Reviews

The language of fictional television brings together linguistic methodology and popular culture in the form of television. The book consists of two major parts. While the first part (Chapters 2–4) describes general aspects of fictional television like the communicative context, genre, audience, and dialogue; the second part (Chapters 5–9) concentrates on various facets of character identity in fictional television.

In Chapter 2, the author starts out by stressing the relevance of television in our culture which may also be regarded as the main motivation behind analyzing the language of television. As a next step, fictional television is characterized by several features, among them the discourse situation with the audience as overhearers, the multimodality with respect to both the characters and the product, and the imitation of reality. The chapter also introduces different formats as possible sources for investigation like shots, scenes, episodes, seasons or series as a whole.

Chapter 3 is used to establish the genre of *dramedy* which contains elements of (soap) drama and comedy. The concrete example used to illustrate the idea of this genre, and the base for most of the further analysis, is the series *Gilmore Girls* (Warner Brothers, 2000–2007) which is introduced in terms of character and plot lines at this stage. The focus on *Gilmore Girls* is due to several reasons, among them the general success of the series and its immense popularity; but there are also practical reasons for the linguistic analysis such as the availability of transcripts and of all episodes on DVD. The previous impact on academia can be measured by two edited volumes which hardly contain linguistic analyses, however. The most important reason for investigating dialogue in *Gilmore Girls* seems to be the crucial role of dialogue in the series itself. The two remarkable aspects of the dialogue are the fast pace and the high number of

intertextual references. The second part of Chapter 3 serves to establish the target audience of the product and contains the first quantitative analysis. The advertising text at the back of the DVD cover is used as a mini-corpus (858 words) because it is this text which directly addresses the audience. After a description of the most frequent words (many of them references to characters and setting), the focus is on evaluative language. The author first introduces the concept of evaluative language and then presents the results for the DVD cover corpus. The relevant parameters of evaluation are emotivity, affect, and expectedness (in that order). While emotivity is almost exclusively positive, affect is more mixed. The term 'zig-zag prosody' is used to characterize the prevalent pattern. The combination of positive and negative events (*hearts break and mend*) creates the idea of drama. The audience and its responses to the series are also directly addressed in the DVD cover texts (*fans, groupies*) which turns watching TV into a social activity rather than an in individual experience.

Chapter 4 concentrates on the dialogue within television. Bednarek starts out naming established linguistic features of television dialogue and illustrates them with examples from different programs (*Sex and the City, Friends, Star Trek*, etc.): conventions of stage dialogue, the use of stock lines, avoidance of unintelligibility, rather short turns, and other features differentiating television dialogue from naturally occurring conversation. For the analysis of dialogue in *Gilmore Girls*, Bednarek uses fan transcripts which are arguably more exact than subtitles. The transcripts are turned into two corpora using WordSmith; one contains the dialogue only (GiGi; ca. 1.1 million words), the other also contains additional information including the names of speakers (TS-GiGi; ca. 1.3 million words). The Santa Barbara Corpus of Spoken American English and the Longman Spoken American English Corpus are used as reference corpora. Bednarek looks at frequencies of both words and n-grams in the dialogue and then describes and explains over- and under-representations in *Gilmore Girls*. The importance of interaction between characters and *talk* in *Gilmore Girls* is stressed by this analysis. The frequent occurrence of emotive interjections which are addressed in more detail in the second part of the book is also mentioned at this point.

In Chapter 5, Bednarek approaches the concept of character identity from different angles, mostly from literature and media studies, and draws attention to the lack of research from the field of linguistics. Possible features for character identity are authorial cues (e.g. character names), explicit cues (e.g. explicit information about self or others, or implicit cues (e.g. conversational structure, lexis, accent, etc.). For the analysis of character identity, the corpus is split into subcorpora; each subcorpus consists of the utterances of one specific character.

These corpora can be seen as representing a linguistic thumbprint of each character as they comprise every utterance which one particular character has produced in the course of the series. The main analysis focusses on Lorelai who is the main character of the series both in terms of turns (>30,000) and in terms of tokens (>300,000). To identify what is specific about a character's speech, Bednarek uses the concept of *keyness*, i.e. statistical 'unusuality', when contrasting the node corpus (a single character's utterances) to the reference corpus (either all other characters' utterances or just one other character's utterances). Lorelai is clearly distinguished from other characters by the words she uses – from some characters more than from others as can be seen from Table 5.3. As a next step, Bednarek sorts the key words/n-grams according to their likely discourse functions: some are references to relationships (*mom, dad…*) or environment (*the inn, school…*), but by far the biggest group is emotional language (*wow, great…*). The conclusion is drawn that scriptwriters are – at least intuitively – aware of the differentiating power of language.

Chapter 6 focusses on the aspect of emotionality in language use that was shown to be especially important in the previous chapter. Bednarek uses the cover term 'expressive character identity' to refer to emotional, attitudinal and ideological aspects of scripted character identity. Bednarek demonstrates that a wide range of resources could function as expressive features. In this chapter, the author investigates the construal of expressive character identity on a micro level only, namely through the use of emotive interjections. Unlike other expressive features, interjections are a rather closed set which makes it easy to search for them in a large-scale corpus. The analysis shows that the characters differ in their use of interjections quantitatively and qualitatively: Lorelai uses interjections most frequently (37.9 per 10,000 words), while her rather reserved father, for example, uses significantly fewer (5.7). The characters are also distinguished by which interjections they use; Lorelai and her daughter use *wow* and *oh my god* fairly often while Lorelai's father uses *for X's sake* most frequently. A general result is that similar characters (e.g. married couples with similar attitudes) also seem to use similar interjections. On the other hand, each character has what Bednarek calls *signature interjections* that are specific to this character; i.e. this character uses the interjection more frequently than other characters and more frequently than all other interjections. In Lorelai's case, the signature interjection is *ugh*, which she uses 4.67 times per 10,000 words. As a next step, Bednarek uses examples to demonstrate what happens if characters break the norm of their own expressive identity, e.g. by using an interjection they would not normally use. The norm-breaking is explicitly addressed by the characters in the examples. At the end of this chapter, the author widens the perspective to fic-

tional television in general and applies the concept of a signature interjection to characters from other series: the signature phrase of the character Sawyer in the series *Lost* would be *son of a bitch*; that of Homer Simpson in *The Simpsons* would be *d'oh*.

After remaining on the micro-level of only interjections by means of a corpus analysis in Chapter 6, the author takes on a holistic stance in Chapter 7. She analyses one particular emotional scene (a break-up) in detail and takes into account multimodal resources. The chapter contains a transcript of the entire scene combined with drawings to illustrate shots. After a short description of the setting (an American coffee shop) and non-diegetic music, the focus is on the interaction of the two characters, Lorelai and her boyfriend Max. The general impression is that Max is much less emotional than Lorelai, who is upset by the breakup. This general impression is explained by a detailed description of both verbal and non-verbal behaviour. In terms of verbal behaviour, Bednarek shows that Lorelai uses more evaluative and emotional language than Max in the given scene. In terms of non-verbal behaviour, hand/arm gestures, head movements, gaze, and facial expressions throughout the given scene are described and analyzed. Again, Lorelai makes use of all these resources more frequently and with more variation. As a third step, Bednarek returns to verbal behaviour and describes how the two participants work together on expressive sequences, namely joking sequences, blame sequences and the actual breaking-up sequence. In all sequences, Lorelai is shown to be the more cooperative participant who gives preferred responses. In sum, this chapter shows how expressive identities are construed multimodally.

Chapter 8 investigates shared attitudes or ideologies as a further aspect of expressive identity. Bednarek defines ideology as shared values and belief systems without the notion of power playing a role. The ideology under scrutiny is the attitude towards meat-eating. Television series have been found to debate and negotiate cultural norms before (e.g. feminism in *Sex and the City*), but not with respect to meat-eating practices. This neglect is named as one of the reasons for an investigation into this ideology, along with the important ethical dimension, the role of culinary preferences for a character's identity, and the importance of food in the series *Gilmore Girls*. As a first analytical step, Bednarek uses examples to demonstrate how characters in the series express their attitudes towards meat-eating. Then she demonstrates how ideologies can be investigated in a corpus by lexical items (e.g. *vegetarian, meat*) and their concordances (negative and positive evaluations, e.g. *delicious, love*). It is important to note that the evaluations of food can occur in the same turn as the lexical item or across turns. The first result of the analysis is the raw frequency

of the lexical items in questions: terms relating to meat-eating are by far more frequent than terms relating to vegetarian or vegan eating practices. The frequencies of the meat-types (e.g. *beef, roast, chicken*) mentioned in *Gilmore Girls* actually mirror American consumer behaviour: red meat is mentioned much more often than white meat and the consumption per capita in the US is also much higher. After discussing the raw frequencies of mentions, the author turns to the evaluations found in the concordance analysis. Bednarek starts with the marked cases, occurrences of *vegetarian, tofu*, etc. Vegan and vegetarian food or characters are either portrayed as unlikeable or strongly associated with the Other, e.g. a Korean immigrant. Consequently, the audience is not invited to share the attitudes of Vegetarians or Vegans. Unlike the occurrences of vegetarian/vegan food, the majority of meat-occurrences are neutral and not evaluated; the implication is, however, that people order food they like – even if they do not explicitly evaluate it. Bednarek calls this portrayal of meat-eating *naturalization*, as this is part of a normal, unchallenged behaviour of the majority of characters. Finally, the possible impact of the ideology presented on the audiences is discussed. The relationship between media texts, ideology, and viewers is deemed as extremely complex, however, so it is not clear how far bonding with likable characters goes and whether viewers take over all the presented ideologies.

After the concluding chapter, the book ends with a very detailed and helpful appendix providing many of the tables which were regarded too complex to include in the text, e.g. lists with the most frequent n-grams in *Gilmore Girls* or transcript conventions.

I like the way new concepts and methodologies are introduced step by step, even such fundamental concepts as corpora. This didactic approach makes it conceivable to use this book as medium for instruction. There is only one minor deviation from this strategy: the terms 'mode pur/vécu' are used on p. 123 without further explanation; four pages later, the terms are used again and this time they are explained.

In Chapter 5, Bednarek repeatedly refers to her own forthcoming paper on the stability of character identity which I found somewhat intrusive, especially as it appeared (to me) that the project was well beyond the stage of a desideratum and that the author already knew the answers to the questions posed; so why not include them in this book?

In Chapter 7, the characters' non-verbal behaviour is referred to in detail for the first time, though not as the behaviour of the characters but as the behaviour of actors. This made me wonder: does it matter that they are actors? Every reader should at this stage be aware that the discussion is about fictional charac-

ters anyway. And do not real people act as well, when it comes to their emotions? How do we know that the acting in the given break-up scene is done by the actor Lauren Graham and not by the character Lorelai Gilmore trying to hide how upset she is?

For her analysis of the ideology of meat-eating practices in Chapter 8, Bednarek gives a number of very convincing reasons. She does not disclose, however, whether there is a personal motivation or interest in the topic. In terms of the ideology of meat-eating, the author of this book is either part of the meat-eating majority or part of the vegetarian/vegan minority. Her stance in the discussion on the impact of this ideology on audiences and the bonding with characters rather suggest that she does not eat meat. For the linguistic analysis it is of course irrelevant; but in terms of scientific integrity, may it not have been better to reveal her personal position on this issue?

On the whole, my questions concern only minor problems. Monika Bednarek has produced a highly valuable book, combining elements from stylistics, media studies, cultural studies and sociology with linguistic corpus analysis. Her book should be on the bookshelf of everybody who is dealing with the language of fictional television, be it as researcher, scholar, lecturer, or student.

Jim Feist. *Premodifiers in English. Their structure and significance*. Cambridge: Cambridge University Press, 2012. 288 pp. ISBN 978-1-107-00086-5. Reviewed by **Hilde Hasselgård**, University of Oslo.

The monograph *Premodifiers in English* "sets out to explain the nature and arrangement of premodifiers in English nominal phrases by relating their order to their meaning and syntax and to other areas of language" (p. 1). The study is qualitative and based on a large collection of attested examples from a variety of sources, including corpora. The collection of examples has been incidental rather than systematic, so there is no quantitative investigation built into the study. The theoretical framework is eclectic, but relies to a great extent on Halliday (2004) for its functional outlook, on Cruse (2004) for the semantic description and on Quirk *et al*. (1985) for the description of premodification in terms of 'zones'.

The bulk of the book (Chapters 2–7) is concerned with an account of the grammar and semantics of premodified noun phrases in present-day English. This is followed by a chapter on the diachronic development of noun premodifi-

cation from Old English to the present. Chapters 9 and 10 discuss the model of premodification in the light of psycholinguistics, discourse, language acquisition, morphology/phonology and grammaticalisation theory. The concluding Chapter 11 is followed by an extensive reference list and an index.

The descriptive model is briefly presented in Chapter 2. Like Quirk *et al.* (1985: 437, 1337), Feist views premodification in terms of *zones*, each of which is associated with distinctive semantic and grammatical features. While Quirk *et al.*'s zones are simply numbered (I–IV), Feist's have descriptive labels: Reinforcer, Epithet, Descriptor and Classifier. The phrase *your actual tinny round percussion instrument* contains all four of them. These zones are similar, but not identical to those of Quirk *et al.*; this is particularly true of the Epithet and Descriptor categories. Each zone may contain one or several words, or none. Importantly, the order of zones constitutes a grammatically set order of premodifiers. (A similar, acknowledged, account of premodification is that of Bache (1978), who outlines three functional classes of premodifiers, corresponding roughly to (i) Reinforcers, (ii) Epithets + Descriptors and (iv) Classifiers.)

Chapter 3, "Semantic explanation of unmarked order across the zones", outlines the following dimensions of meaning, largely based on Cruse (2004): referential (naming), descriptive ('objective', concerned with true and false, perception and conceptualisation), expressive (emotive and attitudinal), social (including dialect, sociolect and register) and grammatical (constructional, modificational and intensifying). When applied to the description of the zones of premodification it emerges that Classifiers have referential and constructional meaning; Descriptors have nonscalar descriptive meaning; Epithets have scalar descriptive or expressive meaning; and Reinforcers have grammatical, reinforcing meaning (p. 74). A syntactic explanation is offered in Chapter 4, which concludes that the zones are syntactic units. The scope of each zone is wider than the one following it. Within each zone, modifiers may be coordinated by means of a conjunction or punctuation. It should be evident even from this brief synthesis that the syntactic explanation does not rely on formal syntactic theory, but is close to semantics.

Chapter 5 takes a closer look at the Classifier zone, with examples of multiple classifiers such as *a British 2-inch brass electronic oil-pressure gauge*. Interestingly, NPs with multiple classifiers are compared to clauses and analysed in terms of processes, participants and circumstances, following systemic-functional grammar (Halliday 2004). A distinction is made according to whether the nominal head denotes a participant (e.g. *Iranian 16th-century brass boat-shaped vessel*), a process (e.g. *overnight on-site explosives storage*) or a circumstance (e.g. *sound travel time*). Premodifiers in participant-headed NPs are said to fol-

low an order of so-called 'qualia' (Pustejovsky 1995), namely origin – dimension – constituency – type – function. Process-headed NPs resemble clauses most, and their classifiers are parallel to participants and circumstances. Circumstantial relations are found to precede participants; moreover, circumstantial extent precedes location, and actor precedes goal. Thus *government farms buy-up* denotes a process in which *government* is actor and *farms* is goal. In circumstance-headed constructions, a modifier denoting process occurs next to the head, and can be preceded by a participant and one or more circumstances. Since the descriptive model in this section draws so much on systemic-functional analysis, it is somewhat curious that no mention is made of nominalisation as a type of grammatical metaphor (e.g. Halliday 1998), which produces the type of noun phrases headed by a process, circumstance or attribute. In fact, Halliday (1998: 191) describes how a clause is turned into a noun phrase by "metaphoric reconstrual", by which the original process is reconstrued as 'thing' (=head), and as a consequence circumstances and participants become deictic, epithet, classifier and qualifier.

In contrast to the Classifier zone, the order of modifiers within the Epithet and Descriptor zones is free (Chapter 6). However, certain tendencies may be observed, pertaining to style, prominence, convention, time order and experience (p. 153). The order of zones, on the other hand, is determined by rules of grammar. Thus, the order is considered marked if a modifier is placed in a position other than its usual zone (Chapter 7). Marked order typically changes the meaning of a modifier: for instance in *a strange, chemical, putrid smell*, 'chemical' takes on the meaning of an Epithet (rather than a Descriptor) because it is co-ordinated with other Epithets. Markers of gradability may also move modifiers that typically function as Descriptors to the Epithet zone. An apparent breach of the order of zones is taken to change the zone membership, and thus the meaning of a modifier; hence the two occurrences of 'young' in *young impulsive over-curious young woman* have different meanings – the first as an Epithet ('youthfully foolish') and the last as a Descriptor ('not yet old') (p. 160).

Chapter 8 aims to provide a historical explanation of premodifier order. The account of the Old English period is based on secondary sources, while the later periods seem to be at least partly based on the author's own observations. The main argument is that the order of premodifiers in Old English was determined chiefly by word class membership: the change into functional modifier types started in the Middle English period, although order was still largely determined by syntax. In Early Modern English, multiple premodification became more widespread, and semantics became more important for the order of modifiers.

The order found in present-day English was fully established in the Late Modern English period (p. 200).

Chapter 9 provides a review of selected studies of psycholinguistics, discourse structure (topic-comment, theme-rheme), first language acquisition and morphology and phonology that lend support to the explanatory model of premodification offered in the book. It is difficult to evaluate this part of the discussion, as the density of references to previous research is so high, and without having read the studies, one cannot be certain that terms and concepts are used in similar or even compatible ways. The last chapter before the conclusion further argues for the advantages of the proposed model in relation to previous descriptions. The author criticises previous accounts for taking a too narrow view of premodification; in particular he advocates the usefulness of zones rather than word classes in explanations of premodifier order. The fact that premodifiers can change their meanings or acquire new senses as they move across zones can be considered a type of grammaticalisation. Feist argues that the process of grammaticalisation need not follow the standard direction from content word to inflectional item, but can also go the other way (pp. 230 ff.). This part of the argumentation remains unconvincing, though, and it seems that much of what is termed 'degrammaticalisation' or 'rise in rank' (p. 232) could be more usefully regarded as reanalysis (e.g. Traugott and Trousdale 2010: 33 ff.). The book ends with a strong claim (p. 255): "Of English premodifiers, and particularly of their order, Cruse wrote (2004: 302): 'Various partial explanations have been put forward, but none is comprehensively convincing.' The principles outlined in this section summarise what is intended to be a comprehensive and convincing explanation."

By and large, I found the book a stimulating read, particularly the chapters that describe the model. However, the study is not without its weaknesses. One has to do with the use of theory and secondary sources. The approach is eclectic, from a wide range of sources, which makes a lot of sense; however, sometimes it may seem that the reading of the selected sources has not been accurate enough. For example, the account of Halliday's processes and participants, used to explain the order of classifiers, contains a couple of unnecessary mistakes, such as setting the number of process types to three rather than six or citing the relational *he felt ill* as an example of a mental process (p. 110). Furthermore, Quirk *et al.* are criticized for their account of the features of adjectives (1985: §7.2) without mentioning that their §7.3 begins with "However, not all words that are traditionally regarded as adjectives possess all of these four features".

Since 'social meaning' is discussed occasionally in the book, one misses more consistent attention to register and style. According to Quirk *et al.* (1985:

1338) "premodification is an area of English grammar where there is considerable variation among the varieties of the language" – by which they presumably mean regional and register varieties. This brings me to another point of criticism – putting on my hat as a corpus linguist – which concerns the empirical basis of the study. As mentioned above, the data collection has been incidental, in the author's own words: "I have in effect examined all the nominal phrases I have met in five years of looking for them in research, and in meeting them incidentally in general reading." (p. 3). The British National Corpus and the Corpus of Contemporary American English "have often been used to check proposed explanations" (p. 3). The cited examples are sometimes constructed or amended, though these have been marked typographically. This type of material works well enough for establishing a model for the analysis of premodification. However, the model is not really put to the test by being exposed to non-selected material and the reader has no way of checking what may have been left out of the account. Although noun phrases with multiple premodification are relatively rare, searches for various tag sequences in large corpora give plenty of material, as evidenced by Wulff (2003), which is, however, restricted to the order of adjectives.

One of the rare quantitative claims made in the book is that Reinforcers (e.g. *utter, pure, sheer*) are rarely used with other modifiers (p. 65). While there is probably no reason to doubt that the claim is correct, it is also a hypothesis that can easily be tested for instance in Bill Fletcher's "Phrases in English" (PIE, a search interface for the BNC). Searches for *sheer* and *pure* followed by another adjective before the noun (at a minimum frequency of three) gave some interesting data suggesting that where a Reinforcer occurs with another premodifier, the second one is a Classifier, e.g. *sheer hard work, sheer common sense, sheer physical strength, pure economic loss, pure alpine style*. However, this point is not picked up in the discussion of Reinforcers (pp. 65 ff.). Another finding from a PIE search for three adjectives followed by a noun is that the majority of 4-grams comprised or contained multiword-units, often names (e.g. *Royal Scottish National Orchestra*) or technical terms (e.g. *severe acute renal failure*). Admittedly, this may not be so much an oversight on the part of the author as a matter of focus: his aim is to explain the internal order of premodifiers, not to account for the function of the noun phrase in context.

The fact that the order of individual modifiers may be conventional is mentioned occasionally in the book, particularly in connection with the free order inside the Epithet and Descriptor zones. Many corpus linguists will probably miss a discussion of collocation, not least the idiom principle vs. open choice (Sinclair 1991: 109 ff.). Feist's proposed model seems to rely on open choice, a

kind of 'slot-and-filler model' although the zones are invested with "meaning and powers of their own" (Feist 2012: 228). A future study might explore how the claim that "a language user has available to him or her a large number of semi-preconstructed phrases that constitute single choices" (Sinclair 1991: 110) tallies with the proposed zones.

With the exception of some extremely citation-heavy passages (especially in Chapters 9 and 10), the style of the book is clear and readable. There are useful summaries at the end of each major section as well as at the end of chapters. The author's great enthusiasm for his topic is visible at every turn. Readers of the *ICAME Journal* may take the strong claims made about the model of pre-modification zones as a challenge – the book presents a rich set of hypotheses to be tested against corpus data.

References

Bache, Carl. 1978. *The order of premodifying adjectives in present-day English*. Odense: Odense University Press.

Cruse, Alan. 2004. *Meaning in language. An introduction to semantics and pragmatics*. Second edition. (Oxford textbooks in linguistics). Oxford: Oxford University Press.

Fletcher, William H. Phrases in English. (Search interface to the British National Corpus). <phrasesinenglish.org> (14 Jan. 2013).

Halliday, Michael A.K. 1998. Things and relations. Regrammaticising experience as technical knowledge. In J. R. Martin and R. Veel (eds.). *Reading science. Critical and functional perspectives on discourses of science*, 185–235. London and New York: Routledge.

Halliday, Michael A. K. 2004. *An introduction to Functional Grammar*. 3rd edition, revised by Christian M. I. M. Matthiessen. London: Arnold.

Pustejovsky, James. 1995. *The generative lexicon*. Cambridge, MA: MIT Press.

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

Sinclair, John. 1991. *Corpus, concordance, collocation*. Oxford: Oxford University Press.

Traugott, Elizabeth Closs and Graeme Trousdale. 2010. Gradience, gradualness and grammaticalization: How do they interact? In E. C. Traugott and G. Trousdale (eds.). *Gradience, gradualness and grammaticalization*, 19–44. Amsterdam: John Benjamins.

Wulff, Stefanie. 2003. A multifactorial corpus analysis of adjective order in English. *International Journal of Corpus Linguistics* 8 (2) 245–282.

Marianne Hundt and **Ulrike Gut** (eds.). *Mapping unity and diversity world-wide: Corpus-based studies of New Englishes* (Varieties of English Around the World G43). Amsterdam: John Benjamins, 2012. 294 pp. ISBN 978-90-272-4903-6. Reviewed by **Andrea Sand**, Trier University.

When the volume *Comparing English worldwide: The International Corpus of English* appeared in 1996, work on various ICE corpora was well underway, but the possibilities these corpora would offer were still more a vision than a reality. Nevertheless, Greenbaum (1996: 10) writes:

> As the parallel corpora become available, new possibilities open up for rigorous comparative and contrastive studies. I envisage the search for typologies of national varieties of English: first-language versus second-language English, British-type versus American-type English, African versus Asian English, East African versus West African English. Researchers might explore what is common to English in all countries where it is used for internal communication, demonstrating how far it is legitimate to speak of a common core for English or of an international written standard.

The volume under review presents research based on 21 ICE subcorpora, six from the Inner Circle, 15 from the Outer Circle, many of them still in the process of compilation or recently released. Some of the more recent additions to the ICE family with a focus on second-language New Englishes have already been documented under the header 'ICE Age 2' in the *ICAME Journal* (No. 34, 2010) and as in the case of the movie with the similar name, we can expect a number of sequels in the future, as more data becomes available, as the editors of the volume under review also suggest in their introduction (p. xiii).

The first five chapters deal with tense, aspect and modality features across ICE corpora. "'Off with their heads' Profiling TAM in ICE corpora" by **Gerold Schneider** and **Marianne Hundt** presents a largely corpus-driven method of studying tense, aspect and modality on the basis of POS-tagged ICE corpora with the help of a chunking software (the random field chunker *carafe*) which yields beheaded verb groups, such as *going to V* or *must have Ved* which are then ranked according to a frequency measure. Using data from Great Britain and New Zealand to represent the Inner Circle and India, Fiji and Ghana to represent the Outer Circle varieties (cf. Kachru 1985), Schneider and Hundt show that there are indeed differences in the TAM profiles of Inner and Outer Circle varieties, but that it is also important to complement statistical evidence with

qualitative analysis, especially in the case of smaller corpora in which individual documents can skew statistical results because of stylistic and content-related factors which cannot be controlled even in the most careful compilation process. **Peter Collins** and **Xinyue Yao** look at four sets of "Modals and quasi-modals in New Englishes", namely *must, should, will* and *shall* as well as *have to*, *have got to, be going to* and *want to* in a range of ICE corpora, covering Inner Circle and Outer Circle varieties from different parts of the world, comparing their results to recent research on the development of these (quasi-)modals in British and American English. As ICE-USA is not yet completed, data from available corpora of spoken and written American English were used. Their results show noticeable differences between speech and writing, as quasi-modals tend to be more frequent in spoken genres. As was to be expected, American English shows the highest frequencies for quasi-modals, while British and New Zealand English, as well as the Outer Circle varieties, lag behind. A clear correlation between the ratio of quasi-modals to modals in the New Englishes and their evolutionary status in terms of Edgar Schneider's evolutionary model (cf. Schneider 2007) could not be established. As their findings are entirely based on normalised frequencies, trends with regard to the development of the different modal functions could not be established. Such differences are taken into account in a study on "The diverging *need (to)*s of Asian Englishes" by **Johan van der Auwera, Dirk Noël** and **Astrid de Wit**, based on ICE corpora from Hong Kong, Singapore, the Philippines and India and other corpora of British, American and Hong Kong English. They find significant differences between British and American English, on the one hand, in which *need* is associated with negative polarity and *need to* is generally used with positive polarity, and the individual Asian varieties under analysis on the other hand. The divergences between the Asian varieties are also quite remarkable, leading to individual profiles for each variety. A generalisation from the developments in British and American English for Englishes worldwide is thus not possible at this stage. A similar trend towards individual patterns across New Englishes also comes to the fore in the contribution on "*Will* and *would* in selected New Englishes: General and variety-specific tendencies" by **Dagmar Deuber, Carolin Biewer, Stephanie Hackert** and **Michaela Hilbert** based on conversation and interview data from six Outer Circle varieties from Asia and the Caribbean. A quantitative and qualitative analysis of the frequencies and functions of *will* and *would* as well as other future markers reveals a number of shared tendencies (e.g. overall low frequencies of *would*), but also individual patterns for each variety. However, the varieties from the Caribbean (Trinidadian, Bahamian and Jamaican English) are more similar to each other than the different Asian varieties. Future research on

the basis of more text types and varieties is needed to ascertain whether this is really due to their similar sociolinguistic background. The final contribution in this section deals with "Progressives in Maltese English: A comparison with spoken and written text types of British and American English" (by **Michaela Hilbert** and **Manfred Krug**), based on subsections of ICE Malta (press reportage and editorials, face-to-face conversations) and questionnaire data. While the frequencies of the progressive in the Maltese data do not differ significantly from their British English equivalents, there are marked differences with regard to the uses of the progressive, for example with stative verbs. In addition to that, the differences between spoken and written usage are significantly greater in the Maltese data than in British or American data. This evidence would support a classification of Maltese English as an ESL variety.

The following three chapters are concerned with verb complementation and particle verbs. **Marco Schilk, Tobias Bernaisch** and **Joybrato Mukherjee** are "Mapping unity and diversity in South Asian English lexicogrammar: Verb-complementational preferences across varieties". Based on data from large web-derived newspaper corpora and the ICE corpora from India, Sri Lanka and Great Britain, they analyse the complementation patters of the verbs CONVEY, SUBMIT and SUPPLY. The individual patterns of variation between the varieties under analysis lead the authors to the conclusion that labels such as 'South Asian Englishes' should be used with caution as the varieties spoken and written on the subcontinent display considerable differences with regard to their preferences for specific complementation patterns. **Gerald Nelson** and **Ren Gontao** examine "Particle verbs in African Englishes: Nativization and innovation" on the basis of a web-derived corpus of Ugandan English comprising a number of different text types. Their results show a strong influence of text type on the frequency of particle verbs, as has been shown previously for British and American English. They also stress the importance of particle verb innovations (i.e. particle verbs not attested in two recent dictionaries of phrasal verbs) in the process of structural nativization of New Englishes. **Lena Zipp** and **Tobias Bernaisch** trace "Particle verbs across first and second language varieties of English" based on smaller ICE subcorpora from three Inner and six Outer Circle varieties. In those cases where the ICE data proved to be too small, follow-up analyses with the help of the Google Advanced Search of the web were undertaken. While they were able to identify certain regional clusters for the uses of particle verbs with *up*, the stylistic and text-type specific distribution patterns also turned out to be very important in the overall results of the analysis. Their study is thus an excellent example of the collection's title, pointing out unity as well as diversity in the study of New Englishes.

Two chapters are devoted to individual clause types, namely relative clauses and specificational cleft sentences. **Ulrike Gut** and **Lilian Coronel** report on "Relatives worldwide", taking into account ICE corpora from Nigeria, Jamaica, Singapore and the Philippines. Their results show similarities between the four varieties with regard to the types of constructions used and varying frequencies across text types, but individual preferences with regard to pronoun choice or absolute frequencies of relatives across varieties. **Christian Mair** and **Claudia Winkle** report on an on-going "Change from *to*-infinitive to bare infinitive in specificational cleft sentences" which has been attested previously in British and American English. Based on data from ten ICE corpora from different parts of the world, they show that the Inner Circle varieties tend to pattern similarly to British and American English, while much intervarietal divergence can be seen in the Outer Circle varieties, possibly due to different input from British and American English. The general trend of the development, however, to move from *to*-infinitives to bare infinitives, appears to hold for all varieties of English under analysis.

The final chapter by **Nicole Höhn** investigates quotative *BE like, go* and *say* in ICE-JAM and ICE-IRE as the spoken data for both corpora was collected during roughly the same collection periods (1990–2005), which is important in the context of features reporting to be spreading in Englishes world-wide, as it also allows for some diachronic comparisons. Considering all factors revealed by previous research on innovative quotatives, Höhn shows that *be like* is used increasingly in Irish and Jamaican English, while *go* is only attested in ICE-IRE. While *be like* is favored by female speakers and for internal dialogue in Jamaican English, it is more likely to be used by men in first-person contexts in the most recent collection period in Irish English. The predictions of previous researchers could thus only be partially supported by the data from the two ICE corpora.

The individual contributions to the volume differ greatly with regard to the number of features or number of corpora under analysis, and with regard to methodological and analytical differentiation. In those studies, in which mere frequency analyses are supplemented by qualitative analysis or the effects of factors such as text type, speaker, or different morpho-syntactic contexts, it becomes clear that a real understanding of the development of the New Englishes requires linguistic studies as fine-grained as those of long established standard varieties such as British or American English. The question of unity or diversity across varieties is really one of perspective – depending on which features you study, which data you select and how you analyse it, British and American English will also display a great degree of similarity (i.e. unity) or an

astonishing degree of diversity. It is time for the New Englishes to be recognised as varieties of English that deserve the same kind of attention as Inner Circle varieties. The research based on the ICE corpora presented in the present volume will certainly add to a more detailed picture and hopefully inspire more research along these lines. With regard to the ICE project, we can state that despite the limitations of one million word corpora, Sidney Greenbaum's vision is slowly becoming a reality.

References

Greenbaum, Sidney. 1996. Introducing ICE. In S. Greenbaum (ed.). *Comparing English worldwide: The International Corpus of English*, 3–12. Oxford: Clarendon.

Kachru, Braj B. 1985. *The alchemy of English: The spread, functions and models of non-native Englishes.* Oxford: Pergamon.

Schneider, Edgar W. 2007. *Postcolonial English: Varieties of English around the world.* Cambridge: Cambridge University Press.

Merja Kytö (ed.). *English corpus linguistics: Crossing paths* (Language and computers: Studies in practical linguistics 76). Amsterdam, New York: Rodopi, 2012. 255 pp. ISBN 978-90-420-3518-8. Reviewed by **Martina Bredenbröcker**, University of Paderborn.

The volume *English corpus linguistics: Crossing paths,* which was edited by Merja Kytö, reflects the characteristics and potential of corpora for linguistics and – relating to the book's title in its literal sense – for interdisciplinary studies as well. The ten contributions assess the merits of corpus linguistic research with a view to their possible transfer to various disciplines other than linguistics.

The first part of the book ("Setting the scene") focusses on more fundamental issues. **Anne Curzan** addresses the question of how methods and insights from English Corpus Linguistics can be applied to other sub-disciplines of English such as e.g. literary and language study. But what do corpus linguists have to offer that could be interesting for other fields of research? Curzan's argumentation is as follows: corpus linguists' chief interests and core competence lie in analysing electronic collections of text. Nowadays, more and more

English literary and non-literary texts are being digitized in projects like EEBO (*Early English Books Online*) and ARCHER (*A Representative Corpus of Historical English Registers*), so that they constitute large databases of electronic texts. Why not use this common ground for collaboration and cooperation? Curzan (p. 13ff.) names several objectives for possible academic synergy:

1. Studies of language history: literary scholars are interested in the historical context of their literature and therefore could consider corpus-based study of language history as part of that context.

2. Work on registers and genres: corpus linguists can offer a different take on the diachronic studies of formal features in scientific texts or political discourse.

3. Study of collocations: a literary perspective could add interesting new concepts to that long standing issue in corpus linguistics.

4. Keyness: Curzan (p. 10) introduces an idea from American author Tony Morrison to describe another possible field of co-operation – examining the 'Africanist presence' in American literature – for which the corpus linguistic notion of keyness seems a ready-made technique. She ends her article with a demand for building databases and enhancing them with tools and methodologies that enable both literary **and** linguistic scholars to not just read the texts, but also to analyse data electronically.

Charles F. Meyer makes the case for a dual approach in textual analysis: inspired by the grammatical descriptions produced by the grammarians of the 'Great Tradition' (e.g. Jespersen, who used a more philology-oriented textual analysis), he combines this with a modern approach using software tools to analyse electronic corpora. Although he clearly sees the unprecedented possibilities corpora offer in synchronic and diachronic analysis of language structure, he also fears that by relying on the automated analysis alone, it might be the case that the scholar is taken too far away from the text as the computer provides a false sense of security. Consequently, Meyer claims that linguists should go "back to the future", meaning to "complement automated analyses of corpora with the more philologically analyses conducted by earlier grammarians" (p. 24). Meyer presents among others a computer-derived study of gapping phenomena, i.e. a type of coordination ellipsis as shown e.g. in the sentence taken from ICE-GB (S2B-037 #88:1:A) where the auxiliaries *have been* are deleted: "*It says three hundred and twenty civilians have **been killed** and more than four hundred [   ] injured*" (p. 31, brackets given by Meyer). In order to minimize the extent to which an automated corpus analysis might be flawed – despite more and more advanced corpus annotation – he opts for restricting searches so that a manual examination is still manageable. He concludes with the advice that scholars doing corpus research should pay heed to the notion that every corpus

has its limitations and that only when being aware of these, reliable interpretation of data is possible.

Part two of the volume, which is dedicated to investigations on present-day English, begins with an article by **Stig Johansson**. He presents three multi-lingual corpus studies carried out on the OMC (*Oslo Multilingual Corpus*), concentrating on English, although the corpus comprises texts from several other languages too. The analyses deal with spatial linking, expressions of possibility, and expressions of habituality. The first study examines the close cognates *her* (from Norwegian) and *here* (from English). Although they are dictionary equivalents, they turn out to have different functions. Whereas in Norwegian, *her* is not only used as an adverb of space but also as an anaphoric discourse marker, the English language does not use *here* in that way, but rather shows a preference for participant continuity in indicating cohesion. The second study analyses modal auxiliaries (e.g. *can* and *may* and their Norwegian counterparts) and their role in requesting, giving, and reporting permission. In the third study, Johansson looks into expressions of habituality in three languages, when comparing the use of English *used to*, Norwegian *pleie* and German *pflegen* and finds quite striking differences. Finally, he sums up the advantages of multi-lingual corpora for "contrastive linguistics in new key" (p. 46), i.e. the fact there are correspondences and that they can be interpreted linguistically. Thus totally new insights are possible: opportunities for crossing paths with lexicology, translation studies, and language pedagogy are given.

Geoffrey Leech, Nicholas Smith, and **Paul Rayson** address two topics related to stylistics. First, they present recent diachronic research on stylistic change carried out on the Brown family corpora. Secondly, they (synchronically) compare a Virginia Woolf text with a reference corpus regarding stylistic norms. They used the Brown family corpora, which are in many respects comparable corpora, for their diachronic study of style in published British English. These were sampled at roughly 30 year intervals (in the 1930s, 60s, 90s): the B-LOB, LOB and FLOB. For the authors, style is "a particular way of using the language or […] expressing meanings" (p. 70). Style is necessarily coupled with frequency and therefore measurability; hence a comparison of linguistic characteristics with a reference corpus is possible. The evidence yielded by their first study can be stated in three keywords: an on-going trend towards colloquialization, densification and democratization (the latter not being part of this paper). To exemplify the tendency towards spoken norms of usage in written English ('colloquialization'), they investigated the increased frequencies of the use of the passive voice, pied-piping, *upon*, and the *not*-contraction. The evolving need

to express more semantic content by using less word material ('densification') is demonstrated by the extended usage of the *s*-genitive and noun-noun-sequences.

In the second part of their study, the authors use Rayson's software tool, *Wmatrix*, to compare a short literary text with regard to lexical, semantic, and grammatical characteristics in order to trace individual author styles. As none of the available corpora seemed ideal at the time of the study, the authors compared Virginia Woolf's short story "The mark on the wall" with three different corpora: one containing texts from 1917, written by female fiction writers, one compiled of general fiction texts from the B-LOB corpus and a third very general one of published British English from the beginning of the 20th century. The methodology used for the study involved the extraction of key features using *Wmatrix* and an analysis of frequency, word class, and semantic domains. Two different quantitative techniques were applied: the counting of occurrences per million words and the keyness measured in terms of the significance ratio of Log Likelihood. The results illustrate the potential of this methodological approach to the stylistic analysis of literary texts. It can be viewed as a pioneering way to use corpus linguistic techniques, because the computer allows a kind of research that would be impossible if carried out manually.

Part 3 ("Focus on early English") opens with **Laurel J. Brinton**'s paper on historical pragmatics, which is split into three parts: a review of historical corpora including the OED (*Oxford English Dictionary*) and the Helsinki Corpus; a description of problems that arise when trying to apply corpus linguistic methods to historical pragmatics; and a case study showing the development of the comment clause *as you say*. Brinton argues that although the quotation bank of the OED is not a corpus in its strictest sense, corpus linguists doing diachronic studies can still benefit from it, as it comprises 2.4 million quotations from Old English until the present day. The other corpus used for this study is the Helsinki Corpus, which contains texts until 1710, but which is fairly small with only about 1.5 million words. Thus, Brinton recommends exploiting other sources as well, depending on the period to be analysed, such as e.g. the Chadwyck-Healey corpora. Brinton applied a mixed approach, combining quantitative (primarily frequency counts) and qualitative methods. The study of historical pragmatics is often problematic per definition, since research on pragmatics is usually carried out by investigating oral discourse, and there is simply no data for the earlier periods. Secondly, pragmatic markers are often ambiguous in meaning and thus cannot be clearly identified. Brinton then presents a case study which investigates the pragmatic functions of the comment clause *as you say* across the history of English. In present usage, it is a content disjunct and therefore functions as a pragmatic discourse marker. It is either used to highlight information or as

an interrogative sentence tag. She sets out to test various hypotheses with regard to the historical source structure, none of which can be proven valid. Instead it appears that a whole mixture of structures, such as the main clause *you say*, the relative/adverbial *as you say*, as well as *you say* following a fronted element supported the development of the comment clause *as you say.* All in all, Brinton recommends adopting a corpus approach to historical pragmatics, as it can help to shed light on the diachronic study of discourse markers, even though some standards usually applied, such as having an appropriate and large data set, are not entirely met.

Claudia Claridge's paper can be briefly summarised under the label "lexicography meets corpus linguistics" (p. 134). Claridge examines the usage and the treatment of transferred senses of a set of body part terms in three Early Modern English (EModE) dictionaries from the 18th century. In a second step, these data are compared with data from three small historical corpora with regard to the following questions: which senses are recorded in the dictionaries, and do they match those attested in the corpora? Among the dictionaries at hand are Samuel Johnson's (1755) *A dictionary of the English language* and Nathan Bailey's (1730) *Dictionarum Britannicum.* They include, apart from a common vocabulary, also terms taken from literary and specialised sources, whereas the selected corpora consist of texts taken from more private and colloquial registers. They add up to just over three million words, covering the span from 1560 until 1791. Claridge focuses on a set of body part terms such as *head, face, eye, leg,* and *foot* and differentiates between their literal meanings (referring to the actual anatomical part) and all other meanings (which are seen as transferred senses). Furthermore, the extent to which EModE dictionaries included semantic concepts such as polysemy or other forms of meaning extension, i.e. metonymy, or metaphor is also examined. In conclusion, considerable overlap can be found in the treatment and occurrence of the body part terms in the two sources, although the dictionaries vary in their degree of overlap. Nevertheless, what is of even more interest is where they differ: the dictionaries obviously advocate a particular usage to support a stylistic ideal in a prescriptive approach to lexicography. The corpus approach adopted by Claridge helps to detect this bias by comparing the evidence gained from the dictionaries with the picture of everyday usage as attested in the corpora. A positive by-product resulting from this study is a recognition of the contribution even small historical corpora can make to the study of historical phraseology, as they yield fair amounts of collocations and other multi-word units. This field of research has until now been underrepresented.

Thomas Kohnen explores the role of text-linguistic and discourse-functional features in another comparatively underrepresented genre, namely vernacular prayers of Late Medieval and Early Modern English. Although prayers are by nature directed to God or a saint and therefore unidirectional, they still show features of interactive and oral conversation. They have not been investigated until now, possibly because of their mostly private and individual character. The author has compiled a corpus of vernacular prayers from the 16th and 17th century whose three divisions add up to a total of 257,000 words. They were found in a collection of prayers made for lay people called *Book of hours*, comprising devotions asking for safety, salvation, pardon, and protection. The study's focus is on 1st and 2nd person personal pronouns, explicit performatives such as *pray, beseech, entreat*, and patterns of address (designations of God, e.g. *Lord*) – all features typical of oral discourse. All of them were found in the data in high frequency, which leads Kohnen to the conclusion that prayers can be viewed as belonging to oral and interactive language use. Some evidence gained from the BNC by investigating the address term *Lord* in recent prayers shows similar findings. For the future, Kohnen envisages more studies looking into various points of comparison between prayers and other text samples representing spoken conversation and discourse from diverse genres such as dramas, plays, charms, and trials.

In the abstract and the introduction of **Ian Lancashire**'s article, the author writes about *semantic deviation*. However, in his actual paper *semantic derivation* or *semantic drift* is addressed. Lancashire looks into the EmodE semantic derivation of three instances taken from Shakespearean plays, namely *the pricking of a witch's thumb* in Macbeth, the name *Aron* in Titus Andronicus and *acting* in Julius Caesar. Lancashire argues that, as no monolingual dictionaries from that period exist, Shakespeare and contemporary writers were less restricted in creating neologisms or examples of extended or transferred meaning. Yet some of these examples were short-lived and did not become lexicalised; hence, it is nowadays difficult to fully understand the allusions Shakespeare wanted to include. To help us bridge the gap of four centuries, the author strongly recommends using online diachronic text collections such as EEBO (*Early English Books Online*) and especially his own corpus, LEME (*Lexicons of Early Modern English*), which offers additional lexicographic information taken from various dictionaries. In summary, his findings are surprising: by taking into account corpus-derived evidence, he interprets the phrase *the pricking of a witch's thumb* as a kind of torture carried out in Scotland to extract confessions of suspected 'witches', a sense even the OED does not mention. Secondly, Lancashire refers to an older study of his to prove that the villain's name *Aron* in 'Titus Androni-

cus' is derived from the name of an English weed, usually called *wake robin*, which shares quite a few characteristics with Shakespeare's anti-hero. In his third case study, the true meaning of the participial noun *acting* in the tragedy 'Julius Caesar' is investigated. The author claims that neither Shakespeare's editors nor the OED have until now interpreted the word correctly. By using the above-mentioned means of diachronic lexicography, he has discovered that Shakespeare "transferred to his new noun 'acting' a specific sense of the old noun 'act'" (p. 191). To sum up, the diachronic corpora discussed are not only valuable resources from a linguistic point of view but are also important for historical or literary disciplines.

Matti Rissanen starts his article by giving an overview of English historical corpus linguistics over the last forty years and the contribution it has made to pragmatic and discourse-based analysis. He then lists available corpora according to the period they cover, from Old English to present-day English, with a view to possible links to interdisciplinary studies, e.g. socio-historical research. His thorough case study of the connective *provided that* from Middle to present-day English illustrates this particular use of corpora by tracing its occurrences in the Helsinki Corpus and a fair number of other multi-genre and specialised sources such as *The Penn-Helsinki Parsed Corpus of Middle English* and the *Zurich English Newspaper Corpus*. He comes to the conclusion that *provided that* has never been entirely grammaticalised, maybe because of its ability to occur separately or because of the high frequency of the verb alone. Rissanen ends on the note that, although the introduction of corpus linguistic methods has revolutionised the analysis of language fundamentally, the human brain is still salient when it comes to interpreting the data which has been gained by computers and software tools.

Elizabeth Closs Traugott discusses the assumption that a subset of 'bridging' contexts is a key factor in morphosyntactic change. She presents various approaches to these contexts with a view to pragmatics and semantics. Her aim is to find evidence from empirical data in order to identify them as essential and distinct stages in the process of grammaticalisation. She presents two studies on syntactic constructions. One deals with *be going to* and the other with a subset of pseudo-clefts with *what* and *all*, as in "*What/All I did was (to) voice support for her*" (p. 222). In the case of *be going to*, Traugott is able to trace evidence for a short-lived stage in the process of grammaticalisation that could be interpreted as bridging a gap between the 'normal' and the altered usage of the structure. Her second example however, the development of pseudo-clefts, yields no evidence of a bridging or critical context. Thus these contexts do not seem to be necessary phases for grammatical features on their way to being conventiona-

lised. However, she takes care to point out that her results might be different if larger co-texts or different sources were chosen for the analysis.

The present volume *English corpus linguistics: Crossing paths* constitutes a strong argument that the field of corpus linguistics has much to offer other disciplines, be it in terms of data, methodology, or a linguistic perspective on their topic of research. However, these paths do not have to be one-way-streets: the articles have also shown intersections and areas where corpus linguists can profit from fruitful interdisciplinary exchange.

Manfred Markus, Yoko Iyeiri, Reinhard Heuberger and **Emil Chamson** (eds.). *Middle and Modern English corpus linguistics. A multi-dimensional approach* (Studies in Corpus Linguistics 50). Amsterdam and Philadelphia: John Benjamins, 2012. 287 pp. ISBN 978-90-272-0355-7. Reviewed by **Paula Rodríguez-Abruñeiras**, University of Santiago de Compostela.

Middle and Modern English corpus linguistics. A multi-dimensional approach is a collection of papers which underlines the relevance of corpus linguistics for the study of Middle and Modern English. The articles were originally presented at *MMECL: Middle and Modern English Corpus Linguistics*, an international conference held in Innsbruck (Austria) from 5th to 9th July 2009. The volume, which opens with an introductory section by the editors, is divided into four thematic blocks. Except for Part I, which contains two studies, each block contains five articles. The use of corpora is the connecting thread for all the papers included in the volume. The importance of corpora is foregrounded in Part I, where some problems and peculiarities of corpus compilation are discussed, but it is in Parts II, III and IV that corpora are used as the main source of data for the various analyses presented. Parts II and III deal with the historical analysis of different linguistic phenomena in Middle (ME) and Modern English (ModE), and Part IV is concerned with dialectal variation.

The two papers in Part I are an overt defence of the use and benefits of corpus linguistics. In the first of these, "*Can't see the wood for the trees?*: Corpora and the study of Late Modern English" (pp. 13–29), **Joan Beal** indicates some of these benefits of using corpora, such as the ability to make linguistic research more thorough and comprehensive by providing statistical information on the phenomenon under study. She notes the advances which the "corpus revolution" (p. 13) has brought to the understanding of Late Modern English (LModE). She

also refers to the eighteenth and nineteenth centuries as the "Cinderellas of English historical linguistic study" (p. 13), in that linguistic changes during these centuries were often seen as minor and unimportant, arguing that such a misconception was made on the basis of a wrong approach. In previous stages, changes in the language occurred due to the evolution of the system itself, whereas in LModE they originated mainly in "the interplay of different varieties" (Strang 1970: 78–79). Indeed, in this era linguistic changes seem to be "statistical in nature" rather than "categorical losses or innovations" (Denison 1998: 93). Beal's paper closes with some optimistic words about the creation of a corpus of eighteenth-century phonology, which would bring an end to the scant attention that historical phonology has thus far received.

In the second paper of Part I, "Spelling variation in Middle English manuscripts: The case for an integrated corpus approach" (pp. 31–45), **Stefan Diemer** discusses some of the drawbacks and deficiencies of current ME corpora. The main problem is the lack of consistency with regard to spelling, a result of the lack of standard orthography at the time. Although aware of the difficulties in accessing some manuscripts, Diemer advocates the integration of manuscripts in the corpora, which would provide linguists with extra information about texts under scrutiny. The use of decoration, for example, would indicate that the manuscript was commissioned by a rich patron, and the abbreviation of a word ending could reveal potential areas of phonetic loss.

Part II opens with a study into the origin and development of numerals in English. In "The development of compound numerals in English Biblical translations" (pp. 49–57), **Isao Hashimoto** analyses the use of three different types of numerals, exemplified by *one and twenty* (used from Old English (OE) until Early Modern English (EModE))*, twenty and one* (a possible but not very frequent option in ME and EModE) and *twenty-one* (the present-day English (PDE) form) in a series of religious texts. Rissanen (1967: 30–32) attributes the intermediate stage of the construction to the "demands of poetic diction" (p. 49), but Hashimoto concludes that such a construction is also influenced by Hebrew and Latin translations. With this contribution, Hashimoto focuses attention on the impact that languages may have on each other.

In the next paper, "The complements of causative *make* in Late Middle English" (pp. 59–73), **Yoko Iyeiri** offers a thorough account of the different types of complements which the causative verb *make* could take in the fifteenth century, namely *that*-clauses, (*for*) *to*-infinitives, and bare infinitives. Of these three types, the former is only occasionally used, whereas the other two variables compete throughout the fifteenth century. Iyeiri's data confirm that the grammaticalisation of causative *make* plus a bare infinitive was not completed

by the turn of the century. As a matter of fact, they showed a rather stable distribution according to different linguistic contexts.

Grammaticalisation is also a central in **Tine Defour**'s "The pragmaticalization and intensification of *verily, truly* and *really*: A corpus-based study on the developments of three truth-identifying adverbs" (pp. 75–92). The history of these three forms runs in parallel over time. In their earliest occurrences in ME, they were used as mode adjuncts with a rather limited scope, but during the ME and especially EModE periods, *verily, truly* and *really* widened their scope and developed new meanings as emphasisers and disjuncts. Findings reveal that *really* is the most common in PDE, especially as an intensifier, whereas *verily* is infrequent and even somehow archaic, probably because it has always been associated with religious contexts. In turn, *very* is used instead of *verily* as an intensifier. Defour's article illustrates a process of grammaticalisation where semantic, pragmatic and syntactic changes cooperate to an almost equal degree.

In the paper by **Sylwester Łodej**, "Concept-driven semasiology and onomasiology of CLERGY: Focus on the lexicogenesis of *pope, bishop* and *priest*" (pp. 93–108), the influence of extralinguistic reality on language change becomes apparent. By using a corpus of comic drama texts, Łodej considers the different meanings which the words *pope, bishop* and *priest* have acquired from the beginning of the Protestant Reformation in the sixteenth century. This historical event is a decisive turning point in the development of these three forms, which gradually acquire secular references. This secularisation reflects a social disdain towards religious figures at the time.

The final paper in Part II, **Hans-Jürgen Diller**'s "*ANGER* and *TĒNE* in Middle English" (pp. 109–124), analyses linguistic competence in the use of the synonyms *anger* and *tēne* and words derived from them. Semantically, the two nouns were very close in the Middle Ages, both of them meaning 'anger' and 'sorrow', which as Diller points out is rather surprising seen from the perspective of modern psychology, given that 'anger' is considered as an active emotion, whereas 'sorrow' is seen as passive. Diller's data show a sudden change in the frequency of these two nouns: whereas *tēne* is the most frequent form between 1100–1400, *anger* sharply increases in frequency and becomes the more common choice at the end of the Middle Ages. *Tēne* eventually disappears in the EModE period.

Namiko Kikusawa opens Part III with "The subjunctive vs. modal auxiliaries: *Lest*-clauses in Late Middle English prose texts" (pp. 127–139). The focus of her analysis is on linguistic competence in the use of subjunctive forms and modal verbs in *lest*-clauses in the fourteenth and fifteenth centuries. She concludes that the use of these devices depended on various factors, both linguistic

(e.g. adverbial *lest*-clauses are prone to the use of subjunctive forms, whereas complement *lest*-clauses prefer modal verbs) and extralinguistic (different text-types favour the use of one device or the other; for example, whereas fiction and religion show a higher use of subjunctive forms, biographies of saints, history and romance favour the use of modals. As far as letters are concerned, no significant difference was found between the two types of clauses).

In his contribution, **Tomohiro Yanagi** provides "Some notes on the distribution of the quantifier *all* in Middle English" (pp. 141–155). He considers different factors which may condition the position of the quantifier *all* with respect to the element which it modifies. If the modified element is a noun phrase, *all* almost always precedes it, whether the construction is a subject or an object. If it is a pronoun, *all* rarely follows it if they function together as an object, but it may precede or follow the pronoun without any significant difference when they function as subject. The use of "*all*-pronoun" or "pronoun-*all*" (p. 141) in subject position is influenced by the grammatical person of the pronoun and the type of clause where it occurs: third person pronouns and main clauses show a preference for the use of "*all*-pronoun". Moreover, *all* can be separated from the modified element when they function as subject, but such split is practically nonexistent in object position.

In his paper "Interjections in Middle English: Chaucer's 'Reeve's Tale' and the Corpus of Middle English Prose and Verse" (pp. 157–175), **Hans Sauer** provides an exhaustive inventory of the interjections used in this tale. The study of these elements reveals some noteworthy linguistic features. For example, interjections reflect the different geographic origin of the characters. They also show the impact which French had on English at the time, given that some of those interjections have a French origin. Sauer also pays attention to other important aspects, such as the function of the interjections and their position in the sentence. In order to provide a more valid and accurate analysis, the paper concludes by comparing the data from "The Reeve's Tale" with data from the *Corpus of Middle English Prose and Verse*.

Ursula Lutzky's contribution, "*Why* and *what* in Early Modern English drama" (pp. 177–189) offers a perspective on the use of *why* and *what* as discourse markers in EModE. The period starts with *what* leading the way, but by the end of EModE *why* is by far the more common option. Lutzky notes the similarities in the use of both forms, especially their tendency to come in initial position to mark a change of speakers. But she also establishes significant differences between them. In general terms, she concludes that *what* is more common before questions and *why* before declaratives, with both conveying a nuance of surprise. Moreover, when used in declaratives they may also indicate that the

speaker is going to add new information with the aim of persuading the recipient. Nevertheless, only *why* can have a contrasting or concluding function.

Part III comes to an end with a solid analysis of NOT-contracted and uncontracted forms in the LModE period. **Erik Smitterberg** presents "Colloquialization and NOT-contraction in nineteenth-century English" (pp. 191–206) as the starting point for a more comprehensive study in which other traces of the colloquialisation of the English language might be considered, such as the use of phrasal verbs or of progressive forms. In this article, Smitterberg focuses on drama, fiction and trials. The use of NOT-contractions significantly increases throughout the nineteenth century in both drama and fiction, but not in trials. He also considers some extralinguistic factors which could influence the colloquialisation of the language here (e.g. more people had access to written documents because they learned how to read and write), as well as other circumstances which had an opposite influence, slowing down such processes of colloquialisation (for example, the influence of prescriptivism).

Part IV is also divided into five chapters. The first three papers describe different aspects of Wright's *English dialect dictionary* (EDD), whereas the last two articles focus on the post-EDD era. **Manfred Markus**'s contribution, "The complexity and diversity of the words in Wright's *English dialect dictionary*" (pp. 209–224), is a perfect user's guide for the EDD in which the organisation of the dictionary's entries are explained in detail. Markus hails the benefits of the EDD online for the study of the LModE period, given that it is one of the most comprehensive sources of information for the time-span 1700–1900. He uses examples to illustrate the complexity of the dictionary, which contains information about virtually all levels of the language system, covering not only semantics, syntax, pragmatics, phonetics and spelling, but also traces the time and place of the use of dialectal forms.

Emil Chamson in turn analyses the usefulness of the EDD for historical research in "Etymology in the *English dialect dictionary*" (pp. 225–240). Even though the EDD was not conceived as an etymological dictionary, etymology plays an important role in it, and thus it can be compared to other historical dictionaries, such as the *Oxford English dictionary* (OED). In spite of its limitations (such as the fact that in most cases only immediate etymons and not ultimate sources are identified), Chamson rightly defends the EDD as a solid and reliable source of data for historical research, capable of putting an end to the conception of English as an "etymological orphan" (Liberman 2005: 4).

In "Towards an understanding of Joseph Wright's sources: White Kennett's *Parochial antiquities* (1695) and the *English dialect dictionary*" (pp. 241–256), **Javier Ruano-García** provides a general overview of those works consulted by

Wright in the EDD, thus shedding some light on the documents used as bibliographic sources in the dictionary. Ruano-García examines one of Wright's sources in detail, Bishop White Kennett's glossary to *Parochial antiquities* (1695). This 1695 glossary is on occasions the only source quoted for certain southern and Midland data, whereas other sources (especially John Ray's *A collection of English words not generally used* 1674, 1691) are preferred for northern data.

The last two papers of Part IV leave aside the EDD and deal with other aspects of dialectal studies. In "The importance of being Janus: Midland speakers and the 'North-South Divide'" (pp. 257–268), **Clive Upton** rejects the idea of dialects as definite entities with clear-cut boundaries and defends the transitional character of dialects as part of a linguistic continuum. By considering the English Midlands as a case in point, Upton foregrounds the importance of transition zones, which have thus far been widely underestimated.

Finally, **Christian Mair**'s concern in "...*ging uns der ganze alte Dialektbegriff in eine Illusion auf*: The deterritorialization of dialects in the 20th and 21st centuries" (pp. 269–283) is not only "rural Britain" but also the "Black Atlantic" (p. 272). Mair describes how pidgin and creole languages (e.g. Jamaican Creole) were largely ignored until recently. Things have changed dramatically, though, and at present these varieties are no longer perceived as marginal. In the past, the link between a linguistic variety and its geographical area was strong, but events such as world-wide migration and the presence of dialectal varieties in the media have made that link "tenuous" and dialectal varieties have now become "deterritorialized dialects" (p. 276).

The contributions made by the studies in the present volume to the field of corpus linguistics are numerous and highly valuable. They provide new insights into such disparate areas as corpus compilation, diachronic developments and dialectal variation. Beyond question, all the articles here contribute to a situation in which the marginal status which the ModE period has traditionally had in historical research is now over.

References

Denison, David. 1998. Syntax. In S. Romaine (ed.). *The Cambridge history of the English language. Volume IV: 1776–1997*, 92–329. Cambridge: Cambridge University Press.

Kennett, Bishop White. 1818 [1695]. A glossary to explain the original, the acceptation, and obsoleteness of words and phrases; and to shew the rise, practice, and alteration of customs, laws and manners. In B. Bandinel (ed.).

Parochial antiquities attempted in the history of Ambrosden, Burcester, and adjacent parts in Oxford and Bucks. (A new edition in two volumes greatly enlarged from the author's manuscript notes), Vol. 2, fols. A2–N. Oxford: Clarendon Press.

Liberman, Anatoly. 2005. *Word origins: And how we know them; etymology for everyone*. Oxford: Oxford University Press.

OED= *Oxford English dictionary online.* http://www.oed.com. (November 2012).

Ray, John. 1674. *A collection of English words not generally used*. London: Printed by H. Bruges. http://eebo.chadwyck.com. (November 2012).

Ray, John. 1691. *A collection of English words not generally used*. London: Printed for Christopher Wilkinson. http://eebo.chadwyck.com. (November 2012).

Rissanen, Matti. 1967. *The use of* one *in Old and Early Middle English*. Helsinki: Société Néophilologique.

Romaine, Suzanne (ed.). 1998. *The Cambridge history of the English language*, *Volume IV: 1776–1997*. Cambridge: Cambridge University Press.

Strang, Barbara M. H. 1970. *A history of English*. London: Methuen.

Wright, Joseph. 1898–1905. *The English dialect dictionary*, 6 vols. Oxford: Henry Frowde.

Tony McEnery and **Andrew Hardie.** *Corpus linguistics: Method, theory and practice* (Cambridge Textbooks in Linguistics). Cambridge: Cambridge University Press, 2011. 294 pp. ISBN 978-0-521 83851-1 (hardback) ISBN 987-0-521-54736-9 (paperback). Reviewed by **Pam Peters**, Macquarie University, Sydney.

With its straightforward title, this book belies its very distinctive content among the largish number of textbooks on corpus linguistics that have appeared recently – at least ten in the last two decades, beginning with McEnery and Wilson (1996, 2nd ed. 2001) *Corpus linguistics*, published by Edinburgh University Press. McEnery and Hardie's (2011) volume is the fourth corpus linguistics title to be published by Cambridge University Press since 1998, counting Biber, Conrad and Reppen's *Corpus linguistics: Investigating language structure and use* (1998), Meyer's *English corpus linguistics: An introduction* (2002), and

Hunston's *Corpora in applied linguistics* (2002). What distinguishes McEnery and Hardie's new *Corpus linguistics* is its embedded critique of corpus methodology and practice from the scientific point of view, and the attention it pays to the qualitative corpus research of the Neo-Firthian school emanating from Birmingham under John Sinclair's influence. The range of subdisciplines of linguistics discussed as having affinities with corpus linguistics is also notably wider than in previously published textbooks, including not only cognitive linguistics and construction grammar, but also experimental disciplines such as psycholinguistics and neurolinguistics. Its strength is in not only explaining what corpus linguistics is and does, but establishing the epistemological value of corpus linguistic findings and their contribution or potential contribution to empirical linguistics at large. What follows below are summary accounts of the main features of each chapter, to indicate the range and richness of the book's content.

In Chapter 1 ("What is corpus linguistics?"), readers are introduced to the design issues in creating a sample corpus. The sample of texts is intended to *represent* the language under investigation, because of the impossibility of accessing the whole language. In corpus linguistics this is a fundamental question: how representative the corpus is of the language or register(s) under investigation, for it to be regarded as a valid source of data. The range of texts sampled in the corpus, and the proportions of each, need critical consideration if the corpus is to stand as a microcosm of the language. So neither a 'serendipity' collection of texts of all sizes, nor a monogeneric collection of texts, would count as a representative corpus if the research aimed to describe the language at large. These principles of corpus design are not endorsed by all linguistic researchers, notably those engaged in critical discourse analysis (CDA) research. But they are fundamental (if still debatable) for users of the major sample corpora of the English language, smaller and larger (e.g. Brown, LOB, BNC, COCA). Because the sample corpus is a planned structure, corpus researchers can commit themselves to being totally *accountable* to the data in it (Leech 1992), rather than using it selectively to suit their purposes. They thus allow their findings to be either replicated or falsified, in accordance with scientific practice.

Chapter 2 ("Accessing and analysing corpus data") takes readers to the practice of annotating corpus texts, i.e. marking up linguistic features of the text such as its grammatical constituents, for more sophisticated computer searching. The annotation systems used may be automatic or manual or a combination of the two, typically an automatic part-of-speech 'tagger', which allows for manual editing, and may then support constituency parsing and parse trees. Other kinds of annotation, e.g. of the pragmatic moves in a conversation, are necessarily

manual. Collocational analysis of corpus data is supported by machine concor-dancing, which has steadily developed over the four 'generations' reviewed by McEnery and Hardie. The classical concordance was the key-word-in-context (KWIC) output from 1970s mainframe computers, which by the early 1990s could be run on personal computers, at least on a smaller scale. In the mid- and later 1990s, concordancing software was enhanced (as in WordSmith) to read scripts other than English, and to provide a wider range of tools for lexical anal-ysis, such as calculating keywords. More sophisticated syntactic analysis also came with the third generation parsers, notably the ICECUP software package used for ICE-GB, though it does not work on parsing schemes other than that developed in London for the ICE corpora. The fourth generation distinguishes itself by the building of large web-based interfaces through which users of per-sonal computers can access processed data from very large corpora such as the BNC and COCA (the client-server model).

Chapter 3 ("The web, laws and ethics") reminds us of how the ready acces-sibility of texts on the internet provides a giant source of data for research – except that (a) their provenance and the generic mix is unknowable (unlike those in a planned sample corpus); and (b) the researcher's legal rights to extract texts from the internet for use in his/her own corpus are still unclear (and may consti-tute copyright infringement if used without permission). Linguistic researchers often wish to share their internet-derived data with colleagues (for reasons of collegiality and replicability), but this now leads on to pressure to make it pub-licly available, and the need to clarify all the inherent legal and ethical issues – against the unsettled backdrop of international law on the use and control of material on the internet.

Chapter 4 ("English corpus linguistics") is an account of the substantial British contribution to the establishment of corpus linguistics, with sections on those of University College London, Lancaster University, University of Bir-mingham and University of Nottingham – as well as Université Catholique de Louvain and the University of Northern Arizona. Without the last two one might wonder whether any significant corpus linguistic research had been done outside England, or whether English corpus linguistics (ECL) relates only to anglo-phone countries. But the chapter's introduction justifies its selection while giv-ing space to the role of the ICAME, as a scholarly community that has contrib-uted substantially to the development of corpus linguistic research, and the creation of numerous corpora. Other contributing universities in Europe and the USA are discussed on the back of the six to which the chapter gives high profile, and in Chapter 5.

Chapter 5 ("Corpus-based studies of synchronic and diachronic variation"). This chapter introduces the two most widely used historical corpora of English, the Helsinki corpus of Old and Middle English texts; and the ARCHER corpus (*A Representative Corpus of Historical Registers*), covering the early modern to modern English periods. Apart from using these large-scale historical corpora, corpus linguists can also study micro-diachronic changes in C20 English, thanks to the construction of sample corpora parallel to Brown and LOB of the 1960s, at 30-year periods before and after. However the chapter's chief focus is on register differentiation using quantitative methods. Biber's multidimensional analysis (MD) used in register description is extensively discussed, showing how using a large set of linguistic variables provides more far-reaching discriminations than previous register studies. MD supports investigation of larger language trends such as 'colloquialization', although this does suggest that the variables themselves will eventually need recalibrating, since they were grounded in late C20 assumptions about the contrasts between spoken and written discourse (Biber 1988). Perhaps a more abstract 'feature tree' would provide a language-external way of identifying the totality of variables which contribute to register differentiation in a given language. MD has nevertheless been effectively applied to languages other than English, and would also lend itself to research in variationist sociolinguistics, as McEnery and Hardie point out.

Chapter 6 ("Neo-Firthian corpus linguistics"). This is the longest chapter, with a substantial review of the qualitative approaches to corpus linguistics taken by researchers associated more and less closely with Birmingham University, and with John Sinclair, its professor of modern English from 1965 to 2000. The chapter lays out the central linguistic concerns of the Neo-Firthians, with collocation and colligation, lexical priming, semantic preference and semantic prosody, the idiom principle, and pattern grammar. Though they make use of corpora, the Neo-Firthians distinguish research which is *corpus-based* from the *corpus-driven* (Tognini –Bonnelli 2001). These not-very-transparent terms mark the distinction between research in which corpus data is used to test elements of an external theory (e.g. the variation research reviewed in Chapter 5), and that which avoids any prior assumptions about language structure. This latter becomes doctrine in the most radical Neo-Firthian position, with the claim that you simply "trust the text" (Sinclair 2004) to produce its own theory. McEnery and Hardie express due scepticism about this corpus-as-theory position. But they attend to the various criticisms raised by Neo-Firthians against quantitative corpus-based research and corpus annotation; and they find value in the Firthian concern with collocation, with Hunston and Francis's careful work on *pattern grammar* (1999), and Hoey's on *lexical priming* (2005). These aspects of lan-

guage usage are indeed the intellectual pivot of the book, treated in some depth in this chapter prior to their return in discussion of the convergences within linguistics in Chapter 8.

Chapter 7 ("Corpus methods and functionalist linguistics"). This chapter discusses the positive relationship between corpus-based research and several usage-based approaches to language, gathered here under the heading of 'functionalist' – as opposed to 'formal' approaches such as that of the Chomskyan school. Functionalism then provides a common platform on which to discuss quite diverse subdisciplines of linguistics, including cognitive linguistics and construction grammar, because both are concerned with how meaning is constructed "in the mind or in interaction" (p. 170). A set of functionalist studies of syntax, with more and less systematic use of corpus data, is used to show the concern with larger, non-rule-based constructions on which cognitive linguistics and construction grammar converge. The cognitive grammarian's *collostruction* (i.e. the syntagmatic tendencies of individual words to associate with others from particular grammatical categories) can in fact be traced back to Firth's early (1957) concept of *colligation*. Yet corpus-based research on collostruction and the profiling of *collexemes* (Stefanowitsch and Gries 2003) is highly quantitative, as expressed in the statistical association between ditransitive verbs and their complements. These coincidences between cognitive linguistics and other construction-focused corpus-based research are also groundwork for the ultimate discussion of convergence in Chapter 8.

Chapter 8 ("The convergence of corpus linguistics, psycholinguistics and functionalist linguistics"). Affinities between corpus linguistics and psycholinguistics can be found in the use of word frequency data in controlled experiments into reading and language processing. But they would be enhanced by the use of relevant, up-to-date word frequencies, i.e. not always drawn from written data of the 1960s Brown corpus — and referencing adult language norms when those of children would be more appropriate (Brysbaert and New, 2009). In language processing the transition possibilities for readers are affected by collocational relationships in language, at which point the notions of psychological priming and lexical priming coincide, as Hoey himself pointed out. The parallels between language processing and research on language acquisition in first and second language learning are also evident. Children's learning of "constructional islands" (Tomasello 2003), and linguistic formulae which are quantitatively weighted (Wray 2002), both align as inputs to connectionist research and the role of neural networks in language processing. These kinds of convergence add to those already found between functionalist linguistics and corpus-based Neo-Firthian linguistics, and dependency grammars and valency grammars are

noted as well. In all of them, grammar is found emerging from lexical patterns. Collectively they contribute to "a unified empirical linguistics" (p. 222).

Chapter 9 ("Conclusion"). Here the authors project into the future with a forward look at disciplinary areas in which corpus linguistics could play a larger role than it does currently. There is untapped potential in computational linguistics, e.g. in data-mining for public opinion (otherwise known as 'sentiment analysis'); and in neurolinguistic research, e.g. as to whether the core semantics of polysemous/polygrammatical words such as *hand* are in some way neurally embodied. Here the relative frequencies of the various verb and noun senses of the lemma could serve as an external reference point.

As those chapter summaries show, McEnery and Hardie's book is an important contribution to linguistics at large, both in its critique of corpus methodology, and its exploration of the affinities between corpus linguistics and other subdisciplines. Those who would debate the points of convergence could have a vested interest in keeping apart, as the authors observe (p. 212). The book is certainly not an imperialistic manifesto for corpus linguistics, but designed to show how it can usefully contribute to evidence- or usage-based research, and serve as a mode of triangulation. Corpus linguistics is at its strongest in providing broadly based evidence of language, as rendered in the written medium. It cannot contribute to research on language in phonological form, except through extensively annotated, multilayered alignment of acoustic recordings with transcribed speech.

McEnery and Hardie's *Corpus linguistics* will be very useful as a textbook for advanced undergraduates in pure and applied linguistics, with its stimulating questions and exercises at the end of each chapter. It will be invaluable for postgraduate researchers from any subdiscipline of language study, as they develop methodology for their doctoral project. The book provides an excellent list of references. And with an expanded index (would that it were larger!), readers could more easily locate where all the critical concepts and significant research in corpus linguistics are discussed within the text.

References

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

Brysbaert, Marc and Boris New. 2009. Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and introduction of a new and improved word frequency measure for American English. *Behavior Research Methods* 41 (4): 977–990.

Firth, John. 1957. *Papers in linguistics*. Oxford: Oxford University Press.

Hoey, Michael. 2005. *Lexical priming: A new theory of words and language*. London: Routledge.

Hunston, Susan and Gill Francis. 1999. *Pattern grammar: A corpus-driven approach to the lexical grammar of English*. Amsterdam: John Benjamins.

Leech, Geoffrey. 1992. Corpora and theories of linguistic performance. In J. Svartvik (ed.). *Directions in corpus linguistics* (Proceedings of Nobel Symposium #82, Stockholm 1991), 105–122. Berlin: Mouton de Gruyter.

Sinclair, John. 2004. *Trust the text: Language, corpus and discourse*. London: Routledge.

Stefanowitsch, Anatol and Stefan Gries. 2003. Collostructions: Investigating the interaction between words and constructions. *International Journal of Corpus Linguistics* 8 (2): 209–243.

Tognini-Bonnelli, Elena. 2001. *Corpus linguistics at work*. Amsterdam: John Benjamins.

Tomasello, Michael. 2003. *Constructing a language: A usage-based theory of language acquisition*. Cambridge Mass: Harvard University Press.

Wray, Alison. 2002. *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.

Isabel Moskowich and **Begoña Crespo** (eds.). *Astronomy 'playne and simple'. The writing of science between 1700 and 1900.* Amsterdam and Philadelphia: John Benjamins, 2012. 240 pp. ISBN 978-90-272-1194-1. Reviewed by **Anu Lehto**, University of Helsinki.

The *Corpus of English Texts on Astronomy* (CETA) is a diachronic corpus of astronomical writing from 1700 to 1900. The corpus is the first published part of the *Coruña Corpus: A Collection of Samples for the Historical Study of English Scientific Writing*, which will consist of several sub-corpora of different scientific disciplines such as philosophy and chemistry (see e.g. Crespo and Moskowich 2010). The *Corpus of English Texts on Astronomy* is accompanied by the book *Astronomy 'playne and simple'. The writing of science between 1700 and 1900,* which offers background information on the socio-historical context of scientific writing and contains pilot studies carried out in CETA. The book con-

sists of eleven chapters; the first three chapters introduce the history of astronomy and the compilation of the corpus, while the following sections report linguistic findings that range from pragmatic studies to the analysis of vocabulary, morphology and syntax.

The first chapter by **Joan C. Beal** examines the historical and linguistic situation in the Late Modern English period. Beal illustrates the effects of the Industrial Revolution and urbanisation that initiated new scientific innovations and that formed new dialectal areas. In the second chapter, **Begoña Crespo** describes the history of astronomy that is rooted in Antiquity. In the Middle English era, astronomical writing was based on scholastic reasoning, but the emergence of the Enlightenment shifted the focus to the use of instruments and mathematical calculations and also raised criticism against the prevailing geocentric model. From the seventeenth to the nineteenth century, astronomy concentrated on the physical aspects of celestial objects rather than the motions of the planets. Further, scientific writing developed from the scholastic method into a more author-centred approach after the founding of the Royal Society in 1660. In CETA, the texts vary from theoretical texts to those that disseminate and popularise astronomy to lay audiences. In addition, the authors differ as regards their levels of education and in their relations to modern scientific methods, and all of these aspects are reflected in their writing.

Isabel Moskowich explains in more detail the compilation and structure of CETA. The corpus contains about 410,000 words, distributed over 42 samples. The word count of the extracts has been limited to 10,000 words, and two samples have been included from each decade. The corpus has been divided into eight genres or text types, namely articles, dialogue, essays, lectures, letters, academic treatises, textbooks and the category other (dictionary). The size of the categories corresponds broadly to the number of published works on astronomy in the Late Modern period: the largest categories are textbooks, academic treatises and research articles with 15, 12 and six texts respectively. The remaining five categories are, however, rather small, as they contain only from one to three texts. The corpus texts further vary in their authorship: the majority of the authors in CETA are from England, although Irish, Scottish and American authors are also represented, allowing the examination of sociolinguistic variables. On the one hand, the diversity of genres and authors successfully increase the representativeness of astronomical writing in the corpus. On the other hand, the smallest genres in CETA can pose challenges for analysis; e.g. the category of essays contains one sample from each century, and one of the authors is from Scotland and the other from England. Hence, the diachronic coverage is not

ideal, and possible linguistic differences can originate from dialectal variation or diachronic changes in the genre.

In Chapter 4, **Begoña Crespo** investigates the linguistic changes caused by the Enlightenment by focusing on persuasion. The analysed features are predictive modals (*will* and *shall*), conditional subordinators (*if* and *unless*), necessity modals (*must, should* and *ought*) and suasive verbs (e.g. *agree* and *ask*). In the data, the predictive modals are clearly most common, followed by conditional subordinators and necessity modals, whereas overt expressions of persuasion with suasive verbs are rare. In general, the features of persuasion are most common in genres in which the author is most visible, i.e. in dialogues, essays, textbooks and lectures. There are only minor differences between the two analysed centuries, but predictive modals, for instance, decrease in the nineteenth-century texts, signalling that authorial involvement shifts to object-centred statements.

In Chapter 5, **Isabel Moskowich** surveys the development of nominal style in scientific writing by examining suffixation in attributive adjectives in eighteenth-century texts in CETA. The chapter analyses the suffixes *-y* of Germanic origin and *-al* of Romance origin, as the hypothesis is that genres representing more specialised readership – such as articles and essays – prefer Romance or Latinate vocabulary. The numbers of adjectives and attributive adjectives are low in the data. The adjectives ending with *-al* are more frequent than the Germanic *-y* suffixes, but the suffixes are not distributed according to different readership.

Conzalo Camiña Rioboo further studies the scientific revolution in the eighteenth-century part of CETA by attesting how new nouns are coined by affixation and by compounding. The analysis shows that affixed nouns predominate over simple nouns and that suffixes are much more common than prefixes. Further, suffixes are typically of French, Latin and English origin, while prefixes are of Greek and Latin origin. The bases are mainly French and Latin but English bases are most common with neutral affixes, which do not cause changes to the base (e.g. *-ness*). Diachronically, the numbers of affixed nouns grow in frequency, especially towards the end of the eighteenth century. In relation to genres, essays have notably fewer noun types than the other categories in the corpus.

Pascual Cantos and **Nila Vázquez** examine astronomical vocabulary in Chapter 7. CETA is shown to grow diachronically in lexical density as a consequence of the technical innovations that needed to be named. However, astronomical terminology does not follow this trend; in fact, its density regresses at the turn of the nineteenth century, suggesting that authors use more general vocabulary in the later periods. The use of terminology is most prolific at the

beginning and at the end of the corpus time line, while the middle period is less abundant and shares similar vocabulary with the earlier period.

Inés Lareo analyses complex predicates that have *make* as a collocative in the eighteenth-century part of CETA. The frequencies for complex predicates with *make* show a slight decline diachronically in the material. Geographically, Irish authors use significantly more complex predicates, and the specialised genres in the corpus generally contain more complex predicates. Additionally, simple verbs are favoured over the corresponding complex predicates, and general nouns are noted to be more common than specific vocabulary, as authors aim at disseminating their findings by using plain language.

In Chapter 9, **Bethany Gray** and **Douglas Biber** investigate the nominal discourse style by attesting prepositional phrases followed by non-finite *-ing* clauses (N+prep+V-*ing*). This pattern started to become recurrent after the sixteenth century and it is common in present-day scientific writing. In CETA, the feature increases in frequency, especially in the latter half of the eighteenth century. Further, *of*-phrases are clearly most common as noun post-modifiers, although *in-* and *for*-phrases increase more quickly. The *of*-construction occurs typically with abstract nouns (*method/ways of discovering*) and after a nominalised head noun that is derived from an adjective or a verb. The latter type increases in CETA after 1750, and the pattern is used to express epistemic stance (*impossibility of*).

Francisco Alonso-Almeida shows that hedging is common in the eighteenth-century part of CETA. One of the most prevailing types of hedges in the data are epistemic lexical verbs, and the verb *suppose* is frequently encountered. Within the other lexical hedges, the most frequent epistemic modal verb is *may* and also the epistemic adjective *apparent* and the noun *uncertainty* are recurrent. Hedging is further expressed non-lexically, and the most common strategy is to leave the source of information unnamed, as the authors refer to common knowledge. In general, the frequency of hedges grows during the period.

In the last chapter, **David Banks** analyses thematic structures by systemic-functional linguistics methods and notices that the structures are similar in texts aimed at professional and lay audiences. Topical themes are common in the data, and they usually function either as subjects or circumstantial adjuncts. Textual themes are also frequent, but they vary highly according to authors, while interpersonal themes are rare. Semantic analysis of the topical themes reveals that the themes mostly refer to the object of study and that mathematical themes are especially important in astronomy texts.

In general, the book offers a comprehensive account of astronomical writing during the Late Modern English period. The chapters also show how the cor-

pus can be used for linguistic analysis. Since the focus of the book is on linguistic studies, I found the corpus information to be slightly lacking; e.g. additional information on the different corpus categories, and an overview of the diachronic distribution of genres could be helpful for the corpus user.

The Corpus of English Texts on Astronomy is included in a CR-Rom. The corpus texts have been encoded by using Text Encoding Initiative conventions, and the corpus is accompanied by the Coruña Corpus tool. The tool has three main windows, i.e. a search window, tag search and an info window. The Coruña Corpus tool can be used for basic corpus searches and for the generation of wordlists. However, the corpus tool does not enable keyword or collocation analysis, which would be important, for instance, when comparing different corpora. As mentioned, the tool contains a function that allows one to search for tags, and there is also an info screen that shows the samples with a closer representation of the original layout, making it more convenient to read the corpus texts. The portrayal of the original graphemes and symbols in CETA is specific, since graphemes such as the long s and the italicised long s are presented in the corpus samples similarly to the original texts. The different graphemes do not affect searches since the corpus tool locates all of the different spelling variants at once, but still in the wordlist the spelling variants appear as two different words. In general, the use of the corpus is easy, and the instructions for the corpus tool are simple to follow.

The advantage of CETA is that in addition to running searches on the whole corpus or on a set of individually chosen texts, the corpus texts can be selected by ready-made parameters that can be applied to authors and texts. The parameters allow the corpus user to limit text selection, for example, by publication year, genre and the author's place of education. In addition, each text in the corpus contains a comprehensive metadata file that gives background information on the sample, i.e. the metadata lists details on the author, bibliographic data and description of the text. The possibility of selecting texts by different parameters is very useful, and the background information of the texts ensures that the corpus findings can be analysed within the socio-historical context.

The *Corpus of English Texts on Astronomy* and the *Coruña Corpus* are unquestionably valuable additions to the set of diachronic historical corpora. They offer a more varied selection of historical scientific texts for analysis than is available in the general historical corpora such as the *Lampeter Corpus*, the *Helsinki Corpus of English Texts* and ARCHER. In addition, CETA and the *Coruña Corpus* complement the *Corpus of Early English Medical Writing* that contains scientific writing on medicine from 1375 to 1800. All in all, CETA is an important addition to the study of historical scientific language.

Reference

Crespo García, Begoña and Isabel Moskowich-Spiegel Fandiño. 2010. CETA in the context of the Coruña Corpus. *Literary and Linguistic Computing* 25 (2): 153–164.

Joybrato Mukherjee and **Magnus Huber** (eds.). *Corpus linguistics and variation in English. Theory and description*. Amsterdam and New York: Rodopi, 2012. 258 pp. ISBN 978-90-420-3495-2. Reviewed by **Valentin Werner**, University of Bamberg.

This volume contains a range of papers originally presented at the 31st ICAME conference held at Gießen in 2010. They reflect the conference topic 'corpus linguistics and variation in English' and exemplify various approaches towards English and other languages that are based on or assisted by corpus data. A second focus of the contributions is the discussion of methodological issues and of the impact that corpus-based work may have on linguistic theories and neighbouring fields of research (and vice versa).

After the short introductory chapter by the editors, pointing out that – unlike in many other proceedings – papers are ordered alphabetically rather than thematically, the first paper by **Gisle Andersen** deals with active listenership through backchannelling devices in simulated business negotiations. His study is innovative in that it considers polylogue, i.e. interactions with more than two participants. It relies on transcribed video recordings from the NHH negotiation corpus, where two teams of students act in a fictitious buyer-seller situation. First, he illustrates the various pragmatic functions of the relevant items such as signalling understanding and agreement or support, and exemplifies them amply with corpus data. In the course of his analysis, he identifies specific usage of backchannelling devices in different stages of the interaction and draws attention to the fact that backchannelling is observable both within and across teams. Subsequently, Andersen shows by way of a quantitative analysis that at least one quarter of all tokens occurs team-internally and that active speakers are also active listeners.

The genre of blogs is the focus of attention in the contribution by **Marina Bondi** and **Corrado Seidenari**. They investigate evaluative phraseological expressions as the characteristic feature of this genre and, with the help of key-

word lists and successive collocational analyses, contrast samples of their English and Italian news blog corpus with reference corpora (BNC, CORIS/ CODIS). The authors find a comparatively high share of subjectivity markers (in the guise of the forms *I, my, I'm, am, mi, io, me*) and provide a list of typical collocational patterns. From a semantic point of view, two sequences with subjectivity markers are salient in the data from both languages, namely combinations with a cognitive marker (e.g. *I think, A me sembra*) or with an expression of evaluation (e.g. *I can't believe, Mi dispiace che*). While the paper contains an abundance of data, it could have been desirable to elaborate further on language-specific differences in section 4 and to describe the methodology more explicitly at times.

Doris R. Dant's paper stands at the interface of corpus and applied linguistics. Its findings are intended to lead to a refinement of the *Chicago Manual of Style*, whose prescriptions are tested against evidence from COCA. Dant's point of departure is the alleged arbitrariness of the rules contained in the style guide, which necessitates a comparison of these with actual usage rates in the corpus data. Accordingly, the paper offers tables with relative frequencies of variants for a wealth of items. Dant establishes that the majority of prescriptions reflect actual usage rates (e.g. for *depend on* vs. *upon*). Intriguingly, however, she also finds evidence that suggests a revision or abandonment of some prescriptions (e.g. *anxious about* vs. *to*). In addition, she argues that register differences have to be taken into account, as some variants are common in one register but rare in another (e.g. *sneaked* in academic writing vs. *snuck* in speech and fiction). Whether the results of Dant's investigation will eventually translate to editorial practice remains to be seen.

In the next paper, **Stefan Th. Gries** strongly advocates increased interaction between corpus linguistics and cognitive and psycholinguistic approaches. After outlining the corpus-based vs. corpus-driven controversy, dismissing the prevalent dogmatism in the field (with the odd polemic aside) and showing the importance of cognitive and psycholinguistic aspects in the study of language, he concludes that the ideal to be pursued is a "psycho- and cognitive-linguistically informed usage-based" (p. 49) approach to linguistics. The author continues with an impressive list of examples of findings from the three disciplines that are compatible with each other and indicates areas where corpus linguistics could benefit from theories and models developed and applied in the other two fields. In the following section, Gries details how corpus-linguistic findings may relate to (psycholinguistic) exemplar-based models and vice versa. Furthermore, he emphasises the theoretical added value of this exercise as well as the necessity of multidimensional approaches.

In their joint contribution, **Hans Martin Lehmann** and **Gerold Schneider** describe the application of syntactic parsing in the investigation of the dative-shift alternation. They base their analysis on a parsed 580 million word corpus of spoken and written British and American English from which approximately two million relevant tokens are automatically retrieved. In particular, they consider lexical types as a factor that determines variation and find that preference for double object or prepositional complement constructions is strongly determined by the individual lemma triplets and the occurrence of pronouns in the indirect object position. While the authors recognise potential weaknesses of their data-heavy, fully automated methodology, they assert that relying on large sets of parsed data can complement traditional variationist studies, especially with regard to rare items and the elimination of unwanted effects of highly frequent fixed expressions.

Michaela Mahlberg builds bridges to literary studies by discussing the potential and limitations of corpus stylistics, an increasingly popular methodology to enhance the description of characters in fictional texts. For that purpose she investigates clusters (i.e. repeated sequences of words) and suspensions (i.e. multi-word interruptions of direct speech or thought) containing reference to the body language of two characters in Charles Dickens's novel *Bleak House*. Mahlberg recommends a careful qualitative analysis of concordances to obtain results that would go unnoticed in a simple reading. However, findings should be related to other points in the text where the characters are described. The latter aspect is of particular importance, as she cautions against the danger of overuse, stating that "corpus stylistics methods [are] not a replacement for reading the novel" (p. 88). In addition, she notes that groups of clusters might be more apt for comprehensive characterizations and describes some areas (such as psycholinguistic research and literary criticism) where corpus stylistics might face challenges in the future.

Manfred Markus tackles the issue of using Wright's *English dialect dictionary* (EDD) as a corpus. After briefly outlining the history and contents of the EDD and introducing the functions of its electronic version and its potential for linguistic investigations, he discusses how well the EDD and its citations section fulfil corpus criteria and why Wright's sources are particularly valuable for historical linguists. Unfortunately, Markus's explanation of the complex layout of the EDD Online data is somewhat hampered by the fact that one of the relevant figures (p. 105) is hardly readable. Nevertheless, he succeeds in indicating the various ways in which the EDD may be applied in systematic studies of dialectal spelling variants, word-formation patterns, phonotactics, etc. as well as in pragmatic and sociolinguistic analyses.

By way of a case study on Bislama, a Pacific creole, **Miriam Meyerhoff** surveys language contact phenomena and illustrates how small corpora can assist in variationist sociolinguistic research. The author first familiarises the reader with some general principles of such research as well as with the socio-linguistic context of Bislama. She continues with the results of her multivariate case study of the presence or absence of overt arguments in subject and object position as a feature of substrate transfer from Tamambo. It emerges that transfer is observable to different degrees and that constraints are ranked differently in the languages in contact ("transformation under transfer", p. 126). Subsequently, Meyerhoff relates her findings to a contact study of migrants' acquisition of English, where social constraints also undergo transformation and contextualises her overall findings with some general notes on transformation and avenues for future research.

Hagen Peukert explores how statistical procedures can be used to predict two registers, English child-directed speech (CDS) and adult language. In an extended methodological section he underpins his choice for a logistic regression model and highlights the potential pitfalls in studies of a similar type. The model he develops is based on a random sample of n-grams of phonemes, which are fundamentally different in the two registers, and succeeds in correctly predicting CDS and adult language in more than 95 per cent of cases. The author then refutes in advance a range of possible objections to his approach and to the use of computational methods in linguistic studies, critically reflecting at the same time on related psycholinguistic issues.

The contribution by **Thomas Proisl** is a work-in-progress software presentation of the Pareidoscope, an innovative web-based tool for studying the lexis-grammar interface. The author emphasises the importance of investigating lexicogrammatical patterns (i.e. constructions or structures) and begins with an overview of collostructional analysis, the general principle behind the tool. He continues with a description of the Pareidoscope 'in action', first illustrating possible types of queries and then reporting the findings of two case studies on structural and lexical associations of *feeling(s)*. Next, Proisl indicates some avenues for further improvement of the tool, some of which (e.g. complex queries at the token and chunk level) have since been implemented.

Constructions of another type are explored by **Paula Rodríguez-Abruñei-ras** in a diachronic study of exemplifying markers (EMs). She considers the grammaticalization of *for instance/example* and provides evidence for their etymology and first occurrence from the OED. Subsequently, she draws on various corpus resources (ARCHER, Helsinki Corpus, LOB, FLOB) to test the accuracy of the OED entries and to outline the further development of the items in ques-

tion. She finds a constantly increasing frequency both for the lexical units *instance/example* and for their relative frequency as EMs, eventually outweighing their nominal use. In a qualitative section, the author then discusses the notion of bridging context and exemplifies variants of the EMs and instances where they combine with other EMs (such as *like* or *as*).

Patricia Ronan conducts a corpus-based study of pragmatic uses of *would* in Irish English. Using a spoken and written subsample of ICE-IRL, she finds *would* to be considerably more frequent in speech. After discussing other semantic contexts where it is used, the author concentrates on settings where *would* functions as politeness marker/softener. Ronan shows that the expression of tentativeness is by no means restricted to speech and thus hypothesises that a "pragmatic habit" (p. 169) allegedly only observable in speech has transferred to writing. A comparison with Outer Circle varieties, where *would* seemingly fulfils a similar purpose, reveals some intricate pragmatic differences relating to the type of modality (dynamic vs. deontic) conveyed.

Another diachronic investigation is presented by **Juhani Rudanko**, who examines the emergence of the transitive *out of -ing* pattern on the basis of 126 tokens from COHA. The data yield no sizeable token counts before the 1930s, but show that the pattern can be viewed as established from the 1950s onwards and most often occurs with the verb *talk*. The author proposes that the pattern should be analysed as a "caused motion construction" (p. 179) and illustrates the applicability of a construction grammar approach for the interpretation of his examples, simultaneously showing that the pattern may also appear with genuinely intransitive verbs.

Edgar W. Schneider's paper is located at the intersection of dialectology and historical sociolinguistics and demonstrates the usefulness of corpora of vernacular varieties, exemplified with the help of corpora of black and white Southern US English. Schneider discusses the issue of obtaining reliable vernacular data and concludes that semi-literate letters are comparatively well suited for the task. However, he advises great methodological care (e.g. as to spelling variants) when dealing with textual material of this kind. After outlining some on-going research debates in studies of US vernacular data, he introduces the *Regensburg Corpora of American English Dialects* (ReCAmED) and illustrates their research potential with a number of case studies. His analyses reveal that black and white Southern dialects share a number of structural characteristics and that some of these can be dated as having emerged post-Civil War (which ties in with earlier studies), but that systematic differences along ethnic lines are not in evidence.

The discourse of television news and methodological issues regarding the compilation of a corpus of this discourse type are the focus of **Stefania Spina**'s paper. It starts with the observation that corpora and studies of television language are rare and subsequently describes the structure of the *Corpus of Italian Television Language*. Spina points out some characteristics of the genre of television news, such as its hybrid nature with regard to the spoken versus written dimension and its 'constructedness'. She next lists the various sub-levels of the genre (headline, interview, etc.) and exemplifies them in relation to the corpus annotation. When comparing the different sub-levels, notable differences emerge as to the proportions of parts-of-speech categories and sentence length.

The volume also includes a contribution by **Michael Stubbs**, who calls for a functional theory that is able to integrate findings from a range of lexical studies of textual organization and evaluation. His point of departure is the fundamental difference between texts as "real language events" (p. 223) and abstracted corpora. The chapter revolves around three short (corpus-based) analyses of textual functions of lexis: contribution of word distribution to textual structure, influence of word classes on textual cohesion and segmentation of discourse through phrasal units. Stubbs shows that, if the analyses remain disconnected and their findings are not related back to the texts, all three types of analysis fall short of comprehensively establishing what constitutes textual cohesion.

Elizabeth Closs Traugott takes a look at the importance of corpus-based work for the study of persistent ambiguous contexts in grammaticalization. She first introduces a functional definition of grammaticalization and further elaborates on the relevant notions of persistence and particularly context. As to the latter, she argues for the importance of taking account of "both the semantic and morphosyntactic contexts that appear to have enabled grammaticalization" (p. 236) even after a construction has started to conventionalise. Based on Early Modern English data from the Helsinki Corpus and the Old Bailey Proceedings, the author substantiates her claim with examples of BE *going to* constructions where the motion-with-purpose meaning might be persistent along with the futurate use. She convincingly shows that the larger context around the construction must be considered in order to fully grasp micro-changes in grammaticalization, while language-specific differences also need to be taken into account.

The volume concludes with a register study of premodifying *-ing* participles in the BNC by **Turo Vartiainen** and **Jefrey Lijffijt**. Their main claim is that these participles can be categorised into the classes 'adjectival' and 'verbal' according to morphosyntactic properties, which also has vital repercussions for

corpus annotation in terms of part-of-speech tagging and parsing, two layers of analysis that should be kept apart. They next outline their search methodology, which yields 3,434 tokens from the registers academic prose, conversations, fiction and newspapers. They find that register impacts both the frequency of *-ing* participles in general and the relative frequencies of the verbal and adjectival categories in particular, thus putting some previous analyses into perspective. However, no tests for the statistical significance of these findings are reported.

While all papers offer interesting insights and are highly readable, the formal editing leaves some room for improvement. As indicated above, some of the figures are grainy (e.g. on p. 105) or too small (e.g. on p. 147), resulting in the loss of important detail. There are also a number of typos, among them *i* (read *I*; p. 24), *dot* (read *not*; p. 25), *50%–25%* (read *50%–75%*; p. 31), *Mark Davis* (read *Davies*; p. 176) and *Pergunia* (read *Perugia*; p. 221). Other minor inconsistencies are the unformatted section heading for the summary section (p. 209) and the abbreviation of the Proceedings of the Old Bailey Corpus as both *OBP* (p. 238) and *POB* (p. 236ff.) within the same chapter.

These minor flaws aside, the volume can be seen as a state-of-the-art compendium for English corpus linguists. In particular, the contributions by the plenary speakers, Stefan Th. Gries, Michaela Mahlberg, Miriam Meyerhoff, Edgar W. Schneider and Elizabeth Closs Traugott epitomise the potential of corpus linguistics in all its breadth and depth and strongly encourage researchers to cross boundaries to neighbouring fields in the future. Broad trends emerging from the individual chapters include the increasing reliance on the analysis of constructions and the application of construction grammar. A second (seemingly trivial, but in fact substantial) trend is the need for both general and specialised corpora and the relation of quantitative and qualitative findings to linguistic theory in order to achieve a sound diachronic and synchronic description of English (and other languages, of course). Readers who may have wondered why contributions on non-native Englishes did not find their way into this volume at all are referred to a forthcoming second conference volume in the eVarieng series.

Doris Schönefeld (ed.). *Converging evidence: Methodological and theoretical issues for linguistic research* (Human Cognitive Processing 33). Amsterdam and Philadelphia: John Benjamins, 2011. 352 pp. ISBN: 978-90-272-2387-6. Reviewed by **Gaëtanelle Gilquin**, FNRS – Université catholique de Louvain.

As the title suggests, *Converging evidence: Methodological and theoretical issues for linguistic research* deals with the different aspects that are involved in collecting, analysing and interpreting converging evidence, that is "evidence or findings that converge on one and the same conclusion" (p. 1), in the field of linguistics. Edited by Doris Schönefeld, the volume brings together thirteen papers which were presented at the *Third International Conference of the German Cognitive Linguistics Association* that took place at Leipzig University in September 2008. It is made up of two introductory chapters and three main parts (covering multi-methodological approaches to (i) constructional and idiomatic meaning, (ii) language acquisition, and (iii) the study of discourse), plus a preface and an author/subject index.

In the first introductory chapter (the main introduction to the volume), **Doris Schönefeld** provides a useful overview of the (empirical and non-empirical) methods available to study language, with a focus on intuition/introspection, corpora and experimental data. After commenting on the role of evidence and the type of evidence used in generative linguistics and cognitive linguistics, she makes a convincing case for the need to provide converging evidence. In particular, she argues for an 'empirical cycle', in which hypotheses formulated on the basis of observation help explain the phenomenon under investigation but at the same time lead to further hypotheses which can be tested, and she claims that such a cycle can only benefit from the exploitation of different types of evidence. The chapter ends with a brief presentation of the contributions to the volume. The second introductory chapter, **Gerard J. Steen**'s paper, while devoted to the topic of metaphor and the question of whether metaphor is always a matter of thought, also tackles more general issues having to do with converging evidence. An interesting distinction is made between phenomenological pluralism, which considers different types of evidence drawn from distinct areas of research and corresponding, in effect, to distinct (though related) phenomena of investigation, and methodological pluralism, which by contrast limits itself to different types of evidence drawn from one area of research and, therefore, pertaining to the same phenomenon. Steen illustrates the problematic character of phenomenological pluralism by highlighting the confusion that often exists between metaphor-as-thought in abstract conceptual structure (semiotic

approach) and metaphor-as-thought in actual cognition (psychological approach). This leads him to conclude that "[p]henomenological pluralism is exciting but problematic" and "[m]ethodological pluralism is exciting and attractive" (p. 46) and to suggest that cognitive linguists should look for converging evidence within their own area of research.

The bulk of the papers belong to the first part of the volume, entitled "Multi-methodological approaches to constructional and idiomatic meaning" (pp. 55–246). Constructional meaning is discussed in five chapters, starting with **Thomas Egan**'s analysis of the '*see* x *to be* y' construction. Using data from three large corpora, namely the *British National Corpus*, the *Corpus of Contemporary American English* and WebCorp, Egan seeks to test Bolinger's (1974) intuition-based claim that perceptual verbs like *see* express a conceptual meaning when used with a *to*-infinitive complement (e.g. *I see this to be the next logical step*). As against Noël (2003), who challenges Bolinger's view by showing that *to*-infinitives are also compatible with a perceptual meaning, Egan provides converging evidence that *to*-infinitive complements cause a semantic shift from the perceptual to the conceptual. Out of over 500 (active) instances of the '*see* x *to be* y' construction, only 43 involve perception, and of these, merely three express perception pure and simple; the others are all characterised by the presence of (explicit or implicit) alternatives, a feature that is related to the general schematic interpretation of the *to*-infinitive. The next chapter, by **Gunther Kaltenböck**, looks into clause-initial *I think* followed by a complement clause. In an attempt to determine the syntactic status of this complement-taking predicate (main clause or comment clause), Kaltenböck considers syntactic and prosodic evidence, and convergence (or lack thereof) between the two in naturally occurring data from the British component of the International Corpus of English (ICE-GB). While syntactically a distinction may arguably be drawn between *I think* followed by a *that*-complementiser (which suggests a main clause interpretation of *I think*) and *I think* with no complementiser (which suggests a comment clause interpretation), in terms of prosody both structural realisations of the complement-taking predicate (*I think* Ø and *I think that*) favour reduced prominence, which is an indicator of a comment clause status. To account for this divergence, Kaltenböck puts forward a dynamic model of grammar which claims, among others things, that the *that*-complementiser now functions more as a filler than as a marker of syntactic hierarchy. **Silke Höche** offers a comparison of two constructions that are often described as nearly synonymous, namely *be about to* V and *be going to* V. In an attempt to examine both the diachronic and synchronic dimensions of the constructions, Höche uses the quotation database of the *Oxford English dictionary* as well as some historical corpora (for the

diachronic approach) and data from the *British National Corpus* (for the synchronic approach). Used in conjunction with the author's and (past and present) grammarians' intuitions, this corpus evidence reveals certain differences between the two constructions. The main difference, which is confirmed by a collostructional analysis of the verb slot and an examination of the temporal adverbials occurring in the constructions, is that *be about to* V has developed towards an aspectualising construction, being closer to aspectualisers like *start to* or *begin to* than to futurate-forms like *will* or *be going to*. *Be going to* V, by contrast, exhibits a lower degree of immediacy and is more flexible with respect to the verbs and adverbs with which it associates. The chapter by **Stefan Th. Gries** deals with the English dative alternation, which is used as a case study to compare the efficiency of three statistical methods for the study of syntactic priming in corpora, i.e. "the tendency of speakers to re-use syntactic patterns they have recently comprehended or produced" (p. 143). Relying on a database of almost 3,000 prime-target pairs extracted from ICE-GB, Gries performs three types of analyses at different levels of granularity: at a coarse level of granularity, a simple cross-tabulation of the constructional frequencies in primes and targets; at an intermediate level of granularity, a binary logistic regression which takes into account fixed effects (like the construction of the prime or the distance between prime and target) and their interactions; and at the highest level of granularity, a generalised linear mixed-effects model (GLMEM) which in addition includes random effects (here, the verb lemma of each target and the name of the file in which prime and target occurred). The GLMEM turns out to be the most powerful technique, i.e. the one with the highest classification accuracy, but also the technique that is most in line with cognitive linguists' theoretical assumptions. Furthermore, the study shows that corpora provide evidence for syntactic priming that is convergent with (and sometimes additional to) evidence gathered through experimental work. Like Höche and Gries, **Ad Backus** and **Maria Mos** demonstrate the value of converging evidence through the comparison of two functionally similar constructions. The constructions under study are two constructions in Dutch expressing potentiality, namely V-*baar*, where -*baar* is a suffix added to verb stems to form adjectives (e.g. *breekbaar*, 'breakable'), and *is te* V, a combination of a finite copular verb and an infinitive preceded by the *te* ('to') infinitive marker (e.g. *... is moeilijk te vinden*, '... is hard to find'). Using data from a corpus of spoken Dutch (*Corpus Gesproken Nederlands*) and applying the techniques of collexeme analysis and distinctive collexeme analysis, the authors highlight some of the similarities and differences that exist between the two constructions, among which their (apparent) productivity, a shared preference for transitive verbs, but also a tendency to express assessed

potentiality (for the *is te* V construction) vs. factual potentiality (for the V-*baar* construction). In an attempt to approach the mental representations of these constructions, Backus and Mos then conduct a magnitude estimation task in which participants are asked to assess the acceptability of constructed test sentences with V-*baar* and *is te* V. The results of the experiment largely confirm the findings from the corpus study, which pleads in favour of the "productivity and psychological reality of the constructions" (p. 165).

The next two chapters still belong to the first part of the volume, but instead of constructional meaning they tackle idiomatic meaning. **Alexander Ziem** and **Sven Staffeldt** investigate German somatisms, that is multi-word idiomatic expressions that include a body-part term, with a focus on somatisms that contain the word *Finger*. They propose a two-step procedure to determine the different meanings of somatisms which involves (i) the identification of the meaning aspects that are shared by all uses of the somatism (typically, the argument roles that inherently belong to the verb's valency) and (ii) the identification of meaning variation (as reflected in the slots that do not belong to the verb's valency). This procedure is illustrated by means of a corpus-based analysis of the expression *jm. auf die Finger schauen* ('keep an eye on sb'). Through this analysis, the authors also address the issues of compositionality (is the meaning of somatisms motivated by the semantic potential of the body-part term?) and embodiment (is the meaning of somatisms grounded in human experience?). They demonstrate, among other things, that the term *Finger* in the expression *jm. auf die Finger schauen* refers to some sort of activity, thus exploiting the conceptual metonymy FINGER FOR ACTIVITY. In the following chapter, **Susanne R. Borgwaldt** and **Réka Benczes** set out to compare the word formation strategies that are used by German and Hungarian speakers to refer to novel hybrid objects composed of two identifiable parts (e.g. an animal that is half penguin, half cat) and novel objects with a salient shape (e.g. a lake with the shape of a heart). The results of a novel object naming task, in which participants are asked to name a number of digitally manipulated pictures, show that German speakers predominantly opt for noun-noun compounds (92 per cent, e.g. *Auberginenlatscher*, 'eggplant slippers'), whereas Hungarian speakers' choices display more variety, with 60 per cent of noun-noun compounds (e.g. *kaktusz béka*, 'cactus frog'), 28 per cent of adjective-noun compounds (e.g. *repülő ananász*, 'flying pineapple' for a pineapple bird) and five per cent of blends (e.g. *krokogáj*, a blend of *krokodil*, 'crocodile', and *papagáj*, 'parrot'). This distributional pattern does not seem to be influenced by the semantic domain of the two entities involved in the novel objects (which were any combinations of animals, plants and inanimate objects).

The authors also emphasise the importance of metonymy in word formation, as well as the role played by creativity and humorous effects.

The papers in the second part of the volume look into language acquisition (pp. 247–316). **Rasmus Steinkrauss** is interested in the L1 acquisition of German *was ... für* ('what kind of…') questions and the factors that influence the order of acquisition and the frequency of production of such structures. Steinkrauss investigates a dataset representative of the production of a German-learning boy between the ages of two and three and of the ambient speech, and shows that, as predicted by a usage-based approach, input frequency plays a role in the child's formation of schemas (i.e. mental representations), and hence his production of *was ... für* questions. At the same time, however, the analysis reveals a number of discrepancies between the predictions based on input frequency and the observations made in the production data. These discrepancies are explained by additional factors such as the presence of supporting constructions (which the child has already acquired and which favour the use of the target structure) or functionally equivalent constructions that are more frequent in the input and thus more likely to be reproduced by the child. The paper ends with the author advocating for the simultaneous consideration of several factors. **Silke Brandt** and **Evan Kidd** deal with the L1 acquisition of relative clauses by English- and German-speaking children. Like Steinkrauss, they take into account input frequency to demonstrate the influence of linguistic experience on children's representations of complex constructions. Since corpus data reveal that object relative clauses are more likely to be attached to an inanimate head NP and to contain a pronominal subject (as in *the ball that he just threw*), children are expected to perform better on experimental tasks that include relative clauses formed according to these constraints. This is confirmed by a sentence-repetition task, in which children are asked to imitate several types of relative clauses, and a referential-choice task, in which they have to select the referent (as represented by a small toy) corresponding to the relative clause uttered by the experimenter. In these tasks, children appear to show better comprehension and production of object relative clauses with inanimate heads and/or pronominal subjects. Contrary to the findings of previous experimental studies that did not consider input frequency, no systematic difference was observed between children's processing of subject vs. object relative clauses. Unlike the preceding two chapters which are concerned with L1 acquisition, the chapter by **Nina Reshöft** is concerned with L2 acquisition. The paper relies on Talmy's (1985) typology of motion events, which distinguishes between satellite-framed and verb-framed languages. Whereas the former typically encode manner of motion through the verb and path of motion through satellites (e.g. *run into the house*),

in the latter the main verb expresses the path of motion while other aspects of movement, including the manner of motion, are expressed by means of adjunct phrases (e.g. French *entrer dans la maison en courant*, 'enter the house running'). Reshöft aims to determine whether EFL learners' lexicalisation patterns for describing motion events are influenced by their mother tongue. With this aim in view, she analyses data from the *International Corpus of Learner English* produced by French, Italian and Spanish-speaking learners of English (i.e. with a verb-framed L1) and compares them to data from the *British National Corpus* produced by native speakers of English (i.e. with a satellite-framed L1). Her study shows that the learners tend to transfer the lexicalisation patterns of their L1 to the L2, which manifests itself, most notably, by the low frequency and weak diversity of manner verbs in their EFL production, or the small number of paths when they describe motion events.

In the last chapter, which also corresponds to the third and final part of the volume (pp. 317–348), **Anke Beger** investigates the metaphorical conceptualisations of experts in psychology and laypersons to refer to anger, love and sadness in English. The study compares performance data from psychology guides from the Internet, where people can obtain expert advice about their emotional problems, and elicited data from interviews with randomly chosen people on the topic of emotions. The combination of the two methodologies makes it possible to confirm the initial hypothesis according to which experts and laypersons draw on (partly) different conceptual metaphors when describing emotions, with a tendency among experts to use metaphors as analytical tools that allow for therapeutic reframing (e.g. by claiming that ANGER IS A VEIL, which suggests that the cause of the problem is to be found in some unexposed emotions). At the same time, it leads to a refinement of the hypothesis, as it appears that laypersons may present different conceptualisations depending on their level of emotional involvement; thus, the metaphor LOVE IS A STRUCTURED OBJECT (as exemplified by *building and maintaining healthy relationships*) hardly ever occurs among the laypersons seeking advice on the Internet (who are presumably troubled by the emotion), but is more common among the interviewees who are invited to talk about love outside the discourse of counselling.

While all the chapters adopt a cognitively-oriented perspective, the book will be of interest to corpus linguists for several reasons. The first one has to do with the obvious link that exists between the cognitive usage-based approach and the framework of corpus linguistics (Schönefeld refers to corpus-linguistic methodology as "the most likely form of usage-based approaches" (p. 18); see also Gries and Stefanowitsch 2006, Stefanowitsch and Gries 2006 or Grondelaers *et al*. 2007, among others). In fact, this book and the studies it contains

should encourage researchers to provide corpus linguistics with the theoretical foundation that it is often said to be lacking (see, e.g., Grabe and Kaplan 1996: 46), since indications are given as to how corpus analysis can be combined with theory. Another reason why this book should appeal to corpus linguists is that the focus is on converging evidence "in practice" (p. 24), with most of the papers relying on corpus data and dealing with topics that are dear to corpus linguists such as phraseology or language acquisition. As for the theoretical issues that do arise, they are always clearly defined and explained in a way that is also accessible to non-cognitivists. What adds to the reader-friendliness of the volume is its great internal coherence. It includes a number of cross-references, especially to the two introductory chapters, from which the authors borrow a number of key concepts, such as 'converging evidence' (p. 1ff.), 'phenomenological pluralism' (p. 40ff.) or 'empirical cycle' (p. 9ff.). Paradoxically, however, this feature may at times become a weakness. This is particularly true for the idea of converging evidence. While some papers are perfect illustrations of the multi-methodological approach (for instance, Backus and Mos's combination of corpus analysis and magnitude estimation), others are less representative examples, limiting themselves, for example, to the comparison of empirical data from one corpus with intuition-based claims from the literature or hypotheses formulated by the authors introspectively, which one could see as no more than the application of a traditional corpus-based (top-down) method. As a result, the term 'converging (or diverging) evidence', which all the authors use at least once, tends to become some sort of empty shell (or umbrella term, to put it more positively), covering the testing of hypotheses or claims from previous research (Reshöft), the combination of quantitative and qualitative analysis (Steinkrauss), the use of several corpora (Höche) or the comparison of two languages (Borgwaldt and Benczes). In this respect, some authors' efforts to apply the concept of converging evidence to their own study seem slightly artificial, as illustrated by Steinkrauss who refers to the challenge of having to provide converging evidence "when just one corpus of one speaker is analyzed – a classical case study – and when neither other corpora nor another methodology may be used to gain further evidence" (pp. 253–254). More generally, one may regret that the volume does not exemplify a wider variety of (combinations of) methods, especially from the experimental side (in this regard, Borgwaldt and Benczes's novel object naming task is probably the most experimental approach one can find in the book), and that some of the studies have a relatively thin empirical basis (e.g. the spoken production of one single child, 67 occurrences of a phraseological expression or 43 relevant tokens of a syntactic construction). This, however, is mainly a reflection of the nature of the event from which the book originated,

namely a conference of linguists (not psycholinguists) working within the framework of cognitive linguistics, a theoretical paradigm that has only recently started to employ authentic empirical evidence. Furthermore, none of the above criticisms should detract from the quality of the book and the editor's merit in having brought these papers together and invested time in this project (as is apparent from the authors' acknowledgements, many of which explicitly refer to Doris Schönefeld). Not only does the resulting work show how data from different sources (and, sometimes, different domains) can be combined with each other in linguistic research, but it also demonstrates that the convergence (or divergence, for that matter) of evidence makes it possible to go one step further by confirming results obtained with a different methodology, refining a hypothesis which can then be tested again against the same or different data, correcting claims from the literature or theoretical models developed on the basis of intuition, providing new insights into certain linguistic phenomena that could not have been produced by a single source of evidence, or more simply expanding our understanding of the relations between different types of data and different methodologies (see also Gilquin and Gries 2009). Hopefully, this selection of studies will support a movement, increasingly noticeable in more recent linguistic research, away from methodological monism and towards triangulation of methods, an approach that should be facilitated by today's easier access to data and tools to analyse them, as well as closer collaborations between linguists and researchers from other fields.

References

Bolinger, Dwight L. 1974. Concept and percept: Two infinitive constructions and their vicissitudes. In N. O. Gakkai (ed.). *World papers in phonetics: Festschrift for Dr. Onishi's Kiju*, 65–91. Tokyo: Phonetic Society of Japan.

Gilquin, Gaëtanelle and Stefan Th. Gries. 2009. Corpora and experimental methods: A state-of-the-art review. *Corpus Linguistics and Linguistic Theory* 5 (1): 1–26.

Grabe, William and Robert B. Kaplan. 1996. *Theory and practice of writing. An applied linguistic perspective*. London & New York: Longman.

Gries, Stefan Th. and Anatol Stefanowitsch (eds.). 2006. *Corpora in cognitive linguistics: Corpus-based approaches to syntax and lexis*. Berlin: Mouton de Gruyter.

Grondelaers, Stefan, Dirk Geeraerts and Dirk Speelman. 2007. A case for a cognitive corpus linguistics. In M. Gonzalez-Marquez, I. Mittelberg, S. Coulson and M. J. Spivey (eds.). *Methods in cognitive linguistics*, 149–169. Amsterdam: John Benjamins.

Noël, Dirk. 2003. Is there semantics in all syntax? The case of accusative and infinitive constructions vs. *that*-clauses. In G. Rohdenburg and B. Mondorf (eds.). *Determinants of grammatical variation in English*, 347–377. Berlin: Mouton de Gruyter.

Stefanowitsch, Anatol and Stefan Th. Gries (eds.). 2006. *Corpus-based approaches to metaphor and metonymy*. Berlin: Mouton de Gruyter.

Talmy, Leonard. 1985. Lexicalization patterns: Semantic structure in lexical forms. In T. Shopen (ed.). *Language typology and semantic description. Volume 3: Grammatical categories and the lexicon*, 57–149. Cambridge: Cambridge University Press.

An Van linden. *Modal adjectives: English deontic and evaluative constructions in synchrony and diachrony* (Topics in English linguistics 75). Berlin and Boston: De Gruyter Mouton, 2012. 363 pp. ISBN 978-3-11-025293-4. Reviewed by **Henrik Kaatari**, Uppsala University.

In this book, An Van linden tackles the concept of modality by looking at the under-researched category of adjectives. She does so by looking at adjectives occurring with extraposed clausal complements, represented by *that*- and *to*-clauses. The analysis is based on a set of modal-evaluative adjectives that express dynamic modality, deontic modality and non-modal evaluation. Van linden shows, by looking at both diachronic and synchronic data, that the development of these three different types of meanings follows specific paths that can be linked to the hierarchical nesting of these meanings in a novel conceptual map.

Chapter 2 serves as the back-bone of the book as it leads up to the introduction of a conceptual map (see p. 68 for an excellent figure) upon which all the subsequent analyses hinge. Most importantly, the conceptual map ties together different concepts related to the three types of meanings covered in the study.

First, the most fundamental distinction in the conceptual map is between dynamic modality, deontic modality and non-modal evaluation. The dynamic

meaning expressed by the modal-evaluative adjectives included in the study is that of situational dynamic modality, represented by adjectives that can be paraphrased with 'possible' or 'necessary'. Deontic modality is defined as involving "the assessment of the degree of desirability of a virtual or tenseless SoA [State of Affairs], whose realization is by default in the future, by some attitudinal source" (p. 62). Two things are important in this definition. First, deontic and non-modal meanings are distinguished from dynamic meaning, as dynamic meaning is non-attitudinal whereas deontic and non-modal evaluative meanings have an attitudinal source. Second, dynamic and deontic meanings are distinguished from non-modal evaluation by virtue of 'factuality status'. Both dynamic and deontic meaning are "undetermined with respect to their factuality status, i.e. they are non-factual" (p. 64) and the clausal complements associated with these meanings are termed 'mandative' complements. Non-modal evaluative expressions, on the other hand, are inherently factual, i.e. presented as either positively or negatively factual and are in turn termed 'propositional' complements. These two attributes, the presence/absence of an attitudinal source and the factuality status, are thus essential to the distinction between the three different meanings (see figure on p. 68).

Second, a distinction is drawn between strong and weak adjectives. Strong adjectives (*critical, crucial, essential, indispensable, necessary, needful, vital*) are distinguished from weak adjectives (*appropriate, convenient, desirable, expedient, fit, fitting, good, important, profitable, proper, suitable*) on the basis that strong adjectives express a stronger degree of desirability than weak adjectives. Strong and weak adjectives are also different as regards the degree modifiers they combine with. Strong adjectives are conceived of as bounded, and as such they combine with totality modifiers (*absolutely critical*) whereas weak adjectives are conceived of as unbounded, and as such they combine with scalar degree modifiers (*very important*). The distinction between strong and weak adjectives correlates with the factuality status of different utterances, as strong adjectives are restricted to constructions with mandative complements (i.e. non-factual dynamic and deontic modality) whereas weak adjectives can occur in constructions with both mandative and propositional complements (i.e. deontic modality and non-modal evaluation) (see figure on p. 72).

The diachronic analyses are presented in Chapters 4–6 (Chapter 3 gives a short introduction to the data and methods used) and include data from Old English up to present-day English. Chapter 4 includes four case studies of the diachronic development of four strong adjectives (*critical, crucial, essential, vital*). Van linden identifies two semantic properties that are necessary for the development of dynamic and deontic modality: relational meaning and potenti-

ality. These two semantic properties are linked to each other as relational meaning is a prerequisite for the development of potentiality. Relational meaning is also the semantic condition for complementation to take place as relational meaning serves to relate two concepts. This semantic extension can be manifested by co-occurrence with nouns that denote a part-whole relationship (e.g. *part, property* etc.) or by metaphorical projection and metonymy. Potentiality is required for the development of situational dynamic meaning and potentiality is thus required for entry into the modal-evaluative domain at the lowest level of the conceptual map. In the next phase, the development of deontic modality is facilitated through subjectification, as deontic utterances need an attitudinal source. There is thus an upward development on the conceptual map from dynamic modality to deontic modality and the adjectives are at this point modally polysemous as, when complemented by extraposed *that-* and *to-*clauses, they can express both dynamic and deontic meanings. Two of these four adjectives, *essential* and *crucial* are even found in the highest position on the conceptual map as they, albeit infrequently, are found to express non-modal evaluation. The semantic development of these adjectives thus shows the diachronic applicability of the conceptual map, as all adjectives move upwards across time (i.e. dynamic -> deontic -> (non-modal evaluative)).

Chapter 5 is mainly concerned with the two types of clausal complements and their distribution diachronically. First of all, the distribution of mandative and propositional complements confirms that the conceptual map holds across time, as from Old English onwards strong adjectives are only found with mandative complements whereas weak adjectives are found with both mandative and propositional complements. Regarding the distribution of *that-* and *to-*clauses Van linden finds that with *that-*clauses, the subjunctive became restricted to mandative complements in Middle English and gradually lost in frequency. This is directly related to a rise of the *to-*infinitive, as subjunctive *that-*clauses were in competition with *to-*infinitives and, starting in Middle English, the *to-*infinitive gained in frequency at the expense of *that-*clauses. Van linden explains this development in terms of analogy with verbal matrices as "after intention verbs and manipulative verbs, the *to-*infinitive started to replace the *that-*clause more abruptly and extremely than was found with adjectival mandative constructions" (p. 195).

Chapter 6 concludes the diachronic analysis with a focus on the construction wholes of matrix and complement. By comparing a set of importance adjectives to a set of appropriateness adjectives, Van linden highlights two paths leading the way in the development of propositional complements from mandative complements, again illustrating the validity of the conceptual map. In other

words, non-modal evaluation developed out of deontic modality along two different paths. With the importance adjectives (*essential*, *crucial, important*) propositional complements first showed up in a combined pattern of complementation as secondary complements, typically realized by a *that*-clause following a primary mandative *to*-clause complement as in "*it is important to ensure that you go to someone who is properly trained and experienced*" (p. 215). There was thus a situation in which propositional complements started out as secondary complements in the combined pattern and from there came to be realized as complements in non-combined structures. It is, however, important to note that the combined pattern is still the most frequent environment for importance adjectives with propositional complements. The appropriateness adjectives (*appropriate, fitting, proper*) followed a different path via bridging contexts in which the complements are ambiguous between a mandative and propositional reading. Van linden finds that these bridging contexts cover quite some time before unambiguous propositional complements are found with the appropriateness adjectives, suggesting that the transitional use was productive and paved the way for propositional complements (see the figure on p. 235 for a clear summary of the development of the two classes of adjectives).

Chapters 7 and 8 cover the synchronic part of the study. Chapter 7 gives a short introduction to the data and methods used for the synchronic analysis. Chapter 8 focuses on the constructional wholes of matrix and complements but from a synchronic perspective. The chapter aims at substantiating the synchronic validity of the conceptual map by looking at the three types of meanings and their synchronic relationship. Most importantly, Van linden shows that the three meanings can be subdivided, arguing that some formally distinct constructions constitute partially filled constructions, i.e. "whose meaning is not fully predictable from the component parts" (p. 264). Van linden finds that non-modal evaluative meaning can be subdivided into four specific subtypes, two of which are seen as partially filled constructions. Thus the locative pattern, in which a closed set of adjectives are found complemented by a *to*-clause followed by a locative or associative phrase as in "*it's nice to be here*" (p. 266), is considered a partially filled construction "in which a specific constructional make-up is paired with a particular meaning that cannot be compositionally derived" (p. 266). The same is the case for the 'acquisition of knowledge' construction expressing non-modal evaluative meaning and the mental focus construction expressing deontic modality. Chapter 8 also problematizes the distinction between dynamic and deontic modality leading up to a final synchronic refinement of the conceptual map, presented in two illustrative figures (pp. 312–313). A final conclusion of the book is given in Chapter 9.

In this book, Van linden clearly illustrates the benefit of incorporating both diachronic and synchronic data. The diachronic data explains the emergence of different meanings across time and is crucial for understanding the recent, and ongoing, change of these meanings in present-day English. The most important contribution made by this book is its focus on adjectives and adjectival complementation; as noted by the author, adjectives constitute an under-researched category in the modal domain. In fact, the term 'modal adjectives' (used in the title of this book) seems to me not to have much traction in linguistics (but see Rivière 1983 for an exception) and is as such most welcome, as it illustrates the rightful place of adjectives in the description of modality. Van linden also manages to show why the study of adjectives is important for the understanding of the workings of modality. This is more generally manifested in her novel conceptual map and most clearly in Chapter 4 which tracks the premodal stages of adjectives, showing how the modal meanings developed in the first placed and identifying the necessary semantic properties involved, in this case relationality and potentiality.

A couple of questions can be raised regarding the choice of constructions and adjectives included in the study. One of these is mentioned and defended by Van linden herself (see note 16, p. 45) and has to do with the interaction of the meaning of adjectives and that of the matrix construction. By only looking at extraposed clauses, the author runs a risk that some of the modal distinctions that some adjectives have depending on whether they are complemented by an extraposed clause or not get lost (see e.g. Perkins, 1983: 74 on *apt*). Similarly, since the lexico-semantic associations are so strong, the selection of adjectives becomes extremely important, and it would be interesting to have a larger set of adjectives, especially in the synchronic part as the diachronic comparability obviously presents problems in this respect.

On a more general note, I have to applaud the use of figures in this book. Each chapter is complemented by a number of illustrative figures, which helps to explain the highly complex concepts covered in the book.

All in all, I am very excited about this book, as it brings together the concepts of modal adjectives and complementation in an impressive way and clearly shows that adjectives are important for both of these areas of study.

References

Perkins, Michael. R. 1983. *Modal expressions in English*. London: Pinter.

Rivière, Claude. 1983. Modal adjectives: Transformations, synonymy and complementation. *Lingua* 59: 1–45.

Vander Viana, **Sonia Zyngier** and **Geoff Barnbrook** (eds.). *Perspectives on Corpus Linguistics*. Amsterdam and Philadelphia: John Benjamins, 2011. 256 pp. ISBN 978-90-272-0353-3. Reviewed by **Andrew Hardie**, Lancaster University.

Viana *et al*.'s edited collection *Perspectives on Corpus Linguistics* (henceforth *Perspectives*) is a peculiar book. It consists of a series of what the editors characterise (p. xi) as "interviews", that is, texts in which the contributors to the volume respond to a series of questions posed by the editors. Given the non-interactive nature of the resulting contributions, however, I consider them more accurately described as *essays* written to a set of predetermined headings, rather than interviews *sensu stricto*. The contributors to the volume are **Guy Aston, Paul Baker, Tony Berber Sardinha, Susan Conrad, Mark Davies, Stefan Gries, Ken Hyland, Stig Johansson, Sara Laviosa, Geoffrey Leech, Bill Louw, Geoffrey Sampson, Mike Scott**, and **John Swales**. Each responds to ten questions – seven that are the same in each piece and three devised by the editors with that particular contributor's expertise in mind.

The seven standard questions are basic and generic; since they are so central to *Perspectives*, I will take the unusual step of quoting these headings in full here:

1. Where do you place the roots of Corpus Linguistics? And to what do you attribute the growth of interest in the area?
2. Is Corpus Linguistics a science or a methodology? Where would you situate Corpus Linguistics in the scientific or methodological panorama?
3. How representative can a corpus be?
4. How far should an analyst rely on intuition?
5. What kinds of questions should an analyst think of?
6. What are the strengths and weaknesses of corpus analysis?
7. What is the future of Corpus Linguistics?

(*Perspectives*, *passim*: numbering as per the original)

It should be clear, then, that *Perspectives* is the result of a highly distinctive undertaking. As such it has a number of strong qualities. For example, question 7 stimulates for the most part diverse and challenging responses from the contributors, and in some (not all) chapters the specific material elicited by questions 8–10 is of very great interest. At a more general level, the very act of assembling commentary from such a wide range of corpus linguists, including

some of great renown, within a single set of covers is surely in itself a remarkable achievement. Most of these scholars give, within their responses to the questions, potted summaries of their most important work – material which is nowhere else to be found covered in such neatly condensed form. However, despite the virtues of nearly all of the individual contributions, I found on closing the book that I was dissatisfied with *Perspectives* overall.

One problem is that in many of these essays, very little new ground is covered – in the sense that readers who are already familiar with the key works of the essayist will find little here that they do not know already. Moreover, the overviews typically come up short in comparison to the key works they summarise. For example, there is little in Baker's essay that will not be familiar to those who have read Baker (2004, 2006, 2010), Baker and McEnery (2005), and Gabrielatos and Baker (2008); nor is Gries' essay much of a surprise to those acquainted with publications such as Gries (2010a, b, c), Stefanowitsch and Gries (2003), Gries and Divjak (2009), and Gilquin and Gries (2009). Some other contributors whose work I am personally not so familiar with (e.g. Hyland) also draw extensively on their own earlier work (e.g. Hyland 2004, 2005a, b), so the same comment almost certainly applies. Clearly there are pros and cons to this: as I noted above, having brief summaries of a wide range of earlier work is indeed a virtue, but it is unavoidable that a short essay written to a fixed schema of headings cannot cover the ground as effectively as multiple full papers. The feeling of having seen the content of several chapters done previously elsewhere was a large part of my overall dissatisfaction with *Perspectives*.

A rather greater part, however, stemmed from the impression that on the majority of the points – with the primary exception of question 7 – nearly all the essayists were substantially in agreement with one another. Going through *Perspectives*, then, I got the distinct impression of reading much the same answers over and over again, with some novel material interspersed. (In all fairness I must report this is merely my own impression, and that in a cover quotation from Michael Toolan and the foreword by Ronald Carter both these readers emphasise the variety and diversity of the volume, indicating that this must lie to an extent in the eye of the beholder.) This is probably because most of the questions are rather basic in nature, dealing with fundamental issues about which there is not a lot new to be said, although they remain critical to practice in the field. Asking a group of leading corpus linguists to answer these questions over and over again does not seem a productive use of the opportunity offered by the availability of these contributors. They are bound to give responses which

reflect the established consensus of the field – in part because these people were responsible for establishing that consensus in the first place.

To exemplify this, let us consider question 2 (quoted above), obviously designed to tap into the controversy between the two schools which I usually refer to as *neo-Firthian* and *methodologist* (see Hardie and McEnery 2010; McEnery and Hardie 2012: Chapters 6, 7). And yet, even on a controversial matter, the same set of points emerge again and again in the responses: corpus linguistics is mostly seen as an approach or method(ology), with the proviso that *some* work in the field can be seen as establishing or bearing on a theory of language; whether this means that corpus linguistics is a 'science' depends on precisely what is meant by *science*; but in any case corpus linguistics should not be considered a separate discipline independent of the rest of linguistics. The manner of expression is obviously very varied, but nearly all the contributors end up saying something similar in import to this. The same point could be made for most other questions. Only in a minority of essays is an answer given that is substantially distinct (e.g. Gries provides a very individual take on question 3). Even Swales, who is brought in explicitly to offer a "critical view on the use of corpora" (p. 221), actually ends up in broad agreement with the general thrust established by the other essays; and on question 6, for instance, Swales in fact has *less* to say in criticism of the weaknesses of corpus linguistics than, say, Conrad, Davies or Leech. If a *really* contrary view was desired, perhaps a traditional literary scholar or a formalist syntactician should have been approached!

There is one major exception to the relative sameness of opinion across the volume: Louw's essay. It will come as no surprise to those familiar with Louw's work that his views diverge substantially from the consensus that the other contributors by-and-large converge around. But here, as elsewhere in *Perspectives*, there is little material that the reader could not understand more readily by a perusal of Louw's key works (Louw 1993, 2000, 2003, 2010); I cannot recommend Louw's essay in *Perspectives* as an adequate substitute for a careful reading of those papers.

Louw notwithstanding, the general agreement on the first six questions can be attributed to the editors' decisions on what questions to ask. The choice of very basic questions such as 3, 4 and 5, on which consensus has been more-or-less attained, must surely reflect an imagined audience consisting largely of beginners in the field, since as I have explained the responses offer relatively little for a reader already well-versed in these issues. But is a readership of novices really well served by an array of essays with so much repetition? The editors' choices regarding the contributor-specific questions are sometimes even more

puzzling. For instance, how on earth does one manage to interview Mike Scott and *not* ask him a question about keywords?

A more serious editorial failing is that the editors do not succeed in imposing very much order on this collection of essays. The preface to *Perspectives* explicitly states that it is up to the reader "to arrive at an overall understanding of the topics discussed by piecing together the various approaches and perspectives" (p. xi). This seems unsatisfactory. *Why* must this task be left to the reader – especially if that reader is a novice in the field? An extensive discursive preface could have ameliorated this difficulty, but in point of fact the preface to *Perspectives* does little more than *list* the topics that are addressed by the different general and contributor-specific questions. Another way of offering support to a novice reader would be to group the essays according to content and to order the resulting sections so as to produce an overall narrative that the reader could follow. But instead, the essays are ordered alphabetically by author – or to put it another way, the editors leave it entirely to the reader to navigate the volume. The very final chapter of *Perspectives* is an attempt by Viana, the lead editor, to draw together the strands of the volume – or, to adopt the editors' phrasing, to present "the underpinnings which surge after reading the fourteen contributions" (*sic*; p. xv). This attempt is largely a failure, in my judgement, as the chapter is unable to accomplish much by way of effective overview or summary, resorting in the main to generalities – which become especially vague towards the end. But the problems with Viana's overview are merely symptomatic of the difficulty the editors have had imposing a coherent structure onto the material elicited by the format they chose for the essays.

All in all, then, and in spite of the points of interest to be found in some individual contributions, *Perspectives* does not work overall as a volume and may represent something of a missed opportunity, given the number of what one might describe as 'big names' involved in the project. If the same collection of people had instead been asked to write a few pages on whatever aspect they chose of their own theoretical stance, their view of the current state of corpus linguistics and their opinions on where the field might/should go next, the result would surely have been a much more varied and exciting book. *Perspectives* represents a significant effort on the part of the essayists and a clearly even greater effort by the editors. Unfortunately, reading the final result, it is hard not to conclude that this effort was largely wasted.

References

Baker, Paul. 2004. "Unnatural acts": Discourses of homosexuality within the House of Lords debates on gay male law reform. *Sociolinguistics* 8 (1): 88–106.

Baker, Paul. 2006. *Using corpora in discourse analysis*. London: Continuum.

Baker, Paul. 2010. Representations of Islam in British broadsheet and tabloid newspapers 1999–2005. *Language and Politics* 9 (2): 310–338.

Baker, Paul and Tony McEnery. 2005. A corpus-based approach to discourses of refugees and asylum seekers in UN and newspaper texts. *Language and Politics* 4 (2): 197–226.

Gabrielatos, Costas and Paul Baker. 2008. Fleeing, sneaking, flooding: A corpus analysis of discursive constructions of refugees and asylum seekers in the UK Press 1996–2005. *Journal of English Linguistics* 36 (1): 5–38.

Gilquin, Gaëtanelle and Stefan Th. Gries. 2009. Corpora and experimental methods: A state-of-the-art review. *Corpus Linguistics and Linguistic Theory* 5 (1): 1–26.

Gries, Stefan Th. and Dagmar Divjak. 2009. Behavioral profiles: A corpus-based approach towards cognitive semantic analysis. In V. Evans and S. S. Pourcel (eds.). *New directions in cognitive linguistics*, 57–75. Amsterdam and Philadelphia: John Benjamins.

Gries, Stefan Th. 2010a. Corpus linguistics and theoretical linguistics: A love-hate relationship? Not necessarily…. *International Journal of Corpus Linguistics* 15 (3): 327–343.

Gries, Stefan Th. 2010b. Methodological skills in corpus linguistics: A polemic and some pointers towards quantitative methods. In T. Harris and M. M. Jaén (eds.). *Corpus linguistics in language teaching*, 121–146. Frankfurt am Main: Peter Lang.

Gries, Stefan Th. 2010c. Useful statistics for corpus linguistics. In A. Sánchez and M. Almela (eds.). *A mosaic of corpus linguistics: Selected approaches*, 269–291. Frankfurt am Main: Peter Lang.

Hardie, Andrew and Tony McEnery. 2010. On two traditions in corpus linguistics, and what they have in common. *International Journal of Corpus Linguistics* 15 (3): 384–394.

Hyland, Ken. 2004. *Disciplinary discourses: Social interactions in academic writing*. Ann Arbor, MI: University of Michigan Press.

Hyland, Ken. 2005a. Stance and engagement: A model of interaction in academic discourse. *Discourse Studies* 7 (2): 173–191.

Hyland, Ken. 2005b. *Metadiscourse*. London: Continuum.

Louw, Bill. 1993. Irony in the text or insincerity in the writer? – the diagnostic potential of semantic prosodies. In M. Baker, G. Francis and E. Tognini-Bonelli (eds.). *Text and technology: In honour of John Sinclair*, 157–176. Amsterdam: John Benjamins.

Louw, Bill. 2000. Contextual prosodic theory: Bringing semantic prosodies to life. In C. Heffer and H. Sauntson (eds.). *Words in context: A tribute to John Sinclair on his retirement*, 48–94. Birmingham: University of Birmingham.

Louw, Bill. 2003. Dressing up waiver: A stochastic-collocational reading of the Truth and Reconciliation Commission (TRC). Harare: mimeo. Reprinted [2005] in *Quaderni del CeSLiC. Occasional papers*, Centro di Studi Linguistico-Culturali, Bologna. Available online at http://amsacta.unibo.it/1142/. (23 February, 2013).

Louw, Bill. 2010. "The painting is where the paint is not": Reflections on the "Bootcamp Debate". *International Journal of Corpus Linguistics* 15 (3): 344–353.

McEnery, Tony and Andrew Hardie. 2012. *Corpus linguistics: Method, theory and practice*. Cambridge: Cambridge 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.