

Hilde Hasselgård and **Signe Oksefjell** (eds). *Out of Corpora: Studies in Honour of Stig Johansson*. Amsterdam – Atlanta, GA: Rodopi, 1999. xvi + 361 pp. ISBN 90-420-0515-7. Reviewed by **Jürgen Esser**, University of Bonn.

A festschrift for a scholar who has really influenced a discipline can be expected to give something like the state of the art. In this respect the present volume is a valuable documentation of corpus linguistics at the end of the twentieth century. Many articles not only give a scholarly contribution but also an indication of the personal dedication to Stig Johansson as a person and a linguist.

The organisation of the book is as follows. After an introduction by the editors there are four sections: I. Representing Language Use; II. Grammar and Lexis in English Corpora; III. Contrastive and Translation Studies; IV. English Abroad. The volume ends with a list of Stig Johansson's publications.

Section I opens with a theoretical article on the description of language use by Jan Aarts. He notes that corpus linguistics has broadened the empirical basis of grammatical descriptions, which had been too much biased by the standard, written, educated variety of English. Rather, the study of language varieties as documented in the large composite corpora like the BNC has become the study of language use. Aarts develops a model in which language variety parameters (medium, genre, editing, formality) and discourse parameters (informational structuring, textual editing and interactional execution) make selections for the appropriate lexical and syntactic rules. Their output is ultimately subjected to a correctness filter, which observes the norms of the various varieties. Thus words and constructions typical of spoken English would be filtered out for a written variety. Aarts mentions further interesting points, which can be seen as resulting from many years' experience in syntactic analysis.

Josef Schmied sketches the project of an internet grammar, which has as its basis a three million word English-German translation corpus. It is designed to combine contrastive linguistics and grammar teaching (inductively and deductively) with corpus linguistics and internet teaching.

Another ambitious project making use of parallel corpora is that described by Jarle Ebeling. He describes a model that tries to link an English-Norwegian parallel-corpus with a bilingual dictionary and a grammar. He focuses on the problem that not all the senses of lexical entries can be exemplified by corpus data and, conversely, that not all senses of lexical items in the corpora are explained in the dictionary. His conclusion is that a bilingual dictionary should ideally be based on a parallel corpus, just like monolingual dictionaries are based on monolingual corpora.

In Section II, Matti Rissanen presents a long-term diachronic study of *rather*, which developed from an adverb of manner indicating speed (eg 'at once') into a comparative adverb and finally into a downtoner. His study is not only based on the Helsinki Corpus but also on other historical electronic corpora, which are usefully documented.

A synchronic study of *famously* is given by Per Lysvåg, who investigates the 117 occurrences of the adverb in the BNC. His exhaustive syntactic, semantic and pragmatic analysis could usefully enrich the entries of *famously* in standard dictionaries.

Anna-Brita Stenström analyses the use of intensifiers such as *completely*, *really* and *fucking* in the COLT corpus. Besides syntactic and semantic findings, Stenström notes gender differences, eg that intensifiers are more frequent in the girls' than in the boys' conversations and that girls tend to talk about personal things (people and feelings) while boys prefer to talk about impersonal things.

Göran Kjellmer describes the emergence of the new lexeme *as is*. With an American background it has developed to an expression that can now be used with reference to past time and the plural. Kjellmer also discusses analogical evidence, word class status (adverbial and, particularly, adjectival characteristics), the lack of a word-formation model and the motivation for its emergence. It is an excellent study that shows how language variation and language change can be studied with the help of (synchronic) corpora.

Kay Wikberg has studied *as if* and *as though* in several corpora and gives a general framework for linguistic expressions of comparisons. The fact that these conjunctions occur in imaginative prose much more often than in other genres lead Wikberg to postulate them as style markers for these genres.

Geoffrey Leech offers a preliminary study of vocatives, which form a category largely neglected in grammatical description, and which are difficult to distinguish from terms of address. There is a useful formal, functional and semantic/pragmatic classification. It is taken as a basis for a corpus analysis of some 100,000 words from the Longman Corpus of Spoken and Written English (British and American). Although data collecting proved difficult and somewhat imbalanced, the analysis reflects addressing and naming habits.

Leiv Egil Breivik's paper on the pragmatic functions of existential *there* sentences in the LOB corpus is a convincing example of how corpus linguistic work can challenge theoretical positions that are founded on a limited empirical base. He challenges the findings and generalisations by Fox and Thompson about the proposition-linking function of relative clauses with existential heads. Their purported function is shown to be marginal in Breivik's material.

Gunnel Tottie and Hans Martin Lehmann explore relative clauses in which the relative marker has an adverbial function. They show a large variety of semantic categories of adverbial functions and also the variability of antecedents which appear to be less formulaic than indicated in the literature.

John Sinclair argues that the commonest words in English defy a categorization into the traditional (ten or so) word classes, and that they rather form classes of their own. Their irregular behaviour is, among other things, shown when they occur in more or less fixed phrases (eg *to* and *a* in *come to a head*) where they cannot be paradigmatically substituted. In all, language theory so far does not provide an appropriate framework for the very common words.

Douglas Biber and Susan Conrad report on their study of lexical bundles in conversation and academic prose. Lexical bundles are sequences of (here: four) words that commonly cooccur in a register. Unlike idioms they are not fixed expressions and not structurally complete (eg *...one of the most...*), they are frequent and can be observed with over 100 occurrences per million words. Not surprisingly, the two registers differ in the types of lexical bundles.

Sylviane Granger exemplifies the usefulness of an error-tagged computer corpus. Such a corpus can be a diagnostic tool at various levels of proficiency. Her corpus analysis shows, for example, that tenses should be taught at discourse level.

Pieter de Haan also deals with learner language. He reports on two projects. The first uses factor analysis to show that English writing by Dutch-speaking students is less informative than native English writing. The second shows that the excessive use of adverbs leads to non-nativelike production of written English.

In Section III, the paper by Helge Dyvik on the complexity of translation develops a meta-theoretical model for the assessment of competing approaches to machine translation. (This is a purely theoretical paper with no direct implications for corpus linguistics. In this respect it differs from the papers reported on so far.)

Cathrine Fabricius-Hansen presents an analysis of the German preposition *bei* that is empirically based on a German-English-Norwegian database compiled in connection with the English-Norwegian Parallel Corpus Project. She regards *bei* not as polysemous but as underspecified and vague, the concrete meanings being contextually derived.

Bengt Altenberg offers another study from a (translation) parallel corpus. He investigates adverbial connectors (conjuncts) in English and Swedish. He develops a descriptive framework (mutual correspondence, asymmetrical correspondence, lexical divergence and cross-linguistic subsystems) which not only

shows language-specific properties and contrastive results but also provides for a methodological tool that can be useful in any contrastive analysis.

Another study of translation equivalents is given by Monika Doherty. She investigates *-ing* adverbial clauses in English and their equivalents in German. She provides a number of corpus examples to substantiate her hypothesis that phrasal translations are used for backgrounded adjuncts and clausal translations for foregrounded adjuncts. But also information structuring plays a role. (Unfortunately, the notions of backgrounded, foregrounded and eventuality remain somewhat unclear to the reviewer).

Continuing studies on parallel corpora, Diana Santos examines the pluperfect in English and Portuguese. Although one would think that the relative ordering of time should be something not susceptible to alteration in translation, Santos discovers striking but fairly systematic deviations, due to such factors as aspect, discourse sequence and connectivity.

Karin Aijmer has studied the correspondence of *may/might/could* and their translation equivalents in Swedish and also Swedish *kan/kunde* and their translated equivalents in English. Based on the English-Swedish Parallel Corpus, her comparison reveals that, in a cross-linguistic perspective epistemic possibility, is not a stable notion, and that there are various factors at work which account for the fact that often modals are not rendered by their equivalent modals.

In Section IV, Anne-Line Graedler reports on code-switching in written texts. She uses a monthly Norwegian entertainment guide and some personal letters to find out regularities and functions of the use of English in Norwegian discourse. Important factors are intra- vs intersentential switches, position and genre. It is her merit to have focussed on the conditioning factors of writing in the area of code-switching.

Niels Davidsen-Nielsen deals with the influence of English on Danish and the influx of direct and indirect loans. He discusses the pros and cons of their influence. No doubt, English is a must in Denmark (ca 75 per cent of the programmes broadcast on Danish television are in English), but he has his doubts whether English is also a must in Danish.

In all, the volume is more than just a collection of essays to honour Stig Johansson. It provides up-to-date information on various fields of corpus linguistics, which is continually expanding. There are, in particular, two areas that are well represented in this book and that are gaining ground: the contrastive studies of translations with the help of parallel corpora and the study of registers. It is by no means insignificant that Stig Johansson has lastingly contributed to both.