article claims that neither is the case; the fact that both items overlap in origin as demonstrative expressions enables their use in similar structures, but their overlap in non-standard grammar is apparently caused by divergent and non-linear developments. Her study also illustrates that lexical and grammatical overlap underlines the relationship between lexical priming and grammaticalization.

In "The historical development of the verb *doubt* and its various patterns of complementation", Yoko Iyeiri presents the results of her research into the emergence and later development of *doubt* in combination with different patterns. Her findings show that various and distinct complement clauses have accompanied *doubt* in the past, whereas present-day English *doubt* has a neatly defined pattern, allowing *whether*-clauses in the affirmative and *that*-clauses in the negative. Iyeiri describes how this use of *whether*-clauses after *doubt* is a gradual development in Modern English, as is the expanded use of *if*-clauses after affirmative *doubt*. The paper also outlines how the use of *that*-clauses after negative *doubt* was not completely stable, but after the middle of the Middle English period, *that*-clauses after negative *doubt* became firmly (re-)established.

Lindquist and Levin's contribution "The grammatical properties of recurrent phrases with body parts: The N1 *to* N1 pattern" demonstrates the pervasiveness of idiomatic expressions and discusses how the N1 *to* N1 pattern (e.g. *head to head, eye to eye, cheek to cheek, etc.*) exhibits a range of uses which cannot be understood on the basis of the meanings of its individual parts. The paper also describes how this pattern is being extended to new nouns and how it is undergoing a lexicalization process in which a grammatical structure is evolving and becoming more lexical.

"A corpus-based investigation of cognate object constructions" by Silke Höche argues that cognate object constructions (e.g. *smile a mysterious smile, die a horrible death, etc.*) constitute a 'family of constructions' or 'schematic

network' with lexical material at the bottom and more abstract patterns on top. Höche's data (3,100 examples of cognate object constructions gleaned from the BNC) are examined and interpreted through a collostructural analysis (à la Stefanowitsch and Gries 2003). Here the results show that Cognate object constructions (COCs) make up a family of three different, but related structures. With this more complex description of COCs, she can present a more refined, usage-based account of the different semantic characteristics of the relationship between the object and the verb and also offer an account of how sub-patterns of COCs are either derived from other, well-established transitive patterns or are metaphorical extensions of transitive constructions.

Matthias L.G. Meyer's paper "Revisiting the evidence for objects in English" discusses the importance of functional categories such as 'direct' and 'indirect' object in English. Meyer's major claim is that objects must be passivisable (the so-called 'passivisable object theory') and that constructions that do not allow passivisation (such as those found with verbs like *resemble*) need to be redefined and relabeled. In his conclusion, Meyer suggests several interesting new variations of verb complementation patterns that might better account for the data.

An exploration of intensifiers, such as *very*, *desperately* or *thumping*, underlies a study in complex collocations by Silvia Cacchiani, entitled "Lexico-functional categories and complex collocations: The case of intensifiers". Her findings indicate that occurrence in complex collocations extending to the right (in the form of multiple collocates, e.g. *completely fearless*, and *shameful and blatant and cheap and bad*) are a clue to the extent of grammaticalization, semantic prosody and pattern of intensification, whereas ability to modify intensifiers in collocations extending to the left (or multiple intensifiers, e.g. *very, very, very, very, very slinky, just absolutely ridiculous*) reflects the type and extent of expressivity and involvement of the intensifier.

The next article, Fanie Tsiamita's "Polysemy and lexical priming: The case of *drive*" emphasizes the role of the different semantic associations that polysemous items typically take. She tests Hoey's *lexical priming theory* using data on the two most frequent meanings of the polysemous noun *drive*, either as 'journey in a car or other vehicle' or 'private road leading up to a house'. Her data provide supportive evidence for the claim that common senses of a polysemous lexical item will avoid each others' primings: the semantic associations of *drive* ('journey') involve predominantly categories of DISTANCE/DURATION and EVALUATION; the semantic associations of *drive* ('private road') involve predominantly the categories of SIZE and MATERIAL, and thus the main catego-

ries that these two senses of *drive* draw their semantic associations from are indeed distinct.

In Michaela Mahlberg's paper "Local textual functions of *move* in newspaper story patterns", she explores the inherent structure of hard news stories, seen as "events or situations which are construed as threatening to damage, disrupt or rearrange the social order in its material, political or normative guise" (White 1997: 104, cited by Mahlberg). Mahlberg is able to show that the textual position of the so-called *move follow** pattern (i.e. in which sentence in a paragraph, which paragraph in a text, etc.) is linked to the news value of the article; it highlights the newsworthiness of the article and it places the news item in the context of other newsworthy events. One interesting upshot of this study is that it nicely exemplifies how a functional analysis of lexical items in a text can be carried out with a corpus-based methodology.

The concluding piece in the "Considering the particulars" section is Alison Duguid's paper "Comparing evaluative discourse styles: Patterns in rants and riffs". Using two purpose-compiled corpora, one from selected authors in British broadsheet newspapers who use 'loud' signatures (i.e. authors who have an informal, exaggerated, often outspoken, style), the other from articles with a more reserved tone gleaned from the *Times Literary Supplement*, she compares the salient features of the two corpora as regards priming, lexico-grammatical patterns, textual interaction and the resources of engagement. She discusses how irony in humorous opinion pieces uses (or flouts) the expectations set up by primings at the lexis-grammar interface, sometimes creating 'collocational shocks', which readers may or may not understand, depending on how well they are primed by their familiarity with the text type.

All in all, this new collection of research articles fulfils the goals set out by the editors; regardless of one's stance on the lexis-grammar dichotomy, ELGI reveals it as a convenient fiction whose usefulness is dubious. I found all the articles to be informative, relevant, well-written and, not least, enjoyable to read. The editing is meticulous, the volume is well-organized and it includes a very useful subject/author index. Many researchers (myself included), who routinely tolerate the separation of lexis and grammar out of a perceived need for methodological expedience, will find food for thought in this work. Though some articles will naturally make a stronger impression than others, each piece, in its own way, is a first-rate illustration of how corpus linguistics is broadening our understanding of topics like the nature of traditional grammatical concepts (e.g. the articles by Herbst and Meyer), psycholinguistic theory (e.g. the report by Ellis, Frey and Jalkanen), language variation and change (e.g. Mukherjee's,

Kolbe's and Iyeiri's contributions) as well as text and discourse analysis (as seen in the papers by Mahlberg and Duguid).

Despite the fairly wide scope of the work and relatively diverse theoretical preferences of the authors in this volume, the main unifying theme in ELGI is that the study of language should be usage-based. All the papers convincingly argue that form and meaning are inseparable and that lexis-grammar interdependence should play a more central role in linguistic theory and description.

References

- Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad and Edward Finegan. 1999. *Longman grammar of spoken and written English*. Harlow: Pearson Education.
- Goldberg, Adele. 2006. *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press.
- Hoey, Michael. 2005. *Lexical priming*. London: Routledge.
- Hunston, Susan and Gill Francis. 2000. *Pattern grammar: A corpus-driven approach to the lexical grammar of English*. Amsterdam and Philadelphia: Benjamins.
- Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech and Jan Svartvik. 1985. *A comprehensive grammar of the English language*. London: Longman.
- Sinclair, John M. 1991. *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- Stefanowitsch, Anatol and Stefan Th. Gries. 2003. Collostructions: Investigating the interaction between words and constructions. *International Journal of Corpus Linguistics* 8 (2): 209–243.
- White, Peter. 1997. Death, disruption and the moral order: The narrative impulse in mass-media hard news reporting. In F. Christie and J.R. Martin (eds.). *Genres and institutions: Social processes in the workplace and school*, 101–133. London: Cassell.

Simone E. Pfenninger. *Grammaticalization paths of English and High German existential constructions: A corpus-based study.* Bern: Peter Lang, 2009. xi + 369 pp. ISBN 978 3 0343 0021 6. Reviewed by **Claudia Claridge**, University of Duisburg-Essen.

The study takes a comparative approach, the main aim of which is explaining the nature of existential constructions in High German (HG), and in particular why it does not have *one prototypical* existential construction like Present-Day English (PDE), but instead a variety of co-existing constructions. The book is divided into five parts: besides Introduction and Conclusion, Part 2 (Chapters 5–10) deals with the historical developments of existential constructions in English and German, Part 3 (Chapters 11–14) treats the PDE *there*-construction (ETC), and Part 4 (Chapters 15–18) focuses on the corresponding modern HG constructions, in particular the *es-gibt*-construction (EGC).

Part 1 introduces terminological and theoretical concepts of importance for the study as well as the data used. The ETC as the focal (comparative) anchor of the study is introduced here as consisting of the unstressed, non-deictic and non-referring element *there* and a form of *be*, whereas a greater range of verbs is accepted in the German equivalents, e.g. in the presentational *es*-construction (PEC). Elements like *there* and *es* are treated as expletive subjects in English and expletive topics in German, based on typological/word-order differences between the two languages. Section 2.3 provides the grammaticalization background relevant for the study somewhat too briefly, highlighting the four mechanisms extension (context-induced re-interpretation), desemanticization (bleaching), decategorialization, and (phonetic) erosion, and four contexts or stages (initial stage – bridging contexts – switch contexts – conventionalization) as concepts to be made use of.

In Chapter 3 the data basis of the study is introduced. Instead of making use of standardly available corpora as much as possible, Pfenninger opted for compiling her own collections. For historical German, these consist of ten Old HG texts, 15 Middle HG texts, in each case supplemented by extracts from standard anthologies, as well as the *Bonner Frühneuhochdeutsch Korpus* and six other sources for Early Modern HG. Historical English is represented by the texts in Mitchell and Robinson's *A guide to Old English* (1999) and in Burrow and Turville-Petre's *A book of Middle English* (2000), and by Shakespeare's *Midsummer Night's Dream* and *The Tempest*. The modern data consists of ten fictional English works (from Joyce, 1914 to DeLillo, 1984; cf. list p. 28) and their German translations, as well as the results of an elicitation study (nine decontextual-

ized English existential sentences translated into German by one hundred Swiss native speakers of German). There are various problems with this data compilation. It is certainly not representative, an issue that Pfenninger does not comment on. It seems to be overall a fairly small corpus, but this is hard to evaluate, as unfortunately no word counts are provided. The Middle HG section is restricted to poetry, which is certainly not unproblematic for a syntactic study such as this one. The whole corpus is needlessly biased towards fictional writing. Above all, it seems a dubious undertaking to aim at explaining modern German characteristics (cf. above) with the help of non-original, translated material only. Pfenninger does not even explicitly mention the potential effects of translation on language usage. Without going into more detail, it can be said that the corpus-linguistic foundation of the study is not as sound as it could be. Also, the use to be made of the data in the course of the book does not correspond to what one expects of a corpus study. Mostly, the ‘corpus’ is used as a quarry for examples; there is relatively little frequency data provided and where this is done it consists of raw (not normalized) figures or of percentages, which are impossible to put into perspective. The last short chapter of the introductory part is headed ‘Method’, but it in fact only describes the outline of the book. Thus, Pfenninger does not actually provide *any* information on the methodology used for retrieving and analysing the data.

The historical part of the book starts with a short overview of word order patterns in Old and Middle English and of the development of the indefinite article (Chapter 5). Chapter 6 is about the development of the English ETC in Old and Middle English (OE, ME) according to the headings, but veers off into modern English and German matters at various points, thus leading to a somewhat haphazard and sometimes incoherent structure of the chapter. Section 6.1 (OE) deals with aspects relevant for grammaticalization, namely locative *vs.* expletive uses of *there* in OE, desemantization, and the development of existential *be*. The argumentation here is not always clear (references to explanations to be provided in Section 15.2 are not helpful) and is backed up by few, not always convincing, examples (e.g. 6.4, p. 52). There seems to be a tendency to overuse grammaticalization theory: why *there*-insertion should be an instance of subjectification remains unclear to me. In Section 6.2 on ME it is claimed that *there* is reanalysed as an expletive subject, *be* becomes more frequent, and that phonetic reduction occurs. The description of the latter (pp. 70ff.) is a rather bizarre mix of graphemic and phonetic change, which is not really relevant for the matter.

Chapter 7 is concerned with word order patterns in Old and Middle HG and identifies the proposed adverbs *thô/than* and *thâr*, used for textual and pragmatic functions, as precursors of dummy pronouns. The aim of Chapter 8 is to exam-

ine the Old HG *thâr*-construction, i.e. the extent to which it existed, the time period it was used in and its divergence from the corresponding English form. Pfenninger treats here Stage I (original locative meaning) and Stage II (bridging context). In my understanding, bridging contexts are ambiguous ones which allow both old and new interpretation, whereas she says at the beginning of 8.2 (p. 93) that she will only present “unambiguous instances of the expletive *thâr*”. I fail to see how unambiguous contexts can shed light on the development, in particular also because no frequency ratio of locative uses to bridging contexts is provided. Pfenninger here mentions as functions the use of the construction at story or chapter beginnings and as an enumeration marker. Furthermore, she points out that the Old HG *thâr*-construction is equivalent in form and function to the Old English ETC and underwent a similar grammaticalization process, but what remains unclear is why this German construction was ultimately not successful.

Chapter 9 is devoted to the development of German *es*-constructions, the types of which are listed on p. 127: (i) impersonal *es*-constructions (IEC), such as *es ekelt mich*, *es wird getanzt*, among which the type *man tanzt* is unaccountably included; (ii) presentational *es*-constructions (PEC), for instance *es klingelt eine Glocke*; (iii) *es-gibt* constructions, e.g. *es gibt einen Park auf der anderen Seite der Straße*; (iv) sentences with a personal pronoun, as in *Es (=das Mädchen) hat schönes Haar*, a type that is actually irrelevant here; and (v) extraposed sentences, e.g. *es ist schön, dass ihr hier seid*. Some IECs go back to Old HG, some to Middle HG. Type (v) is attested since Old HG, while PECs and extraposed constructions develop in Middle HG; EGCs have developed in Modern HG (cf. p. 128). The grammaticalization process of *es* is treated here, but it is neither well argued for nor really supported by figures. PECs are being connected to the grammaticalization of V2-order, which made expletive topics necessary, and are said to be developed from sentences introduced by *thâr* and *tho*. Existential constructions with other verbs (e.g. *find*, motion verbs), and in particular with *wesan* are dealt with here. These were fairly common in Old and Middle HG, but were superseded by the EGC. Finally, the development of the EGC is described. This involved various contributing factors or sub-changes, such as (i) the use of *geben* as a three-place and then as a two-place verb; (ii) the switch from personal to impersonal subjects (via abstract subject nouns); (iii) increasing non-referentiality; (iv) and meaning change of *geben* towards first ‘produce’, then ‘lead to’ and finally ‘exist’. The developments described in Sections 9.6 and 9.7 are fairly well presented, but a solid frequency foundation would have made the treatment even more convincing.

Chapter 10 juxtaposes the English and German developments. While six existential Old English constructions were ousted by the grammaticalized ETC in Middle English, diversity of constructions persisted in HG due to semantic and functional differentiation. Also, the expletive form satisfies different constraints in the two languages, namely the V2 constraint in German and SVO constraint in English. Unfortunately, the differences are described here rather than explained. It is also claimed that the Modern HG expletive *es* emphasizes the importance of the verb, whereas English constructions highlight the following NP – but given the fact that the verb is a fairly empty element in German as well, this is not necessarily convincing.

The following four chapters deal with the ETC in Modern English. Extended ETCs come first (Chapter 11), whose definition remains somewhat unclear, as e.g. *there was a candle burning* is accepted as an extension by Pfenninger, but not *there was one condition imposed on them by God*. Extended types are seen as important because of their frequency (three quarters of the data; cf. p. 231); in my opinion, however, this frequency is an artefact of the fictional data used here. Chapter 12 analyses the semantics of non-extended ETCs, identifying the existential type, the presentational type (with verbs other than *be*), the locative and the enumerative types, the last two of which make up 45 per cent of the non-extended data section. Pragmatic aspects of the ETC, notably the given-before-new principle, are described in Chapter 13. The data used in this part of the study sometimes leads to an unfortunate tendency to argue from a contrastive instead of from a language-internal viewpoint; cf. “Only from these verbs (= those used in the German translation, CC) might we be able to tell whether the original English sentence contains a locative ETC or a locative inversion...” (p. 248). Surely that decision should be made on the evidence of the English data alone.

The remaining part of the book focuses on the Modern HG situation. The uses of *es* are highlighted in Chapter 15. *Es* is shown not to be equivalent to English *there* on syntactic grounds. Furthermore, the main contribution of German *es* is identified as that of a style marker, which, however, is not convincingly argued for. In translations of ETCs, both IECs and PECs are rare, which is explained by the verb being in focus in the German constructions in contrast to the English one. Also, PECs are identified as having an event-reading. The fairly diverse Chapter 15 is followed by another somewhat unfocused Chapter (16), which claims to be about syntactic approaches to IECs and PECs, but instead collects various aspects mainly connected to language change. The semantic and pragmatic dimensions of HG constructions are in focus in Chapter 17. With regard to the EGC two semantic groups are identified (*vs.* three for the

English ETC), namely (i) one with purely existential meaning, and (ii) one with a dynamic aspect, with *geben* understood as ‘lead to’, ‘develop’. As to IECs and PECs, a greater range of verbs is possible, which add greater detail and precision to the statement and take over what English expresses by nominal means.

Pfenninger concludes her investigation with three generalisations (p. 324): (i) the development of existential constructions in English and German is linked to the word-order and information structure developments in both languages; (ii) grammaticalization in German was often restricted to a limited number of parameters, thus leaving space for various constructions; and (iii) a higher degree of grammaticalization of English existential constructions led to the loss of competing constructions. It seems to me that this is not much progress given the more than three hundred pages of this book and given previous research, in particular as (ii) and (iii) are left underexplained. To conclude, a very interesting and worthwhile topic has not been treated here in the best possible way. Drawbacks are certainly also the approach taken to data and the presentation and structuring of findings, which is often not done in the most reader-friendly way.

References

- Das Bonner Frühneuhochdeutsch-Korpus*. See Korpora.org, <http://www.korpora.org/fnhd/>.
- Burrow, J. A. and Thorlac Turville-Petre. 2000. *A book of Middle English*. 2nd ed. Oxford: Blackwell.
- Mitchell, Bruce and Fred C. Robinson. 1999. *A guide to Old English*. 5th ed. Oxford: Blackwell.

