

Christian Mair
Englisches Seminar
Universität Freiburg/ Germany
christian.mair@anglistik.uni-freiburg.de

The Corpus-Based Study of Language Change in Progress: The Extra Value of Tagged Corpora

In their untagged versions, the "Brown quartet" of corpora, consisting of the original Brown and LOB corpora and their 1991/ 92 updates compiled at Freiburg ("Frown" and "F-LOB"), have become a widely used resource for the study of ongoing grammatical and stylistic changes in British and American written English. In collaboration with Lancaster, a manually post-edited part-of-speech tagged version of F-LOB was produced which revealed numerous diachronic shifts, both in the frequency of individual tags (such as modal verbs) and tag combinations (such as noun-noun sequences) – see Mair, Hundt, Leech & Smith 2002.

In the absence of a tagged version of the Frown corpus, the results of this work remained necessarily provisional, as the diachronic developments observed in British English could not be compared with trends in American English during the same period. Nor was it possible to provide an integrated account of diachronic developments and synchronic regional variability between the two standards.

The tagged version of the Frown corpus is now available. The present paper will present a global account of variability in tag frequencies in the four corpora, plus detailed results from a small number of case studies of individual phenomena (e.g. genitives and noun-noun compounds). The dominant factor determining the frequency of tags and tag combinations is textual genre and style, but in addition there is some systematic regional variability between British and American English and some directed diachronic development.

Reference:

Christian Mair, Marianne Hundt, Geoffrey Leech & Nicholas Smith, "Short-Term Diachronic Shifts in Part-of-Speech Frequencies: A Comparison of the Tagged LOB and F-LOB Corpora." *International Journal of Corpus Linguistics* 7 (2002): 245-264.