Digital History of Concepts: Sense Clustering over Time

Inga Kempfert[1], Saba Anwar[1], Alexander Friedrich[2], Chris Biemann[1]

[1] *Universität Hamburg,* [2] *Technische Universität Darmstadt*

{5kempfer, anwar, biemann}@informatik.uni-hamburg.de, friedrich@phil.tu-darmstadt.de

Mittwoch, 04.03.2020 15:45–16:30 ESA1 W Foyer

CLP

We present a tool for tracking word senses over time for enabling a digital, data-driven history of concepts, in an ongoing collaboration between computational linguistics and philosophy [FB 2016]. Research in the history of concepts deals with the historical semantics of terms with a special focus on the evolution or change of scientific or political concepts related to them. Special interest is given to the (often polysemic) meaning of basic concepts that are still crucial for the understanding of contemporary cultural, political or scientific world- and self-descriptions (such as "freedom", "power", "life", or "crisis"). A fundamental question here is how relevant contexts and periods of time of such changes can be identified. Our exploration tool allows to visualize the semantics of conceptual terms and their change over time, based on digital text analysis. The visualization shall help conceptual historians to identify and track significant changes of conceptual terms within and across different periods of time in different text corpora as well as to narrow down relevant contexts and source documents for further study.

Distributional semantics and word sense induction enable a data-driven approach to tracking word senses over time. Distributional semantics represents word meaning by their global contexts [MC 1991], allowing us to compute word similarity over large text corpora, such as the Google Books collection, using the graph-based JobimText framework [BR 2013]. Word sense induction creates data-driven hypotheses of coherent paradigms of target words, forming clusters that reflect different word senses. On time slices of time-stamped text corpora, we can access the formation, change and the demise of word meanings [MMM+. 2014].

Our contribution consists of SCoT (Sense Clustering over Time), a web interface to access the different senses of a word as they change over time. The paradigms of the target word are displayed as a graph, where word nodes are connected with edges indicating their similarity. The interface allows for parameterization of the graph creation and display as well as setting time intervals of interest. For the formation of concepts, graph clustering provides an automatic initialization of colour-coded sense clusters, which can be labelled and post-edited by the user, since clustering is known to produce distinctions that are highly correlated, but not necessarily congruent to the user's needs. For the visualization of differences, colour coding is employed to show, for selected time intervals, which paradigms resp. senses are stable, are added or fall out of use. This mode also allows for stepping through the time intervals for visually analysing the dynamics of change. Across both modes, it is possible to pin nodes onto the visualization canvas to ensure visual continuity of senses across time slices. The tool is implemented as a freely accessible web interface that allows locally saving and loading its current state, enabling the interruption and the sharing of sessions, and will be demonstrated live.

References: Biemann, C. & M. Riedl (2013). Text: Now in 2D! A Framework for Lexical Expansion with Contextual Similarity. J Lang Mod 1(1):55–95. Friedrich, A. & C. Biemann (2016). Digitale Begriffsgeschichte? Methodologische Überlegungen und exemplarische Versuche am Beispiel moderner Netzsemantik, Forum Interdisziplinäre Begriffsgeschichte 5(2), 78–96. Miller, G. & W. Charles (1991). Contextual Correlates of Semantic Similarity. Language and Cognitive Processes, 6(1), 1–28. Mitra, S., R. Mitra, S. Maity et al. (2015). An automatic approach to identify word sense changes in text media across timescales. J Nat Lang Eng 21(05), 773–798.