

Reviews

Paul Baker. *Sociolinguistics and corpus linguistics*. Edinburgh: Edinburgh University Press, 2010. 200 pp. ISBN: 978 0 7486 2736 3 (pbk). Reviewed by **Miriam Meyerhoff**, University of Auckland.

Baker starts this volume by observing “the absence of a book to date which details a corpus-based approach to *sociolinguistics*” (p. 1). Therein lies both the challenge and the hazard of a book like this. It is intended as a teaching text – the first in a new series, Edinburgh Sociolinguistics – targeting more advanced undergraduate and postgraduate students. This is a tricky audience because you have to steer an appropriate path between the subject-specific jargon that students at this level can be expected to know and the jargon they will not know. In *Sociolinguistics and corpus linguistics* (SLC), this challenge is exacerbated because the text is attempting to bridge two fields which have not always had much to say to each other. To the extent that they have, it is thanks to Baker’s research where he models the usefulness of combining quantitative analyses of large corpora with qualitative or even micro-analysis of texts. SLC shows Baker’s background: it is a fluent and engaging “book about corpus approaches to language in use” (p. 146), enthusiastic about the potential in corpus linguistics (CL), frank about its limitations, and a fine model of how CL can be deployed in a virtuous circle with more detailed analysis of texts, e.g. discourse analysis and conversation analysis. However, notwithstanding the title, it has more to say about CL than about sociolinguistics (SL).

Full disclosure: I am a sociolinguist who works with corpora, rather than a corpus linguist who looks at language in use. Hence, I am probably more inclined to find fault with the sociolinguist side of the book. Baker does a reasonably good job of explaining how sociolinguists engage with their corpora, but I am not sure whether this book will elucidate the major differences between SL and CL methods. Students will certainly come away with a very impoverished notion of SL.

I have thought quite hard about the audience SLC might benefit most. On the one hand, corpus linguists by and large seem to be unaware (or unconcerned, I am not entirely sure which) about the SL debates over social determinism and relativism, the appropriateness of generalisation given different types of data collection and different modes of analysis. In my experience, even where CL does document the (linguistic or social) details of the contexts in which particular variants occur, there seldom seems to be an urge to connect these findings with linguistic or social theory in the way most SL does.

On the other hand, sociolinguists seem to view CL as both arcane and naïve: there are super-large corpora from which the users stand some distance because they have not personally collected them; searches that focus almost entirely on lexical items or forms that are readily automated, i.e. the most frequent or overt variants (omitting null/zero realisations of a form); results analysed without linguistically nuanced consideration of the context; the practice of excluding outliers from further consideration of results. The last three points violate the sociolinguist's 'principle of accountability', namely that all and all possible tokens of the variable must be considered. Baker introduces the principle of accountability but does not explore its implications for CL practice in this book, a point I will elaborate below.

The structure of the book is as follows. Chapter 1 (Introduction) provides key terms, CL tools, corpora, annotation, central CL concerns such as *frequency*, *concordances*, *collocation*, *keywords* and *dispersion*. Aspects of its structure suggest that the chief purpose of SCL is to persuade sociolinguists of the merits of CL (e.g. myth busting about CL, reassurances to sociolinguists about the mysteries of annotation) rather than to inform CL of how they can engage more meaningfully with SL.

Chapter 2 (Corpora and sociolinguistic variation) shows how corpora can be used to explore the distribution of lexical items according to social factors and introduces some key notions in SL such as *age-grading* (individuals use different forms at different points in their life; the community patterns alike in this) versus *generational change* (individuals maintain the same form over their lifespan; the community norms change). The relationship between different patterns of variation and change across individuals and the community (cf. Sankoff and Blondeau 2007) is not discussed. This is important because these different profiles of variation and change are not simply theory; there are practical implications for how we explore data. Later in the book, Baker notes the difficulty of determining whether you are really seeing evidence of change when you have compared data across only two points in time – with only two data points it is not obvious whether you are seeing evidence of age-grading or generational

change. This is precisely why sociolinguistic studies of change prefer to have a minimum of three age groups of speakers for their analyses, but this is not explained for the student. Finally, in this chapter, there is brief mention of phonetic variation. The option of maintaining time-aligned spoken corpora is acknowledged but the various options for this are not detailed in the manner that the different online, written corpora are introduced elsewhere in the book.

Chapter 3 (Diachronic variation) explores changes between Brown and Frown, LOB, FLOB and BLOB (before LOB) and a corpus of 2006 materials that Baker had assembled (perhaps in response to critiques of relying on the BNC and LOB corpora as sources of ‘contemporary’ British English, e.g. Cameron 2009). In this chapter, the student is introduced to the possibility of considering clusters of features (e.g. involved versus informational styles (Biber 1988), and to the possibility of interpreting quantitative results with qualitative information, e.g. the evaluative co-text surrounding *girl* and *boy*, and external factors such as social change and marriage patterns to interpret changes in address terms.

Chapter 4 (Synchronic variation) raises some crucial points in the comparison of different regional corpora, e.g. the ICE family. This includes whether we are actually looking at ‘the same’ form when we compare it in British, Singapore and Indian Englishes, also the problems with considering the absence of forms. Because this chapter highlights some basic problems with reconciling SL and CL, I will return to them in more detail below.

Chapter 5 (Corpora and interpersonal communication) provides a fine overview of how CL and conversation analysis (the detailed analysis of turn-by-turn sequencing in texts) can be used effectively together. Here, the level of detail needed in a transcription is discussed, as are sources of spoken corpora. It also raises some larger questions, such as if one decides to conduct an analysis of a corpus that considers (i) frequency of a form, (ii) its collocates, and (iii) the (conversational) function of each token, what theory of language is this committing you to? It does not deal in detail with the potential problems of using CL techniques with corpora prepared for conversation analysis (CA). CA encourages the use of non-standard spellings, eye dialect forms etc.

Chapter 6 (Uncovering discourses) builds on one of Baker’s most significant contributions to the field of CL, the use of corpus data to infer ideologies and to critique the social order. Here, Baker makes trenchant connections with linguistic theory, from critical discourse analysis, to psycholinguistic priming. The key terms introduced in Chapter 1 are used effectively and there is a helpful reflection on traps and issues with using corpora in these ways.

Chapter 7 (Conclusion) has more substance than most concluding chapters do. It assumes the student reader has by now achieved a degree of proficiency that will enable them to build and use a corpus of their own. There is an appendix with examples of constituent and semantic tagging systems.

The book is generally well-written to the target reader level. It assumes that the student will know next to nothing about maths and statistics (it is probably right to err on this side of caution). There is patient explanation of very basic measures: how to normalise frequency per million words (p. 20), how to compare percentage change between corpora (p. 62), how to calculate the chi-square for a 2 x 2 table (pp. 35–36) and similarity measures (pp. 91–93). More complex statistics (cluster analysis, multi-factorial analysis, Dice coefficient) are mentioned but there is no explanation of how they work. I think this is appropriate to the level.

The book suffers somewhat from two biases in the sources it draws on. The development of corpora and corpus analysis outside the UK are not discussed, e.g. the Linguistic Data Consortium at Penn (including the Santa Barbara Corpus), the early use of cluster analysis in sociolinguistics by Horvath (1985), Janet Holmes' exposition of the lack of a direct form-function correspondence for numerous pragmatic particles in the 1980s (Holmes 1995, for a review). In addition, there are few references to the extensive tradition of quantitative, variationist SL in North America, or the more corpus-based research characteristic of functional syntax there. This second bias cuts to the heart of the issues sociolinguists will have with this book.

Consider, for example, the researcher who is interested in exploring the distribution of *that* as a complementiser (*She realised that/Ø it was raining quite hard; I think that/Ø I better leave now*). Baker's discussion of this alternation (p. 76) notes that finding the zero complementiser cannot be automated (unless you have a syntactically well-tagged corpus – and syntactic tagging so as to include zeroes is in itself potentially problematic. Baker does not discuss this but the tagging must proceed on the basis of some predetermined theory of grammar). He recommends that it be done by searching for the high frequency verbs that are most commonly followed by *that*. However, this makes a variationist sociolinguist nervous for a number of reasons.

First, it violates the principle of accountability. This means that any findings can only apply to the three or four most frequent verbs investigated, thereby weakening any sense one might have of understanding what complementiser deletion/insertion does in the syntactic system more generally.

Second, the practice ignores the fact that in some contexts *that* is obligatorily present or absent (Huddleston and Pullum 2002; Torres Cacoullos and

Walker 2009 – indeed, no-one should be allowed to analyse *that* deletion in English without starting from the comprehensive analysis of their 2009 paper). Variationists will not include such contexts in their analysis *because they do not vary*. We will only understand what the alternation between *that* and \emptyset does (structurally or socially) if we restrict the analysis to genuine loci of variation.

Third, applying this methodology to corpora gathered at different points in time (as discussed in SLC) implies acceptance of something that variationists know as the *Constant Rate Hypothesis* (CRH, Kroch 1989). The CRH proposes an alternative model of variation and change to the wave model (Bailey 1973). The wave model (and more recently, exemplar models, Bybee 2006) assumes that changes happen fastest (go to completion first) in most favouring environments. Different factors might define what makes a most favouring environment (e.g. lexical frequency, structural markedness). By contrast, the CRH proposes that the actuation of a change may ‘favour’ certain contexts more than others, but that the diffusion or incrementation of the change proceeds at a constant rate within each context. That is, some contexts may start out with relatively low frequencies of the change (e.g. the main verb *realise* is more likely to have *that*, less likely to have \emptyset) and some start with relatively high frequencies of change (e.g. the main verb *think* is more likely to have \emptyset and less likely to have *that*). This relationship is maintained (remains constant) throughout the change – the most favouring contexts always have a greater probability of \emptyset and the less favouring contexts will always have a lower probability of \emptyset , and the probabilities will remain roughly the same as the change works its way through the system. I think only this model of change would make it reasonable to study *that* deletion by extracting the most frequent verbs from corpora collected at different points in time. If your understanding of language change is more akin to the wave model, then sampling the most frequent verbs and the most favouring contexts will give you an artificial sense of when a change is complete.

Finally, the focus on only the most frequent contexts may mask other strong linguistic constraints. For example, Torres Cacoulios and Walker (2009) show that there are very strong effects relating to whether the subjects of the matrix and subordinate clauses are co-referential or not. The strength of a constraint like this may not show up if only the most frequent and most favouring contexts for *that* deletion are selected for study.

At the start, I posed the question of who SLC is for. On balance, I think its audience is principally teachers of a corpus linguistics course. For practicing sociolinguists, well-versed in the field, it makes an excellent introduction to CL (sociolinguists might like to consider how CL has been using dispersion plots and incorporating this into their software for some time, rather than trying to

reinvent the wheel by creating their own representations of dispersion and co-occurrence). For quantitative sociolinguists, it offers a good model of how to teach some of the basic statistics students will need (cf. Walker 2010). However, sociolinguists may be frustrated by the points at which the text fails to reflect on important methodological and epistemological differences between SL and CL. Despite the stated and implied aims of convincing sociolinguists of the utility in CL methods, it falls short of placing CL in the context of variationist SL especially.

Corpus linguists are unlikely to be troubled by this. For them, the text provides a clear and concise progression through CL practice. It offers a consistent and principled distinction between corpus-based and corpus-driven analysis and is well-organised pedagogically: important terms, methods and corpora are revisited from different perspectives. For the readers of this journal, I recommend the book as a teaching text and as a taster of how CL can interact with other branches of linguistics.

References

- Bailey, Charles J. 1973. *Variation and linguistic theory*. Washington, D.C.: Center for Applied Linguistics.
- Biber, Douglas. 1988. *Variation across speech and writing*. Cambridge: Cambridge University Press.
- Bybee, Joan L. 2006. From usage to grammar: the mind's response to repetition. *Language* 82: 711–733.
- Cameron, Deborah. 2009. Review: Paul Baker, *Sexed texts: Language, gender and sexuality*. *Journal of Sociolinguistics* 13: 547–552.
- Holmes, Janet. 1995. *Women, men and politeness*. London: Longman.
- Horvath, Barbara. 1985. *Variation in Australian English: The sociolects of Sydney*. Cambridge: Cambridge University Press.
- Huddleston, Rodney and Geoffrey K. Pullum. 2002. *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press.
- Kroch, Anthony. 1989. Reflexes of grammar in patterns of language change. *Language Variation and Change* 1: 199–244.
- Sankoff, Gillian and H el ene Blondeau. 2007. Language change across the lifespan: /r/ in Montreal French. *Language* 83: 560–88.

Torres Cacoullos, Rena and James A. Walker. 2009. On the persistence of grammar in discourse formulas: A variationist study of *that*. *Linguistics* 47: 1–43.

Walker, James A. 2010. *Variation in linguistic systems*. London: Routledge.

Ana Frankenberg-Garcia, Lynne Flowerdew and Guy Aston (eds.). *New trends in corpora and language learning*. London and New York: Continuum, 2011. 304 pp. ISBN 978-1-4411-5996-0. Reviewed by **Fanny Meunier**, Université Catholique de Louvain.

The volume, prefaced by Lou Burnard, is a collection of selected papers given at the eight Teaching and Language Corpora conference (TaLC) held in Lisbon in July 2008. The editors have divided the fifteen contributions to the volume into three main sections: section one is entitled “corpora with language learners” and focuses on the use of corpora by teachers and learners; section two is called “corpora for language learners” and deals with tools for language learning; the third section, “corpora by language learners”, contains contributions describing corpora produced by second/foreign language learners.

In Chapter 1, Tono presents several corpus-based applications in English language teaching in Japan. Those applications include the implementation of a user-friendly web-based corpus query system, the creation of a corpus-based English conversation TV programme, and the development of corpus-based English materials (including computer-game apps). Whilst this article is a personal narrative taking place in one of the most technologically-aware countries (Japan), it nevertheless shows the impressive impact that corpora can have. To give only one example, 100-go – a 10 minute English conversation TV programme focussing on 100 core lexical items – is broadcast 25 weeks per year and has over 1 million viewers each year.

In the second chapter, Charles examines the practical implications of using hands-on concordancing in large mixed-discipline EAP writing classes (of predominantly non-linguist students) to study rhetoric. One of the discourse functions addressed during the course is presented in the paper, viz. criticizing the work of other researchers. The students are required to perform specific searches whose output is then further analysed in class. The analysis of the evaluation

questionnaire given to students at the end of the 12-hour course provided positive results overall (in terms of usefulness of the approach, user-friendliness of the procedure, and future use of corpora by learners) and this, despite the fact that the course was taught to a large, heterogeneous group, with irregular attendance, and limited class time. What the author does not stress enough, perhaps, is the paramount importance of the qualitative preparation work she did: viz. the specification of a useful and coherent search pattern that can readily be used by learners (e.g. *but*, *however* and *though* when *19** is found in the context) and which undoubtedly constitutes an invaluable asset to the course.

In Chapter 3, Kettemann shows how corpus analysis can be used in cultural studies. He presents the corpus-based analysis of an alternative youth culture (the Emos), characterised by introversion and withdrawal from an outside adult world perceived as unsympathetic, misunderstanding and demanding. Here again, teacher-input is required prior to the subsequent analysis made by students. Students are given four documents to analyse: a list of lexical verbs following first person *I* and categorised semantically into 12 categories (e.g. communicative processes: *ask*, *write*, *tell*, *defend*; style: *wear*, *look*, *pierce*; physical contact: *push*, *stab*, *kiss*, *touch*, etc.), a keyword comparison between the Emo corpus and another corpus of teenage language (the COLT corpus), a list of the top fifty content words in the Emo corpus, and finally concordances of *alone*, *lonely*, and *on my own*. Kettemann then presents the student-centered exploratory research activities done in class with the students. In terms of pedagogical evaluation, only 25 per cent of the students completed the questionnaire, but the author nevertheless recommends the use of corpora to complement other forms of teaching in cultural studies. Whilst such an approach, essentially based on wordlists and frequencies, might seem appropriate for very specific sub-cultures such as the Emos', it might however prove insufficient to extract more subtle aspects of culture.

After the use of corpora in discourse and culture studies presented in the two previous chapters, Kübler shows how corpora can play a role in the teaching of pragmatic translation in France. She first lists some of the reasons behind the lack of use of corpora by translators and then shows how corpus-use can be beneficial to pragmatic specialised translation, a translation process taking into account the readership's reception of the translation and pays particular care to the issues of intent, genre and register, terminology, phraseology, and semantic prosody. She illustrates, with the help of concrete examples of tasks, how different types of corpora can be used for different task types ranging from pure translation tasks, to the identification of genres and registers, or to the discovery of the importance of semantic prosody in language.

In the last chapter of section one, Kazubski presents IFAConc, an online concordancing environment containing three main interfaces: a core search interface, a searchable record of past work (called ‘History’) and a special site containing tutorials, tasks and hyperlinks (‘Resources’). After a presentation of the initial pedagogic and theoretical inspirations that have led to the development of the concordancing environment, the author provides the results of several pilot studies. Although few students evaluated the tool, it was generally assessed as useful in the English for Specific Academic purposes pilot study. In the second pilot study, i.e. English for General Academic purposes, the students had diverging opinions about the user-friendliness of IFAConc, although here again most students found the tool useful. As IFAConc is in continuous use, the results of the assessments will lead to the implementation and/or modification of some of its features.

Liu, Wible and Tsao, in the first chapter of section two, explain how the use of novel statistical methods can lead to the automatic correction of identified miscollocations. The approach they use for automatic correction of verb plus noun miscollocations is innovative and markedly different from others, as it relies solely on the concepts of collocation clusters and intercollocability initially proposed by Cowie and Howarth (1996). Liu, Wible and Tsao use the limitations on intercollocability within a collocation cluster (which are often the cause of miscollocations produced by learners) as a means for linking miscollocations back to the correct collocates. They have built a hybrid model – combining their ‘overlapping nouns’ (ON) method with mutual information (MI) – which yields very encouraging results. Although the authors state that the issue of the automatic recognition of miscollocations is still left unresolved, they believe that their work can also have implications for automatic detection of miscollocations, especially when verb and noun pairings belonging to a collocation cluster show a much lower MI than the other pairings in the cluster.

Chapter 7, written by Coccetta, is the only paper of the volume to address multimodal corpora, and more specifically multimodal functional-notional concordancing techniques. After a brief presentation of the Padova multimedia English corpus, the author presents the scalar-level approach used in the study. Four levels are adopted to analyse language functions: whole text, phase, sub-phase and utterance. Those levels are particularly useful as language functions often extend beyond utterance level. The author then introduces multimodal concordancing and the way it relates to the scalar-level approach. Clear illustrations are provided of how to annotate language functions and notions, and to access the semiotic modalities of multimodal corpora that complement the transcribed oral data (such as tone of voice, stress pattern, facial expressions, or pos-

tures). Suggestions for multimodal DDL activities addressing functions and notions are also presented in the article.

In Chapter 8, Curando Fuentes presents a content-based machine translation project (CBMT) whose aim is to produce machine translation (of written English into Spanish) which reduces human effort and cost. The two key resources of the CBMT system are an English-Spanish fully inflected dictionary containing mapping tables for English to Spanish translations and, second, a mega Spanish corpus collected using web-crawling techniques. Although the exact procedure is not included in the article, the author explains that the dictionary and corpus are linked via different database tables including information on co-occurrence, collocations, colligations, semantic associations and text-related colligations. Whilst the existence of such databases is undoubtedly crucial in fostering automatic translation, the case study presented afterwards did not, at least in my case, make things much clearer in terms of how exactly the CBMT system works. The potential for LSP teaching is also stressed by the author in the last section of the article.

In the next chapter of section two, Warren shows how the ConcGram (Greaves 2009) programme can be used to automatically retrieve phraseological units that go beyond simple N-grams and include lexical bundles displaying constituency (e.g. *work hard* but also *work so hard* or *work very very hard*) and positional variation (e.g. *work hard* and *hard work*). Warren then presents examples of the three main types of phraseological associations that can be extracted by ConcGram, viz. meaning shift units, collocational frameworks and organisational frameworks, and discusses the importance of promoting phraseological variation in language learning and teaching.

In Chapter 10, Widmann, Kohn and Ziai introduce SACODEYL (System Aided Compilation and Open Distribution of European Youth Language), a web based system created for the assisted compilation and open distribution of European teen talk in the context of language education. Widmann *et al.* describe the development and illustrate the features of the SACODEYL search tool. It should be added that this tool is rather unique as it is one of the very few existing corpus tools created to specifically meet the needs of secondary school language learners and teachers (such as topic based searches, authentication, multimodality, or constructivist learning approach).

The third section starts with an article which shows how findings from learner corpora can help inform assessment of foreign language speaking skills within the Common European Framework of Reference. In this excellent article, Osborne demonstrates that quantitative (e.g. speech rate, length of utterance) and qualitative (e.g. lexical range, accuracy) measures of fluency applied to cor-

pora of English, Italian and French L2 speakers do not, when taken separately, provide a reliable measure of proficiency. It is only when those measures are taken together (as a bundle of features representing fluency) that they can help differentiate various levels of proficiency. The author also addresses the possible automation of some of these measures as a useful way of obtaining rough indicators of fluency. He also pleads for the definition of fluency bands (where minimum and maximum values are provided for each band) at different CEFR levels.

In Chapter 12, De Cock reports on a study of attitudinal stance in native and learner speech (more specifically in the French, German and Chinese sub-corpora of the Louvain International Database of Spoken English Interlanguage, LINDSEI). She focuses on evaluative/emotive descriptors, such as *good*, *wonderful*, *bad*, *terrible*, etc. The data selected (informal interviews) are particularly well-suited to such analysis as they contain numerous expressions of personal feelings and attitudes. The author's study confirms an overall bias towards positive items in academic speech (a bias also pointed out by other researchers). De Cock then studies preferred patterns of use (e.g. predicative vs. attributive, or proportion of evaluative sentential relative clauses) and identifies potential candidates for inclusion in ELT materials. In the next chapter, Nesi presents the British Academic Written English corpus (BAWE). She defines the design matrix (4 levels of study, 4 broad disciplinary groupings and a total of 28 disciplinary fields), the markup conventions (including the marking up of 13 'genre families' such as case study, critique, literature survey, research report, etc.). The 6.5 million word corpus is an ideal resource to study the development of novice to more proficient academic writers, to investigate specific genre families, but also to discover discipline specific or cross-disciplinary features of academic written English.

The last two chapters of the volume adopt a more socio-cultural approach to the analysis of second/foreign languages. Chapter 14 deals with the impact of culture on the use of stance exponents as persuasive devices. Hatzitheodorou and Mattheoudakis show that Greek university students writing in English use rhetorical techniques which are different from those used by their American peers. The learner and native corpora are examined through the lens of cross-cultural research and contrastive rhetoric. The authors, using an adapted version of Biber and Finegan's (1989) and Hyland's (2005) frameworks of stance, show that Greek advanced learners of English prefer boosters and attitude markers to hedges. The authors explain that those tendencies may be attributed to a variety of factors including L1 and L2 instruction, L1 transfer of writing style and cultural features, and developmental features of L2.

As for the last chapter of the book, it is devoted to the comparison of discourse features in two corpora of academic English papers: one written by established Portuguese academics and the other by English native speakers. McKenny and Bennet argue that the different epistemological frameworks that underlie the scholarly discourse of highly proficient non-anglophone cultures (such as wordiness, redundancy, tendency towards abstraction or figurative language) may jeopardise these writers' chances of being published in international journals. This seems to be particularly true for the discourse of the humanities which, according to the authors, vary dramatically between cultures. The authors first present 10 key differences (overuse or underuse of specific language features) between the two corpora and then focus more specifically on lexical choices and collocational patterns. They explain that such studies help raise the awareness of the contingent nature of academic discourse conventions and may also help promote epistemological pluralism in academia.

Despite the fact that the division into three main sections can sometimes be questioned (Chapter 5 could have been included in the section on tools; the BAWE corpus included in section 3 is not a learner corpus per se), and that separate and more fleshed out subject and author indexes would have been welcome, the volume does present a very attractive range of articles and papers. In the preface to the book, Lou Burnard questions the evolutionary path of the community of researchers/teachers making use of corpora in their teaching practice, mentioning the danger of stagnation and the over-representation of academic contexts. Whilst it is true that academic contexts are clearly over-represented – be it in terms of applications or types of corpora analysed – I would argue that Tono's and Widmann *et al.*'s articles should inspire further applications of corpora to lower levels of proficiency in secondary school contexts. Another asset of the book is that it shows how corpora can be used in many different domains of applications including some which have, so far, been slightly underrepresented in the literature, such as translation studies, cultural studies, or work with multimodal corpora. Many of the studies in the volume also demonstrate how, to quote Kettemann (this volume, p. 45), the use of corpora can lead to “a pedagogical reconceptualization” of some disciplines.

References

Biber, Douglas and Edward Finegan. 1989. Styles of stance in English: Lexical and grammatical marking of evidentiality and affect. *Text* 9 (1): 93–124.

- Cowie, Anthony P. and Peter Howarth. 1996. Phraseological competence and written proficiency. In G.M. Blue and R. Mitchell (eds.). *Language and education*, 80–93. Clevedon: Multilingual Matters.
- Greaves, Chris. 2009. *ConcGram 1.0.: A phraseological search engine*. Amsterdam and Philadelphia: John Benjamins.
- Hyland, Ken. 2005. *Metadiscourse: Exploring interaction in writing*. London: Continuum.

Gaëtanelle Gilquin. *Corpus, cognition and causative constructions*. Amsterdam: John Benjamins, 2010. xvii + 326 pp. ISBN 978-90-272-2313-5. Reviewed by **Jürgen Esser**, University of Bonn.

The book is a corpus linguistic study of ten periphrastic causative constructions as exemplified by *he made me laugh* or *I had my hair cut*. The theoretical and methodological coverage is ambitious. On the one hand, the study proposes to show “how corpus linguistics and cognitive linguistics can be fruitfully combined into a single, integrated theoretical model” referred to as the “Corpus-Cognition Integrated model” (p. 2). On the other hand, the corpus data are also combined and compared with experimental data from elicitation tests.

After a short introduction, Chapter 2 attempts (somewhat long-windedly) to clarify the theoretical status of corpus linguistics (if there is any) and its relation to other linguistic theories. What it boils down to is that corpus linguistics shares with cognitive linguistics that it is usage-based, i.e. interested in performance. The difference is, however, that while cognitive linguistics is a mentalistic theory (like Chomsky’s generativism), corpus linguistics provides new methodologies which lead to new theories. As often in science, new tools (here: the computer) have led to new results and new theories. In linguistics this is particularly true of frequency, collocations, and the interrelation of syntax and lexis.

Chapter 3 deals with data collection. For her investigation the author selected a sample of roughly ten million words from the BNC from the genres (written) academic prose and (spoken) conversations/discussions. The author used the *BNCweb* platform, which allows not only for searching for word forms but also for tag sequences with varying numbers of intervening word forms. The constructions that were searched for are:

- [X CAUSE Y V_{to-inf}]
- [X GET Y V_{to-inf}]
- [X GET Y V_{p(ast) p(participle)}]
- [X GET Y V_{pr(esent p(participle))}]
- [X HAVE Y V_{inf}]
- [X HAVE Y V_{pp}]
- [X HAVE Y V_{prp}]
- [X MAKE Y V_{inf}]
- [X BE *made* V_{to-inf}]
- [X MAKE Y V_{pp}]

The author had to apply ‘repair mechanisms’ to improve recall and precision rates. In a post-editing process homonymous sequences which were not periphrastic causative constructions had to be weeded out. One of the advantages of the carefully chosen methodology is the exhaustiveness of the analysis, which yielded also irregular cases like *if I’d been made do it* (without *to*). As to be expected, the frequency of the constructions differed considerably with [X MAKE Y V_{inf}] occurring 114.4 times per one million words and [X MAKE Y V_{pp}] only 3.2 times (p. 48).

As indicated above, the author carried out various elicitation tests. The idea was that “corpus data provide the basis for descriptive adequacy” and that “elicitation data go some way towards giving the model some psychological plausibility” (p. 57).

While the reviewer appreciates the combination of a corpus-tested top-down approach and a corpus-driven bottom-up approach in the exploitation of corpus data, the role of elicitation data in relation to corpus-tested and corpus-driven procedures remains unclear to him (cf. Figure 2 p. 58). It is difficult to see what it is that establishes the *integration* of the data in the proposed “Corpus-Cognition Integrated model”.

In Chapter 4 the periphrastic causative constructions are explained in terms of Langacker’s cognitive linguistic concept of the ‘action chain’. An expression like *Jealousy caused the Queen to kill Snow White* is analysed as consisting of a causer (*Jealousy*), causative verb (*caused*), causing event (*Jealousy caused*), causee (*Queen*), effect (*kill*), patient (*Snow White*) and caused event (*Queen [to] kill Snow White*). The expression *The Queen caused Snow White to die* is analysed as having only causer, causative verb and causee, but no patient (p. 66). (This analysis may be accepted by some but challenged by others.) 68 per cent of the causative constructions in the author’s data correspond to a three-participants action chain and 32 per cent to an action chain with only two participants.

Various tables are presented, explained and exemplified with corpus data to show: the length of the action chain; the number of participants; gapping (ellipsis) of the causing event; identity and ambiguity between patient, cause and causee. As most noteworthy results the author mentions “the absence of an explicit causing event in most *get* and *have* constructions” and “the frequent identity between the CAUSER and the CAUSEE in [X *GET* Y V_{pp}]” (p. 95). These two findings could be confirmed by the elicitation data.

Chapter 5 proposes to explore the syntax and semantics of causative constructions. It is committed to the idea that a difference in form corresponds to a difference in meaning. This old tenet has become a cornerstone of cognitive linguistics and equally of corpus linguistics. Gilquin does not describe meaning differences of the ten periphrastic causative constructions (resulting from context and morphological form) within a cognitive linguistic framework. She rather opts for statistical analyses of some 50 features that pertain mainly to causer, causative verb, causee, effect and patient. With the help of Chi Square analysis she finds out, for example, that [X *CAUSE* Y V_{to-inf}] appears to be associated with present tense and present participle and that it prefers generic participants. (The caption ‘tense’ as used in Table 20 on page 107 is not felicitous for the morphological categories there listed.) Unfortunately, the list of features in Table 17 (p. 100) is hard to associate with the features used for the frequency analyses in Tables 18 to 42. The gain of the elicitation tests is limited: “a number of syntactic and semantic regularities brought to light by the corpus analysis are also found in the elicitation data” (p. 139). However, elicitation is useful when corpus evidence is sparse or lacking altogether (cf. p. 142).

Chapter 6 attempts to define a causative prototype on the basis of frequencies in corpora. The author considers three models of prototypical causation as suggested in the literature: iconic sequencing, the billiard-ball model and direct manipulation. Iconic sequencing means that the sequence of linguistic elements should mirror the sequence of the event categories. The billiard-ball model describes the transmission of energy by a physical causer to a physical causee. The model of direct manipulation describes clusters of interactional properties concerning the causer, the patient and the event.

According to the axiom that prototype elements are more frequent than others, Gilquin carries out frequency counts of the ten periphrastic causal constructions. Her analysis questions the validity of various theoretical models of prototypical causative constructions.

In a next step the three models are evaluated in terms of the elicitation tasks. The result is not surprising because “the proportions of prototypical constructions in the elicitation data are higher than in the corpus data” (p. 161). This kind

of discrepancy has been described in the literature and one of the conclusions is that “corpora and intuition draw on different pools of linguistic knowledge” (p. 165).

Chapter 7 deals with lexical co-occurrence in causative constructions. The author extracted her relevant hits from the *BNCweb* in order to search with *WordSmith Tools* the collocational behaviour of her ten constructions. This is one of the most interesting parts of her study. For example, she qualifies the assumption that the verb *cause* generally involves a negative semantic prosody. Influenced by the nature of the selected genres, Gilquin discovers that *cause* can be found more often with neutral semantic prosody in technical and scientific contexts. Although based on comparatively small numbers, she presents interesting analyses of the semantic prosodies of her ten constructions. There is, for example, a semantic prosody of effort and difficulty connected with *get* and one of service with *have*.

Chapter 8 opens further new perspectives on causative constructions. The aim is to find out more about the verbs in the effect slot of causative constructions. To this end, the verbs were tagged for sense, for example, *go_OPERATE*, *go_BECOME*. With the help of statistic procedures the author is able to show the distinctive lexical units (rather than lexemes!) for each construction. For example, [X *MAKE* Y V_{pp}] is much more strongly attracted to the lexical unit *know* in the sense of ‘be familiar with’ than ‘have information’. The insightful results are well summarised in Table 71 (p. 220).

In Chapter 9 the influence of variety dimensions on the choice of causative constructions is explored in statistical terms. We learn, for example, that the construction [X *CAUSE* Y V_{to-inf}] is the one whose ratio between speech and writing is the highest, amounting to 1:12 (p. 226). Furthermore, statistical analyses between the two mediums are carried out on the basis of the syntactic and semantic features discussed in Chapter 5, the semantic prosodies discussed in Chapter 7 and the preferred effect-expressing lexical units discussed in Chapter 8.

Chapter 10 is devoted to causative constructions in foreign language teaching. Gilquin drew on the International Corpus of Learner English in order to compare the use of causative constructions as used by learners of English with the use observed in her BNC sub-corpus. The contrastive analysis focuses on three aspects of learner writing: under- and overuse of causative constructions in learner writing, their syntactic misuse of and preferences for lexical units. The analyses show that learners have difficulties with the idiomatic use of causative constructions. Moreover, their presentation in grammars is generally only fragmentary. Therefore, a highlight of the book is the proposal of how the chapter on causative constructions in a grammar book could be rewritten. The proposed

grammatical unit offers a unified account, including syntax, semantics, style and lexis.

The text of the book ends with the conclusion in Chapter 11. There are appendices showing the elicitation test and the distinctive and most frequent syntactic and semantic features of causative constructions.

Judging from the methodology and the analysis of the results, the book is a valuable contribution to corpus linguistic research. The question remains, however, whether the proposed *integration* of cognitive linguistics and corpus linguistics is more than lip service. The reviewer finds it also difficult to understand the clear-cut distinction between a corpus-tested approach in Chapters 4 to 6 versus a corpus-driven approach in Chapters 7 to 9. What, for example, is the difference between frequency counts to account for prototypicality (Chapter 6) and frequency counts to account for semantic prosody (Chapter 9)? In any case, it is clear that Gilquin has made an important study in the fascinating field of corpus linguistics.

Yufang Ho. *Corpus stylistics in principles and practice: A stylistic exploration of John Fowles' The Magus.* London: Continuum, 2011. 272 pp. ISBN: 978-0-8264-26178. Reviewed by **Brian Walker**, University of Huddersfield.

John Fowles novel *The Magus* was first published in 1966 and then revised and republished in 1977. Ho's book represents an investigation of the stylistic differences between the two editions of the novel, using a combination of stylistic analysis and corpus tools and techniques.

Chapters 1 and 2 form Part I of the book, and set the scene for the analysis of the two versions of *The Magus*. Ho first highlights the features and principles common to stylistic and corpus linguistic approaches that can be usefully combined in the emerging field of corpus stylistics. Importantly, she emphasises the need for the careful combination of quantitative and qualitative analyses. Ho sets out the research questions that guide her study, and describes in appropriate detail the stylistic theoretical frameworks that she uses, including a clear rationale for why the theories are relevant to the investigation. Ryan's (1991) possible worlds theory is of particular importance to the research and is used by Ho to describe the worlds projected in both versions of the novel, and how they contribute to plot development and characterization. Ho goes on to discuss the lin-

guistic markers important to the construction of fictional worlds with reference to Doležel's (1998) four modal systems, Simpson's (1993) modal grammar, and McIntyre's (2006) cognitive model of point of view. This is followed by a discussion of Short's (1996) checklist for linguistic indicators of point of view and Biber and Finegan's (1989) list of stance markers for 'affect' and 'evidentiality'. Both lists are important for the corpus strand of the investigation.

Part II of the book comprises Chapters 3 to 8, and describes the comparative analysis of *The Magus*. Chapter 3 provides a clear and thorough synopsis of the novel, which gives the reader sufficient detail to understand the analyses in subsequent chapters. The chapter also includes analyses of short extracts from the novel that help to demonstrate how the character of Nicholas (the main protagonist and I-narrator) changes. Other important characters of the novel are also discussed. An overview of the revised version of *The Magus* follows, which describes changes to the language, theme, characters, and point of view. Ho discusses the comments made by critics concerning the revisions, and concludes the chapter by saying that such commentaries are based on intuitions or limited analyses, and are therefore insufficient to draw conclusions from. The following chapters endeavour to address that shortfall.

In Chapter 4, the corpus tools TESAS/Crouch (a tool for measuring text reuse in journalism) and WCopyfind (a tool for detecting plagiarism) are introduced with clear and detailed descriptions of how these tools work and the output they generate. The tools show that the majority of the changes in the revised edition of *The Magus* are in Part II of the novel, which comprises Chapters 10 to 67. On one hand, the results are quite promising and provide some analytical purchase, but, as Ho explains, there are features concerning content and style that cannot be captured by these tools. Ho suggests that a combination of micro-level qualitative stylistic analysis and macro-level quantitative linguistic analysis is necessary, which she carries out over the next three chapters.

Chapter 5 pursues the micro-level strand of Ho's investigations and contains detailed linguistic analyses of three short extracts from the two versions of the novel. The extracts are taken from chapters that are shown to have differing levels of similarity by TESAS/Crouch and WCopyfind, and are chosen because of their importance to the plot. The analyses confirm some of the observations made by critics: the second version of the novel is more explicit, containing more contextual information. They also generate two new hypotheses about the new version of *The Magus*, namely that Fowles (i) shifts the focus of the narrative to the I-narrator's internal world, and (ii) uses more figurative language.

Chapter 6 tests out hypothesis (i) using WMatrix (Rayson 2008) and is part of the macro-level analysis of the novel. Ho uses the semantic analysis that

WMatrix is able to provide, because this most obviously relates to Ryan's (1991) possible worlds. Guided by the previous TESAS/Crouch and WCopyfind analysis, Ho compares the 22 most dissimilar chapters (which account for 99% of the changes) from the two versions of the novel. Using the same chapters from the first version as the comparator, she identifies those semantic categories that are significantly under-used (negative key-concepts) and over-used (positive key-concepts) in the revised chapters. The negative key-concepts, of which there is only a handful, are dealt with first, and point to a decline in the detail of descriptions of objects and movements in the text actual world. The over-used categories present the psychological viewpoint of the I-narrator and relate to his hypothetical and inferential speculations being more prominent. These findings therefore support hypothesis (i).

Chapter 7 tests hypothesis (ii), claiming that Fowles uses more figures of speech in the revised version of *The Magus*. Ho sets out to extract and compare (quantitatively and qualitatively) metaphors and similes from the two versions of the novel. This is no small task, but it is carried out rigorously and systematically. Ho first introduces the concord facility of WordSmith Tools (Scott 2003), which she uses for the extraction task, before describing previous corpus approaches to analysing figurative language, and the associated issues of metaphor identification and extraction. Ho starts with similes and utilizes those findings to search again for metaphorical expressions. Focussing again on Part II of the novel, similes are detected by looking for connectives that tend to be found in such constructions, such as *like*, *as*, *seem* and *similar to*. Results from the search are assessed manually for their validity as similes and are quantified. From these results a set of pervasive core metaphor groupings are found, including PLAY, GAME, JOURNEY and EDUCATION. Further searching is then carried out, this time looking for concords containing source domain vocabulary, derived from collocational information from the simile concordances. The results from this search are again manually checked, and any non-metaphorical entries removed. From this, the amount of metaphors and similes relating to the core metaphors identified by Ho are shown to be used many times more in the revised version of the novel, thus supporting the second hypothesis.

The metaphors are analysed qualitatively in order to show that, in the later edition, they interact to reinforce the theme and the structure of the novel. The analysis here provides the necessary balance to the quantitative results, and demonstrates very clearly why and how the qualitative step should be done. The same can be said for Chapter 8, which brings together and assesses the results of all the analyses and discusses further the stylistic differences of the new version of the novel.

Finally, Chapter 9 assesses the tools and techniques used for the analyses and makes suggestions for further research. Ho also considers the contributions made by her study which are worth repeating here. Firstly, it contributes to the stylistic appreciation of *The Magus*, thus attending to the gap identified in Chapter 3. Secondly, it demonstrates that corpus tools and techniques can be usefully applied to stylistic analysis, and thirdly it develops a corpus methodology for analysing figurative language.

Without a doubt, Ho's work provides much food for thought with regard to corpus and stylistic analytical techniques and their application, and I think, in particular, it shows how those interested in cognitive approaches to literary analysis might usefully engage corpus tools. Ho's study also shows how the sorts of stylistic analyses that are usually carried out on short texts/extracts can be extended, using corpus tools, to larger, complete texts. At the same time, though, Ho's meticulous approach to the study demonstrates that corpus tools alone do not make an analysis systematic. It must be carefully controlled by the analyst. So too should the balance between quantitative and qualitative analysis, something which Ho points out in Chapter 1. Ho achieves this and in doing so provides us with a fine example of corpus stylistic research.

References

- Biber, Douglas and Edward Finegan. 1989. Styles of stance in English: Lexical and grammatical marking of evidentiality and affect. *Text* 9 (1): 93–124.
- Doležel, Lubomir. 1998. *Heterocosmica: Fiction and possible worlds*. Baltimore and London: Johns Hopkins University Press.
- McIntyre, Dan. 2006. *Point of view in plays: A cognitive stylistic approach to viewpoint in drama and other text-types*. Amsterdam: John Benjamins.
- Rayson, Paul. 2008. *WMatrix: A web-based corpus processing environment*. <http://ucrel.lancs.ac.uk/wmatrix/>.
- Ryan, Marie-Laure. 1991. *Possible worlds, artificial intelligence and narrative theory*. Bloomington and Indianapolis: Indiana University Press.
- Scott, Mike. 2003. *WordSmith Tools*. Oxford: Oxford University Press.
- Short, Mick. 1996. *Exploring the language of poetry, plays and prose*. Harlow: Longman.
- Simpson, Paul. 1993. *Language, ideology and point of view*. London: Routledge.

Susan Hunston. *Corpus approaches to evaluation. Phraseology and evaluative language.* London/New York: Routledge, 2011. 199 pp. ISBN 978-0-415-96202-5. Reviewed by **Monika Bednarek**, University of Sydney.

Evaluative language, as Susan Hunston and others have long pointed out, is important to human interaction for a variety of reasons: it pervades human behaviour and discourse, it expresses subjectivity, construes relationships between speaker/writer and audience, and organises discourse (e.g. Thompson and Hunston 2000; Bednarek 2006). This book brings together Hunston's work on status, phraseology and local grammar, with the overall aim to explore evaluative language through the lens of corpus linguistics (in particular, research on phraseology). Hunston has a lot of expertise to draw on, given her long history of working in the area of language and evaluation, including her own research as well as supervising others' (e.g. my own PhD research while I was a visiting scholar at the University of Birmingham 2003–2004).

Corpus approaches to evaluation is divided into nine chapters, including the conclusion. The first chapter ("Evaluative language, phraseology and corpus linguistics") introduces the book and in turn briefly discusses evaluative language, corpus linguistics and phraseology. Hunston claims that, despite the cumulative and implicit nature of evaluation, corpus linguistics is well suited to analyse evaluative language and can provide more than just "adding quantitative detail" (p. 4). In particular, she argues that "phraseology as broadly defined plays a number of roles in the study of evaluative language" (p. 7). Despite this focus on corpus linguistics and phraseology, Chapters 2 and 3 both deal with evaluation from a discourse perspective, as Hunston acknowledges the research done in this area.

Thus, Chapter 2 ("Appraisal, stance, evaluation") discusses approaches outside corpus linguistics that have explored evaluative language. Using an elegant solution, Hunston here chooses to focus on points of consensus, rather than trying to tease out in detail similarities and differences between these approaches. These points of consensus are that:

1. Evaluation is subjective as well as intersubjective;
2. Evaluation is tightly linked with value systems that may be ideological;
3. There are a wide range of linguistic means that indicate evaluative meaning;
4. Evaluation "is both contextual and cumulative" (p. 14);
5. The elements involved in evaluation include a source (who evaluates), as well as someone or something that is evaluated;
6. It is difficult to reliably identify evaluation or non-evaluation.

Four discourse approaches to evaluative language are discussed: systemic-functional linguistic Appraisal theory; Hunston's own work on status, value, relevance; research on stance (a term used both in corpus linguistics and in sociolinguistics); and work on metadiscourse. While additional concepts such as *evaluation* (à la Labov), *subjectivity* and *affect* – which are part of overviews of evaluative language in other books (Bednarek 2006; Englebretson 2007; Jaffe 2009) – are not discussed here, what is welcome is that Hunston is very open to these text-based frameworks, noting that corpus linguistics cannot replicate close-reading analyses. This is a refreshing attitude, which sees the value in different methodologies. As Hunston notes, corpus linguistic studies simply “answer a different set of questions” (p. 24).

Chapter 3 (“Status in written texts and multi-modal texts”) goes on to discuss one of the concepts introduced in Chapter 2 in more detail, the notion of (epistemic) status, which was first identified in Hunston's PhD research (Hunston 1989) as one of the functions of evaluation. Status has to do with the reification of texts and propositions “by assigning them an epistemic status” (p. 25), for example as hypotheses, assumptions, or facts. The chapter introduces both the resources for marking status (such as modals, evidentials, reporting expressions in language and – following Kress and van Leeuwen 2006 – colour, detail of context, degree of abstraction, brightness etc. in images), and discusses the connections between status and attribution (e.g. Sinclair 1987) as well as the two systemic functional linguistic concepts of modalization and Engagement. Crucially, Hunston argues that Engagement works well with particular kinds of texts such as opinion writing, whereas status works better in analysing academic research. As Hunston notes, the disagreement between the two notions (Engagement and status) “arises at least in part from the kind of language that gave rise to the theories of evaluative language being used” (p. 49). This is an extremely important point, I would say, which needs to be taken into account by any theoretical framework that is built ‘bottom-up’ from analyses of particular text types. Such frameworks need to remain open to modifications and, in addition to simply applying them, considerable energy needs to be extended to modify them, too, as more and different data are analysed. It would be interesting now to compare more recent research that uses Appraisal theory (including Engagement) to investigate academic writing (e.g. Hood 2010) with Hunston's and other corpus linguistic approaches to evaluation in academic language.

After the emphasis on discourse approaches in Chapters 2 and 3, the following chapters all deal with quantitative and corpus approaches. Chapter 4 (“Evaluation, quantity and meaning”) reviews what can broadly be termed the ‘quantitative’ literature on evaluative language. The focus is first on work on

stance by Biber and Hyland (who both use the term *stance*, although they include different linguistic resources under this concept) and on computational linguistic *sentiment analysis* (the automatic detection of opinion). The latter will be particularly useful for linguists who have yet to discover related research in computational linguistics. The second part of the chapter discusses in more detail phraseological concepts that can be tied to evaluative language, namely *semantic preference* and *semantic prosody*. Again, rather than delving deep into the theoretical and terminological debate in this area, Hunston refreshingly chooses to focus on points of agreement and on the connection to evaluative language, which she summarises as follows:

1. Certain words/phrases collocate with words that have a positive or negative evaluative meaning;
2. Certain words/phrases have affective meanings (e.g. frustration, irritation);
3. Evaluation can be implied by the typical associations of words or phrases even in neutral contexts;
4. Irony or insincerity can arise from the non-typical use of such language.

Following on from the more theoretical overviews of previous chapters, Chapters 5 to 7 offer case studies on phraseology and evaluation. Chapter 5 (“Modal-like expressions”) focuses on phraseology and modality, or rather “modal-like expressions” (p. 66). These are expressions other than modal auxiliaries which express modal meanings, for example *be on the verge of*, *face the prospect of*, *have the job of*, *have a knack of* (pp. 78–79). From a more qualitative perspective, the chapter uses concordance lines to demonstrate that some word forms occur in contexts of modal-like expressions. From a more quantitative perspective, it compares frequencies to show that some verbs co-occur with modal-like meaning more than others and that *wh*-words (e.g. *whether*) strongly attract modal meaning. The chapter also offers an introduction of qualitative (concordances) and quantitative (frequencies) corpus methodologies. This kind of analysis of modality, Hunston argues, is significant because “in looking at the [modal-like] expressions themselves we get an insight into the complexity of evaluative meaning in English discourse; and in looking at verbs frequently modified in this way we see the importance of phraseology to that evaluative meaning” (p. 85). Indeed, this chapter clearly shows the many (phraseological) ways in which modal meaning can be expressed in the English language. As Hunston notes, her studies here “only scratched at the surface of what may be a much more pervasive phenomenon” (p. 91), and her pilot study of the modal uses of verb sequences and modal meaning needs to be developed to incorporate what she calls “more sophisticated” (p. 90) methods of analysis.

Chapter 6 (“Corpus approaches to investigating status”) continues with corpus linguistic case studies of evaluative meaning. The methodology is similar to that applied in Chapter 5, but the focus is on linguistic resources that explicitly indicate *status* (introduced in Chapter 3). The chapter starts by describing a previous study by Hunston (2008) of 11 status nouns (e.g. *idea*, *assumption*, *conclusion*), and their discourse functions before comparing selected nouns and verb lemmas that evaluate something as *assumption* or *discovery* in more detail and also examining the frequent status noun *fact*. These studies give an insight of the ways in which propositions are evaluated commonly in a given corpus. One of the points of this research is the potential to compare this across corpora, giving insight into the social construction of knowledge (e.g. across disciplines or across time). Indeed, the discussion at the end of “how facts travel” is, for me, one of the highlights in this chapter.

Chapter 7 (“Grammar patterns, local grammars and evaluation”) focuses on the ‘local grammar’ approach to evaluation (e.g. Hunston and Sinclair 2000) and discusses this primarily in the context of Appraisal theory and FrameNet (<https://framenet.icsi.berkeley.edu/fndrupal/home>). The chapter first discusses the relationship between a word and its co-text in terms of meaning (where does it reside?) before introducing the theory of pattern grammar (Hunston and Francis 1999) which inspired the development of local grammars. The chapter also describes local grammars and their connection to FrameNet (see also Hunston 2003). Hunston argues that it is *patterns* rather than words that need to be taken as a starting point for analysing meaning and that the level of pattern generality has an impact on the usefulness of local grammars. Further, Hunston uses corpus linguistic methodology to test Martin and White’s (2005: 58–59) grammatical frames for distinguishing Attitude sub-types (Appreciation/Judgement/Affect) – independently also done so by Bednarek (2009). Both Bednarek’s and Hunston’s studies seem to suggest that these frames “can be of some use in distinguishing between types of evaluative meaning, most decisively identifying Affect” (p. 138; see Bednarek 2009: 179) but that there are limitations to their use.

Chapter 8 (“Phraseology, intensity and density”) takes up the concepts of intensity and density/saturation from Appraisal theory, referring to the way that evaluative meaning is distributed throughout a text (saturation/density) and the degree of evaluative meaning expressed (intensity). Hunston argues that both “work together to make evaluation more or less strong in a text” (p. 152) and that there is often redundancy in meaning. Following a more qualitative analysis and discussion of three texts, Hunston continues with an investigation into phrases that function in general to intensify evaluation (e.g. adjective + *almost* /

bordering on / to the point of + adjective/noun) as well as with an explanation of how corpus methods can be used to identify intensifying phrases. This can be done by exploring phrases (with prepositions) that co-occur with negative evaluative words such as *tragedy*. Interestingly, Hunston draws here on a wordlist that is semantically annotated for evaluative meaning (the General Inquirer database) and also used in Sentiment Analysis. Intensifying phrases found using this methodology include *as much as*, *on the brink/edge/verge of*, ADJECTIVE *dose of* and *be greeted with*. These indicate meanings such as proximity and quantity and are frequently figurative (e.g. *a wave of*).

The final chapter of the book is the conclusion, which summarizes its main topics before addressing the implications and applications of Hunston's work. This includes implications for register studies, research into academic disciplines, and computational applications drawing on local grammars.

Overall, this book is indispensable reading for anyone working on evaluative language/Appraisal. It is written in Hunston's typical clear, concise and accessible style, with explicit stating of aims/structuring, manifold examples and expert guidance of readers into complex phenomena. The book draws on key references on evaluation, not just corpus linguistic research. Indeed, one of the admirable features of this book is the open-mindedness and honesty of Hunston's writing – for instance, she openly acknowledges where her research has been, as she calls it, “superseded”, what other approaches are good or better at than her own, and she makes explicit connections to research outside her own. Other researchers in the field would profit from an equally open mindset.

References

- Bednarek, Monika. 2006. *Evaluation in media discourse. Analysis of a newspaper corpus*. London/New York: Continuum.
- Bednarek, Monika. 2009. Language patterns and attitude. *Functions of Language* 16: 165–192.
- Englebretson, Robert (ed.). 2007. *Stancetaking in discourse*. Amsterdam/Philadelphia: Benjamins.
- Hood, Susan. 2010. *Appraising research: Evaluation in academic writing*. Basingstoke/New York: Palgrave Macmillan.
- Hunston, Susan. 1989. Evaluation in experimental research articles. Unpublished PhD thesis, University of Birmingham.

- Hunston, Susan. 2003. Frame, phrase or function: A comparison of frame semantics and local grammars. In D. Archer, P. Rayson, A. Wilson and T. McEnery (eds.). *Proceedings of the Corpus Linguistics 2003 Conference. UCREL Technical Papers* 16, 342–358. Lancaster: UCREL.
- Hunston, Susan. 2008. Starting with the small words: Patterns, lexis and semantic sequences. *International Journal of Corpus Linguistics* 13: 271–295.
- Hunston, Susan and Gill Francis. 1999. *Pattern Grammar: A corpus-driven approach to the lexical grammar of English*. Amsterdam/Philadelphia: Benjamins.
- Hunston, Susan and John Sinclair. 2000. A local grammar of evaluation. In S. Hunston and G. Thompson (eds.). *Evaluation in text: Authorial stance and the construction of discourse*, 75–100. Oxford: Oxford University Press.
- Jaffe, Alexandra. 2009. Introduction: The sociolinguistics of stance. In A. Jaffe (ed.). *Stance: Sociolinguistic perspectives*, 3–28. Oxford: Oxford University Press.
- Kress, Gunther and Theo van Leeuwen. 2006. *Reading images: The grammar of visual design*. 2nd edition. London/New York: Routledge.
- Martin, James and Peter White. 2005. *The language of evaluation: Appraisal in English*. Basingstoke/New York: Palgrave Macmillan.
- Sinclair, John. 1987. Mirror for a text. Unpublished manuscript, University of Birmingham.
- Thompson, Geoff and Susan Hunston. 2000. Evaluation: An introduction. In S. Hunston and G. Thompson (eds.). *Evaluation in text: Authorial stance and the construction of discourse*, 1–27. Oxford: Oxford University Press.

Yoko Iyeiri. *Verbs of implicit negation and their complements in the history of English*. Amsterdam: John Benjamins, 2010. xv + 223pp. ISBN 978-90-272-1170-5. Reviewed by **Hendrik De Smet**, University of Leuven/Research Foundation Flanders.

Most historical studies of change in the English system of complementation track the rise or demise of one or more complement types (Fanego 1996; Rudanko 1998; Los 2005; De Smet 2007). In *Verbs of implicit negation and*

their complements in the history of English, Yoko Iyeiri changes the perspective, approaching the history of the English system of complementation through an elaborate case study on a cluster of semantically related verbs and the various complement types they appeared with over time.

The group of verbs at issue consists of eleven ‘verbs of implicit negation’, including *forbid*, *refuse*, *forbear*, *avoid*, *prohibit*, *prevent*, *hinder*, *refrain*, *fear*, *doubt* (as well as *question*, when meaning ‘doubt’) and *deny*, all of which “imply negation in English” (p. 1). Their status as covertly negative items appears from the fact that in earlier stages of the language they are found to combine with *that*-clauses with expletive negation, as in (1).

- (1) They did *prohibit that* no man shoulde ... sell openly ... wine of Candie or Spaine. (p. 81)

Using the quotation database of the electronic *Oxford English dictionary*, supplemented by data from the British National Corpus, the BROWN-family of corpora, the Helsinki Corpus of English Texts and The Bible in English on CD-ROM, Iyeiri traces the development of each of these verbs, with specific attention to the major changes in their patterning with different complement types.

The history of the verbs of implicit negation generally reflects what Vosberg (2006) has called ‘The Great Complement Shift’, whereby *to*-infinitives developed at the expense of *that*-clauses and gerunds subsequently developed at the expense of *to*-infinitives. For example, *that*-clauses following *prohibit* as in (1) above were ousted by *to*-infinitives as in (2) and later *to*-infinitives were ousted by gerunds as in (3):

- (2) The reading of history *prohibyteth* reprouable persons *to do* mischeuous dedes. (p. 84)
- (3) The London Government Act of last year ... *prohibited* women *servng* as alderwomen or councillors on borough Councils. (p. 85)

Against this general background, however, Iyeiri demonstrates that the complement shift is considerably more complicated than the name implies, and the main contribution of her study lies in revealing the various factors that complicate complement shifts.

For a start, more constructions are involved than the three main patterns introduced above, and so there is much more variation and competition. The rise of gerunds is complicated by the appearance of gerunds introduced by *from* as in (4) (or, very occasionally, other prepositions), which often compete with simple gerunds as in (3) above. Among the latter, there is further competition between gerunds with a possessive subject and with an oblique subject.

- (4) The Assistant District Attorney tossed out the charges saying that although the law *prohibits* Klansmen *from appearing* in public, the faces of the 24 were visible. (p. 85)

Similarly, the history of *that*-clauses is not only marked by competition with *to*-infinitives, but also by variation between *that*-clauses with and without overt *that*, by variation between *that*-clauses with and without expletive negation, by occasional variation with *whether*-clauses, *if*-clauses and even open *wh*-clauses as in (5), and by the rise of *lest*-clauses and *but*-clauses as in (6). With respect to the latter, Iyeiri argues that *but*-clauses have for some time served specifically as a way of avoiding *that*-clauses with expletive negation.

- (5) Some *doubted* how far such volage expressions inferred treason, being but *lubricum linguæ* (p. 141)
- (6) I do not *doubt* but you want constant every-day debaters. (p. 151)

Finally, for some verbs, Iyeiri proposes that the function of *that*-clauses has been (partly) taken over by epistemic parentheticals as in (7). An exception here, however, is presented by the parenthetical *God forbid* as in (8), which is shown to have fossilised at an early stage, to have later developed into the last stronghold of *that*-clauses with *forbid*, but in the end to have briefly merged back with the small residue of remaining *that*-clauses following *forbid*.

- (7) Take my armour of quickly, 'twill make him swoune, *I feare*. (p. 131)
- (8) But *God forbyd* that I shulde synne so vnto the LORDE ... (p. 36)

Another complication affecting the overall picture of the Great Complement Shift is that, where complement types compete, the resultant variation is constrained by a variety of factors. For example, Iyeiri points out the relevance of Rohdenburg's (1996) Complexity Principle to the development of gerunds with and without *from* following *prevent* (invoking the principle also to argue that from a historical point of view the variant without *from* cannot have been derived from the variant with *from*, as some synchronic analyses maintain). She also points to the effect of *horror æqui*, which not only gives rise sporadically to an unusual choice of complement (e.g. using a *to*-infinitive following *avoiding* to prevent direct consecution of two *-ing*-forms) but also seems to bias the forms the main verb takes (e.g. with language users generally steering clear of the form *avoiding* because *avoid* is normally followed by another *-ing*-form). Another surprising constraint on variation relates to the origin of a given verb, with verbs of Romance stock (e.g. *avoid*) being generally found to be more open to innova-

tive patterns than the verbs of Germanic stock (e.g. *forbear*), presumably because with the latter older uses are more firmly established.

Finally, complement shifts do not necessarily bring about the disappearance of a pattern. For many verbs variation lingers on and in some cases the expected directionality of a particular change is even reversed. For example, it is argued that *that*-clauses following *deny* first declined under the pressure of the *to*-infinitive but later recovered. Similarly, it is shown that with *fear*, *that*-clauses with unexpressed *that* initially advanced on the variant with overt *that*, but in Present-day English the trend seems to have changed.

The overall picture that emerges, then, is one in which the history of each of the verbs examined is to some extent unique. That is, while each of the verbs is affected by the same changes taking place in the English system of complementation, each verb still responds to these in its own way. The timing for the rise of *to*-infinitives and gerunds differs from verb to verb, the pace and eventual outcome of developments varies, and even within a single verb there may be variation as to how a given change unfolds. The most striking illustration here comes from *fear* and *doubt*, which undergo complement shifts to a different extent in negative and affirmative contexts.

As such, Iyeiri's work must be praised for showing the full complexity of diachronic complement shifts. However, although the study is on the whole very thorough and produces interesting (at times surprising) insights in the workings of complement shifts, it also has a few weaknesses – though it should be added that these do not detract from its general merits, as outlined above. The choice of the *Oxford English dictionary* as main data source may be criticized, as the nature of the data makes it hard to compare frequencies across periods and the collection of quotations is (from a corpus linguistic point of view) essentially unprincipled. Smaller but better-balanced and better-documented data sources, such as the Helsinki Corpus of English Texts, the Penn-Helsinki Parsed Corpus of Early Modern English or the Lampeter Corpus could perhaps have been used more systematically to check historical trends at least for the most frequent patterns. At the same time, one must also sympathize with the author here, and recognize that a larger-sized corpus of (especially) Early Modern English is on many Anglicist's wish list.

At the more theoretical level, Iyeiri may underestimate the value of the semantic approach to complementation, which argues that different complement types serve essentially different functions (e.g. Bolinger 1968; Wierzbicka 1988; Langacker 1991). She invokes the critiques by, among others, Curme (1931), Quirk (1974) or Noël (2003) on the semanticist approach, and is probably justified in pointing out that the very occurrence of complement shifts contradicts

the idea that different complements (exist to) serve different functions (p. 195). Further, the fact that a group of more or less semantically similar verbs is found to undergo quite diverse developments does not speak in favour of the semantacist view. Nevertheless, some findings could have been linked to semantic explanations.

For example, Noonan's (1985) distinction between dependent and independent time reference could partly account for the ease with which a non-finite clause (which lacks temporal and modal grounding) can replace a finite clause, since some matrix verbs automatically specify the temporal and modal status of the event in the complement clause, whereas others do not. *Avoid*, for instance, predicts that the event in the complement clause is (at the time of avoiding) unfulfilled. Therefore, this information need not be redundantly encoded in the complement clause, and so a non-finite clause will be unproblematic. By contrast, *doubt* readily combines with states of affairs in the past, present or future and so makes no predictions as to the temporal status of its complement, which therefore still needs to be marked in the complement clause, restricting the potential for non-finite complements. That *avoid* and *doubt* respond differently to the Great Complement Shift is therefore not entirely surprising.

As another example, when a verb allows both finite *that*-clauses and subject-controlled non-finite clauses, competition is only possible to the extent that the subject of the *that*-clause is typically co-referential with the subject of the matrix clause. Here too the semantic potential of a given complement type codetermines its likely role in complement shifts; yet the opportunity to involve semantics in the account here has been missed.

That said, *Verbs of implicit negation and their complements in the history of English* presents an excellent piece of research that will be useful to anyone studying complementation (in English and in other languages) or other areas of competition between grammatical variants in language history. It demonstrates very convincingly the intricacy of change in the system of complementation, specifically pointing to the lexical differences that (in this case) strongly affect competition processes in grammar.

References

- Bolinger, Dwight. 1968. Entailment and the meaning of structures. *Glossa* 2: 119–127.
- Curme, George O. 1931. *Syntax*. Boston: Heath.

- De Smet, Hendrik. 2007. *For...to*-infinitives as verbal complements in Late Modern and Present-day English. Between motivation and change. *English Studies* 88: 67–94.
- Fanego, Teresa. 1996. The development of gerunds as objects of subject-control verbs in English (1400–1760). *Diachronica* 13: 29–62.
- Langacker, Ronald W. 1991. *Foundations of cognitive grammar*. Vol. 2. *Descriptive application*. Stanford, CA: Stanford University Press.
- Los, Bettelou. 2005. *The rise of the to-infinitive*. Oxford: Oxford University Press.
- 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 and change*, 347–377. Berlin: Mouton de Gruyter.
- Noonan, Michael. 1985. Complementation. In T. Shopen (ed.), *Language typology and syntactic description*. Vol. 2, 42–140. Cambridge: Cambridge University Press.
- Quirk, Randolph. 1974. *The linguist and the English language*. London: Edward Arnold.
- Rohdenburg, Günter. 1996. Cognitive complexity and increased grammatical explicitness in English. *Cognitive Linguistics* 7: 149–182.
- Rudanko, Juhani. 1998. *Change and continuity in the English language. Studies on complementation over the past three hundred years*. Lanham, MD: University Press of America.
- Vosberg, Uwe. 2006. *Die große Komplementverschiebung. Außersemantische Einflüsse auf die Entwicklung satzwertiger Ergänzungen im Neuenglischem*. Tübingen: Gunter Narr Verlag.
- Wierzbicka, Anna. 1988. *The semantics of grammar*. Amsterdam: John Benjamins.

Fanny Meunier, Sylvie De Cock, Gaëtanelle Gilquin and Magali Paquot (eds.). *A taste for corpora* (Studies in Corpus Linguistics 45). Amsterdam: John Benjamins. 2011. xv + 295 pp. ISBN 978-90-272-0350-2. Reviewed by **Martina Bredenbröcker**, University of Paderborn.

This volume is a tribute and a special gift to Sylviane Granger on her 60th birthday. The title already suggests that the editors' aim lies in giving readers the opportunity to refine their own taste for corpora. The book consists of eleven chapters which were written by highly renowned specialists in the field of corpus linguistics. In their articles they shed light on the various fields in which corpus linguistics can be applied: The first two chapters deal with two key topics, frequency and contrastive analysis. The improvement of academic writing skills is the subject of Chapters 3 and 4, whereas Chapters 5 and 6 address software tools for the automatic analysis of learner corpora. In the next three chapters various themes related to corpus linguistics are handled, namely corpora for English as a Lingua Franca, learner knowledge of phrasal verbs and standard and non-standard varieties of the New Englishes. The final two chapters dedicate themselves to the role computers and corpora can play in the development of lexical and lexicographical resources for language learners.

The opening chapter is written by Geoffrey Leech on "Frequency, corpora and language learning". He divides frequency into three different kinds: raw, normalized, and ordinal frequency and considers the latter to be the most useful when dealing with language learning because it allows certain aspects to be prioritized above others.

Starting with a look back at the history of corpus linguistics, Leech addresses the changing role that frequency has played over the years. He then moves on to recent developments in frequency studies and concentrates on their importance for ELT (English Language Teaching), including such subdivisions as the preparation of teaching resources and lexicography. He poses the critical question of whether the equation 'more frequency = more important to learn' still holds. His answer is a strong yes, though with certain qualifications (e.g. the problem of overuse effects). When he looks into the frequency of word combinations, he touches on the lexis-grammar-interface and claims that phraseological aspects should be included in the learning process. He also stresses the benefit teachers of English would have if frequency information on grammatical categories and constructions were included in syllabi.

To conclude, Leech addresses challenges such as combining a corpus linguistic approach with a cognitive linguistic approach. He raises the question of

whether there is a lack of corpora that would exactly fit the needs of language learners, although he acknowledges the vast number of corpora that are potentially useful for ELT. However, he concludes that the most important task to tackle is the development of longitudinal corpora of native speakers and of non-native learners of English (cf. LONGDALE) to reach the optimum in ELT.

The article by Hilde Hasselgård and Stig Johansson opens with a historical overview about interlanguage studies before the arrival of computer corpora. They move on to sketch the developments of computerized learner corpora such as ICLE (International Corpus of English) and computer-aided interlanguage analysis like CIA (Contrastive Interlanguage Analysis). CIA is an essential feature of ICLE, which allows quantitative and qualitative comparisons between not only L1 and L2, but also different varieties of L2.

CIA presents us with some significant findings regarding the lexis, grammar and discourse of non-native speakers of English. Not surprisingly, their vocabulary is less varied and the degree to which their writing is influenced by informal spoken language is high. The authors point out the key role computer corpora play in contrastive analysis and sketch recent developments, such as the compilation of the English Norwegian Parallel Corpus (ENPC) and the creation of the Integrated Contrastive Model. In this paper, Hasselgård and Johansson have examined the use of *quite* and *I would say* in different L1 groups of ICLE and *seem* from only a Norwegian perspective. They conclude that the connection between learner data and contrastive data is rather complicated, because what originally would have been attributed as L1 transfer in their study had to be revised in the light of contrastive analysis.

As many corpus linguists before, Hasselgård and Johansson address the challenge of transferring results of corpus research into the practice of English language teaching. This transfer can prove difficult because there are many factors at play, such as the role of the target language and a possible reciprocal effect of teaching. So they agree with Leech's opinion that a longitudinal corpus like LONGDALE will yield invaluable insights into the process of language learning in the future.

In Chapter 3 JoAnne Neff van Aertselaer and Caroline Bunce examine the development of academic literacies by concentrating on the use of reporting verbs and four types of lexical devices for evaluation. Students exchange programs like ERASMUS have boosted student mobility considerably, and therefore university curricula nowadays emphasize argumentative competence in English academic writing.

For their study, they compiled a small corpus consisting of texts produced in an academic writing (AW) class by Spanish university students. Their corpus

data was then compared with the Spanish subcorpus of ICLE (SPICLE) to measure the effectiveness of the AW course layout. In order to do so, they “found it necessary to draw up a series of guidelines or ‘can do’ descriptors to make explicit the required structural and rhetorical features to be learned” (p. 67).

The authors collected essays from the beginning and the end of their AW course with the intention of being able to assess the students’ progress and the effectiveness of their courses. The SPICLE corpus provided a picture of Spanish EFL student’s essays without the benefit of this form of explicit teaching of writing skills. One focus of the study was the development of lexical devices for stance taking like *it is obvious that...* and the other one was on the examination of the use of reporting verbs (e.g. *conclude*).

In summary, Neff van Aertselaer and Bunce argue that academic literacy can be improved by means of data driven learning (DDL) to support the student’s use of text-internal and -external features and by providing them with individualized exercises. The detailed ‘can do’ statements proved helpful for a valid comparison between course-initial and course-final texts. They also helped shifting the teachers’ focus from merely marking the students’ errors to a measurable improvement in their discourse competence.

In Chapter 4, Christopher Tribble compares the written assignments of university students with published texts of renowned authors, drawing on Biber’s (2006) account of lexical bundles. They are defined as “the most frequently occurring sequences of words [...] they are usually not idiomatic in meaning, and they are usually not complete grammatical structures” (p. 87), such as *do you want to*.

Since the beginnings of corpus-informed language teaching and DDL, it has always been a crucial question which language model to present to students. Questions of authenticity played an important role in this discussion. Hence, Tribble opts for a corpus of – what he calls – ‘exemplar’ texts (here: a collection of journal articles from *Applied Linguistics*) which can be seen as expert performances. In total, he draws on five data sets of different sizes for his study of differences and similarities in the usage of lexical bundles. He shows that only a relatively small number of lexical bundles occurs in every examined corpus and that the absence of core academic framing markers (e.g. *the extent to which*) in the learner corpus indicates a need of support for students in this field. In brief, Tribble postulates to align ‘exemplar’ corpora as closely to the needs of the learners as possible. He acknowledges the importance of learner corpora for finding out about under-, over- and misuse of lexical bundles in order to develop useful curricula and learning material to support master level students who need to write English essays for academic purposes.

In their paper (Chapter 5), Paul Rayson and Alistair Baron present a novel application of a hybrid approach to the detection of spelling errors in learner data. Error tagging is one of the key issues of computer learner corpus research since it helps to gain insights into the factors at play under which learners of different mother tongues make special kinds of spelling mistakes. As error tagging was mainly done manually or semi-automatically at most, it has always been a time-consuming task.

This paper represents the recent trend to transfer knowledge from NLP to corpus research in order to improve the automatic analysis of learner language, especially the tagging of spelling mistakes. The authors designed the Variant Detector (VARD) software to find spelling variants in historical corpora as a pre-processor for NLP and other corpus tools. Their current aim is to examine its potential applicability to the automatic detection and correction of learner spelling errors. The newer version VARD2 has the ability to be trained and is therefore applicable in other contexts where spelling variations occur. This interactive tool is complemented by DICER (Discovery and Investigation of Character Edit Rules), which extracts letter replacement rules and analyses their frequencies. This data set can subsequently be used to further improve the accuracy of VARD.

The present study examines the occurrence of spelling and morphological mistakes in three subsets of learner corpora with Spanish, German and French as L1 (ICLE). A manually corrected corpus served as a yard stick in the DICER tool for measuring the extent to which spelling differed. The results showed that the accuracy rate was high and that the recall values significantly improved after the training course for VARD.

Yet, as Rayson and Baron point out, further research is required in this field by taking into account contextual patterns, e.g. different language backgrounds. In the long run, a fruitful partnership between corpus research and NLP can further contribute to the mutual improvement of both.

In Chapter 6, Scott Jarvis discusses data mining techniques to detect mother tongue-related patterns in samples of L2 writing. To begin with, he distinguishes between two types of data mining: the unsupervised and the supervised classification. The first one is often referred to as ‘clusterer’; it identifies text clusters with similar textual features. The second one, also called ‘classifier’, is a sort of interactive software tool which is able to learn what the relevant classes are and is then directed to discover patterns in the data.

Jarvis concentrates in his paper on the role of supervised classifiers for second language research using learner corpora. According to him, their merit is that they “highlight the way in which multiple language features work together

in the language use of individuals sharing particular background characteristics” (p. 130). He then gives an overview of the major types of classifiers in relation to the kinds of algorithms they are based on, e.g. *Linear Discriminant Analysis* (LDA). To face the challenge of having to decide which classifier is best suited, Jarvis follows Kotsiantis’ (2007) advice to rely on the majority vote of an ensemble of classifiers.

After sketching previous studies of L1 detection, Jarvis moves on to describe his research project. He examines the ability of twenty classifiers based on various algorithms to identify L1 related patterns in the use of 722 n-grams in L2 essays written by learners with 12 different mother tongues (taken from ICLE). He finds that the best performing classifiers for his task are Linear Discriminant Analysis (LDA), Sequential Minimal Optimization (SMO), Naïve Bayes Multinomial (NBM) and Nearest Shrunken Centroids (NSC) with only little difference between them in their ability to learn the n-gram patterns.

To sum up, Jarvis envisages the development of classifiers which perform cross language comparisons and retrieval as something which could improve the ability of classifiers to discover direct effects of one language on another. Thus, we can expect more interesting findings here in the future.

Anna Mauranen opens Chapter 7 with emphasizing the importance of empirical corpus research for discovering new facets of language as such and for the practical application in the area of language learning and teaching. In her paper, she concentrates on the comparison of learner corpora which consist of data produced by second language learners with lingua franca corpora (ELF) which contain data produced by non-native speakers who use English as a contact language. Therefore she distinguishes between second language acquisition (SLA) and second language use (SLU), even though it is obvious that the roles can be alternating. The differences between SLA and SLU can be subdivided according to social, cognitive and interactive parameters.

Social contrasts include the diverse cultural backgrounds ELF users obviously have, whereas learners often share the same L1 and are therefore socially orientated towards the new language. Hence, English speaking countries constitute the target set for them, but to achieve that target is for most beyond reach. According to Mauranen, this perception of being deficient communicators has to be changed because there exist far more ELF users than native speakers of English, and their potential influence on the English language is considerable. Their playing central roles in international companies, politics and science could predispose them to holding the key to the future of English. She considers as the most likely field of change the one of phraseology and examines in a case study the phraseological frame *-ly speaking* (e.g. *frankly speaking*) in two corpora, the

MICASE (Michigan Corpus of Academic Spoken English) and ELFA (English as a Lingua Franca in Academic Settings). She finds frequency shifts in the ELF data and argues that this will eventually affect English usage.

One of the most important cognitive differences is that ELF users focus their efforts on making sense; for them content rules over form. From an interactional point of view the situation is very diverse because discourse norms are not given but must be negotiated by participants. The described contrasts are also reflected in the different principles of ELF corpus compilation. Mauranen concludes that the main dissimilarity between SLA and SLU corpora is the role of language: the object of study versus language as a means of achieving particular objectives in real environments. But, on the other hand, they are alike as regards the key principle: they both collect data from speakers using a non-native language. Consequently, an exchange between these two fields of corpus research is of mutual interest.

In the next chapter, Norbert Schmidt and Stephen Redwood present us with an in-depth study of productive and receptive learner knowledge of phrasal verbs (PVs). In terms of their frequency, PVs constitute a key feature of spoken and written English. However, learners and teachers of EFL often consider them as difficult to acquire. The authors aim firstly to explore the link between PV frequency and their productive and receptive learnability and secondly, whether the mode (spoken/written) has any influence on that. Additionally, they take into consideration various extra-linguistic factors, such as the gender and age of their participants. The authors compiled a test which consisted of a productive part (a cloze test with first letter prompts) and a receptive one (multiple choice options). After the tests had been completed, correlations were carried out using the BNC as a reference corpus. The findings are somewhat surprising as they show that the extent to which frequency and learnability are related is rather low and that there is virtually no difference with regard to the mode of the data. However, there seems to be a disparity concerning productive and receptive knowledge. To acquire receptive mastery, the learner's exposure to PVs does not have to be as intensive as when developing productive knowledge. Schmidt and Redwood validated their results with the Corpus of Contemporary American English (COCA).

Very interestingly, there was a considerable amount of high frequency PVs which nevertheless had low scores in the learner data. The authors offer a number of possible reasons for this phenomenon, e.g. non-corpus-based teaching material and one-dimensional exercises. Using a questionnaire, Schmidt and Redwood looked also into various extra-linguistic aspects and their importance for learning PVs and found that real life experiences of the English language

(e.g. extensive reading) had positive effects. The authors end their paper on the positive note that their participants had a sound knowledge of the most frequent English PVs, even though a lot more research into this field has yet to be done.

At the beginning of his paper (Chapter 9), Christian Mair provides a brief historical overview of corpus linguistics' contribution to the research field of English varieties. Corpus research has been done in the classical varieties British English and American English, but also in the standard varieties of the New Englishes. Mair's present focus is on non-standard, vernacular varieties such as the Corpus of Cyber-Jamaican (CCJ).

His working definition of New English is "any postcolonial variety of English which is undergoing the process of endonormative stabilization and standardization that Standard British and American English [...] have completed" (p. 211). The International Corpus of English (ICE) project has been studying the different varieties of English all over the world but concentrates very much on Standard English. Mair regrets the fact that, therefore, vernacular languages like Jamaican creole (JC) are not included, although they are based on a distinctive grammar and phonology. Nevertheless, for the present study he makes use of the Jamaican part of ICE (ICE-JA).

In his research, Mair examines Jamaican English (JE) and JC in the CCJ, a large corpus of diasporic Jamaican web posts, which was compiled at Freiburg University and amounts to about 16 million words. The compilation followed certain criteria, such as the inclusion of numerous participants of diverse backgrounds and a broad range of topics. At the moment individual speakers' profiles are being compiled and analyzed so that new insights are yet to come.

Mair's findings illustrate that e.g. the basilect variants of the *going-to* future are overrepresented in computer-mediated communication in comparison to traditional writing and – remarkably – face-to-face contact taken from ICE-JA. Mair offers several explanations for this phenomenon, but for him the key role in lowering the speakers' inhibition to use more JC variants is played by anti-formality. The special environment in a web forum seems to enhance a Jamaican tendency to use language in a playfully creative and – above all – anti-formal way. This trend is shown at different levels of language, even in spelling. Other applications of the CCJ lie in sociolinguistic questions and in the investigation of the "globalization of vernacular languages" (p. 229). The author discusses the spreading of vernacular features, e.g. the quotative *be like* or the use of African loanwords which, according to the OED, belong to the inventory of *exotica*. Although they are mostly used in the specialized discourse of African literature, one finds several occurrences in the CCJ. Mair argues that this surprising discovery could contribute to a study from a language and globalization perspec-

tive. In his conclusion, the author makes the case for the compilation of web-derived multilingual corpora, as they are the research tool of choice for studying language in a globalized and multilingual reality.

In Chapter 10, David Wible and Nai-Lung Tsao look critically at the role of corpora in the development of lexical resources for language learning, and the following quotation summarizes their view quite aptly: “[There is] a gap between the sorts of knowledge language learners need, on the one hand, and the sort of thing a corpus is, on the other” (p. 251). Their line of reasoning is that learners need to fulfill too many prerequisites to successfully use a corpus; e.g. they have to specify a pattern in their query to find multi-word expressions. But often L2 learners are not even aware that there might be a pattern. Consequently, Wible and Tsao focus in their article on presenting a lexical resource which could supplement corpora and dictionaries in providing a knowledgebase which overcomes this dilemma.

Traditionally, linguists have extracted lists of n-grams from corpora to research patterns of word behaviour. However, the authors claim that n-grams are only one-dimensional in the sense that they do not take into account the relation between expressions like e.g. *consider him lucky* and *consider yourself lucky*. This lack of the paradigmatic dimension is also the reason why n-grams are always represented in lists. In contrast to this linear format, Wible and Tsao aim at developing a more-dimensional and navigable web which would better resemble the neural web. Drawing on the field of collocation research, they consequently launched the lexical knowledgebase StringNet, which works with hybrid n-grams instead. It allows relations and similarities among them to be discovered. Moreover, they implemented a concordancer through a web-based search interface called LexChecker, which allows detecting patterns in word behaviour, e.g. colligations. The authors provide access to StringNet through a browser-based tool which detects collocations on the current webpage and presents them in a dropdown menu. Therefore, it can highlight multi-word expressions which the users may never have thought to look up and thus helps bridge the gap between the learners’ needs and what conventional corpus-based learning material fails to provide.

In the last chapter of *A taste for corpora*, Michael Rundell and Adam Kilgarriff review the development of the creation of dictionaries from the early beginnings of computer-aided lexicography to the present day. Their focus is on how machines have taken over more and more tasks from the humans involved. In their view, the start of the COBUILD project marks the kickoff for modern lexicography because it was the first time a systematic corpus-based methodology was applied. Therefore it represents a paradigm shift in that field. The main

characteristic of the 80s and 90s was the growing importance and availability of computers in many disciplines. For linguistics, this improvement assisted the advent of larger corpora with advanced corpus query systems which made an efficient data analysis possible. However, Rundell and Kilgarriff take care to point out that in those days the automation of processes was still confined to very few areas of lexicography. The development since 1997 is marked by the widespread access of the internet. The 21st century saw the development of enormous web-based corpora which proved useful for lexicographers working on general purpose dictionaries. In the authors' view the often-criticized web corpora can stand their ground in comparison to standard references like the BNC. In the following, the main tasks necessary for creating a dictionary are listed and assessed with regard to its progress of automation, e.g. the compilation of headword lists.

Providing systematic information about lexico-grammatical word behaviour in the form of e.g. collocations or colligations proved a great step forward in the process of automation of lexicographical tasks, because these so-called word sketches allow users to find the distinct senses of a word. Throughout their article the authors concentrate on showing how computerization has not only eased the labour-intensive work of lexicographers – or ‘drudgery’ as Samuel Johnson called it – but also contributes to the reliability and consistency of corpus-based dictionaries. Although they envisage a further fruitful collaboration between the fields of lexicography and computational linguistics, they have to admit that “[a]utomated lexicography is still some way off” (p. 279). In any case, the work of lexicographers is currently changing from selecting and copying samples from corpus sources to validating and evaluating the choices made by the computer.

To honour Sylviane Granger and her groundbreaking work in the field of learner corpora, all papers in this volume have a strong focus on (learner) corpora and the transfer of results to the field of language learning. The great strength of this book is its diversity, since the selection of articles range from the creation and analysis of corpora, over the use of software tools in research and pedagogical applications, to the development of computer-related methodologies. Thus, the volume presents a state of the art account of Granger's research domain. Interested readers do indeed get a very good taste of corpus linguistic studies in (learner) language research.

References

- Biber, Douglas. 2006. *University language: A corpus-based study of spoken and written registers*. Amsterdam: John Benjamins.
- Kotsiantis, Sotiris. 2007. Supervised machine learning: A review of classification techniques. *Informatica Journal* 31: 249–268.

Joybrato Mukherjee and **Marianne Hundt** (eds.). *Exploring second-language varieties of English and learner Englishes. Bridging a paradigm gap* (Studies in Corpus Linguistics 44). Amsterdam and Philadelphia: John Benjamins, 2011. vi + 222 pp. ISBN 978-90-272-2320-3. Reviewed by **Sandra Mollin**, University of Heidelberg.

The title of the reviewed edited volume refers to the paradigm gap that Sridhar and Sridhar (1986) identified as existing between research on New Englishes on the one hand and research on Second Language Acquisition (SLA) on the other. In the 25 years since, this paradigm gap has remained unbridged, with World Englishes researchers refraining from taking SLA theories and concepts into account since they do not wish to see these new varieties equated with learner English, and with SLA researchers neglecting the special case of new varieties, rather focusing on individual language learning in classic instruction or immersion settings. However, as the editors, Joybrato Mukherjee and Marianne Hundt, write, “since both learner Englishes and second-language varieties are typically non-native forms of English that emerge in language contact situations [...], it is high time that they were described and compared on an empirical basis” (p. 2), for the mutual benefit of both lines of research. The present volume begins to bridge the paradigm gap by presenting contributions that expressly unite both paradigms. Most contributions originate from a workshop held at the ISLE-1 conference in Freiburg in 2008; yet the editors manage to avoid presenting a collection of only loosely connected papers, as is often the case with conference proceedings. Rather, by providing the contributors with lead questions, some or all of which were to be addressed in their paper, the editors achieve a good degree of coherence in the volume. These lead questions include how a distinction can be made between variety-specific features and learners’ errors (even if they are structurally identical), what the similarities and differences are in the

development of new varieties and learner varieties, whether and how classical tripartite models of English users or English-speaking communities can be upheld, and how corpus linguistic methods may be applied to describe and model ESL (English as a second language) and EFL (English as a foreign language) varieties.

Rather than presenting the contributions to the volume in alphabetical order of the first authors' names, as in the volume itself, they will be summarized here in broad thematic groups, depending on which approach the authors take in trying to reconcile the two paradigms of New Englishes and SLA research.

The largest number of contributions uses corpus linguistic methods to compare selected New Englishes and learner Englishes. The first of these, by Sandra Götz and Marco Schilk, focuses on formulaic language, comparing the frequency and range of 3-grams in spoken British English, spoken Indian English and English spoken by advanced German learners of English. Götz and Schilk find that, while the ENL and the ESL varieties do not differ significantly in their frequencies of recurring 3-grams, the learners use a significantly smaller number of these, pointing to an important potential difference between institutionalized varieties and learner English: since the English of learners is functionally restricted compared to that of ESL-speakers, it lacks a full formulaic repertoire.

Marianne Hundt and Katrin Vogel in their contribution investigate one specific feature that has received a relatively large amount of attention in both research on New Englishes and learner English: the progressive. They adduce corpus evidence for three ENL varieties and five ESL varieties and compare these to figures for different learner groups of English to see whether non-native speakers do, as reported, overuse the progressive. Their results are surprising in that there is no clear division between the ENL, ESL, and EFL groups as far as frequency of the progressive is concerned. In particular, the result for New Zealand requires an explanation since it has a far higher use of progressives than the other ENL varieties, close to the figures for the learner Englishes. Hundt and Vogel suggest that this may be the effect of ESL and EFL speakers influencing ENL speakers in New Zealand – a finding that, if it can be corroborated, would be revolutionary indeed. In a qualitative analysis, they also find an important difference between ESL and EFL usage of the progressives: ESL speakers creatively transfer the progressive to new stative verbs, while learners stick to the prototypical uses that they have been taught.

A further contribution that uses corpus linguistic means to compare Englishes from all three Kachruvian circles is the one by Benedikt Szmrecsanyi and Bernd Kortmann, even though they take a non-conventional perspective in corpus linguistics, that of the typologists. Szmrecsanyi and Kortmann, rather

than focusing on just one linguistic feature, classify varieties in terms of their profile of syntheticity versus analyticity. They define the degree of syntheticity as the frequency of bound grammatical markers, and analyticity as the frequency of free grammatical markers. In their results, the New Englishes (as in five ICE corpora) and learner Englishes (as in eleven ICLE-subcorpora) emerge as two very distinct groups, with learner English generally being more analytic than New Englishes, and New Englishes generally being more synthetic than learner English. These overall tendencies can be traced back to a number of grammatical features: learners have markedly higher frequencies of pronouns (possibly as a result of an overuse of finite subordinate clauses), negators (possibly as a result of gaps in the lexicon which require paraphrases), auxiliary *do* (related to the use of negative paraphrases) and auxiliary *have* (possibly as a result of an overuse of the perfect), while there is a lower frequency of inflected verbs. These features could well be taken up in future research. The typological profiling approach could thus serve as a first bird's-eye view of the differences and similarities between varieties, which more qualitative studies could then investigate in more detail.

Interestingly, the other contributions in the collection do not attempt to bridge the paradigm gap between research on ESL and EFL by comparing varieties from both sides of the putative fence, but take different approaches. For example, both the contributions by Biewer and by Hilbert consider a number of New Englishes only, but interpret the results in the light of SLA concepts and processes, while the paper by Gilquin and Granger interprets findings on different learner Englishes in terms of a cline of exposure to the target language, which could accommodate New Englishes as well. To begin with this latter approach, Gaëtanelle Gilquin and Sylviane Granger focus on one feature that is well known to cause learners of English difficulties: the preposition *into*. They are able to show in a series of meticulous analyses regarding the use of the preposition syntactically, semantically, and phraseologically that the three learner Englishes discussed (ICLE-Dutch, ICLE-French, and ICLE-Spanish) typically form a cline, with the variant that is characterized by the highest degree of exposure to English for its speakers, Dutch, coming the closest to the British English benchmark, and Spanish, with the smallest degree of exposure, being the furthest away from it, with French falling in between. The fourth variant they consider, Tswana English, falls out of this picture since, as an emerging New English, it is subject to variables other than mere exposure to the target language. In the different analyses, it occupies different positions along the cline.

Carolyn Biewer focuses on modal auxiliaries in newspaper corpora representing several New Englishes (Fiji, Samoa, Cook Islands, Singapore, Philip-

pires and Ghana) as well as three control ENL varieties (British, American, New Zealand). In general, she finds more variability in the frequencies of individual modals among New Englishes than among native-speaker varieties as well as a certain restriction of the modal repertoire in the new varieties. The similarities in patterns of preferences for modals in New Englishes are explained in terms of a number of principles from SLA theory. For example, the 'transfer to somewhere principle' (only unmarked features of the target language are transferred into the interlanguage) is evoked to explain the fact that deontic meanings appear to be overused in the ESL-varieties. Similarly, differences in the preference for one modal at the expense of another between the varieties is explained in terms of the 'shortest path principle' (if there is variation in the target language, the interlanguage will select only one variant, namely the one with the closest correspondence in the mother tongue).

Michaela Hilbert considers interrogative inversion in three varieties of English: the new varieties of Indian English and Singapore English as well as Irish English, an ENL-variety with a contact history. The developmental sequence in SLA regarding interrogative inversion is well known, and Hilbert sets out to test whether the non-standard inversion patterns found in her three varieties can be compared to learners' errors. In her analysis of the syntactic contexts in which non-standard inversion occurs, however, she finds that this is not the same phenomenon in all three varieties. For the inversions found in polar interrogatives in Irish English, only transfer is a plausible explanation, while Hilbert sees an intriguing unitary explanation for embedded inversion in all three varieties as well as in learner SLA: formulaicity. These inversions occur only for a small number of subject+verb combinations, especially *it* or *there* plus forms of *to be*, suggesting reproduction of a chunk in this context rather than rule overgeneralization.

Two further contributions in the volume address two Englishes whose status is still disputed, so that their analysis may shed further light on the potential boundaries between ESL and EFL: Cyprus English and Black South African English. Christiane Bongartz and Sarah Buschfeld report on the first results of a larger study of the status of English in Cyprus, integrating the collection and analysis of a corpus and a survey of language attitudes. They present a number of features which may turn out to be Cypricisms and propose to map the use of these features onto a spectrum of sociolinguistic variables in order to make their spread visible. In general, Bongartz and Buschfeld emphasize that we can only speak of a new variety if features are communal and stable, and accompanied by a process of identification in the community. Bertus van Rooy also suggests two criteria to distinguish between learners' errors and conventionalized innova-

tions, namely grammatical stability and acceptability. Two features of Black South African English are analyzed: the progressive used with stative verbs and the use of *can be able to*, showing that they are stable and systematic features of the variety, rather than an unsystematic violation of an external norm. In addition, Van Rooy reports on acceptability studies and anecdotal evidence to the effect that the features are accepted by large parts of the speech community. The same goes for a third example presented, the combination of *enable* with bare verb in East African English. Both the papers by Bongartz and Buschfeld and by Van Rooy in this volume therefore draw attention to an important extralinguistic factor in the distinction between varieties and learner Englishes, namely the social acceptance of new features, as it is also mentioned in the classic New Englishes literature (e.g. Kachru 1992: 56, Moag 1982: 12, Schneider 2003: 249). Unfortunately, this is neglected in most other contributions to the volume, possibly because it requires corpus linguistic methodology to be combined with extensive acceptability studies.

I have kept for last the only purely theoretical contribution to the collection, Ulrike Gut's paper focusing on the study of linguistic innovations in varieties of English. Gut reviews previous empirical studies in the field to the result that they are frequently naïve as far as SLA theories go, typically stating only that innovations may be traced back to mother tongue transfer. She presents a comprehensive survey of cross-linguistic influence in SLA and from this, she offers a new model for the evolution of New Englishes. The main tenets that are going to be news to researchers in the New Englishes paradigm are that cross-linguistic influence is unlikely in many New Englishes, since this is rare with typologically unrelated languages, and that indirect transfer is more likely than direct transfer, since the latter predominantly occurs only in the early stages of language acquisition. Following from these insights, Gut suggests that future research focusing on structural innovations needs to take into account whether the same innovations also occur with New Englishes with typologically similar or different substrates and whether explanations other than cross-linguistic influence may be responsible, such as internal regularization processes and external, attitudinal factors. She thereby outlines a best practice model for future research that no longer relies on the cover-all explanation of mother tongue.

The volume concludes with a brief discussion forum, in which the editors sum up the most important issues that the contributions raised, with comments made by some of the authors. All in all, the editors are to be congratulated on bringing this topic to the fore and on recruiting such thought-provoking contributions. The volume is indispensable for scholars in the field of New Englishes, but will also prove interesting to researchers of learner English and generally to

corpus linguists interested in variety comparisons. The paradigm gap has received a tentative bridge in this volume.

References

- Kachru, Braj B. 1992. Models for non-native Englishes. In B.B. Kachru (ed.). *The other tongue. English across cultures*. 2nd ed., 48–74. Urbana and Chicago: University of Illinois Press.
- Moag, Rodney. 1982. English as a foreign, second, native and basal language: A new taxonomy of English-using societies. In J. Pride (ed.). *New Englishes*, 11–50. Rowley, Mass.: Newbury House.
- Schneider, Edgar W. 2003. The dynamics of New Englishes: From identity construction to dialect birth. *Language* 79(2): 233–281.
- Sridhar, Kamal K. and Shikaripur N. Sridhar. 1986. Bridging the paradigm gap: Second language acquisition theory and indigenized varieties of English. *World Englishes* 5(1): 3–14.

Päivi Pahta and **Andreas H. Jucker** (eds.). *Communicating early English manuscripts*. Cambridge: Cambridge University Press, 2011. 312 pp. ISBN 978-0-521-19329-0. Reviewed by **Melanie Borchers**, University of Duisburg-Essen.

The four parts of the volume focus on combinations of codicology, mediality and discourse studies of manuscripts ranging from the 14th to the 19th century and their communication with 21st-century readers. The papers present various approaches to a wide variety of genres and thus offer manifold perspectives on and of the texts under investigation. It is their diversity, i.e. the creative spirit of each individual article, that pays tribute to the importance of pragmatic analyses of historical discourse. And while this collection of texts is not a *festschrift*, the contributions read as if they were a well-mixed and, nevertheless, wonderfully harmonious birthday bouquet for Irma Taavitsainen.

The first five chapters concentrate on authors, scribes and their audiences, which means that they focus on the respective perspective within and towards the manuscript(s) at hand. The first contributor, Thomas Kohnen, investigates

the role shift in the communication in a 15th-century commonplace book. According to Kohnen there are four distinct shifts in the roles of the original senders and addressees that individual texts within the commonplace book might be fitted into: receiver-receiver, composer-receiver, receiver-sender and receiver-scribe. He presents four convincing exemplary classifications and thus demonstrates how these shifts trigger the multi-functionality of commonplace books with respect to their constative as well as performative text functions (p. 21). This classification which needs further investigation on a larger text basis in order to generalise Kohnen's claims is, however, an innovative contribution to genre studies and might prove beneficial to them.

In line with Coleman (1996), Gabriella Del Lungo Camiciotti provides evidence against the commonly accepted assumption that book printing changed an oral to a visual culture. By means of analysing the rhetorical strategies in two mid-15th-century manuscripts from the religious domain, she concentrates "on how the relationship between author and audience is built" (p. 27). The "audience-oriented devices clearly point to the use of written text for reading, not merely as script" (p. 36) and thereby propose another argument in favour of Coleman's assumption that England remained an oral culture. She considers instead the shift from an oral-based culture to silent reading decisive for the "reconceptualisation of the book" (p. 37). By enlightening upon the discursive space of the author (e.g. narrator-in-the-text and meta-discourse) Del Lungo Camiciotti conclusively argues in favour of a more complex shift than widely believed.

In the third chapter Patricia Deery Kurtz and Linda Ehsam Voigts provide us with a detailed description of a scattered 16th-century manuscript that was dismembered during the 1980s. MS Bute 13 once represented a "unified and purposeful compendium for astronomical and astrological calculation and prognostication" (p. 54). Its unique Middle English and Early Modern English as well as its Latin texts contribute to the compilation's importance. With their reconstruction of the original manuscript Deery Kurtz and Ehsam Voights thus contribute to our understanding of early scientific knowledge discourse.

Maurizio Gotti and Stefania Maci propose a reinterpretation of the linguistic variation in MS Digby 133. Whereas former research considered the variation merely incompetent and careless inconsistencies on the part of the *Mary Magdalene* scribe, the present contribution provides evidence in favour of the fact that the variation is due to poetic effect. While for pious characters the scribe is most conservative, the most innovative diatopic as well as diastratic variations characterise pagan discourse; i.e. they represent the "result of a struggle between Good and Evil, expressed in words" (p. 62). By taking an unbiased

approach to the apparent arbitrary variation in the 17th-century hand, Gotti and Maci uncover systematic indicators of social status and moral values that marvellously display the author's original linguistic design.

The last paper in this part focuses on reporter-reader interaction in the proceedings of the Old Bailey Corpus 1674–1834 and the truth value attached to it. Elizabeth Closs Traugott examines the interaction with the audience in London's central criminal court during the 17th and 18th century. While reporters first drew highly subjective and sensational accounts of the proceedings, thereby constructing their audience and a commercial enterprise alongside, the later proceedings display more objective interaction. According to Closs Traugott it is the Mayor of London's 1679 regulation to "provide a 'true, fair and perfect narrative'" (p. 80) that is assumed to have brought about the shift towards neutral accountability. She hence allows insights into the text type of court proceedings and thereby adds to the mosaic on the only recently compiled Old Bailey Corpus.

Merja Stenroos and Martti Mäkinen open the second part of the collection, which focuses on communication through handwritten correspondence. By means of a linguistic investigation of dialectal and pragmatic evidence within the correspondence between the Welsh rebel, Gruffuth, and his English adversary, Lord Grey, during the Welsh Uprising (1400-1414), Stenroos and Mäkinen re-establish Gruffuth's status from "strengest thiefe of Wales" (p. 99). Like in the above-mentioned contribution by Gotti and Maci, an unprejudiced approach to the correspondence of the two historical figures rather affirms the contrary (e.g. avoidance of face-threatening acts). The article thus presents one of the extraordinary instances where a neutral and factually-based linguistic analysis of the original documents causes a socio-historical re-evaluation of historical events.

The next contribution also investigates the letter genre. Minna Palander-Collin and Minna Nevala study the interplay of 1st-, 2nd- and 3rd-person reference with particular regard to the construction of macro- and micro-society; i.e. the social and the individual factors respectively. By providing evidence from Nathaniel Bacon's 16th-century correspondence within the contexts of family, business and county administration, the contributors relate their findings to Layder's (1997) sociological domain theory. They survey how reference places a person within a social framework (e.g. social rank and gender as well as their relation towards their correspondent). Their method being firmly established in sociolinguistics, Palander-Collin and Nevala contribute further evidence in favour of the hypothesis that language use mirrors social standing.

In her investigation of two 17th-century letters by the diplomat and poet George Stepney, Susan M. Fitzmaurice addresses the intriguing topic of literary collaboration. One of the letters was addressed to Stepney's publisher and another to his friend poet Matthew Prior. Both letters illuminate "the highly interactive, and indeed sociable, nature of literary labour in the late seventeenth century among the members of a social network" (p. 131) and hence "may be more revealing of how close friends approach openness to criticism and the desire for editorial advice in private than the letters published as ostensible models of communication" (p. 132). Fitzmaurice thus allows the present collection to flourish with decisive details about our knowledge of 17th-century (literary) social networks.

With the help of several selected business letters from the Corpus of Nineteenth-century Scottish Correspondence, Marina Dossena examines the printed and handwritten elements and their use in business writing. According to her findings, the letters display face considerations and politeness that are not merely due to linguistic choice. It is also the physical appearance that contributes to the communication between the sender and the addressee and hence to the different degrees of personalisation. Dossena's investigation offers outstanding insights into the interplay of layout and linguistics and contributes important details to the varied text type of letter writing.

By choosing the *Romaunt of the Rose*, Graham D. Caie moves on to the third part, which focuses on the transition from manuscript to print. While comparing MS Hunter 409 as a potential manuscript to Thynne's print edition, Caie performs masterly detective work. Although it is not the first study of its kind, Caie's thorough codicology study closely examines linguistic alongside layout indices (e.g. transcription errors as well as column marking) that genuinely confirm the potential manuscript as the origin of the printed text. Besides adding to our knowledge of the shift from manuscript to print, editorial methodology is the major prime of this contribution.

On the basis of five 17th- and 20th-century play-texts Jonathan Culpeper and Jane Demmen study the causal effects of printing on the discursive structure and techniques. They conclusively interpret three extraordinary findings. First, due to the so-called cutting technique, a manuscript practice that supplied each actor solely with their own passages, the 17th-century plays show longer turns with less turn-taking. Second, due to less repetition, they are lexically richer and thus less cohesive. And third, the earlier texts confirm that by means of address actors do not only learn about their cues but also about their social relation to their dialogue partners. While certainly also other aspects (e.g. the question of

subgenres) might have played a role, Culpeper and Demmen convincingly present the medial effects of manuscript vs. print as one of the key factors.

The next two chapters are linked by their choice of domain, medicine. By means of quantitative evidence, Päivi Pahta, Turo Hiltunen, Ville Marttila, Maura Ratia, Carla Suhr and Jukka Tyrkkö compare three different 15th-, 16th- and 17th-century parallel versions of Galen's seminal *Methodus medendi* in the light of the textual tradition of works of medicine and their vernacularisation (e.g. the information structure). Not only do the contributors demonstrate the importance of the medium alongside socio-cultural changes, they also highlight some of the domain-specific peculiarities of the advancing vernacularisation in medicine.

Douglas Biber, Bethany Gray, Alpo Honkapohja and Päivi Pahta rather consider a grammatical issue in medical writing, namely that of *on* and *in* prepositional phrases. As there is historical evidence that points to a "fundamental shift in the discourse style of academic writing" (p. 199), the contributors study the use in specialised medical treatises in manuscripts in Middle English Medical Texts as opposed to first print versions in Early Modern English Medical Texts. While the concrete meaning of the two prepositions is most common in Middle English, the 16th and 17th centuries produce evidence of an extension to a more abstract meaning, the one prevalent in present-day written academic discourse.

Jeremy Smith and Christian Kay complete this section with their examination of six editorial interpretations of William Dunbar's poem "The Taking" and thus highlight the role of the third party in the communication between creator and receiver, the editor. The six versions range in the style of mediation between poet and reader from critical editions to conservative editions and provide evidence of the commonly acknowledged fact that the editorial process is always biased by interpretation, which itself is bound to individual and socio-historical factors and thus always represents transformations of the original.

In the fourth and last part of the collection of papers, the two contributions focus on manuscripts and their communicating characters. Andreas H. Jucker adds an investigation of greetings and farewells between fictional characters in the *Canterbury Tales*. His findings are three-fold: "considerable variability" (p. 233) and creativity, rather little formulaicity and a wide range of non-obligatory features (e.g. well-wishing). According to Jucker, the "great diversity" (p. 239) is not only due to the fact that all social classes are present in the *Canterbury Tales*, the inventory is even richer due to the fact that we find "characters both from realistic settings in fourteenth-century England and in fictional settings" (p. 239). Jucker's case study provides considerable insights into Middle English speech acts and their phraseology.

In contrast to the preceding contribution, Leena Kahlas-Tarkka and Matti Rissanen present a descriptive analysis of the newly re-edited Salem Witchcraft Trials and hence real historical characters as to the strategies the defendants chose to either deny guilt or confess to witchcraft. The contributors make a first close observation on the correlation of different uses of discourse strategies (e.g. use of figurative language, sarcasm or verbosity) and the defendant's attitude of response (e.g. factual vs. aggressive) with the defendant's success in saving their life in trial.

In a great number of different approaches the papers manifest the ways in which the investigation of written historical evidence adds to our understanding of early English communication and its development. Apart from outstanding contributions to codicology and mediality studies, *Communicating early English manuscripts* offers methodologically sound case studies of historical discourse.

References

- Coleman, Joyce. 1996. *Public reading and the reading public in late medieval England and France*. Cambridge: Cambridge University Press.
- Layder, Derek 2003 [1997]. *Modern social theory: Key debates and new directions*. London/New York: Routledge.

Martin Weisser. *Essential programming for linguistics*. Edinburgh: Edinburgh University Press, 2009. 184 pp. ISBN 978-0-7486-3856-7. Reviewed by **Christen Geisler**, Uppsala University.

Perl is an ideal programming language for corpus linguistics, since it makes it relatively easy to write short, but powerful programs, or 'scripts', that can carry out very complicated searches, count grammatical features and mark up texts in corpora. *Essential programming for linguistics* is a welcome introduction to Perl programming for corpus linguists. The book comprises 13 chapters and two appendices. The reader gets acquainted with the Perl language, from basic programming concepts to eventually writing a graphics-based KWIC-concordance program.

Chapter 1 introduces the reader to the Perl language and explains the programming environment on no fewer than three different platforms: Microsoft Windows, Apple's Mac OS X, and Linux. Chapters 2 through 4 present the essentials of programming in Perl. These introductory chapters provide a good general background on elementary concepts, and the pace of the chapters is relatively well suited for beginners.

Reading input and writing output are clearly treated in Chapter 5. It is unfortunate that there is no description of inputting several files (so-called 'globbing'), since corpus linguists often work with a number of input files in identical format. Instead, the reader is briefly introduced to the rather dangerous manipulation of directories (the reader is warned twice about fiddling with directories). Here, it would have been better to show how to process multiple input files, such as all text files in a directory.

Chapter 6 covers so-called 'regular expressions', which can be regarded as the core of Perl programming for researchers working with textual data. Regular expressions constitute a very advanced syntax for finding, comparing and replacing strings in text. The treatment of regular expressions in Chapter 6 is in my opinion too short and too fast, with too few illustrative examples. This chapter would benefit from a clearer presentation of this very central programming concept. The presentation could also have been more extensive, with plenty of simple examples to demonstrate the power of regular expressions.

Various types of search-and-replace operations are exemplified in Chapter 7, and we are also shown how to write a simple KWIC-concordance program (Exercise 24). The concordance program works well, and it is not difficult for a beginning Perl writer to modify it to suit his or her special requirements, which is after all the main advantage of writing your own programs. Chapters 8 and 9 cover different ways of sorting data, and the reader learns how to write a program that produces both an alphabetically sorted vocabulary list and a frequency list. One of the strengths of the book is the practical use of hashes (or 'associative arrays') in the sorting programs. Hashes are one of the more useful features of Perl for text manipulation. Chapters 6 through 9 cover the most essential tasks in corpus linguistics.

Chapter 10 introduces the writing of subroutines in Perl programming, and Chapters 11 and 12 deal with object-oriented programming and GUI (graphical user interface) programming, respectively. These three chapters race through the various topics at such great speed that a beginner will probably feel somewhat lost.

Each chapter in the book includes various types of exercises with brief instructions, and the corresponding Perl code is placed in an appendix. In many

instances, however, the instructions are far too sparse to enable the reader to be able to write the more complex programs. Instead, one would probably jump to the appendix, simply type in the program in an editor and then execute it. Since the book is clearly targeted towards beginners, it would have been preferable if the shorter exercises had been incorporated in the body of the text and commented on by the author. As the book is organized now, readers have to flip back and forth between the particular chapter and the corresponding Perl code in the appendix.

In the blurb, the book is advertised as a “gentle” introduction to Perl programming. However, some of the later chapters are far from gentle and even very difficult to grasp. In addition, the book is designed with a very compact layout, which adds to the denseness of the content at times. A linguist with no prior programming experience will probably find the final chapters (Chapters 10 through 12) almost incomprehensible. In many instances, the author provides a small number of examples that are too complex for beginners. That said, Chapters 1 through 9 provide a very good overview of learning Perl.

In conjunction with *Essential programming for linguistics*, a reader wishing to learn Perl should also consult a general introduction such as Schwartz *et al.* (2008), or make use of the abundance of information about beginning Perl programming on the Internet, or buy one of the many ‘teach-yourself’ books available for beginners.

All in all, *Essential programming for linguistics* serves as useful introduction to Perl programming. A simple concordance program, which the user can easily modify, and a sorting program for vocabulary and frequency counts are very important products of the book for any corpus linguist with no or very little programming experience.

Reference

Schwartz, Randal L., Tom Phoenix and Brian D. Foy. 2008. *Learning Perl*. Sebastopol, California: O’Reilly.

