Leiv Egil Breivik and **Angela Hasselgren** (eds.). *From the COLT's mouth ... and others'. Language corpora studies: In honour of Anna-Brita Stenström* (Language and Computers series 40). Amsterdam and New York: Rodopi, 2002. x + 260 pp. ISBN 90-420-1479-2. Reviewed by **Joybrato Mukherjee**, University of Bonn.

Anna-Brita Stenström, linguist laureate of the present festschrift, has exerted an enormous influence on – and, what is more, pioneered the development of – several linguistic fields. The versatility of her research over the past decades can perhaps best be thought of as occupying three widening circles: (1) corpus-based language description; (2) the exploration of spoken discourse; (3) the compilation and analysis of teenage-language corpora. Despite the enormity of the task that Leiv Egil Breivik and Angela Hasselgren must have been facing in compiling a festschrift that would do justice to the breadth of Anna-Brita Stenström's body of work, they have succeeded in putting together a volume with fifteen papers in which "every paper is stamped – directly or indirectly – by something Anna-Brita has done, created, said or written" (p. x).

In his opening paper, Jan Aarts deals with the one crucial question that many corpus linguists probably avoid asking themselves: does corpus linguistics exist? Jan Aarts starts off by clearly defining corpus data as distinct from introspective, informant and anecdotal data. However, he argues that it is not this type of data as such that turns corpus linguistics into a linguistic discipline in its own right. Rather, it is the prominence that is given to real language data and real-data language models in corpus work that distinguishes it from other linguistic fields, such as theoretical linguistics in general and Chomskyan linguistics in particular. Also, the fact that two fundamentally different methodologies are now competing for the best utilisation of corpus resources (i.e. the 'corpus-based' approach vs. the 'corpus-driven' approach) seems indicative of the existence of corpus linguistics. In a wider setting, this illuminating discussion of the status of the subject makes it clear, in my view, that many other issues of corpus-linguistic theory also warrant a much more detailed treatment in the future, e.g. the role of intuition and frequency in the description of grammar.

Karin Aijmer and Bengt Altenberg are concerned with the translations of spoken discourse items that have no clear-cut equivalents in the target language. Drawing on the English-Swedish Parallel Corpus, they show, for example, that in such cases (e.g. the translation of Swedish *nämligen* into English and English *now* into Swedish) zero translation is a prevailing translation strategy. Another strategy that may come into play in the field of cross-linguistic non-equivalence

is compensatory translation, which is also analysed quantitatively and qualitatively by Aijmer and Altenberg. The paper is rounded off by some examples of misleading translation correspondences which abound in translations of English terms of endearment (e.g. *darling, honey*) into Swedish. In general, this paper is a good example of the usefulness of the data provided by reciprocal parallel corpora for an in-depth and empirically sound analysis of the translations – and, one is tempted to add, the translatability – of specific linguistic forms.

The so-called double-copula construction (e.g. *my own view is is that…*) is at the heart of Gisle Andersen's contribution. By drawing on data obtained from the British National Corpus (BNC) and the world-wide web, he describes different ways in which this typically spoken construction is used and provides first insights into its surprisingly wide distribution across various text types. Since the double copula can no longer be regarded as a marginal phenomenon and is clearly on the increase in the internet, Andersen is certainly right in envisaging in his concluding remarks a detailed analysis of this construction within the framework of grammaticalisation.

Syntactic features of spoken English are also described by Pieter de Haan, though from a wider perspective. Specifically, he puts to the test the widely held view that spoken English displays a strongly verbal (or clausal) character, while written English is believed to be of a more nominal kind. On the basis of data from the BNC Sampler CD-Rom with its one million words, he draws the conclusion that this contrastive view of spoken and written English is a gross oversimplification in that the number of nominalisations (e.g. *inform > information*) seems to depend on a complex interaction of medium and register. In particular, the informational content and the abstract style of a text impinge on the extent to which nominalisations can be found in the text at hand. Thus, the nominal character of language is shown to be strongest in informative writing, less striking in context-governed speech and imaginative writing and weakest in informal speech. Despite the caveat that the database is relatively small (as pointed out by de Haan himself), this paper definitely opens up new perspectives for future corpus-based research into the influence of specific spoken and written registers on the verbal and nominal quality of texts.

Eli-Marie Drange offers an overview of a slang survey conducted among Scandinavian teenagers. The survey, which is part of the Nordic Teenage Language project, shows, for example, that words originating in English (e.g. *party*) have already become an institutionalised part of Norwegian teenage slang. It comes as a surprise, though, that also languages such as Berber have contributed words (e.g. /*spa*/ meaning "good") that are frequently used by Norwegian teen-

agers. The slang survey thus captures some interesting recent trends in teenage slang that stem from contact with various languages.

By drawing on the English-Norwegian Parallel Corpus, Thorstein Fretheim and Stig Johansson scrutinise the semantics and pragmatics of the Norwegian concessive marker *likevel*, which turns out to have a surprisingly wide range of English correspondences (e.g. *after all* and *anyway*). Also, a substantial number of zero translations of *likevel* can be found. One important conclusion that the authors draw from the close inspection of the data is that there is a strong inter-relation between the position of *likevel* and its preferred translations. Since the corpus is analysed bi-directionally, the authors are able to provide for a detailed discussion of various correspondences and their semantic and pragmatic implications. For example, sentence-initial *likevel* (positioned before the subject and the finite verb) tends to be translated into link expressions (e.g. *even so, nonetheless*) while sentence-final *likevel* strongly corresponds to the conclusive expression *after all*. Also, some English correlates of *likevel* are only used as translations (e.g. *anyway*), while others occur both as source and translation (e.g. *after all*).

Angela Hasselgren explores native English and non-native Norwegian teenagers' use of so-called smallwords such as *all right*, *okay* and *kind of* in spoken language. Capitalising on the framework provided by Relevance Theory, smallwords are interpreted in terms of the macrosignal and the microsignal that they represent. On the basis of an in-depth quantitative analysis, the differences between native and non-native speakers are then systematised along two dimensions: repertoire and fluency. For example, native teenagers frequently use *right* as a turn-taking signal, while non-native teenagers tend not to have this smallword at their disposal. This is but one example of a smallword which should be included in a basic repertoire of smallwords not only because of its basic communicative function but also because it increases the overall fluency in learner language.

From a contrastive analysis of the Corpus of London Teenage Language and the Corpus of Oslo Teenage Language, Ingrid Kristine Hasund draws the conclusion that English *like* and Norwegian *liksom* are pragmatic particles that fulfil very similar functions in spoken discourse in the two languages. In particular, both *like* and *liksom* indicate the speaker's epistemic stance to his/her utterance by signalling different kinds of discrepancy (i.e. between the speaker's thought and the content or form of an utterance, or between what is reported and what is really said). Both particles may also be regarded as similar metalinguistic comments (in that they signal informality, solidarity and in-group identity) and as sharing the same social stigma.

John Kirk proves that Anna-Brita Stenström's highly influential model of discourse analysis can be easily and readily applied to various types of naturally occurring spoken discourse. He gives six examples of transcribed excerpts from natural speech and dramatic dialogue which are fully annotated by carrying out an Stenströmian analysis (i.e. by indicating the various kinds of acts, moves and exchanges that she distinguishes).

Göran Kjellmer provides ample testimony of the fact that there are surprises in store for the unbiased linguist not only when it comes to teenage language. By searching various corpora as well as the internet, he delves closely into the seemingly awkward articled form *the Britain*. Not only does he show that this form is not at all a marginal phenomenon, but he also offers some language-structural explanations for its emergence, most importantly an analogy that may be increasingly drawn between non-articled *Britain* and the near-synonymous, institutionalised and articled form *the UK*. Once again Kjellmer succeeds in drawing our attention to an interesting linguistic phenomenon, sketching out an alluring step-by-step analysis, and thus shedding light on the possible beginning of a diachronic change in miniature.

The interdependence between lexical choices and genre distinctions is the basis of Magnus Ljung's paper. Specifically, he analyses the keywords in the 1997 issues of *The Times* and *The New York Times* and compares them with the keywords in the general corpora FLOB and FROWN. The WordSmith-based analysis shows, for example, that the vocabulary in *The New York Times* is more remote from FROWN (representing standard American usage) than the vocabulary in *The Times* (as compared to FLOB representing British English usage). However, particular newspaper genres (e.g. special interest news) display a larger extent of overuse of words than others, which is obviously to do with the contents they cover.

In a similar vein to Aarts, Dieter Mindt devotes himself to a basic and important theroretical issue: what is a grammatical rule? He gives many examples from the field of morphology, syntax, lexico-grammar and semantics, in which the three most frequent options account for 95 per cent or more of all cases at hand (e.g. the word classes of negator, adverb and personal pronoun in between a modal verb and an infinitive). The examples show that the distribution of possible realisations of a grammatical phenomenon approximates to the exponential function of decay. Consequently, a grammatical rule tends to cover some 95 per cent of all instances of the phenomenon at hand (i.e. the "core area"), while the remainder is usually not covered by the rule: it is this remainder of exceptions to the rule that Mindt calls "the compost of the language". Here we find both realisations that are on the decline (and that may become

obsolete) and newly emerging forms (that may become institutionalised). In general, I would contend that Mindt's empirical and dynamic redefinition of the notion of grammatical rule (and of what so far has been regarded as exceptions) might turn out to have a huge impact on future grammars.

The formal adposition *notwithstanding* can be used both prepositionally and postpositionally in English, and it is the distribution of these two variants that David Minugh puts into perspective. To this end, he analyses various corpora of present-day English and newspaper CD-Roms. That many instances of *notwithstanding* can be found in newspaper corpora is not surprising, given the formality of this linguistic form. However, it turns out that the postpositional form is most frequent in written American English. Also, there is a clear correlation between the length of the governed noun phrase and the preference for the prepositional/postpositional use of the adposition: very long noun phrases often co-occur with prepositional *notwithstanding*, while short noun phrases are also admissible with a postpostional use of the adposition.

The use of relativisers after *same* in present-day English is scrutinised by Gunnel Tottie and Hans Martin Lehmann. Their quantitative analysis of the spoken section of the BNC and the 1999 issues of *The Times* yields some interesting results: for example, the relativiser *as* (e.g. *at the same time as you were doing the mail*) is much more frequent in speech than in writing. Tottie and Lehmann also cast light on the range and the extent of the various functions (e.g. subject, direct object, adverbial) that all relativisers co-occuring with *same* in the corpus data (e.g. *as*, *that*, zero) fulfil. The qualitative analysis pays particular attention to the relativiser *as* and identifies factors that turn out to favour its choice (e.g. adverbial function) as well as variables that disfavour its use (e.g. subject function). The paper concludes with a brief discussion of some functional and diachronic reasons for the grammaticalisation of *as* as a relativiser after *same* (e.g. its etymological roots in *eall swa > also > as*).

In the final paper, Anne Wichmann digs into the attitudinal functions of English intonation by identifying and interpreting relevant passages in the British component of the International Corpus of English (ICE-GB) in which speakers explicitly refer to different tones of voice. Items searched in ICE-GB include, for example, attitude, tone and sound. Despite the relatively rare occurrences of such comments on other speakers' tone of voice, Wichmann sketches out a first systematisation of the attitudinal functions of prosody. For example, labels such as "arrogant" or "patronising" that speakers assign to their interlocutors' tone of voice indicate a specific and intonationally generated interpersonal stance towards the addressee. The methodology that Wichmann puts into operation certainly provides a very useful framework for future research into the

largely uncharted territory of attitudinal functions of prosody, although the author is absolutely right in pointing out that a "systematic account of attitudinal intonation remains elusive" (p. 260).

As usual, the present festschrift aims at two overlapping target audiences: (1) all linguists in general that are interested in corpus linguistics, spoken English and/or teenage language; (2) Anna-Brita Stenström in particular. As for the first group, the present festschrift provides them with both a mixed bag (as far as the range of topics is concerned) and a goody bag (as far as the interesting papers are concerned). Anna-Brita Stenström may see it as an impressive celebration of the stamp that she has put on various fields of linguistic research – and rightly so. It is a pity, however, that quite a few blunders have gone unnoticed in the making of the volume under review, for example **differerence* (p. 14), the missing bibliographical entry for Scheurweghs (p. 14/16), **compensatoy* (p. 19), **en enormous amount* (p. 68), **on the occasion of sixtieth birthday* (p. 101) and the mismatch between the years of publication that are given for Galtung and Ruge's paper in the text (1965, p. 188/195) and in the bibliography (1973, p. 196). Despite these blemishes, the present volume indisputably illustrates the wide-ranging impact of corpora on linguistic theory, methodology and language description as well as Anna-Brita Stenström's place in the corpus-linguistic enterprise.