

## Reviews

**Sylviane Granger, Jacques Lerot and Stephanie Petch-Tyson** (eds.). *Corpus-based approaches to contrastive linguistics and translation studies*. Amsterdam-New York: Rodopi, 2003. 219 pp. ISBN 90-420-1046-0. Reviewed by **Erik Smitterberg**, Stockholm University.

With the advent of computerized corpora, a new subdiscipline of linguistics usually referred to as “corpus linguistics” has developed. One of the characteristics of this subdiscipline is a rigorously empirical methodology.<sup>1</sup> Corpus linguistic methods have also increasingly influenced other disciplines where the use of empirical data is considered important. Two such disciplines, which have a number of additional concerns in common, are Contrastive Linguistics (CL) and Translation Studies (TS). *Corpus-based approaches to contrastive linguistics and translation studies* reflects not only the impact of corpus linguistic methods on these two disciplines but also the extent to which the use of empirical data has brought them closer, despite their differences regarding the aims of research. A volume such as that reviewed here thus has the potential to alert CL practitioners of advances in TS and vice versa, and also to raise CL and TS scholars’ awareness of the potential of exploiting corpora.

After a Preface, the volume is divided into three sections: Theoretical Approaches (three contributions), Corpus-based Case Studies (six contributions), and Cross-linguistic Tools and Applications (four contributions). The first article in Section I, by Sylviane Granger, briefly describes the development of CL and TS. In her account of the two disciplines, Granger considers both similarities, e.g. the increased reliance on empirical data, and differences, such as the use of different labels for the same types of corpus. In a useful survey of corpora used in crosslinguistic studies and the uses to which such corpora can be put by CL and TS specialists, Granger also tries to remedy the problem of different labels by presenting a system common to both disciplines. Furthermore,

Granger recognizes limitations of corpus-based approaches, e.g. the lack of suitable corpora and/or the difficulty of (semi-)automatic retrieval: owing to these limitations, she claims that “[w]hat matters is the use of solid empirical data, whether electronic or not” (p. 23). While this is certainly true, it might be added that, even in cases where manual retrieval of data proves necessary, there are still advantages to using electronic, publicly available standard corpora, as this increases the comparability and replicability of studies. Granger highlights the need for better software tools, the potential importance of multilingual corpora to teaching, and the advantages of more cross-disciplinary co-operation.

The topic of Stig Johansson’s contribution is the meeting of CL, broadly defined, and corpus linguistics. Like Granger, Johansson begins with an overview of the history of CL, after which he focuses on the role of multilingual corpora in this discipline; this discussion, which also addresses several relevant issues in corpus selection, compilation, and design, takes Aijmer and Altenberg (1996) as its starting-point. First, comparing languages may reveal facts about these languages that a monolingual investigation would not produce. Second, translation/parallel corpora can be used to highlight patterns of cross-linguistic equivalence by charting translation equivalents. Third, using translation corpora enables researchers to uncover both specific, source-language-induced features and general features of translated texts. Fourth, Johansson lists a number of practical applications, including materials design for the training of translators. Finally, he lists a number of future challenges, such as developing usage-based grammars and dictionaries. Owing to the many different topics covered in separate subsections, Johansson only provides a brief overview of each topic, but his contribution is still valuable as an introduction to the potential of corpora for CL specialists.

Sara Laviosa’s contribution addresses the potential role of corpora in TS, and more specifically the relation between Descriptive Translation Studies, as outlined by Toury (1995), and Corpus-based Translation Studies. She defines the latter as “the branch of the discipline that uses corpora of original and/or translated text for the empirical study of the product and process of translation, the elaboration of theoretical constructs, and the training of translators” (p. 45). After discussing Descriptive TS, she points to its similarities to Corpus-based TS, such as the use of large numbers of authentic data and a probabilistic rather than prescriptive perspective. Laviosa also considers differences between the two branches: for instance, unlike Descriptive TS, Corpus-based TS is not concerned with translators’ decisions, only with texts. Such differences notwithstanding, Laviosa considers the two branches similar enough to benefit from each other’s achievements. Of particular interest is perhaps Laviosa’s suggestion

that the use of corpora may have an impact on theory-building in Descriptive TS. She hypothesizes that it may no longer be necessary to make theoretical a priori postulations about what is possible in translation once large-scale corpora that enable the observation of real-life behaviour are available. This suggestion implies that the use of corpora has the potential to change certain aspects of TS profoundly.

The contributions to Section II are arranged along an approximate continuum from CL to TS; while not all of them are based on electronic corpora, every study draws on empirical data. The first case study, by Kristin Davidse and Liesbet Heyvaert, concerns the middle construction (also known as the mediopassive or activo-passive) in English and Dutch, as in the English example *Broiler rack removes easily*. Davidse and Heyvaert start out from two existing treatments of the middle construction, referred to as “the ‘ergative’ approach” and “the ‘transitive’ approach”, respectively. The authors use corpus data as well as accounts of the middle construction in Dutch to demonstrate that neither analysis is complete. The contrastive perspective is thus used to enhance the description of an individual language (cf. Johansson’s contribution above). The authors propose an analysis of the middle construction that centres on “*the constructional link between a non-agentive (patientive or circumstantial) Subject and an active VP*” (p. 63 [emphasis original]), which then enables them to consider different transitivity types within the same description. They also claim that the non-agentive subject of middle constructions is presented as conducive – or not conducive – to a process, and that this subject-centred meaning is modal in nature (it is suggested that reference to specific properties of the subject is implied). The authors’ analysis is interesting. They also point to one of the limitations of corpus-based approaches: as the middle construction is infrequent, even a large corpus of English such as the COBUILD corpus need not provide scholars with attestations of all variants.<sup>2</sup>

André Hantson’s contribution is chiefly devoted to two types of nominal clause: English gerund clauses and Norwegian *det(te)* + infinitive/*at*-clause constructions. The article is divided into two main parts. In the first part, Hantson argues that the distinction between action nominals (e.g. *their slow counting of the votes*), gerunds (e.g. *their/them counting the votes so slowly*), and participial clauses (e.g. *Reading a book, I heard a shout*) should be maintained, although a limited number of hybrid constructions can be found. Hantson then turns to an examination of Norwegian *det(te)* + infinitive/*at*-clause constructions and their English equivalents, drawing for data on a manual corpus consisting of Norwegian translations of Agatha Christie novels. Based on an analysis of the 61 occurrences found and their English equivalents, the author claims that the Nor-

wegian construction is more nominal than the English gerund, as both gerunds and more nominal English original constructions correspond to translated Norwegian *det(te)* + infinitive/*at*-clause constructions. The author also suggests that cases where a *det(te)* + infinitive/*at*-clause construction instead corresponds to, for instance, a *that*-clause do not undermine this claim, as the Norwegian pattern “then has a more nominal feel about it than the English original” (p. 82). Although the author supports his claim with a few corpus examples, this type of reasoning seems potentially dangerous, as corpus evidence in the form of translation correspondences is given less weight when it does not support the author’s hypothesis (*that*-clauses are in fact the most frequent English construction translated by a *det(te)* + infinitive/*at*-clause construction). Hantson further supports his claim by showing that 47 out of the 61 Norwegian examples occur in the highly nominal subject position. This is an important observation, although 61 instances are perhaps too few to permit definite conclusions.

The next article, by Mieke Van Herreweghe and Myriam Vermeerbergen, focuses on a comparison of Flemish Sign Language (FSL) and Dutch, and specifically on what may be termed FSL interference in written Dutch produced by deaf school-aged children. After a valuable survey of the linguistic situation of Flemish deaf children, the authors analyse linguistic material that consists of a retelling of an animated cartoon in FSL and written Dutch, respectively.<sup>3</sup> The analysis reveals several areas where FSL features appear to influence the informants’ written Dutch. Much of this influence appears to be due to FSL having spatial grammatical mechanisms that are absent from written languages. For instance, a signer may assign a locus, i.e. an area in space, to a non-present referent in a narrative; gestural reference to this locus can then function like features such as pronouns and arguments of verbs like GIVE. The deaf informants had problems with these features in their written Dutch, often avoiding pronouns and failing to specify arguments. Written Dutch grammar is thus not substituted for the spatial grammatical features of FSL, and the result is that their compositions resemble what the authors refer to as “a transliteration of some kind of *degrammaticalized* Flemish Sign Language” (p. 102 [emphasis original]). Van Herreweghe’s and Vermeerbergen’s approach is novel and has important pedagogical implications; quantification of their results would have added further to the value of their study. While the authors’ qualitative analysis shows clearly that the problems exist, their study does not, for instance, tell the reader which of the problems are the most frequent.

Hélène Chuquet’s contribution focuses on the French *imparfait* and its equivalents in English translations of French literary and journalistic texts. Chuquet claims that, while suitable translations of *l’imparfait* can always be found,

the English options often entail either gain or loss in translation. She aims to show that *l'imparfait* and the English simple past share more features than has often been assumed in previous research. Chuquet does this by means of close readings of a number of corpus examples. As examples of gains and losses in translations discussed by Chuquet, translations with past progressives may involve a gain in aspectual explicitness, while some translations with the simple past result in a loss of stylistic effects such as a subjective perspective. Nevertheless, Chuquet argues that the simple past and *l'imparfait* are similar in that they “share an indeterminacy that makes them both context-dependent and, as a consequence, enables them to take on a wide variety of shades of meaning” (p. 118). Thus, the simple past can be used to retain the notion of ambivalence inherent in some realizations of *l'imparfait*, although the type of ambivalence expressed need not be the same. Chuquet's discussion is qualitative, and she refrains from drawing quantitative conclusions, also mentioning the non-representative nature of her material (p. 107). As Chuquet points out, her material includes both French texts that are translations of English originals (p. 119n.) and French texts where she herself or colleagues of hers have supplied the translations (p. 107). Since the object of study is English translations of the French *imparfait*, it would have been methodologically safer to use a corpus consisting exclusively of French originals and their English translations, with the latter being produced independently of the author. Nevertheless, Chuquet's close readings of examples seem plausible, and it is to the author's credit that she makes such methodological limitations explicit.

Jonathan Charteris-Black's article adopts a quantitative perspective. The contribution is devoted to problems involved in the translation of idioms between English and Malay, and to pointing to the value of corpus-based work in this regard. The author begins by problematizing matters of corpus availability and comparability relevant to such an undertaking; phraseological comparability, in this case the issue of what can be classified as an idiom in the two languages under scrutiny, is also discussed. Charteris-Black uses prototype theory to arrive at a definition of idioms that is based on both conceptual criteria (e.g. the use of figuration) and linguistic criteria (e.g. syntactic and lexical stability). Among other things, Charteris-Black claims that corpora can help to specify the frequency of surface forms and to clarify the connotative values of the words that make up idioms (for instance, *hati* 'liver' connotes positively but *kaki* 'foot' negatively). The quantitative investigation focuses on idioms containing a body part (the idioms were selected from dictionaries). Charteris-Black distinguishes five different categories of correspondence depending on the degree of similarity of English and Malay with regard to surface lexis and/or

conceptual metaphor (the conceptual basis is given precedence in translation). The model is clearly outlined and illustrated with examples and their normalized frequencies, but occasionally the conceptual metaphor is not wholly clear to me. For instance, it is claimed (p. 131) that the conceptual metaphor for the idiom *black sheep* is DARKNESS IS SECRECY, where I would rather have expected DARKNESS IS BAD (DARKNESS IS SECRECY perhaps being more suitable for an idiom such as *dark horse*).<sup>4</sup> Overall, however, Charteris-Black's contribution identifies many uses of corpora in idiom research, such as enabling a focus on the most frequent ones and identifying cross-linguistic similarities and differences in conceptualizations (e.g. English NEW IS GOOD vs. Malay NEW IS BAD). It also points to the importance of corpus compilation.

The final case study is by Tiina Puurtinen; its topic is nonfinite constructions in original and translated Finnish children's literature. Four types of nonfinite constructions are considered, all of which can typically be replaced with a finite-clause variant. Based on criteria such as information/lexical density and grammatical redundancy, the nonfinite constructions are claimed to be more cognitively difficult than their finite counterparts. Given the difficulty of retrieving the relevant nonfinite constructions automatically, most of the data were retrieved manually from the corpus, which covers the years 1940–1998, with further subperiodization. All periods investigated reveal a higher frequency of nonfinite constructions in translated than in original Finnish texts, and the difference is statistically significant. The possible reasons for this difference mentioned by Puurtinen include a desire for formal equivalence to the source language (as nonfinite *-ing* clauses are common in written English), nonfinite clauses being a general feature of translated Finnish, and possible advantages of using nonfinite clauses. However, she focuses on the relation between her results and postulated translation universals. These universals stipulate that, compared with the language of non-translated target-language texts, the language of translated texts should be lexically and syntactically simple, contain a great deal of explicitness, be characterized by an exaggerated use of features characteristic of the target language, and favour concrete rather than abstract (senses of) words. Puurtinen's results seem not to conform to these postulated universals. Finally, Puurtinen addresses the advantages and drawbacks of corpus-based TS. Her contribution is well-reasoned, and she is also careful to hedge her claims and conclusions.

Section III begins with Philip King's contribution, entitled "Parallel concordancing and its applications". King describes and compares two parallel concordancers for Windows, Paraconc and Multiconcord, and discusses the uses to which such programs can be put. Parallel concordancers work with aligned ver-

sions of the same text (typically—but not necessarily—in different languages, e.g. an original and a translation), and present both the results of the retrieval procedure as a concordance for one of the texts and the corresponding sentence in the other text, thus enabling a comparison of, for instance, translation equivalents. The comparison of the features of these two programs is valuable, as such a survey may enable researchers with little knowledge of concordancing to decide which program is more suitable for their needs. It would have added further to the account if the author had included screen captures from both programs instead of focusing on Multiconcord alone in this regard; more space is devoted to Multiconcord than to Paraconc in other respects also.<sup>5</sup> The fact that King addresses limitations of this type of software clearly adds to the value of his contribution. The actual and potential applications listed include materials design, studies of translator behaviour, lexicography, and grammar studies. The combination of a description of relevant software and a discussion of its applications is likely to be appreciated by scholars who have little previous experience of parallel concordancing.

The second contribution, by Lynne Bowker, focuses specifically on the role of corpora in translator training. The author looks at two types of corpora: corpora created by translators, which may resemble learner corpora in the field of second language acquisition, and corpora created for translators, such as parallel corpora.<sup>6</sup> The combination of these two into comparable corpora is also discussed. Corpora created by translators may consist of student output: among other things, they can be used to examine a given section in several student translations and thus to identify weak spots, or to chart students' progress longitudinally. Corpora created for translators, in contrast, do not comprise translations, but are designed to help translators. Bowker points out that translating texts which contain language for special purposes requires specialized knowledge of the target language and of the subject field that may even go beyond that of native speakers. She shows that students who had used target-language corpora as translation resources generally produced more accurate translations of such texts than did students who had used traditional resources such as dictionaries. The combination of these two types of corpus produces what Bowker considers a special kind of comparable corpus, which can be used to assess student translations, and to research the translation process.

Natalie Kübler's and Pierre-Yves Foucou's contribution also has pedagogical implications regarding language for special purposes. It centres on difficulties that French learners of English face regarding the use of English verbs in computer science texts, and on how the use of corpora enables the authors to describe usage more comprehensively and to prepare teaching material. As dic-

tionaries of technical English often focus on nouns, little information on the use of verbs is provided, although verbs may cause learners more difficulty; in addition, dictionaries in this field quickly become obsolete (a point also made by Bowker above). The authors use Internet texts containing specialized and non-specialized English as their material. Their reference corpus also exists in French versions, which enables an investigation of French equivalents of English verbs. Problem areas for French learners addressed by Kübler and Foucou include new verbs created from nouns (some of which have no French equivalent), specialized uses of polysemous verbs that also exist with a different meaning in general use, and structural differences between English and French. The authors show that the frequency distribution of English and French verbs is not identical, as different verbs in one language may correspond to different uses of the same verb in another; moreover, a technical English verb need not have a stable French equivalent. The authors demonstrate that the use of bilingual corpora can facilitate the search for translation equivalents and improve the description of the English verb system for French learners. For example, the use of concordances makes it easier to find out which context tends to prompt which translation equivalent, and which syntactic structures are used with a given verb. The contribution makes a strong case for using corpora to a greater extent in language learning, both as reference tools and as the basis for exercises. However, it would perhaps have benefited from one more reading round, which might have improved aspects of the structure of the contribution, such as repetition (for instance, the reader is told that dictionaries of technical English focus on nouns in three places within the space of eight pages), and removed some remaining language mistakes (e.g. “with regards to” for *with regard to* on p. 185; the sentence “[c]omparing general corpus with specialised corpus for non-specialised verbs showed up frequency differences for different uses”, which is difficult to process, on p. 202; and “look for equivalences” for *look for equivalents* on p. 203).

The topic of the final article in the volume, by Elizabeth Dawes, is a dictionary project. Based on a corpus of newspaper and magazine texts, Dawes’s project concerns a dictionary of idioms which are attested in both the French of France and that of Quebec, and which have two or more idiomatic variants. Taking a contrastive approach, Dawes aims to uncover differences between the two regional varieties and to make a contribution to lexicographical analysis. Each variant of an idiom “is classified according to a typology of idiomatic variability” (p. 207), which enables a more systematic treatment of the idioms than that found in dictionaries; both variation and fixation are thus regarded as relevant to the description of idioms. Each idiom is seen as an abstract model, with each of

its different manifestations being a variant. A variant, in turn, is defined as “a conventional phrase with a fixed form whose meaning has to be learnt” (p. 210), a definition which is also elaborated; variants of the same idiom typically share at least one lexeme, but need not have the same meaning. The variants are coded in a relational database whose underlying typology is also changed to reflect the addition of new variants. Dawes illustrates various types of variants where the two regional varieties differ, such as morphological, syntactic, and lexical variants (e.g. *la cerise sur le gâteau* attested in France and Quebec vs. *la cerise sur la sundae* attested in Quebec only). Dawes’s approach gives an impression of thoroughness, although her article would have benefited from the inclusion of some quantitative data: for instance, it would have been very interesting to see how frequent the different types of idiomatic variability are in relation to one another in the data she has gathered so far. Dawes concludes that Quebec idioms are characterized by a conservative tendency as well as by innovations based on English loans and neologisms (e.g. *sundae* in the example above).

In sum, *Corpus-based approaches to contrastive linguistics and translation studies* shows clearly that CL and TS scholars would both benefit from increased co-operation. The editors’ suggestion that better corpora and software tools are needed is also borne out by the individual contributions, many of which are both insightful and novel in their approach. In addition, many of the contributions demonstrate that both CL and TS stand to gain from making use of corpus-based approaches to language analysis in their respective disciplines. However, some contributions also seem to indicate that the tenets which inform corpus linguistics as regards data selection and quantification have perhaps not yet been adopted fully by CL and TS practitioners. The editors’ choice to publish the volume in a TS series rather than one aimed at corpus linguists thus seems sound, considering that TS and CL scholars will probably learn more from the volume than will corpus linguists. In general, the structure of the volume is logical, and although there are language mistakes in some contributions, they are usually few and do not detract much from the overall impression of the volume. This collection will be of great value for CL and TS researchers alike; it is to be hoped that it will be followed by others as the two fields develop further.

### Notes

1. Rigorous empirical investigations of the language of principled collections of texts were of course carried out before the advent of corpus linguistics also. However, characteristics such as the importance placed on the collection of material and data, and on the statistical exploitation of the data, still

justify treating corpus linguistics as a separate, and recent, subdiscipline of linguistic studies.

2. For this reason, the authors chose to complement their corpus data with examples invented by native speakers. However, one option might have been to consider using Internet data culled from, for instance, a Google search of relatively reliable domains instead.
3. The 99 Dutch compositions analysed were produced by school-aged subjects, and the FSL signed narratives by eight signers between the ages of 8 and 36.
4. Nor is it clear why *black sheep* and some other idioms, such as *black market*, were included in the study, as they do not appear to contain a human body part in either the English or the Malay version.
5. King states (p. 161) that Multiconcord is used to illustrate the applications because he is more familiar with that program. However, it would have been valuable to see an applied discussion where the same case study is carried out in parallel fashion using both programs.
6. The author also discusses a difference between corpora and text archives: text archives are considered to be more random collections of texts from which stratified corpora can be extracted in accordance with particular research questions. In the interest of brevity, I will limit my discussion to the actual corpora extracted.

### **References**

- Aijmer, Karin, and Bengt Altenberg. 1996. Introduction. In K. Aijmer, B. Altenberg, and M. Johansson (eds.). *Languages in contrast. Papers from a symposium on text-based cross-linguistic studies, Lund 4–5 March 1994*, 11–16 (Lund Studies in English 88). Lund: Lund University Press.
- Toury, Gideon. 1995. *Descriptive translation studies and beyond* (Benjamins Translation Library 4). Amsterdam and Philadelphia: John Benjamins.

**Stefan Thomas Gries.** *Multifactorial analysis in corpus linguistics: A study of particle placement* (Open Linguistics Series). London: Continuum, 2003. xii + 226 pages. ISBN 0-8264-6126-3. Reviewed by **Christer Geisler**, Uppsala University.

This monograph investigates the placement of particles of transitive phrasal verbs, as in *Mary picked up the book* and *Mary picked the book up*, where the particle *up* can either precede or follow the direct object. On the basis of previous research, Gries identifies a number of features claimed to influence the position of the particle. A subset of these features is then tested statistically using both univariate (involving one variable at a time) and multivariate statistics (several variables). The study is corpus-based in that the analyzed examples come from the British National Corpus (BNC).

Chapter 1 serves as an introduction to the study, outlines the work, and includes a brief historical treatment of phrasal verbs. The variables/features that previous research has identified as affecting particle placement are described in Chapter 2. Chapter 3 is a short discussion of the overall aims of the study. Chapter 4 introduces the author's own *Processing Hypothesis* governing the choice of particle placement. The corpus data from the BNC are introduced in Chapter 5, and here the author describes the way in which the data were coded for statistical analysis. Chapter 6 presents the main statistical analyses, which are discussed further in Chapter 7. Various theoretical approaches to syntactic variation are elaborated on in Chapter 8, including a discussion of the position of quantitative analysis in linguistic research. The monograph concludes with a summary in Chapter 9.

The author's *Processing Hypothesis* is central in the study. For phrasal verbs, it can be summarized as: the construction with the particle before the direct object (construction<sub>0</sub>) will be chosen if the direct object requires a great deal of processing effort, whereas the construction with the particle after the direct object (construction<sub>1</sub>) will be chosen if this direct object requires little processing effort. From the listener's point of view, unidentified referents require more processing effort than identified referents.

The results of the quantitative part are presented in Chapter 6. Two types of statistical analyses are used: monofactorial (univariate) methods involving one variable at a time (Section 6.1) and multifactorial (multivariate) discriminant analysis using several variables (Section 6.3). The majority of the variables belong to four classes: phonological, morphosyntactic, semantic, and discourse-functional. Interestingly, in the monofactorial analyses, the morphosyntactic

variables come out as the strongest ones, followed by some discourse-functional variables. More specifically, the complexity, length, and the lexical form of the direct object are three important variables affecting particle placement. The most important discourse-functional variables include ‘last mention of the referent of the direct object’, ‘times of preceding mention’, and ‘distance to last mention of the referent of the direct object’. In other words, the complexity of the direct object as well as the degree of previous mention and topicality of the referent of the direct object prove crucial for the placement of the particle. Commendably, for the multivariate testing, Gries primarily uses discriminant analysis, which is a multiple regression method when the dependent variable is a discrete classification variable. The purpose of discriminant analysis is to find an equation which can predict the appropriate category (here construction<sub>0</sub> or construction<sub>1</sub>). The results of the discriminant analysis indicate that construction<sub>0</sub> is chosen if the direct object NP is long, lexical, indefinite, and construction<sub>1</sub> is chosen if the NP is simple and pronominal, and if it has been mentioned in the preceding linguistic context. Gries particularly emphasizes the discourse-functional variables, and especially those that involve preceding discourse. In other words, if the referent of the direct object has been mentioned frequently or recently in the preceding context, construction<sub>1</sub> is favoured. New referents, however, are generally introduced with indefinite and complex NPs and would consequently need more processing effort according to the author’s *Processing Hypothesis* and in those cases construction<sub>0</sub> is favoured.

As a second step in the discriminant analysis, by using so-called discriminant scores, one can predict whether an observation belongs to construction<sub>0</sub> or construction<sub>1</sub>. Convincingly, in Gries’ study, between 70 to 81 per cent of the examples are correctly predicted as belonging to either of the two constructions. Moreover, one valuable follow-up analysis involves the concept of prototypical constructions (Section 7.1). Based again on the discriminant scores, prototypical examples of the two types of construction can be identified. Gries illustrates this with an instance of a marked discriminant score that was correctly predicted as taking construction<sub>0</sub>, as in (1).

- (1) ... take up erm an interest or activity which will channel them  
into other activites (= Gries’ example (75))

In example (1), the verb-particle construction is *take up*, which is followed by an indefinite direct object NP containing a postmodifying relative clause *an interest or activity which will channel them...* From a strictly statistical point of view, (1)

represents a prototypical example of construction<sub>0</sub>; again, note the connection between syntactic complexity, indefinite reference and type of construction.

In both the univariate and the multivariate analyses, the morphosyntactic variables are important variables. In Chapter 8, however, Gries argues that it is possible to prove the existence of a causal relationship between the discourse-functional variables and the morphosyntactic ones (Section 8.3). That is, the syntactic complexity of the direct object NP is largely dependent on the information status of its referent, so that given referents are generally syntactically less complex and new referents are syntactically more complex. The most important overall finding is that particle placement is largely dependent on discourse-functional factors.

The study offers a wealth of interesting new facts about transitive phrasal verbs. However, clarifications and further attention paid to presentation would have been of help to the reader. For example, the presentation of variables/features to be subjected to statistical analysis is difficult for the reader to follow. Considering the complexity of the investigation, the study would have benefited from more extensive editing. Somewhat confusingly, the variables to be used in the statistical analyses in Chapter 6 are spread out over four chapters (Chapter 2, 4, 5, and 6): names of variables differ in the text, some variables are discarded in one chapter, and others are introduced in a later chapter. Sometimes the variables are italicized to simplify identification; yet at other times they are not. It is unclear in Chapters 2 and 4 whether the variable 'Production and planning effects' will be used and what it includes exactly; in Chapter 5, however, this variable is included and it is also exemplified for the first time. Particularly in Chapter 6, some discourse-functional variables are not mentioned in full but only as abbreviations: the variable NM (Next Mention) is not explained anywhere on p. 92, but is subsequently referred to as 'this variable'. Some variables are called something other than their semi-mnemonic code: for instance, ACTPC is used in the text for the variable 'Distance to last mention of the referent of direct object' (DTLM, p. 90). The same goes for the variable 'Distance to next mention of the direct object's referent', which is suddenly referred to as CLUSSC in the text (p. 93). It would have helped the reader if each variable had been assigned a mnemonic code immediately after its introduction in Chapter 6. There are no BNC references in the cited examples: the simple three-letter code from the BNC filenames would have sufficed after each example. Furthermore, most examples are short and simple, and more linguistic context would have been useful.

Despite the above editing problems relating to the naming and the presentation of variables, the present study is nevertheless an important contribution to our knowledge of transitive phrasal verbs. The author should be highly commended for his quantification of features treated in previous research and for the testing and coding of especially the discourse-functional features. Apart from shedding light on what linguistic variables affect the use of the two constructions, the study could also be very useful to beginning graduate students in the process of building statistical datasets from corpus data.

**Susan Hunston.** *Corpora in applied linguistics*. Cambridge: CUP, 2002. 254 pp. ISBN: 0521 80583 (paperback), 0521 801710 (hardback). Reviewed by **Ilka Mindt**, Universität Würzburg.

Susan Hunston fulfils what the title of her book promises: she focuses on corpus work in the broad field of applied linguistics. She considers aspects of language description by looking at the use of real language in the form of corpora and discusses a wide range of studies which have been undertaken with the help of corpora. Within the field of applied linguistics she covers subjects as diverse as dictionaries and grammars, ideology and culture, stylistics, language variation, and language teaching.

*Corpora in applied linguistics* directs attention to five areas:

1. General topics in corpus linguistics as well as the design and purpose of corpora (Chapters 1 and 2)
2. Methods in corpus linguistics such as the analysis of concordance lines and frequency lists, collocation, and corpus annotation (Chapters 3 and 4)
3. Different applications of corpora (Chapter 5)
4. Corpora and language teaching (Chapters 6 to 8)
5. The impact of corpora on applied linguistics (Chapter 9)

Each of the five areas will be discussed in the following. I will then move on and consider how Hunston goes about presenting her arguments.

Under the heading “Introduction to a corpus in use” **general topics** in corpus linguistics are dealt with. Hunston focuses on three fields of data processing: frequency, phraseology, and collocation. She starts by presenting corpus data in the form of frequency lists and groups of concordance lines which are then explained and analysed. In this chapter key terms of corpus linguistics such as *type*, *token*, *tag*, *parse*, and *annotation* are briefly defined. The introduction finishes with a short evaluation of the impact and usefulness of corpora in language research.

Chapter 2 elaborates on “The corpus as object: Design and purpose” (p. 25). Hunston discusses the size of a corpus, its content, its representativeness and the aspect of permanence of a corpus. She then goes on to point out how corpus investigation can be tackled, depending on one’s research aims and the selection and choice of corpora.

The analysis of concordance lines is a word-based **method in corpus linguistics**. This aspect is dealt with in Chapter 3. Great detail is given to a thorough analysis of language data with the help of KWIC-concordances. Hunston explains the concept of collocation by providing in-depth analyses of many examples, mostly taken from the Bank of English. She not only looks at the meaning of different words in certain contexts but also shows how patterns of usage reflect distinctions in meaning. The difficulty of finding probes – sets of words or expressions that cannot easily be detected otherwise (e.g. *something + Adj + about + him* as in *something different about him* (p. 62)) – is raised and exemplified.

In contrast to the word-based approach in Chapter 3, Chapter 4 covers category-based methods. This is achieved by going beyond the concordance line. The main topic in this chapter is the statistical calculation of collocations as well as corpus annotation. Statistical methods in the analysis of collocations go beyond the concordance line in so far as all instances of a word and its neighbours can be accounted for, whereas it is simply impossible for a human being to explore collocations in thousands of concordance lines by hand. Hunston explains the main differences between three statistical methods for measuring collocation: t-scores, scores of mutual information, and z-scores. She goes on to emphasise the uses of collocational information, namely getting clues as to the different meanings of words as well as the dominant phraseology of a word.

The probabilistic concept behind corpus annotation when assigning tags to words is briefly explained in Chapter 4. Hunston moves on to parsing and describes other ways of corpus annotation, such as semantic annotation. She finishes by critically examining the advantages and disadvantages of annotated corpora versus plain text corpora. This brings her back to word-based and cate-

gory-based approaches in corpus analysis. She concludes that corpus investigation ought to combine the three techniques of analysis outlined in Chapter 3: frequency, phraseology and collocation.

**Different applications** of corpora in the field of applied linguistics are covered in Chapter 5. Diverse areas such as dictionaries and grammars, ideology and culture, translation, stylistics, forensic linguistics, and the use of corpora as a help for writers are considered.

The topic **corpora and language teaching** is dealt with in Chapter 6 under three main headings: language description, general applications, and specific applications. Language research with the help of corpora has led to a new and sometimes radically different description of language. Hunston offers two approaches to language description that challenge traditional views. The first approach follows John Sinclair's arguments expressed in his book *Corpus, concordance, collocation* (1991), which essentially looks at language as phraseology. Hunston explains the key features of this approach, namely the relation between pattern and meaning, the idiom principle and the open-choice principle, as well as lexis and grammar. The second approach to language description focuses on language variation. Hunston presents research results on different aspects of language variation such as sites of variation (region, gender, social group, register), parameters of variation (word frequency, word meaning and use, feature frequency), and co-occurrence of variation (clusters of language features). Her main aim is to show that linguistic entities (e.g. grammatical features across registers) are used to account for language variation rather than those which are referred to in traditional descriptions (subject, object, word).

Chapter 7 covers three "general applications" of corpora in language teaching. Firstly, Hunston considers data-driven learning as one possibility of using corpora in language teaching. Data-driven learning refers to situations in which learners investigate language through concordance lines. The challenging effect of data-driven learning is described as "consciousness-raising" (p. 184). A second possibility of applying corpora in language teaching is by reciprocal learning. In a reciprocal learning situation two language learners with different mother tongues are paired and help each other learn their language. A parallel corpus can aid this type of learning, usually through the presentation of one language feature in language A and its translation(s) in language B. Thirdly, Hunston discusses the notion of a "lexical syllabus" (p. 189) in which "the central concept of organisation is lexis" (p. 189). The chapter closes with a critical evaluation of "challenges to the use of corpora in language teaching" (p. 192).

Specific applications of corpora and language teaching are dealt with in Chapter 8. Among these are discussions of the use of corpora in the field of

English for academic purposes (EAP), the influence of corpora in language testing, and the information to be gained from learner corpora for the purpose of language teaching.

**The impact of corpora** on applied linguistics is addressed in the last chapter. Hunston answers the question “What difference have corpora made to the working lives of applied linguists?” (p. 213).

Hunston’s first-class book is a refreshing combination of theoretical considerations and real language data. She starts every chapter with a clear outline of its contents and then approaches the topics in question, usually along the following lines: first, the matter is explained, for instance what a concordancer is (p. 39) or the distribution of meanings across one form – e.g. the present perfect. Then, in some cases, she refers to studies which exemplify the particular point under discussion, always followed by examples from these studies to illustrate her argumentation; for example the most frequent meaning of the present perfect refers to the indefinite past (p. 99). In other cases Hunston gives examples of real language data – mostly drawn from the Bank of English – which are analysed and explained in detail. These examples very frequently take the form of concordance lines: e.g. a KWIC-concordance of randomly selected lines for the word *critical* (p. 40). The results of the analysis are then related to the subjects being discussed, for instance, what is observable from concordance lines (p. 42ff.) or the finding that the corpus-based frequency information about the present perfect is not at all reflected in course books (p. 99).

Throughout her book Hunston advocates the importance of looking at real language. This is also supported by the studies she cites. Most of them start their investigation from the language rather than from a theory of language; that is, they follow an inductive approach. Hunston is aware of competing methodologies in the field of corpora within applied linguistics. Many of her chapters close with an appraisal of the subject matter discussed. These include, for example, the limitations of corpus use (Chapter 1); possible problems with the interpretation of concordance lines (Chapter 3); or “challenges to the use of corpora in language teaching” (pp. 192–193).

All the chapters of this outstanding book can be read independently of each other. This allows for many different applications at university and school. Chapter 3, for example, concentrates on the analysis of concordances and can be used for undergraduates in linguistics. Chapter 7, which introduces data-driven learning could be used as a starting point for devising new methods in language teaching at university or even at school. Apart from the possibility of employing chapters independently, there are also links between the chapters. The tendency of words to occur together is addressed in Chapters 1, 3 and 6, focusing in turn

on different areas. Chapter 1 just gives a brief introduction to collocations, Chapter 3 elaborates on collocation, phrases and variation, whereas Chapter 6 uses these ways of exploring language to end up with a description of language. Another example of the connection between chapters is the notion of local grammar which is raised in Chapters 4 and 6, thereby connecting methods in corpus linguistics with aspects of language description.

Susan Hunston's excellent book *Corpora in applied linguistics* is very well written and easily comprehensible. It offers a wealth of insight into the area of applied linguistics and its strength lies in exhaustively concentrating on corpus use and thereby on the empirical study of real language. *Corpora in applied linguistics* can be highly recommended to everyone who is interested in investigating the English language.

### **Reference**

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

**Pepi Leistyna** and **Charles F. Meyer** (eds.). *Corpus analysis. Language structure and language use* (Language and Computers 46). Amsterdam and New York: Rodopi, 2003. v + 288 pp. ISBN 90-420-1036-3. Reviewed by **Bernard De Clerck**, Ghent University.

This volume comprises a selection of papers which were originally presented at the Third North American Symposium on Corpus Linguistics and Language Teaching held on 23–25 March 2001 in Boston, Massachusetts. On the basis of extensive corpus analysis of one or more of the linguistic corpora now available, each paper disentangles and illuminates some aspect of language structure and language use. A welcome development in this collection of papers is the growing attention that is being paid to those registers, especially in spoken language, that have not received ample attention so far and the comparisons that are being made between different varieties of a language (British and American English) and the obvious and more subtle differences between spoken and written data of a certain genre or register within these varieties. By zooming in on the use and

the characteristics of a particular linguistic feature within a single dialect or register and the range of similarities and differences within and between them, each of these papers clarifies one of the complex intricacies of language use. Students and scholars will find this volume to be a valuable source of knowledge on corpus, socio- and applied linguistics, discourse analysis, pragmatics and the ways in which these studies can improve language teaching.

The book opens with a series of papers that analyse speech and writing in an academic context. The article by John M. Swales and Amy Burke investigates evaluative adjectives and their intensifiers in a portion of the MICASE corpus (the Michigan Corpus of Academic Spoken English) and compares these findings with those in the Hyland corpus (a collection of research articles). Their results suggest that adjectival evaluation in spoken academic discourse is much more prevalent, but not quite as polarized (i.e. at positive or negative extremes) as originally envisaged. The survey provided in this article shows that evaluative adjectives in academic speech seem to be part of a larger component of “various appraisal resources that interact to establish an evaluative mood or tone that ebbs and flows as the speech event progresses” (Swales and Burke 2004: 16).

Anna Mauranen uses the MICASE corpus to explore the organising and socialising role of ‘evaluative metadiscourse’ (or discourse reflexivity, e.g. ‘*That’s a good question*’, ‘*The fundamental point is*’, ‘*It is important to emphasize*’, etc.) in an academic spoken context. She more specifically focuses on items related to argumentation (‘*argue*’, ‘*claim*’, ‘*observe*’) in evaluative contexts and the roles they play in organising ongoing discourse both in a linear way (indicating order and cohesion) and a hierarchical way (indicating importance, establishing and reorganising knowledge structures). The author concludes that some of our socialisation into the academic skills of argumentation/evaluation appears to take place fairly explicitly through evaluative metadiscourse which, interestingly, tends to be predominantly positive. Mauranen, however, also raises the question to what extent similar socialisation processes might also work with negative evaluation, or, if not, whether negative evaluation gets expressed by other means than metadiscursive ones. It turns out that metadiscourse indeed occurs in more negative contexts as well, such as expressing criticism and disagreement, but the expressions are less repetitive than positive ones and are not easy to associate with consistently negative contexts or negative collocates. Negative expressions also, perhaps predictably, tend to be hedged. So although they were found, the earlier observation that they are much more hidden from view was supported by the corpus data.

The article by John Flowerdew investigates the potential register-specific features of the use of signalling nouns, i.e. abstract nouns such as *process* or *dif-*

*ference*, which are common in academic discourse and which provide various kinds of cohesive links within texts. The paper examines to what extent the use of these nouns varies across two different academic genres: textbooks and lectures. The research is based on a corpus that consists of transcribed recordings of an undergraduate lecture course in biology and a corpus that is made up of the relevant sections of the prescribed textbook upon which the lectures were based. The most striking finding is that the total usage of signalling nouns is much more prevalent – twice as frequent – in the textbook than in the lecture. The register specificity of signalling nouns is identified within Halliday's three situational parameters of field, tenor and mode. While signalling nouns may be widespread across all expository prose, the evidence presented in the article indicates that there is a strong correlation between, for example, the choice of signalling nouns and the rhetorical acts which are prevalent in a given register. In addition, distinctive features can be identified not just in the choice of signal, but also in the choice of modifier. Furthermore, Flowerdew also points out that there may be a high degree of frequency variation between speech and writing within the same register as well.

Douglas Biber's paper presents a new Multi-Dimensional (MD) analysis of university spoken and written registers. Based on the assumption that linguistic features do not randomly co-occur in texts but together serve a functional purpose, these dimensions can be used to show how genres vary, as reflected by their use of different linguistic features, in order to accomplish generic goals. Biber's study is based on the analysis of the TOEFL 2000 Spoken and Written Academic Language Corpus (T2K-SWAL), which represents a range of spoken and written registers typical of university life. On the basis of a factor analysis, Biber identifies four major dimensions of variation within this register: (1) oral vs. literate discourse, (2) procedural vs. content-focused discourse, (3) narrative orientation and (4) academic stance. The major patterns of variation among university registers are briefly described with respect to these dimensions. For instance, non-narrative styles predominated in engineering and the natural sciences; narrative styles in education, the humanities and the social sciences. Biber's overview highlights the distinctiveness of language use in university settings, pointing to the need for basing language materials and instruction on empirical descriptions of the target registers.

The next two papers apply Biber's multi-dimensional analysis to other discourse types. Upton and Connor use the multidimensional analysis technique to create a linguistic profile of the genre of philanthropic direct mail letters on the basis of a corpus of 316 direct mail letters from 108 organizations across five philanthropic fields. Their findings show that philanthropic direct mail letters

are indeed a unique genre, displaying a pattern of linguistic features that are quite distinct from other genres. What makes this genre all the more interesting is that it contains some counter-intuitive features. These include the fact that these letters are more like academic expository texts than like personal letters; they have a strong information focus as opposed to the involved, interpersonal features the authors expected to see; they are mostly expository in structure, only sprinkled with narrative tales; and they tend to be highly polished, closely edited texts, which runs counter to the impression they attempt to give as quickly penned, chatty letters. Upton and Connor, however, also admit that the method of multidimensional analysis of linguistic features cannot fully explore one of the most crucial features of promotional discourse: its persuasiveness. The very essence of these texts is to persuade the audience in the most effective manner possible. Some of the ways to achieve such persuasion, however, are not covered by the dimensions in the MD approach. Features related to achieving the communicative goal of these letters need to be identified through other analyses. Such analyses conducted using the sample have included the study of rhetorical “moves” (Upton 2001), metadiscourse (Crismore 2001), and persuasive appeal (Connor 2001).

The paper by Christer Geisler investigates variation in the personal letters of men and women in A Corpus of Nineteenth-Century English (CONCE) and also uses the multi-dimensional framework developed in Biber (1988). The results suggest that men’s and women’s language use differ on several dimensions. Men’s writing is characterised by significantly higher frequencies of informational and abstract features such as passives, word length, type-token ratio and prepositions. Men also use noun phrase elaboration to a greater extent than women, in the form of adjectival premodification or clausal postmodifiers. Two argumentative-persuasive features, prediction modals and suasive verbs (such as *ask* and *demand*) are prevalent among male letter writers. Letters written by women, on the other hand, contain higher frequencies of features marking involved, situated and non-abstract style, such as private verbs, emphatics, stranded prepositions, and *that*-deletion. Geisler also finds that, diachronically through the nineteenth century, men and women develop both similarities and differences. Women’s personal letter writing becomes less narrative and less involved while men’s writing becomes more narrative and more involved. Both genders tend to use less elaborated reference towards the end of the period, the writing of men and women becomes markedly less persuasive/argumentative, and writers seem to develop a slightly more personal/non-abstract style. In terms of the extent of diachronic change, women’s language consistently changes more than men’s language: in other words, women do not only show greater

changes, but push further ahead in their new writing styles. In addition to this research, it would be interesting to investigate if and to what extent the gender of the intended recipient(s) of the letters affects the style used by the author. Do women (or men for that matter) write differently when writing to the same or the opposite sex? If statistically significant differences could be found in this respect, this would imply that any observation on women's or men's language should have to be placed within the larger frame of the communicative sender-message-receiver model, with both sender and receiver as important parameters, influencing the composition of the message.

Susan Fitzmaurice continues the historical thread with an analysis of 'stance markers' in a corpus of early eighteenth-century English letters of writers associated with the essayist diplomat Joseph Addison. She conducts a corpus linguistic analysis of the relative occurrence of modal auxiliaries and lexically explicit stance expressions with the first person subject to explore the grammatical realization of speaker involvement in epistolary discourse. A first question Fitzmaurice addresses concerns the extent to which the grammar of personal letters reflects the semantic domains of attitudinal and epistemic stance. She concludes that early eighteenth-century epistolary discourse is marked by grammatical constructions that convey both types of stance in explicit ways. Especially favoured are verb constructions with the attitudinal verb *hope*, more than epistemic verbs *think* and *know*. The second question posed concerns the extent to which constructions that are demonstrated to mark stance explicitly in present-day English also fulfil this function in early eighteenth-century English. The study yields two main findings. The first is that stance is implicitly marked by modal auxiliaries that convey epistemic and affective subjective perspectives on what is being talked about. The second finding is that the letters exhibit the use of explicit syntactic constructions that are used to mark stance. Close analysis of the letters demonstrates that modal expressions combine with stance verbs to convey a somewhat modified sense of the writer's perspective on what is being talked about, or his or her attitude towards his or her addressee. Further examination of the kinds of grammatical constructions favoured by the stance expressions indicates that verbs like *think*, *hope* and *believe* appear to favour zero-marked complement clauses with first person subjects, whereas *know* favours *wh*-complement clauses.

Kristen Precht also studies stance markers, but she does so in a 100,000-word corpus of spoken British and American English, excerpted from the Longman Corpus of Spoken and Written English. She generates stance marker frequencies through the computer programme StanceSearch, which automates the identification of stanced lexical items occurring in particular grammatical

frames. Four categories of stance markers are examined: affect (marking emotion and attitude), evidentials (marking certainty, doubt and commitment), amount (marking hedges and emphatics), and modality (modal verbs). Similarities are found in American and British conversations in stance category and part of speech use. For one, there is a strong relationship between part of speech and stance category: affect is expressed with adjectives and verbs, amount is adverbial and evidentials are verbal. The main differences are in lexical choice. The British conversations have lower frequencies than American conversations in emotion-expressing affect markers, first-person verbs which express emotive affect, and emphatics. The American conversations have lower frequencies for modal verbs. The results suggest that cultural variations are not based on differences in stance categories as a whole, but rather on subtle lexical differences. A comparison of the use of two verbs, *love* and *hate*, for example, highlights the ways in which the stance markers are used in different ways in the two dialects. For both items, the specific contexts in which the items are used differ culturally. The verb *love* is used in three contexts in both American and British conversation: for humans, for enjoyment and for a specific inanimate object. However, on the basis of her corpus data, the author observes that the British do not seem to use *love* for generic inanimate objects with any regularity (e.g. *I love the smell of onions* occurred in AE but not in BE). A similar pattern is evident with the use of *hate*. Both the Americans and the British use *hate* in breaking bad news and for inanimate objects, but the corpus shows only Americans using it for humans (*God, I hate you for that*). This pattern suggests that stance marker differences need to be examined on a case-by-case basis in order to pinpoint the differences between the dialects. Frequent use of *love* in a generic sense or *hate* for humans may be a contributing factor in Americans appearing to be more direct in expressing emotions. Kirsten Precht also critically remarks that, despite the fact that the StanceSearch programme may be able to point to where such differences lie, it is of less help in identifying such subtle differences as uses with human and inanimate topics. Evaluating the nature of stance differences between American and British English requires a great deal more qualitative analysis of such data in order to be understood thoroughly. To pinpoint where cultural variation lies, Precht rightfully suggests that these lexical differences must be examined more closely. While qualitative analysis is indeed necessary, it is also advisable to analyse larger corpora when comparing two varieties of English. The findings in this paper are based on a 100,000-word corpus and seem to suggest a rather clear-cut difference between the two varieties in the use of generic inanimate objects following *love* and human referents following *hate*. However, queries on the multi-million BNC seem to indicate that there is some

overlap in use between British English and American English. Both constructions with *love* followed by generic inanimate objects (e.g. *I love the feel of this drummer*) as well as *hate* followed by human referents were attested in British English (e.g. *I hate the smell of moth balls and refuse to use them*). It is possible that the perception of Americans appearing to be more direct in expressing emotions is not only triggered by the kinds of referents following *hate* and *love*, but also by a higher frequency of these referents or even by an overall higher frequency in the use of verbs of emotions, as was also suggested in the article. An in-depth comparative analysis of the BNC and the ANC, for example, could shed more light on these issues.

Three papers focus exclusively on spoken English. The paper by Michael J. McCarthy and Anne O’Keeffe looks at the use of vocatives across two corpora: the 5-million word Cambridge and Nottingham Corpus of Discourse in English (CANCODE) and a 55,000-word corpus of radio phone-in calls. One hundred vocatives are sampled from the CANCODE corpus, using only informal, casual conversations among intimates, friends and close associates. All vocatives (n=232) were extracted from the radio data. The vocatives in both datasets are classified according to the contexts in which they occur. The contexts themselves are categorised under headings connected with topic and turn management, face concerns, general relational concerns, humour/badinage and summons. In the article, the authors compare both the distribution over the two datasets as well as the position of the vocative in the speaker turn. A comparison between both kinds of data reveals a tendency for vocatives to be used more in the management of phone calls, and turn-taking, topic management and face concerns in the radio data. It also shows a preference for vocative initial position, rather than final, while the casual conversation data shows a reverse pattern.

The paper by Hongyin Tao examines the Switchboard Corpus and the Cambridge University Press/Cornell University of Spoken American English to investigate the grammar of constituents occupying the first positions in speaker turns in English. The author finds that turn-initial elements in English are overwhelmingly lexical in nature. Moreover, the data shows that not only are these turn initiators lexical, they also tend to be syntactically independent. Tao goes on to suggest that one of the designing features of the grammar of turns in English involves a short free form of some sort, and that English can be considered a turn-initial language in grammaticalizing turn signals (as opposed to Japanese for example).

Yeager-Dror, Hall-Lew and Deckert use the Switchboard Corpus and the Presidential Debates Corpus to analyse variation in prosodic strategies. They more specifically focus on the influence and relative importance of cognitive

and interactive determinants, as well as the importance of register (Biber 1995), stance and footing (Goffman 1981) for the intonational choice of prominent or reduced *not*, especially when the speaker is doing a repair (Sacks 1992). Apart from showing that a quantitative study of prosodic variables is indeed possible on the basis of corpus data, this paper also shows that register, stance and footing influence a speaker's prosodic choices. The study finds that, if we consider Biber's Dimension 1 (the variation from informative to interactive situations), negatives, which carry important information, will, all things being equal, be prosodically prominent more consistently in informative situations. However, the paper also shows that stance and footing must be distinguished from each other: when speakers are in an interactive situation which requires a non-supportive/adversarial stance, the repair negatives – and even the purely informative but not face threatening negatives – will be prosodically prominent more consistently than they are when the interactive stance is not adversarial, and the repair tokens will be more consistently prominent than informative tokens. In addition, the distinction between purely informative, interactively supportive, and remedial footing must also be taken into consideration. This paper thus nicely provides evidence that, while linguistic choices are theoretically shaped by the cognitive needs of the hearer, in interactive situations social concerns predominate over such cognitive 'needs'.

The final papers deal with corpus analyses in various contexts. Fonseca-Greber and Waugh study subject pronouns in the Corpus of Everyday Conversational European French to demonstrate that their usage is different in conversational French as compared with standard written French. Analysis of the corpus shows that the subject clitics (especially the first and second person, and third person to a certain extent) have become grammatical prefixes. In addition *nous* has almost completely disappeared as a subject clitic and has been replaced by *on*. At the same time, the analysis shows that the use of *on* for 'one' is much less frequent than before, which leads to a reversal of the basic/marginal relation in its meaning, such that the meaning 'we' occurs in by far the majority of its uses, and 'one' is now only a marginal meaning. There are, however, vague uses of *on*, which could be interpreted as either 'we' or 'one' – thus showing the path of change from the one to the other both diachronically and synchronically. This indefinite meaning is shown to be shifting over increasingly to *tu* (and only to a very small extent to *vous*), so much so that in the corpus, *tu* seems to have two basic meanings, split almost 50-50 between 'you' and 'one'. This is inherently an unstable situation and probably predicts more changes to come. Fonseca-Greber and Waugh emphasise that more corpus work is needed for a fuller understanding of spoken European French. This paper once more reflects the

importance of corpus-based research for the development of learning materials. If we expect students to develop any real, pragmatically appropriate, communicative proficiency in a language – in this case French – then the results of corpus-based research of this kind should be reflected in reference works and textbooks.

In the paper by Meyer et al. the World Wide Web is evaluated as a linguistic corpus. The paper demonstrates that, despite the limitations of commonly available search tools and despite the fact that the Web's size and the particular kinds of texts on it are difficult to estimate, it can nevertheless yield valuable information. Although frequency information generated by search engines must be interpreted with caution, such information is 'suggestive' and can give a sense of which linguistic usages are common and which are not. In addition, the examples that can be found on the Web are valuable for establishing common patterns of usage. To Meyer et al. the challenge for corpus linguistics in the future will be to develop tools that will not only help linguists find linguistic constructions on the Web, but which will also enable them to locate these constructions within particular genres. This also requires – and is at the same time a strong appeal to – Website builders to make consistent use of traceable metatags when creating a Website.

Robert Bley-Vroman investigates the acquisition of English multiple *wh*-questions, in particular the relationship of frequency to grammaticality judgements in English native speakers and in advanced Japanese learners of English. Using the techniques of corpus linguistics, Bley-Vroman wants to attest existing research in second language acquisition which speculates that the grammatical systems of learners may be heavily affected by the relative frequency of grammatical structures, whereas native speakers, in contrast, would be better able to extract abstract principles from the input and are therefore less dependent on frequency. In this point of view, native speakers will readily accept a structure if it follows a principle, even if it has seldom been encountered, while non-native speakers may reject it, if it is rare. On the basis of an acceptability judgement task, using the Bank of English Corpus, Robert Bley-Vroman finds that, with regard to the acquisition of *wh*-questions, the ratings of native speakers of English clustered according to principled grammaticality, while the ratings of non-native speakers clustered according to frequency, providing confirmation for the hypothesis that native speakers operate more on principle, while non-natives operate more on the basis of what they have often heard.

The article by Juhani Rudanko focuses on sentential complements of the verbs *pressure* and *prevent* in present-day English. It is shown how each verb selects two types of complements that are similar, for instance with respect to

control properties and the semantic roles involved. However, it is argued on the basis of data from the full Bank of English Corpus that it is possible to identify the specific characteristics of the complementation patterns involved. In the case of *prevent* and *pressure*, the use of corpus data is essential, as it makes it possible to identify variation in the complements of these verbs in both British and American English. Rudanko found some noticeable differences in their use between British and American English: in American English, sentential complements following *prevent* are typically of the *from* –*ing* type (e.g. *I prevented her from leaving*), whereas in British English, it is possible to use only the bare present continuous (e.g. *I prevented her leaving*).

The constraints of a volume with a multitude of papers like this one are obvious and almost inevitable. Many authors have to summarize their findings and the way they have been achieved in terms of methodology and corpus tools, which often does not pay full respect to the richness of their research and the effects it can have on language learning. The same can be said for the job done by the reviewer. Due to limitations of space, there has not been enough room in this review to engage in a critical dialogue with each of the articles or to formulate suggestions for further research in each case. Nevertheless, it is hoped that this review to some extent reflects the richness of the articles included in *Corpus analysis. Language structure and language use* and the intrinsic value they have for language teaching and learning and linguistic description in general.

## References

- Biber, Douglas. 1988. *Variation across speech and writing*. Cambridge: Cambridge University Press.
- Biber, Douglas. 1995. *Dimensions of register variation*. Cambridge: Cambridge University Press.
- Connor, Ulla. 2001. Persuasive appeals in direct mail letters. In T. Upton (Chair). *How to make a better pitch: Analysing philanthropic discourse through corpus linguistics*. Colloquium conducted at the annual meeting of the American Association of Applied Linguistics (AAAL), St. Louis, MO.
- Crismore, Avon G. 2001. Rhetorical roles of metadiscourse and pronouns in fundraising letters. In T. Upton (Chair). *How to make a better pitch: Analysing philanthropic discourse through corpus linguistics*. Colloquium conducted at the annual meeting of the American Association of Applied Linguistics (AAAL), St. Louis, MO.
- Goffman, Erving. 1981. *Forms of talk*. Oxford: Blackwell.

- Sacks, Harvey. 1992. *Harvey Sacks' lectures on conversation*. Blackwell: Oxford.
- Upton, Thomas A. 2001. A moves analysis of philanthropic direct mail letters. In T. Upton (Chair). *How to make a better pitch: Analysing philanthropic discourse through corpus linguistics*. Colloquium conducted at the annual meeting of the American Association of Applied Linguistics (AAAL), St. Louis, MO.

**Norbert Schlüter.** *Present perfect: Eine korpuslinguistische Analyse des englischen Perfekts mit Vermittlungsvorschlägen für den Sprachunterricht*. Tübingen: Gunter Narr, 2002. xiv + 374 pp. ISBN 3-8233-4949-X. Reviewed by **Joybrato Mukherjee**, University of Giessen.

Norbert Schlüter's book, a slightly revised version of his PhD thesis, addresses a topic which has already spawned a vast literature; many linguists have tried to offer systematic and unified accounts of the forms, functions, meanings and the overall status of the present perfect. In many regards, the present study deviates from existing approaches to the analysis of the present perfect and, thus, offers fresh insights into the way the present perfect is used in natural language. What renders the book under review particularly interesting is the attempt, as the subtitle indicates, to derive from a corpus-based description of the present perfect new suggestions for teaching the present perfect in the EFL classroom. The present study consists of 12 chapters, a list of references and an appendix.

Chapter 1 provides a general introduction and sketches out the overall structure of the book. It also specifies the object of inquiry in the present study, which concentrates on typical cases of the present perfect. For example, verb phrases in the present perfect with a preceding modal verb (e.g. *must have bought*) and infinitival constructions (e.g. *To have done so would have been petty*) are left out of consideration.

In the second chapter, Schlüter distinguishes between various types of grammar that have been discussed in linguistics and language pedagogy. Of particular importance is the concept of a didactic grammar ("didaktische Grammatik"): as suggested by Mindt (1981), a didactic grammar provides a language-pedagogically relevant, comprehensive and quantitative description of language use on

the basis of a corpus which is taken to represent the target language. A didactic grammar is not identical with – but a prerequisite for – a pedagogical grammar (“pädagogische Grammatik”), which is geared towards specific learners’ needs. Schlüter’s aim is to develop a didactic grammar for the description of the present perfect which can serve as a basis for pedagogical-grammatical descriptions.

In the third chapter, Schlüter presents and systematises a wide range of previous studies of the present perfect. While some linguists argue that the present perfect has one basic meaning with slight variations in different contexts (i.e. the “monosemic view”), most studies draw on a “polysemic view” with more than one discrete and contrastable meaning of the present perfect, ranging from two-reading accounts to five-reading accounts. Schlüter also provides a detailed overview and critical review of previous empirical and corpus-based studies of the present perfect (with a particular focus on Biber et al.’s (1999) and Mindt’s (2000) grammars) as well as attempts to integrate the analysis of the present perfect into a general framework of tenses in English. From the review of previous studies and their shortcomings, Schlüter draws the conclusion that a functional analysis of the present perfect should not posit any basic meanings or functions of the present perfect beforehand but ought to abstract them *a posteriori* from a detailed analysis of large amounts of actual instances of the present perfect in corpora by grouping those instances into the same category that are characterised by the same “variables”.

These variables are presented in Chapter 4, which also presents the corpus data that are used in the present study: a selection of texts from the London-Lund Corpus (“CONV”, 113,023 words); a selection of texts from the non-fictional text categories of the Lancaster-Oslo/Bergen (LOB) Corpus (“LOB S”, 133,495 words) and the Brown University corpus (“BROWN S”, 133,769 words); a selection of texts from the fictional text categories of the LOB corpus (“LOB F”, 239,242 words) and the Brown corpus (“BROWN F”, 235,671 words). Schlüter also analyses corresponding subcorpora that are based on the British National Corpus: “BNC C” (conversation, 1,062,251 words), “BNC S” (non-fictional writing, 998,679 words) and “BNC F” (fiction writing, 999,071 words). The instances of the present perfect in these data have been coded according to a multitude of variables, including, for example, structural variables (e.g. sentence structure), grammatical variables (e.g. aspect), semantic variables (e.g. point of reference of the verb phrase), functional variables (e.g. resultative vs. non-resultative meaning) and variables of temporal specification (i.e. the choice of a temporal specifier such as an adverbial of time in the same clause).

The two following chapters provide an impressive wealth of data on the frequency, structure and syntax (Chapter 5) and the meanings and meaning variants of the present perfect (Chapter 6). Both chapters have fresh insights into the actual use of the present perfect in general and quantitatively significant trends in particular to offer. For example, not only does the corpus analysis confirm that present perfect passives (e.g. *These chaps have been called into existence...*) occur much more frequently in the non-fictional texts than in the fictional texts and that the present perfect progressive passive (e.g. *they've all been being paid*) occurs extremely rarely, but it also shows for the first time on an empirically sound basis that the present perfect is linked to specific “discourse functions”: it tends to occur in affirmative statements (> 80% of all cases), in negative statements (> 12% of all cases) and in affirmative questions (6% of all cases); in contrast, it rarely occurs in negative questions, affirmative exclamations and negative exclamations. Among many other things, Schlüter also reveals that in more than 96 per cent of all cases the point of reference of the present perfect is identical with – or includes – the point of speech. Thus, he draws the conclusion that the use of the present perfect for points of reference that are different from the moment of speech are to be neglected in the EFL classroom. On the basis of the clear quantitative trends that show in the corpus analysis in Chapters 5 and 6, Schlüter abstracts away two basic meanings (or functions) of the present perfect: (1) “indefinite past”, which refers to actions or events that took place in the past and are concluded at the point of speech/reference (e.g. *He has served as a border patrolman and was in the Signal Corps of the US Army*); (2) “continuative past”, which refers to actions or events that started in the past and continue until the point of speech/reference (e.g. *I've worked in universities for nearly ten years now*). Both basic meanings have two variants: “single acts/events” and “multiple acts/events”. The entirety of all instances of the four variants in the most frequent sentence structures (i.e. simple or main clause), in the most frequent discourse functions (i.e. statements) and in affirmative contexts constitutes the “core area” of the present perfect (comprising 1,418 instances in total), on which the analysis in the three following chapters focuses.

Chapter 7 provides a detailed analysis of the co-occurrence of various variables for which all instances were coded according to the coding scheme sketched out in Chapter 4. Whenever characteristic clusters of these variables recur in the data, the instances are interpreted as displaying a specific reading (“Lesart”) of one of the four variants of the present perfect (e.g. the “accomplishment reading” of the resultative, durative, non-agentive indefinite past, encoding a single act/event as in *She's written a novel apparently*). In Chapter 8,

Schlüter analyses the additional temporal specifications that are attested with the present perfect and offers various lists of temporal specifiers and their frequencies in all variants of the present perfect. One of the central findings in this context is the independence of temporal specifications from genre distinctions and the variation between American and British English. It transpires that across all genres and varieties, approximately one third of all instances include an adverbial of time as a temporal specification (e.g. *for* + *temporal NP*, *never*, *now*). Chapter 9 gives a thorough and quantitative account of the verbs that tend to be used in the present perfect in general and in the various genres, varieties and variants of the present perfect in particular. Schlüter compares his findings with Biber et al.'s (1999) data whenever possible (although in many regards his analysis is much more fine-grained) and produces numerous verb lists that will serve as valuable points of reference for EFL materials and textbook designers.

The next two chapters pick up on the concept of a didactic grammar à la Mindt (1981). In Chapter 10, Schlüter deduces from his corpus findings a corpus-based and language-pedagogically relevant description of the “core area” of the present perfect; the focus here is of course on those forms, functions and meanings of the present perfect that occur frequently and should therefore be given priority in EFL teaching. In Chapter 11, Schlüter starts off from a critical review of existing approaches to the teaching of the present perfect in two major textbook series that are widely used in German grammar schools (i.e. *Learning English: Password Green* and *English G 2000 A*). In the light of the shortcomings of these approaches when compared to the actual use of the present perfect in natural language (e.g. the lack of distinction between single acts/events and multiple/acts/events in the two textbooks), Schlüter offers new, corpus-based suggestions as to how to teach the present perfect, including concrete definitions and examples for learners of English at secondary level. He also suggests a specific sequencing of present perfect variants, i.e. when to teach which forms and functions in grammar schools (e.g. to start off from the indefinite past with single acts/events in the second year and to introduce the present perfect progressive towards the end of the third year). This chapter should thus be seen as a corpus-based and learner-oriented operationalisation of the didactic grammar developed in Chapter 10. The book is concluded by a summary and some prospects for future research (Chapter 12), a list of references and an appendix in which all the BNC texts are listed which make up the subcorpora “BNC C”, “BNC F” and “BNC S”.

In general, Schlüter's book is a goldmine for everyone in descriptive and applied linguistics as well as in EFL teaching who is interested in the present perfect. On the linguistic level, Schlüter finds out many new things about the use

of the present perfect by taking into account a wide array of variables and analysing them in combination. From a language-pedagogical perspective, his study is testimony to the immediate relevance of corpus work to foreign language teaching. This said, a few critical remarks seem to be in order. To begin with, I feel that the author at times gets carried away by the sheer amount of data and forgets to discuss his findings in a wider setting, linking them, for example, to descriptive frameworks and models with which the results could most fruitfully be combined. It is a pity, for example, that no mention is made of the ongoing discussion between proponents of a corpus-based and a corpus-driven approach to language description and grammar writing in Chapter 3. In Chapter 6, the only source for the notion of “situation type” that the author gives is Quirk et al.’s (1985) *Comprehensive grammar of the English language*; the entirety of studies in functional grammar, where the abstraction of types of situation or events is a key issue, is ignored (e.g. Jackson 1990; Halliday 1994). Also, one would have liked to know whether the notion of “core area” which is introduced in this chapter is linked to the old concept of core in Prague-school linguistics (or, for that matter, to the more recent notion of prototypicality). Another example in this context is the lack of any reference to models of a unified lexicogrammar in Chapter 9; it is strange that Schlüter does not offer any pronouncements on what his findings on the associations between individual verbs and the present perfect imply for an integrated lexicogrammar. In the language-pedagogical chapters, too, I would contend that the author does not realise the full potential of what his findings might suggest. I would have welcomed a few comments, for example, on whether his results are a plea for an integrated teaching of grammar and the lexicon, on whether frequency of occurrence should be the only criterion for the selection and sequencing of forms and functions to be learned and on whether authentic native-like language use should remain the target norm for learners of English in German grammar schools. The language-pedagogical innovations that Schlüter offers have their appeal, no doubt – but it all seems to me very much “linguistics applied” (cf. Widdowson 2000) in disguise. EFL materials and textbook designers will most certainly benefit from the present book, but in the language-pedagogical context, questions like the ones mentioned above will have to be addressed as well before the present perfect (or any other grammatical structure) can be taught and learned differently.

Schlüter provides the reader with many figures and tables. These visualisations of his findings are very useful in order not to get lost in the overwhelming vastness of the aspects of the use of the present perfect that he tackles. Only very few typing errors can be found in Schlüter’s book (e.g. \**Krusinga* on p.21, \**SBCASE* on p. 65, \**Bosten* on p. 361). The only major shortcoming from a for-

mal point of view is the lack of an index. This will make it difficult for any reader to browse through the book and to search for specific information on, say, a particular form or meaning of the present perfect.

To conclude, Schlüter's book is a meticulous piece of corpus-based research which paints a very comprehensive and detailed, functionally differentiated and frequency-sensitive as well as language-pedagogically relevant picture of the present perfect. Notwithstanding the above reservations with regard to the lack of linguistic theorising and to some language-pedagogical issues, the book is a must for anyone in English linguistics who is interested in the way the present perfect is used in reality. It should also prove to be a valuable resource for ELT materials and textbook designers as well as language teaching professionals and practitioners. One can only hope that the present study will find an international target audience in spite of its being published in German.

### **References**

- Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad and Edward Finegan. 1999. *Longman grammar of spoken and written English*. Harlow: Pearson Education.
- Halliday, M.A.K. 1994. *An introduction to functional grammar*. Second edition. London: Arnold.
- Jackson, Howard. 1990. *Grammar and meaning: A semantic approach to English grammar*. London: Longman.
- Mindt, Dieter. 1981. Linguistische Grammatik, didaktische Grammatik und pädagogische Grammatik. *Neusprachliche Mitteilungen* 34(1): 28–35.
- Mindt, Dieter. 2000. *An empirical grammar of the English verb system*. Berlin: Cornelsen.
- Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech and Jan Svartvik. 1985. *A comprehensive grammar of the English language*. London: Longman.
- Widdowson, Henry G. 2000. On the limitations of linguistics applied. *Applied Linguistics* 21(1): 3–25.

**Anna-Brita Stenström, Gisle Andersen and Ingrid Kristine Hasund.** *Trends in teenage talk. Corpus compilation, analysis and findings* (Studies in Corpus Linguistics 8). Amsterdam-Philadelphia: John Benjamins, 2002. xi + 228. ISBN 90-272-2278-9. Reviewed by **Marianne Hundt**, Ruprecht-Karls-Universität Heidelberg.

Variationist research is a largely adult-focused discipline, as Roberts (2002: 333) points out. The volume under review is therefore a welcome addition to the field of child language variation. As the subtitle indicates, the eighth volume in the series *Studies in Corpus Linguistics* reports on the compilation of a corpus of London teenage language (COLT) and presents some results. Chapter 1 deals with corpus compilation – the more technical details of recording, transcribing and tagging. Chapters 2 and 3 give background information on the corpus, namely the social background of the speakers (Chapter 2) and an overview of the topics that the teenagers talk about and their domains of language use (Chapter 3). The analysis focuses on slangy language or ‘slanguage’ (Chapter 4), reported speech (Chapter 5), non-standard language (Chapter 6), tags (Chapter 7) and conflict talk (Chapter 8). The main results are summarized in the concluding chapter. The appendix gives information on transcription conventions of the orthographic transcript, an example of a personal data sheet and an index.

In the recording of the data – approximately 500,000 words of spontaneous conversations – the teenage subjects were given maximum control. They recorded their speech in different situations (including peer group talk, class room interaction or family conversation); background information on the recordings was collected in a ‘conversation log’ (p. 5). As a result, both the quality and quantity of the recordings from individual recruits vary considerably, and the resulting corpus is unbalanced with respect to the variables age and gender (p. 7). The transcribed material was tagged for parts of speech with the CLAWS tagset and part of the data (25%) analysed prosodically (p. 9f.). One of the problems in the research of teenage language is categorisation of the subjects into social classes – it is quite common for adolescents simply to be categorised according to their parents’ social background. The approach taken for the COLT data was slightly different: subjects were classified according to (a) their residential borough, (b) their parents’ occupation and (c) their occupational status.<sup>1</sup> The social stratification of the subjects in the corpus turned out to be evenly distributed across the three social groups (high, middle, and low) (p. 21). The language produced by teenagers (age groups 10–13, 14–16 and 17–19) makes up the largest proportion of the corpus;<sup>2</sup> the language of adults is represented

because of the different environments where the data were recorded (including school and family). Information on the social variable ‘ethnicity’ was not systematically collected as part of the personal data survey but assigned in a post hoc fashion on the basis of impressionistic information of the field worker and meta-comments from the subjects. As a result, the rather crude ethnic labels (‘white’ vs. ‘ethnic minority’) have to be taken with a pinch of salt. Part One of the book – the description of the corpus – ends with a qualitative analysis of the topics teenagers talk about in the recordings, both in peer group interactions and in mixed aged groups (school and family talk). Not surprisingly, it turns out that one of the central concerns of teenagers is their identity which, in turn, depends largely on their peers (p. 29). The conversations in COLT thus corroborate Taylor’s (2001: 298) position that adolescents need “to look to each other to develop their own standards and codes of behaviour”.

The second part of the book – the analysis of the data – partly reports on previous findings, including publications that are not easily accessible.<sup>3</sup> In what follows, I will comment on selected results only.

The authors’ definition of slangy language or ‘slanguage’ is rather wide, ranging from slang words, dirty words and swearing to so-called ‘small words’ (hedges etc., p. 65). While the results on individual items, i.e. the actual slang words that are en vogue, are likely to be subject to a rapid rate of change, the patterns of variation that Stenström et al. found in the London teenage language from 1993 are likely to be more stable. The finding that male adolescents use slang expressions (especially dirty slang and swear-words) more frequently than female teenagers, for instance, corroborates intuitions about gender-specific language behaviour.<sup>4</sup> Likewise, the observation that kids attending school in a working class area (Tower Hamlets) used slang – especially dirty slang – more frequently than those attending middle class or upper class schools, does not come as much of a surprise, either. Not all of the results are in line with common stereotypes, however. The authors remain strangely silent, for instance, on the finding that the relative frequency of swear-words in Hackney (all recruits in this school borough were from ethnic minorities) and Hertfordshire (all subjects attended a private school and mostly belonged to the highest social class) is rather similar at around 4.9 and 4.8 swear-words per 1,000 words, respectively (p. 82). Similarly, no explanation is offered as to why the female subjects in the 17–19 age group use swear-words more frequently than their male peers (p. 81f.). One might suspect that these results have to be attributed to the size of the sub-corpora – generalisations across gender, for instance, are likely to produce more robust results in the COLT corpus than those that are based on cross-tabulations of both gender and age. Another curious result of the study is that, in a

comparison of teenage usage with adult usage (BNC), the adults turned out to make more frequent use of swear-words. It will also be interesting to see whether future studies will confirm the result that the use of vague expressions (*like, thingy*, etc.) does not correlate with a specific gender or social class.

With story-telling as an important conversational activity that teenagers frequently engage in, the study of reported speech provides an interesting field of study. The authors subsume mimicry as a subtype of reported speech – a strategy most frequently used by younger (male) teenagers (p. 113f.). An interesting result emerges from the study of the quotative marker *be like* which is less frequently used in London teenage language than in American English, but also less often than in data collected in York (p. 117f.). An interesting innovative feature in London teenage language is the use of *go* as a reporting verb which mainly introduces direct speech in the historical present (p. 119ff.). The use of this reporting verb was also found to have a strong correlation with mimicry and sound effects (p. 123f.). Young working class girls from ethnic minorities are most likely to use it as a reporting verb (pp. 125–128).<sup>5</sup>

The chapter on non-standard grammar compares the results from COLT with Cheshire's (1982) findings in Reading in a checklist fashion (p. 134). Rather surprisingly, the authors make the sweeping claim that “[n]ot much seems to have changed during the ten-year period, which shows that grammatical features are fairly stable” (p. 133). The mere presence or absence of a feature, however, does not automatically imply that we are dealing with a stable situation, as change usually goes hand in hand with a shift in frequencies. For comments on individual features, see the more informed critique in Anderwald (2004: 69). The major part of Chapter 6, however, is devoted to the use of intensifiers. Comparison with data from the BNC showed that teenagers use a different set of intensifiers, among them also words like *bloody* and *fucking*. But they also use *enough, well, real* and *right* in this function. These are used more frequently by boys than girls, whereas girls had an overall more frequent use of intensifiers. The use of *well, enough* and *right* as adjective intensifiers is, according to the authors, a revival of older patterns, but the question remains whether they “have led a dormant existence until the late 20th century” (p. 158) or whether they survived in some regional or social dialect from where they are taken up by the London teenagers.

With the exception of *eh*, teenagers use tags more frequently and with a wider range of functions than adults. The authors also point out that there are no clear sociolinguistic patterns for tags as such, but that “each item must be considered separately” (p. 191).

The study of conflict talk in the COLT data largely corroborates intuitions and earlier studies: ritual conflict is more frequently used by speakers of a lower social class and more frequently among boys than girls; whenever girls engage in aggressive verbal behaviour they tend to be from an ethnic minority and working-class background (p. 196f.). Statements on the topics of aggressive verbal behaviour are slightly contradictory: on p. 205, the authors claim that only boys engage in sexual insults about the opponent's mother, whereas girls' playful arguments about sex concern each other's deviant sexual behaviour. In the chapter summary, however, girls are reported to participate in the typically male 'your mum' insults (p. 208).

All topics discussed in this volume are illustrated with numerous and sometimes fairly long quotations from the corpus. The result is that the reader is left with a good 'feel' of the data – one of the strongest points of the book. Among the weak points are the rather numerous typographical errors and other editorial mistakes (the Introduction is not mentioned in the table of contents, for instance, the number of words for each social class in figure 2.2 (p. 21) is missing, and Quirk et al. (1985) is referred to as a dictionary on p. 135).

One problem that research on teenage talk will probably never be able to overcome is the ephemeral nature of this variety – the data reflect usage of the early 1990s. Whereas the individual items of slang, reporting verbs or intensifiers may be subject to a rather quick rate of change, the sociolinguistic patterns that were observed are likely to be more stable. *Trends in teenage talk* is therefore a valuable addition to the field of child language variation.

### Notes

1. For the subjects from a boarding school, the school borough was taken to be the residential borough “[...] since adolescents are more likely to identify with classmates than with parents and are thus likely to adapt to the language of their network of friends in these conversations, rather than to the norms of their parents and people in their residential area” (p. 20). One could argue that the same classification should also have been applied to the other subjects.
2. The figure given on p. 19 is 85%, but the percentages in figure 2.1 add up to 94%; the figure given on p. 19 for the adult group is 6% which implies that the 94% in figure 2.1 is probably the correct percentage.
3. One example would be Stenström's (2000) article on slang and 'slanguage' that appeared in a volume published by Kossuth Lajos University Press in Debrecen (p. 219).

4. Interestingly, the data indicate that verbal abuse in COLT is mainly used in single-sex conversations. The authors conclude that “the boys tend not to confront the girls with verbal abuse and vice versa” (p. 81).
5. For the chi-square tests in this chapter, the values rather than the significance levels are given (e.g. p. 122f.) – this is done without quoting the df which makes it difficult to interpret the results.

### **References**

- Anderwald, Lieselotte. 2004. ‘Dialectology and sociolinguistics.’ *Year’s Work in English Studies* 83 (covering work published in 2002): 59–74.
- Cheshire, Jenny. 1982. *Variation in an English dialect*. Cambridge: Cambridge University Press.
- Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech and Jan Svartvik. 1985. *A Comprehensive grammar of the English language*. London: Longman.
- Roberts, Julie. 2002. Child language variation. In J.K. Chambers, P. Trudgill and N. Schilling-Estes (eds.). *The handbook of language variation and change*, 333–348. Oxford: Blackwell.
- Taylor, R.L. 2001. Adolescent peer group language. In R. Mesthrie (ed.). *Concise encyclopedia of sociolinguistics*, 297–303. Amsterdam-New York: Elsevier.