

ISLE Workshop:
Grammatical Information in Dictionaries of English

Abstracts

Philip Durkin, Oxford English Dictionary

**Historical dictionary senses and theories of diachronic semantic change:
How approaches to indeterminacy and complexity in historical lexicography can
interface with advances in understanding semantic change**

Although seldom conceptualized in precisely these terms, the editing of a major historical dictionary involves a process of daily practical examination of instances of semantic change. The numbered senses in a historical dictionary such as the *Oxford English Dictionary* inevitably allocate the evidence for semantic variation into discrete units, which can appear monolithic and categorical, demarcated from one another by having distinct definitions, and also by aspects of the dictionary's structure and presentation, such as a distinct sense number, and some amount of presentational "white space" between each sense. The assumed diachronic pathways of development are flagged by the arrangement of senses into sequences and sub-sequences, by use of branches, sub-senses, and other structural arrangements. Nonetheless, historical lexicographers are keenly aware of how often they encounter uses that strain against such categorical assignment to a single dictionary sense, and that such data often has a close connection with processes of semantic change. *OED* entries frequently (and increasingly) highlight such tensions with explicit notes, which may for example:

- Draw attention to particular examples (i.e. dictionary quotations) which show aspects of more than one of the defined senses (as distinct from examples which are straightforwardly ambiguous to interpret, but where only one of the defined dictionary meanings is likely in fact to have been intended). Very often such notes concern the earliest evidence for a particular *OED* sense.
- Note explicitly how one of the defined senses is in origin a contextual reinterpretation of an existing sense.
- Note how a particular sense may have arisen by a different route from that implied by its place in a numbered sequence, or that multiple pathways of development are likely to have converged.

This paper will examine a selection of such instances from *OED* entries, and explore the similarities they reveal between the methodologies and assumptions of historical lexicographers and current approaches to semantic change in historical linguistics (notably the invited inferencing theory developed by Traugott and Dasher, and the diachronic prototype semantics developed by Geeraerts). The paper will contend that such similarities can be shown to exist, even though there is often little direct contact between the work of historical lexicographers and that of researchers in historical semantics as a subfield of historical linguistics. Indeed, it will be a central aim of this paper to identify areas where such contact might be particularly fruitful for both groups, particularly in exploiting the data and narratives in historical dictionaries for testing and developing theories of historical semantic change.

Marc Alexander and Fraser Dallachy, University of Glasgow

Historical Semantics in the 21st Century: Semantic Annotation Using the Historical Thesaurus Semantic Tagger Abstract

Corpus semantic studies are impeded by the requirement to search for strings of written characters rather than meanings in texts. This especially poses difficulties when a string – what most people would think of as a ‘word’ – such as *set* may have hundreds of possible meanings, and where each of those meanings may be expressed by a number of different words (such as *pattern*, *collection*, or *group*). Research into semantic analysis of text has led to the creation of semantic tagging programmes which attempt to add a layer of metadata to electronic text in the form of meaning ‘tags’. Where a single orthographic word has multiple meanings, tagging programmes such as the UCREL Semantic Analysis System (USAS) have employed thesauruses as a means of resolving ambiguity.

The SAMUELS project (Semantic Annotation and Mark-Up for Enhancing Lexical Searches) follows in the footsteps of the USAS tagger, advancing the state of the art through integration of the *Historical Thesaurus of English*. The *Historical Thesaurus* classifies each meaning of each word in the English language into a semantic hierarchy of sense relationships. Using this highly fine-grained structure, the SAMUELS consortium of researchers working at the Universities of Lancaster, Glasgow, Huddersfield, and Central Lancashire developed the Historical Thesaurus Semantic Tagger (HTST) – the first semantic tagging software pipeline which is capable of employing the full structure of the *Historical Thesaurus* as well as a set of ‘thematic’ categories which allow users faster access to subjects of everyday importance.

This chapter will describe the processes underlying the HTST before demonstrating its application in a case study on the semantically tagged Hansard corpus which consists of textual records of all the speeches made in the UK Houses of Parliament between 1803 and 2005. This case study will focus on the advantages to employing the thematic category set in investigating the behaviour of the semantic field of ‘Goodness and Badness’ in the Hansard data. It will investigate the prevalence of positive and negative evaluation across time and in the different Houses of Parliament, and identify the words and categories which collocate with these evaluations. In this way, it aims to provide an overview of what semantic fields are discussed as ‘good’ or ‘bad’ in these two centuries of UK parliamentary debate.

Sara M. Pons-Sanz, Cardiff University

The *Gawain*-Poet's Terms for SPEECH in the Context of the Gersum Project

This paper will introduce the aims and methodology of the three-year AHRC-funded *The Gersum Project: The Scandinavian Influence on English Vocabulary* (www.gersum.org), led by Dr Richard Dance (University of Cambridge) and myself. The project aims, on the one hand, to provide a consistent and clear typology for the classification of Norse-derived loans in English in terms of the evidence that we have for their etymological identification; and, on the other hand, to investigate the process of adaptation of the terms into the English vocabulary by studying the semantic and stylistic relationships between the Norse-derived terms and other members of their lexico-semantic field.

The paper will focus on the lexico-semantic field of SPEECH as represented in *Sir Gawain and the Green Knight*, one of the texts in our corpus, as a case study for the two aims of the project, although more attention will be paid to the analysis of the make-up and stylistic uses of the field. Thus, after presenting a classification of the terms belonging to this field in terms of their etymology, the paper will highlight the differences between Norse- and French-derived terms their distribution amongst the various levels of semantic specialisation within the text, as well as the registers and connotations they are associated with. The results of this analysis can be extrapolated to further our understanding of the significance of Norse-derived terms in our corpus and Middle English texts in general. The paper will show that it is fundamental to go beyond the mere word-count, an approach taken sometimes in studies centred on dialectal identification; we need to establish the make-up and use of a particular lexico-semantic field in order to assess the significance of Norse impact.

Justyna Robinson, University of Sussex, and Seth Mehl, University of Sheffield

Linguistic DNA and new ways of exploring conceptual variation and change

Accounting for change in concepts and ideas has been a methodological challenge for many relevant disciplines, such as philosophy or history. In linguistics, this issue has been approached through the exploration of usage of words across time. On the one hand, the changing meaning of a word can be traced across time (semasiological perspective) providing access to, for example, shifting social evaluations of a given idea (*villain* 'villager' > 'bad character'). On the other hand, concepts can be explored through changes in the choice of names for a given idea at different points in time (onomasiological perspective).

Approaching conceptual change in the semasiological and onomasiological manner requires prior knowledge on changing ideas and words associated with them. Thus, one of the key challenges is to find bottom-up ways of investigating conceptual change and discovering relationships among words and ideas that exceed human intuition. Precisely this challenge is the premise of the Linguistic DNA project. The Linguistic DNA is a three-year AHRC-funded collaborative project which aims to investigate concepts and conceptual change in early modern England. This is done by developing new computational methods to analyse big data.

This article aims to assess the contribution of the Linguistic DNA methods in investigating conceptual change. We present case studies which outline the manner in which our computational approach works and how it challenges more traditional ways of investigating meaning change. The three areas of investigation presented in this paper are informed by the specific research questions pursued by the three research themes of the Linguistic DNA. Research theme 1 examines historical and social contexts of conceptual change. Research theme 2 analyses lexical semantic relationships within conceptual structures. Research theme 3 explores lexicalisation pressure, using categories from the Historical Thesaurus of English.

Kris Heylen and Dirk Geeraerts, KU Leuven

Of *Hereticks* and *Schismaticks*: Showcasing the Diachronic Analysis of Conceptual Construal in the Nephological Semantics Project

Background

The Nephological Semantics Project investigates how distributional modelling (see Lenci 2018 for a recent survey) can be deployed in the study of language variation and, more specifically, in the analysis of the semantic phenomena involved in linguistic variation. Further developing QLVL's long-standing research programme in Cognitive Semantics and Cognitive Sociolinguistics, the project's three work packages all analyse variation as an interaction between formal, semantic, and extra-linguistic (lectal) factors, but each with a different focus: WP1 looks at lexical variation and semasiological structure; WP2 analyses constructional, morpho-syntactic alternations; and WP3 considers lexical variation on an aggregate level for lectometric purposes, i.e. the measurement of overall con- and divergences in onomasiological preferences between language varieties (or *lects*). Distributional modelling and co-occurrence-based vector representations of semantics serve a double purpose in the project: Methodologically, they are a statistical tool to include, and control for, semantic factors in large-scale constructional alternation studies (WP2) and lectometric (WP3) analyses. On a more fundamental level however, distributional models are also the basis for a data-driven investigation of the phenomenon of *meaning* itself and how it emerges in, and from, contextual usage. In this talk, we focus on this latter application and, more specifically, on how distributional models are used in WP1 to analyse the semasiological structure of lexemes that represent alternative and changing *construals* of a concept through time. As construal is commonly defined as the alternative conceptualisation or categorization of the same state of affairs in reality (Verhaegen 2010), it entails an onomasiological perspective. Following Geeraerts (2016), we distinguish between *direct* and *indirect* construal as types of onomasiological salience: In direct construal, the competing lexicalizations themselves foreground a perspectivization through their taxonomic level, etymology or figurative usage. In indirect construal, lexicalizations become associated with a particular view or connotation through their frequent use in specific contexts. It's for the detection and analysis of this type of frequency-induced onomasiological salience that distributional modelling is particularly well suited.

Case study

For this workshop, we look at the alternative construals of PERSON WITH DEVIATING RELIGIOUS BELIEFS as a 'key' concept undergoing significant change in 17th Century English. The analysis is based on two time samples from the EEBO-TCP-corpus¹ with texts from 1649 (the end of the Second Civil War) and from 1699 (ten years after the Glorious Revolution and in the later reign of William of Orange), demarcating half a century during which the attitudes towards religious non-conformity changed considerably. From the online *Historical Thesaurus of the Oxford English Dictionary* (Kay et al. 2017), we selected the two

¹ Preliminary data from EEBO-TCP phase 2, lemmatized to a single spelling and POS-tagged with MorphAdorner, were kindly made accessible by the LDNA project team at the University of Sheffield for the duration of the first author's research stay in the Spring of 2016.

person reference nouns for religious non-conformist² that had a substantial number of occurrences in both time samples: heretick(s) (n=1298) and schismatick(s) (n=186). On the one hand, the terms' different etymologies lead to different *direct* construals that more or less remain stable over time: Whereas *hereticks* foregrounds the non-conformists' deviant beliefs (heresy), *schismatics* highlights their contribution to a breach of unity in the church (schisma). On the other, different indirect construals come about by the terms' association through usage with different contexts and these are highly variable over time. Using distributional models on the type and token level (cf. Heylen et al. 2015), we model the typical contexts and occurrences of *heretick* and *schismatic* in both time periods and, using token cloud visualisations³ (Heylen et al. 2015), we show, amongst other shifts in usage, that, whereas both words occur in contexts of political conflict in 1649, *schismatic* becomes the preferred term here in 1699. In other words, by 1699, referring to a religious non-conformist as a *schismatic* not only comes with a *direct* construal of “threatening the unity of the church”, it also activates, by virtue of frequent association, an *indirect* construal as “causer of political conflict”, which is a connotation that by 1699 is largely absent in the use of *heretic*.

References

- Geeraerts, D. (2016). Entrenchment as onomasiological salience. In: H.-J. Schmid (ed.) *Entrenchment and the psychology of language learning: how we reorganize and adapt linguistic knowledge*. Mouton De Gruyter: 153-174.
- Heylen, K., T. Wielfaert, D. Speelman, D. Geeraerts (2015) Monitoring polysemy: Word space models as a tool for large-scale lexical semantic analysis. *Lingua* 157. 153-172
- Kay, C., J. Roberts, M. Samuels, I. Wotherspoon, and M. Alexander (eds.). (2017). *The Historical Thesaurus of English, version 4.21*. Glasgow: University of Glasgow.
<http://historicalthesaurus.arts.gla.ac.uk/>.
- Lenci, A. (2018). Distributional Models of Word Meaning. *Annual Review of Linguistics* 4: 151-171.
- Verhaegen, A. (2007) Construal and Perspectivization. In: D. Geeraerts, H. Cuyckens (eds.) *The Oxford handbook of cognitive linguistics*. Oxford University Press: 48-81.

² Under headings *society > faith > aspects of faith > [heterodoxy|heresy|nonconformity|apostasy|sectarianism]* and including all spelling variants.

³ Hence the project's jocular title, from Greek *nephos* (cloud) + *logos* (study)