

Corpora and non-native speakers: issues of representativeness and methodological approaches.

Anna Mauranen

University of Helsinki

The fastest-growing use of globalised English is among speakers for whom it is not a first language, that is, English used as a lingua franca (ELF). To keep up with the developments of the language in such varying circumstances poses a challenge to research: how can we access reliable data that captures new directions in this expanding use of English? Clearly, corpus methods have a lot to offer in teasing out the big picture and emergent patterning from the bewildering detail that small-scale studies easily drown themselves in. Yet they require a good database if they are to yield good answers. Given that speech in authentic interaction is arguably the best kind of data for investigating what is going on in ELF, a speech corpus is the most interesting alternative to start off with.

ELF has established itself particularly in two important and influential inherently highly international domains: academia and business. Both are high-stakes domains where language plays an important role. This paper looks into the academic sphere, and explores the first ELF corpus, comprising academic speech (ELFA: www.eng.helsinki.fi/elfa/elfacorporus). It will tackle issues of data selection, relevance, and meaningful combinations of analytical methods, also touching upon the controversy of corpus-based vs. corpus-driven approaches.