

Corpus Linguistics in Scotland, Edinburgh 2.12.2016 'Diachrony through corpora'

TALK ABSTRACTS

Benjamin Molineaux (University of Edinburgh)

Spelling variation in historical corpora: A window to sound change

Orthographic variation, which is endemic to non-standard spelling systems, constitutes a major stumbling block for many researchers interested in building historical corpora. Graphemic alternations, however, have long been a treasure-trove for historical phonologists, who attempt to piece together bygone sound-systems and their changes by close examination of scribal spelling practices. This wealth of information has heretofore been explored only in a non-systematic manner, using qualitative and impressionistic methods. In this talk I focus on how the AMC-based [FITS Project](#) has harnessed the potential of non-standard spelling variation, in order to produce a corpus-based resource mapping spellings to sounds in the early history of Scots (1380–1500). I survey the technique of grapho-phonological parsing, which allows each grapheme in the fifteenth-century Scots materials to be mapped to a potential sound-value, triangulating on the distribution of spellings across the corpus, what we know about earlier and later stages of the language, and the typology of sound change. Finally, some examples are provided for the usefulness of quantitative data in exploring phonological change, reconciling the core objectives of corpus methods with a level of linguistic analysis previously dismissed as cumbersome and uninformative.

Robert Truswell (University of Edinburgh)

Diachronic syntax with and without parsed corpora

Large-scale parsed historical corpora, with explicit representation of sentence structure, have revolutionized research into syntactic change, by dramatically speeding up research into the temporal dynamics of a change: questions which, only 20 years ago, could only be answered with years of work can now be investigated in a matter of hours. However, working with parsed corpora presents its own challenges, the most serious of which is scarcity of data. Even for a language like English, with a rich textual record and a long research tradition, the currently available parsed historical material amounts to just a few million words, an order of magnitude smaller than POS-tagged corpora such as the Penn-York Computer-annotated Corpus of a Large amount of English (PYCCLE) or the Corpus of Historical American English (COHA). This problem is compounded in syntactic and semantic change simply because the units of analysis are larger and more diverse than for morphophonological or lexical

change. This talk will cover a case study (from Aaron Eca'y's work) of investigating syntactic change in a very large unparsed corpus, suggestions for making the most of existing parsed corpora, and a discussion of prospects for expanding the stock of parsed resources.

Wendy Anderson (University of Glasgow)

Warp and weft: following lexical threads in diachronic corpora

Using a selection of examples from the semantic field of textiles, this talk will explore the place of metaphorical lexis in corpora of English, Scottish English and Scots, covering the period from 1700 to the present, drawing also on the Historical Thesaurus of English, the Oxford English Dictionary, and [Metaphor Map of English](#). It will focus on the ways in which metaphors create meaning by reflecting the social and cultural contexts in which they are used, and will demonstrate how the various resources can be used to establish a picture of the lexical and metaphorical options open to speakers at particular times. It will also address the knotty issue of 'dead metaphor' (i.e. metaphorical connections that are no longer cognitively 'live' for speakers) and suggests ways in which corpus analysis can help us to determine the status of metaphors.