Introduction

Oliver Mason and Andrea Gerbig

The papers in this volume are all concerned with structural aspects of language in relation to its users in a variety of socio-cultural situations. All papers are based on the assumption that only the use of authentic language data can inform us about the role and function of language, its structure and use. The papers give qualitative and mostly also quantitative analyses of their data, always with respect to the specific 'work' the language does for its users.

The contributors to this volume have all been in critical exchange with Michael Stubbs' leading, corpus-based work on theories of language structure and use and its applications to education, cognition and culture, literature, and politics. Such discussion is of course always ongoing; it involves innovative theoretical thinking, looks at the most recent data which have become available through new technology, and puts in perspective and evaluates established theories and findings. The present papers are all original work providing such new insights. They are all written in honour of Michael Stubbs' outstanding contributions to this discussion which are often of a programmatic nature, paving the way for more detailed work, following up his theoretical leads.

Susan Hunston's short contribution evaluates and honours the work and influence of Michael Stubbs in major fields of linguistics, thereby providing an outlook on directions of future research. This is followed by three papers of a more theoretical orientation, starting with an article by John Sinclair, who had a profound influence on the field of corpus linguistics and on Michael Stubbs himself. Sinclair discusses how imports from other disciplines have shaped linguistics, and points out a number of pitfalls: sometimes imports do not make sense, or are simply misapplied, so that statistical significance tests for example provide a false sense of security where they should not have been applied in the first place. As always, Sinclair is not afraid to be controversial, casting doubt on many assumptions that are usually taken for granted by many researchers.

Continuing in the realm of theory, Robert de Beaugrande attempts to find an answer to how systemic a corpus of English is, investigating a number of systems in the process. He investigates the relationship between text and language, and comes to the conclusion that both of them are in fact systemic.

Wolfgang Teubert then looks at the mental lexicon. Meaning has often been ignored by corpus research, as it is much harder a problem to tackle than lexis or even grammar/phraseology. Teubert argues that there could be a fruitful dialogue between cognitive linguists and corpus linguists, and that Michael Stubbs would be the one person who could facilitate such a dialogue.

Michael Byram discusses issues of language and politics. Commonly, the national identity is based on a common (national) language, but how does it work with a supranational entity such as the European Union? Questions arise here of language policies, both regarding communication and cultural issues. And who
draws up those European policies? In his contribution, Byram echoes early work by Michael Stubbs on the National Curriculum.

The collection further continues with a set of papers tracing linguistic progress in the description and investigation of diachronic language data, again in relation to the role such language had at the time it was used and what influences result for our present views on language. The first of these papers, by Wolfgang Kühlwein, draws connecting lines from lexicological to intertextual to semiotic research, and demonstrates how these can be applied to a language for which we have something we can only dream of for modern languages: a corpus of all known utterances of the language.

Still within the theme of historical work, David Reibel looks into empiricism among early grammarians, working on 'traditional grammar', which is often used by modern corpus linguists as something to distance themselves from. Reibel shows that issues such as what constitutes 'proper English' have been around for a long time.

Andrea Gerbig then shows that both sides of the Saussurean dualism 'synchronic/diachronic' can be studied with a diachronic corpus, touching on issues such as language and the representation of reality/shared cultural knowledge, and language change. Gerbig here exploits the fact that her corpus of travel writing is controlled for topic, and captures the experience of a closely defined sub-group. Gerbig's contribution also provides a link from the historical studies to the phraseological ones.

The following four papers emphasise the phraseological aspect of language, the field in which Michael Stubbs has most recently set new standards in a collection of publications. Function words are routinely being ignored by corpus linguists, on the grounds that they are too frequent or have no meaning. John Sinclair looked at of in Corpus Concordance Collocation (1991), and Naomi Hallan here analyses the uses of out, with a special focus on the use by children of different age bands and the differences there are compared to adults. Unsurprisingly, the picture is more complex as one would have expected; this again shows that there is no substitute for looking at real data.

In the following paper, Bettina Starcke investigates changing discourse prosodies of phrases based around the same nucleus. Starcke finds that the prosody of a phrase relates to its length (effectively its specificity) because the contexts in which the longer phrase is used are more restricted. The shorter variants are also more often used in a literal sense, whereas longer phrases tend to be non-compositional.

Hans Lindquist then compares varieties of English, British and American, and discovers that there are differences between literal and metaphorical uses of a formulaic expression. Even though originating from America, the particular phrase under investigation (to stub one's toe) is now about equally frequent in British English, but is predominantly used literally. Studies such as this can provide useful insights into the development and change of language.

Oliver Mason concludes the phraseology section with a new approach to the description of grammatical structure. Mason uses multi-word units (partly