Abstract for Workshop on *Historical Corpus Linguistics: Methods and Applications*

*Linguistic DNA: Discursive semantics and the quest for the automatic identification of concepts and conceptual change in English 1500-1800*

Susan Fitzmaurice
University of Sheffield

This paper concentrates on models of language change with specific reference to historical semantics, and the challenges of conducting semantic analysis of digital bodies of historical discourse. The project that informs the discussion is Linguistic DNA, a collaboration among linguists, historians and digital humanities developers at the Universities of Sheffield, Glasgow and Sussex. The project uses two large digital collections of early modern texts (EEBO-TCP, ECCO-OCR) to investigate and model semantic change in English from 1500—1800. The goal is to understand the ways in which the concepts (or paradigmatic terms) that define modernity emerge in the universe of Early Modern printed discourse that these corpora represent. Unlike traditional work in intellectual history which involves the investigation of specific cultural keywords or paradigmatic terms intuited as important in the literature of the past, Linguistic DNA deploys a series of computer-assisted language processing techniques to identify concepts in the digital linguistic data of a complex historical database.

This process assumes a theory of semantics that precedes the in-depth study of named concepts (onomasiology) or identified words (semasiology). Our model involves testing language-internal evidence of conceptuality using cultural literacy. The results of the processes of (distant) corpus reading are inspected using close, culturally literate analysis in order to answer research questions about the discourse contexts in which semantic change occurs, prominent polysemies and the relationships among synonyms, and the historical moments in which lexical fields expand or contract dramatically.

In this talk, I outline the epistemology developed to operationalise ‘concept’, and I discuss some of the bottom-up methods the project is currently developing in order to identify concepts in early modern English discourse. I discuss the different procedures used to prepare the data for analysis, the challenges posed by the nature of these databases for semantic analysis, and the implications of this work for the development of historical corpus linguistics.

---

1 The Linguistic DNA project is funded by the Arts and Humanities Research Council (project AH/M00614X/1), and runs till 2018.