Abstracts

Christian Mair and **Marianne Hundt** (eds). *Corpus linguistics and linguistic theory.* Amsterdam–Atlanta, GA: Rodopi, 2000. 395 pp. ISBN 90-420-1493-8. Abstracted by **Peter Grund**, Uppsala University.

This volume contains 30 papers presented at the Twentieth International Conference on English Language Research on Computerised Corpora (ICAME 20), 1999, in Freiburg im Breisgau, Germany. The focus of the volume is on the connection between current practices in corpus linguistics and linguistic theory. This focus is shown in that 'the authors (a) pay extra attention to the wider theoretical and methodological implications of the often very detailed results they present or (b) reflect more thoroughly on the justification of the analytical procedure employed' (p 1). The contributions comprise corpus-based case studies of a variety of linguistic features. Both studies dealing with modern and historical topics, from a synchronic as well as a diachronic perspective, are represented in the volume. The specific areas addressed include, for instance, the use of *gonna* and *going to* in present-day British English; constructions with *all*, *both*, and *half*; *it*-extraposition; the present perfect; causal connectives in Middle English; and idioms.

The articles also comprise discussions of corpus tools and their application. Topics such as tagging corpora with the TOSCA-ICLE tagger, parsing learner corpora, and the CUE-system are taken up.

Simon Philip Botley, **Anthony Mark McEnery,** and **Andrew Wilson** (eds). *Multilingual corpora in teaching and research*. Amsterdam–Atlanta, GA: Rodopi, 2000. 208 pp. ISBN 90-420-0551-3. Abstracted by **Peter Grund**, Uppsala University.

This volume is concerned with multilingual corpora and their use in research and teaching. It is divided into twelve chapters, which report on a variety of

research projects on multilingual corpora. Text alignment is in focus in three chapters. In Chapter 1, M. Oakes and T. McEnery provide a survey of the available techniques for text alignment in bilingual corpora. In Chapter 2, M. Simard et al present the results from testing three techniques of text alignment working on three different linguistic levels: characters, sentences and words. Finally, T. McEnery, S. Piao and X. Xin discuss the research carried out on developing an English-Chinese parallel corpus and the alignment scheme employed (Chapter 12).

A number of chapters deal with how multilingual corpora can be used fruitfully in various teaching and learning contexts. Chapter 3 (written by P. Danielsson and D. Ridings) reports on the development of a parallel corpus at Gothenburg University especially catering for the needs of translation students. In Chapter 6, J. Pearson discusses a similar approach adopted in a terminology course at Dublin City University: she shows how texts in electronic format can be exploited for extracting information about technical terms in the compilation of translation glossaries. Focusing on English and Italian, C. Peters, E. Picchi, and L. Biagini deal with the use of parallel and comparable corpora for foreign language teachers and learners in Chapter 4. In Chapter 7, M. Barlow presents a discussion of different areas in the teaching of English where parallel texts can be exploited. He also introduces Paraconc, a multilingual concordance program. Basing his discussion on the INTERSECT parallel corpus (French-English/German-English), R. Salkie argues for the advantages of SMEMUC (Small and Medium-Sized Multilingual Corpora) for teachers, students and researchers of language (Chapter 10).

Two chapters concentrate on the development of software. In Chapter 5, R. Meyer, M. E. Okurowski and T. Hand introduce the OLEADA multilingual software system, which integrates on-line corpora, information retrieval and language analysis tools. D. Woolls reports on the stages in developing a multilingual parallel corpus concordancer in Chapter 8.

Finally, in Chapters 9 and 11, research carried out on an English-Norwegian parallel corpus (ENCP), and a German-English/English-German translation corpus, respectively, is taken up. S. Johansson and K. Hofland show how the ENCP can be used in studying different linguistic items, and discuss the expansion of the corpus to include other languages. J. Schmied and B. Fink investigate the translation of English *with* into German. On the basis of the results, it is argued that research on parallel corpora is of importance for the areas of lexicography, language teaching and for mapping linguistic change.

Peter-Arno Coppen, **Hans van Halteren,** and **Lisanne Teunissen** (eds). *Computational linguistics in the Netherlands 1997*. Amsterdam–Atlanta, GA: Rodopi, 1998. 239 pp. ISBN 90-420-0504-1. Abstracted by **Peter Grund**, Uppsala University.

Computational linguistics in the Netherlands comprises a selection of papers presented at the 8th CLIN (Computational Linguistics in the Netherlands) meeting, in Nijmegen, the Netherlands, in 1997. The volume contains twelve articles divided equally among three sections: 1) Statistical Models; 2) Language Description with HPSG (head-driven phrase structure grammar); and 3) Towards Practical Application. In the articles found in the first section, *Statistical Models*, the focus is on how 'statistics are used to simulate or model some aspects of language' (pp v–vi). The second section presents discussions of HPSG employed to describe different linguistic phenomena. In the third and final section, *Towards Practical Application*, the contributions deal with document analysis, controlled languages, text generation, and computational morphology.

Frank van Eynde, **Ineke Schuurman,** and **Ness Schelkens** (eds). *Computational linguistics in the Netherlands 1998*. Amsterdam–Atlanta, GA: Rodopi, 2000. 233 pp. ISBN 90-420-0609-9. Abstracted by **Peter Grund**, Uppsala University.

This volume presents articles dealing with current topics in computational linguistics in the Netherlands and Flandres. The contributions comprise selected papers given at the 9th CLIN (Computational Linguistics in the Netherlands) meeting, which took place in 1998 at Leuven University, Leuven. Of the originally 40 presentations, 14 articles have been selected for publication in the present volume. The specific topics addressed in the articles fall into three major categories: statistical methods, syntax and semantics, and practical applications. The last category, practical applications, contains the fewest contributions (three). They deal with teaching application of NLP-tools, NLP techniques in document processing, and evaluation of the NLP components of the OVIS2 spoken dialogue system. The second category, syntax and semantics, is concerned with syntactic and semantic analyses of language, especially in the framework of head-driven phrase structure grammar (HPSG). Five articles are included in this section taking up, for example, parenthetical reporting clauses in Dutch, and

Polish clitics. The third category is the largest, consisting of six articles. The topics range from phonotactics and style adaptation of statistical models to memory-based word sense disambiguation.

María Ángeles Gómez-González. *The Theme-Topic Interface: Evidence from English*. Pragmatics & Beyond New Series 71. Amsterdam-Philadelphia: John Benjamins, 2001. 438 pp. ISBN 90-272-5086-3. Abstracted by **Merja Kytö**, Uppsala University.

The Theme-Topic Interface (TTI) aims to synthetise and revise current approaches to pragmatic functions. In addition, by running multivariate statistical tests, the author analyses a number of constructions in spoken English (the *Lancaster IBM Spoken English Corpus* (LIBMSEC)). This study reveals different discourse strategies related to point of view and speaker subjectivity, and shows the mutually supportive role of linguistic form and discourse function vis-à-vis each other. TTI contains eight chapters divided into three parts dealing with the following areas:

On the Theme-Topic Interface: The first chapter offers a state-of-the-art account of pragmatic functions. Chapter 2 looks at the pros and cons of three approaches (and their off-shoots) as they are defined in this book: *semantic*, invoking such notions as 'aboutness' or 'relevance', *informational*, resorting to the given-new distinction, and *syntactic*, related to the 'staging' of information in (non-)initial position.

Previous studies: A sympathetic critique: These three chapters review pragmatic insights made by the Prague School (PS), Systemic Functional Grammar (SFG), and Functional Grammar (FG). The author's intention is to unravel the confusion that has pervaded these hitherto conflicting functionalist accounts in order to accommodate their rapprochement.

A corpus-based analysis of syntactic Theme in PresE: Part III is geared towards remodelling the pragmatic domain, bringing together views from PS, FG and largely SFG. The category of Theme is identified with the clause-initial transitivity/mood slot and is dissociated from (a) Topic, ie 'what a text/discourse is about', (b) the suprasegmental coding of given and new information, and (c) the morphosyntactic coding of (non)recoverable information. The Theme zone is shown to act as an orientation zone which has both forward-looking and backward-looking potential which tends to display a centripetal array arguing for a scope-based, rather than a parallel organisation of the metafunctions. Besides,

the study identifies the frequencies, formal structure, discourse function and distribution of (un)marked thematic constructions in contrast with seven types of special constructs, namely: clefts and pseudo-clefts, *it*-extrapositions, *there*-presentative constructions, inversions, and right and left dislocations.

All in all, TTI may be used both as a textbook for (under)graduate courses in discourse analysis and as a resource for analysts interested in exploring the use and functioning of natural language.