Semantic micro-dynamics as a reflex of occurrence frequency. A semantic networks approach

Klaus Hofmann[1], Andreas Baumann[1], Anna Marakasova[2], Julia Neidhardt[2] & Tanja Wissik[3]
[1] University of Vienna, firstname.lastname@univie.ac.at [2] Vienna University of Technology
[3] Austrian Centre for Digital Humanities and Cultural Heritage

Keywords: Semantic change, computational linguistics, corpus linguistics

Despite a long tradition of semantic change research (Paul, 1985), many of the factors involved in the process are still poorly understood. In particular, predicting change has remained elusive. This study seeks to contribute to our understanding of the parameters conditioning and influencing diachronic variation in lexical meaning by investigating the link between semantic change and frequency of occurrence.

Inspired by the achievements of recent large-scale studies employing state-of-the-art methods from the field of natural language processing (Hamilton et al., 2016; Tahmasebi et al., 2018), we approach word semantics from a distributional perspective. That is, we assume that a word's meaning is (at least in part) determined by, and can be expressed through, the linguistic context it customarily occurs in. We conceptualize this link in terms of exemplar theory (Nosofsky, 1988; Bybee, 2007). Thus, the semasiological dimension of a word can be thought of as an exemplar cloud around its lexical form, made up of rich, multi-modal memory traces of situations, objects, actions, or, broadly speaking, contexts in which a speaker-listener has experienced the lexeme being applied. Crucially, this also includes its habitual linguistic context (Bybee, 2013).

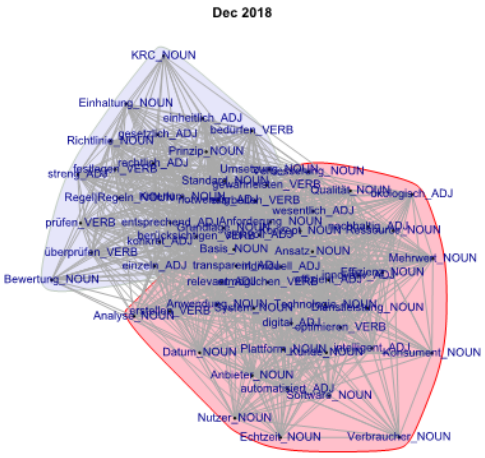

Fig. 1: Semantic network for the item transparent_ADJ, Dec 2018

We use embedding-based semantic networks (Hughes & Ramage, 2007; Akkasi & Snajder, 2021) (Fig. 1) to capture this link between lexical form and usage context - as reflected in a large linguistic corpus - and compare the networks diachronically in terms of their degree of similarity (Horn & Johnson, 1985). Our data come from the Austrian Media Corpus (AMC) (Ransmayr et al., 2017), whose size and high temporal resolution allow us to focus on microscopic variations in usage, measuring network similarity and word occurrence frequencies at monthly intervals.

The results of the regression analysis confirm that high occurrence frequency has a stabilizing effect on the semantic representations of words (Hamilton et al., 2016). We interpret this effect in terms of cognitive entrenchment: the more frequently words occur in the ambient language, the more stably they become entrenched in the minds of speakers alongside memory traces of the usage contexts they occur in.

References

Akkasi, A. & J. Snajder. 2021. Word sense Induction using Leader-follower Clustering of Automatically Generated Lexical Substitutes. *Expert Systems with Applications* 181. doi:10.1016/j.eswa.2021.115162.

Bybee, Joan. 2007. *Frequency of Use and the Organization of Language*. Oxford: Oxford University Press.

Bybee, Joan. 2013. Usage-based Theory and Exemplar Representations of Constructions. In Thomas Hoffmann & Graeme Trousdale (eds.), *The Oxford Handbook of Construction Grammar*, 49–69. Oxford: Oxford University Press.

Hamilton, William L., Jure Leskovec & Dan Jurafsky. 2016. Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change. In *Proceedings of the 54th annual meeting of the association for computational linguistics*, Association for Computational Linguistics. https://arxiv.org/abs/1605.09096.

Horn, Roger A. & Charles R. Johnson. 1985. *Matrix Analysis*. Cambridge: Cambridge University Press.

Hughes, Thad & Daniel Ramage. 2007. Lexical semantic relatedness with random graph walks. In *Emnlp-conll*, 581–589.

Nosofsky, Robert M. 1988. Exemplar-based Accounts of Relations between Classification, Recognition, and Typicality. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 14(4). 700–708. doi:10.1037/0278-7393.14.4.700.

Paul, Hermann. 1985. *Prinzipien der Sprachgeschichte*. Berlin: de Gruyter.

Ransmayr, Jutta, Karlheinz Mörth & Matej Ďurčo. 2017. AMC (Austrian Media Corpus) – Korpusbasierte Forschungen zum österreichischen Deutsch. In C. Resch & W. U. Dressler (eds.), *Digitale Methoden der Korpusforschung in Österreich*, 27–38. Wien: Verlag der Österreichischen Akademie der Wissenschaften.

Tahmasebi, Nina, Lars Borin & Adam Jatowt. 2018. Survey of computational approaches to lexical semantic change. In *Preprint at arxiv 2018.*, https://arxiv.org/abs/1811.06278.