Introduction: Corpus linguistics and variation in English

Joybrato Mukherjee and Magnus Huber

Justus Liebig University, Giessen

The present volume contains a selection of papers presented at the 31st Annual Conference of the International Computer Archive of Modern and Medieval English (ICAME) in Giessen, Germany in May 2010. The overarching conference topic was ‘Corpus linguistics and variation in English’. Although papers and posters from all areas of English corpus linguistics were of course welcome, the conference topic indicated a particular focus at the 31st ICAME Conference on research activities and projects that place special emphasis on the compilation, annotation and – most importantly – the analysis of computer corpora in order to capture the multi-dimensional variability of the English language, depending on a range of factors that cause variation in language, e.g. language contact, genre and style, region and time.

The papers and posters that were presented at the 31st ICAME Conference were characterized by a very high quality and introduced a number of innovative and original research ideas as well as new methodological approaches. Thus, it does not come as a surprise that a large number of the conference contributions qualified for the Conference proceedings. After a strict peer-reviewing process, we were left with some forty papers. We decided to group them into two volumes: The present volume includes all the papers that are based on the five plenary lectures at the 31st ICAME Conference and 13 further papers which combine theoretical considerations and/or methodological innovations with descriptive case studies. The second volume of the Conference proceedings is available electronically in the eVarieng online series, published under the auspices of the Varieng Research Unit of the University of Helsinki, Finland (see http://www.helsinki.fi/varieng/journal/index.html). The eVarieng volume carries the subtitle ‘Focus on non-native Englishes’ as it includes all the papers that are concerned with variation in postcolonial second-language varieties of Englishes, learner Englishes and English as a lingua franca.

As can be seen in the five plenary papers included in the present volume, all our keynote speakers provided the conference participants with very distinct perspectives from various linguistic subdisciplines on the conference topic ‘Corpus linguistics and variation in English’: The paper by Stefan Th. Gries highlights the importance of quantitative-statistical methods for the description of variation in English, picking up on current debates in the corpus-linguistic community; Michaela Mahlberg provides a hands-on practical introduction to the corpus-based analysis of stylistic variation, a research area at the interface between linguistics and literary studies; Miriam Meyerhoff shows that even modest corpora of lesser known languages can be used to shed light on processes of variation and change in contact variants of English such as English-based
creoles in postcolonial contexts; Edgar W. Schneider uses a range of specialised and historical corpora to discuss the evolution of Southern vernacular forms of American English; and Elizabeth C. Traugott makes it clear that corpus resources are indispensable tools for the analysis of historical change in general and grammaticalisation processes in particular. All five plenary papers provide rich and thought-provoking theoretical and methodological discussions above and beyond the actual case studies that they include.

The other papers in the present volume, too, combine theoretical and/or methodological considerations with a corpus-based description of actual language use. They also illustrate the enormous range of corpus-based work in English linguistics. Gisle Andersen investigates the proactive manifestation of listenership in business negotiations. Marina Bondi and Corrado Seidenari analyse the use of evaluative language in blog comments, highlighting significant features of an increasingly important web genre. Doris Dant’s study based on the Corpus of Contemporary American English (COCA) confirms that The Chicago Manual of Style, in spite of being the most influential style guide in the United States, is not always in line with variation in linguistic reality. Hans Martin Lehmann and Gerold Schneider discuss various kernel issues related to dependency treebanks and how they can be used to analyse syntactic variation. Wright’s English Dialect Dictionary (EDD) is at the heart of Manfred Markus’s paper who shows that – and how – the EDD can be used as a corpus. Hagen Peukert discusses statistical ways of making meaningful text-type predictions for a given text on the basis of the analysis of the text-internal distribution of phonemes. New ways of detecting lexicogrammatical routines and degrees of predictability of lexical choices at the level of n-grams are explored by Thomas Proisl. Using various diachronic and synchronic corpora of British English, Paula Rodríguez-Abruñeiras traces the development of the exemplifying markers for example and for instance, which also reveals features of a process of grammaticalisation. The Irish component of the International Corpus of English (ICE) is used by Patricia Ronan to describe and categorise the variable functions of would as a pragmatic softener. The transitive out of -ing pattern in American English is analysed by Juhani Rudanko along the lines of construction grammar. Stefania Spina focuses on methodological issues that are relevant to the compilation of a corpus of television news, e.g. genre distinctions. From a more theoretical perspective, Michael Stubbs looks at the relationship between (naturally occurring) texts and (artificially constructed) corpora, making it clear, inter alia, that texts and corpora relate to fundamentally different conceptions of context. Finally, Turo Vartiainen and Jeffrey Lijffijt use a parsed version of the British National Corpus (BNC) to analyse the heterogeneous group of premodifying -ing participles.

We have abstained from grouping the papers in the present volume into smaller sections, as we felt that this would create artificial topical distinctions. Instead, we have arranged all the 18 papers in a strictly alphabetical order.

We are very grateful to a number of people without whom the present volume would never have materialised. First and foremost, we would like to thank our