

Preface

Lidun Hareide¹, Michael Oakes² and Christer Johansson¹

¹University of Bergen; ²University of Sunderland

We are very pleased to announce the third volume of the BeLLS series, *The Many Facets of Corpus Linguistics in Bergen – In Honour of Knut Hofland*, edited by Lidun Hareide, Michael Oakes and Christer Johansson.

The idea of this book emerged, like many great ideas do, over a beer in the Swedish pub *Naboen* in Bergen. Michael Oakes had just presented his and Meng Ji's new book *Quantitative Methods in Corpus Based Translation Studies* at the University of Bergen, and someone suggested continuing the discussion at a pub. Christer Johansson, being a Swedish native, immediately suggested going to *Naboen*. We were discussing corpus linguistics in general, and we marvelled at the great number of corpus linguistic projects found in Bergen, and the fact that Knut Hofland seems to be involved in most of them. Lidun Hareide, co-founder of BeLLS with Gard B. Jensset, had for a while been looking for someone interested in editing a book on corpus linguistics in Bergen, and asked Michael Oakes and Christer Johansson if they could be persuaded to take on the challenge. "We will, if you join us", said Michael, and so the issue was settled.

We then went on to talk about Knut's role in the corpus linguistic community, and how we continually discover more projects that he has worked on. Michael had just worked with Knut at UNI-digital, and Lidun had worked closely with him on the construction of the Norwegian-Spanish Parallel Corpus. We discussed the fact that Knut had turned 60 in 2010, and although he was celebrated with cakes and speeches at the time, we found it strange that he had not been honoured with something more permanent. We therefore decided that our book would be a belated Festschrift for Knut, and although an open access online publication is in keeping with Knut's vanguard spirit, we decided to explore the possibility of a print on demand solution for BeLLS in order to be able to hand Knut a physical book. This book therefore exists in both an online and a printed version.

Knut Hofland has been a key contributor to computer corpus linguistics throughout its history. Geoffrey Leech in his introduction discusses the particular importance of Knut's work on the Brown family of corpora, and his work at ICAME including the moderation of the popular "Corpora List". The various corpora in the Brown family, all carefully balanced for content, are invaluable for the comparison of the different types of English spoken throughout the world, and for comparing the English spoken across generations.

We are also fortunate to have a contribution by the team at the Computational Linguistics Unit, Uni Research, Knut Hofland, Paul Meurer and Andrew Salway. The unit started out 40 years ago as the Norwegian Computing Centre for the Humanities, and over the years has produced many corpora and language tools, most recently for the analysis of messages posted on social media. In contributing to this article, Knut was "tricked" into writing for his own Festschrift. Most of the research presented in this volume was conducted on resources that Knut has been instrumental in creating.

Jarle and Signe Ebeling open our first section of articles on the theme of parallel corpora. Lining up texts in one language with their translations in another opens up many possibilities for

research in translation, lexicography and language teaching. They describe the Translation Corpus Aligner (TCA) originally developed by Knut Hofland for the English-Norwegian Parallel Corpus. Rosa Rabadán and Marlén Izquierdo describe their work on the P-ACTRES parallel corpus on the translation of “approximate negations”, such as “rarely”, “hardly” and “seldom”, and how this work was made computationally possible by Knut. Carla Parra Escartín describes how to go about encoding a new parallel corpus, in her case the Technical Regulations Information System (TRIS) parallel corpus of German and Spanish. Pedro Patiño describes a parallel corpus of English and Spanish Free Trade Agreements, and its use for the study of domain-specific collocations. Pedro also made use of Knut’s TCA package.

This leads logically on to our second section on domain specific corpora, which looks at the language of specific subgroups in society, particularly those with a shared knowledge of some technical or academic expertise. Ingrid Simonnæs and Sunniva Whittaker describe the TK-NKK corpus, produced in collaboration with Knut Hofland. It contains texts produced by candidates sitting for the translators’ National Accreditation examination, and as such is a valuable resource for research and pedagogical purposes. Gisle Andersen writes about the identification of neologisms in journalistic prose, derived from the Norwegian Newspaper Corpus (NNC) of Andersen and Hofland, which has had a huge impact on corpus research in Norway. Marita Kristiansen also looks at methods of harvesting neologisms from the NNC, specifically terms of financial jargon. Kjersti Fløttum, Trine Dahl, Anders Didriksen and Anje Gjesdal describe the KIAP corpus of academic discourse, and how they used it to examine the standardisation of scientific terminology over time. Annette Myre Jørgensen writes about the COLAm corpus (Corpus Oral de Lenguaje Adolescente) of Spanish teenagers’ language, which would not have existed without Knut’s expertise. Teenagers’ language is of particular interest, as it is highly innovative and a prime mover of linguistic change.

Section three is on the theme of language development. Martha Thunes looks at the semantic property of inalienability in both English and Norwegian, using a subset of the ENPC. Inalienability refers to things which cannot be taken away from the person, such as body parts and kinship terms. Benedikte Vardøy and Margje Post look at the development of English loanwords ending in -ing in Russian, and how they have sometimes drifted in meaning from the original. They provide a useful introduction to the Russian National Corpus and the Integrum database, both valuable resources for linguistic studies.

The editors of this volume have all chosen to write on the theme of statistical analysis of corpora. Gard Jenset and Lidun Hareide, using the Norwegian-Spanish Parallel Corpus produced by Hareide and Hofland, show how patterns of sentence alignment, which vary by translator and genre, can provide insight into the translation process. Christer Johannson contrasts statistical significance (how confident we can be that the use of a linguistic feature varies between corpora) with effect size (how strong is this association). Effect size has often been neglected in corpus studies, but Christer here illustrates its importance with some real-world examples. Michael Oakes and Alois Pichler describe their work in the computational stylometry of Wittgenstein’s writings.

Last but not least, Knut’s colleague Øystein Reigem’s contribution is the speech he delivered at Knut Hofland’s 60th birthday. The speech is a funny account of the development of corpus linguistics since the 1970’s and Knut’s pioneering role in this development, illustrating sides of Knut previously unknown to most of us. This contribution is written in Norwegian.

We hope that this book has helped unite “the many facets of corpus linguistics in Bergen”, and will be a fitting tribute to Knut Hofland’s many contributions to that work, which have influenced corpus linguistics far beyond this town.

Bergen, 2013.