

Semantic shifts in Austrian public discourse: A lexical networks approach

Andreas Baumann¹, Klaus Hofmann¹, Anna Marakasova², Julia Neidhardt²,
Tanja Wissik³

¹ University of Vienna, ² TU Wien, ³ Austrian Academy of Sciences

andreas.baumann@univie.ac.at, klaus.hofmann@univie.ac.at, anna.marakasova@tuwien.ac.at,
julia.neidhardt@ec.tuwien.ac.at, tanja.wissik@oeaw.ac.at

Recent years have seen a marked increase in the application of computational methods for the detection of semantic change (Tahmasebi et al. 2018). In order to add to this evolving field, we present ongoing research from the project *Diachronic Dynamics of Lexical Networks* (Baumann et al. 2019). The two main goals of the project are investigating (1) network-based methods for semantic shift detection, and (2) factors that bring about semantic change.

In terms of its scope, our research is concerned with the integration and analysis of two diachronic corpora: the Austrian Media Corpus, which contains more than 30 years of the journalistic prose published in Austria (Ransmayr et al. 2017), and the ParlAT corpus, which covers 20 years of Austrian parliamentary records (Wissik & Pirker 2018). The use of two parallel corpora allows us to explore (1) to what extent diachronic dynamics of semantic change are sensitive to the language domain, and (2) to what extent lexico-semantic change is driven by individuals (in our case, politicians). The choice of the data is motivated by our (variationist) interest in Austrian German and its public discourse(s). The data is closely tied to the events that have been prominent in Austrian social and political life. As we consider a relatively short time span (20 years), our approach aims to detect subtle shifts in the usage of words. We take this to primarily reflect changes in discourse, which may potentially lead to semantic change in the narrow sense. For example, before 2015 the word *Balkanroute* ‘Balkan route’ mainly occurred in the context of drug trafficking. Since then, it has become one of the most politically charged words in Austria, denoting a major migration route into Central Europe.

Many of the current methods for the automatic detection of semantic change make use of word embedding algorithms (Tahmasebi et al. 2018). These methods come along with their own challenges: they are unstable and sensitive to word frequency and corpus size. Instead, we opt for a network-based approach (Barabási 2016), which allows us to define the exact type of change and to deal with the aforementioned issues since frequency counts are used separately for each network of the target word. For the selected target words we build ego-networks (one for each time slice), cluster them in order to obtain meaningful representations of the words’ usage in a certain time period, and, finally, we use statistical measures to track to what extent words within the clusters change over time. We will discuss our approach in more detail along with some open research questions.

References: Barabási, A.-L. 2016. *Network Science*. Cambridge University Press, 2016. Baumann, A., Neidhardt, J., Wissik, T. 2019. DYLEN: Diachronic Dynamics of Lexical Networks. In *Proceedings of the LDK-PS 2019*. Ransmayr, J., Mörth, K., Đurčo, M. 2017. AMC (Austrian Media Corpus) – Korpusbasierte Forschungen zum Österreichischen Deutsch, 27–38. Verlag der ÖAW, 2017. Tahmasebi, N., Borin, L., Jatowt. 2018. *A Survey of Computational Approaches to Lexical Semantic Change*. Wissik, T., Pirker, H. 2018. ParlAT Beta Corpus of Austrian Parliamentary Records. In *Proceedings of the LREC 2018 Workshop ParlaCLARIN*. European Language Resources Association.