



Semantic Shifts

Fréjus - Destremau
12-23 Sept. 2022

Semantic Shifts: From lexicon to grammar. Diachronic and typological perspectives

Workshops, Methods & Tools 3

Computer-Assisted Approaches to Lexical Typology

Johann-Mattis List & Annika Tjuka

In contrast to purely computational or purely manual approaches, computer-assisted approaches in historical and typological language comparison focus on the interface between quantitative, data-driven research, and qualitative evaluation by experts. In the field of lexical typology, these frameworks are of specific importance, because datasets can be large -- and thus difficult to be analyzed by eye-balling them alone -- while problems involving questions of semantics can still not be sufficiently handled in automatic frameworks alone. An example for the power of computer-assisted frameworks is new version of the *Database of Cross-Linguistic Colexifications* (CLICS², List et al. (2018), <https://clics.clld.org>). In this database, which shows cross-linguistic, automatically inferred patterns of polysemy (or colexifications) across more than 1000 languages world-wide, interactive visualization techniques are used to break down the massive amount of information in such a way that linguists can quickly explore the data in order to find interesting patterns that automatic approaches could not identify.

In the course, we will discuss how computational techniques can be employed in order to filter large amounts of lexical data in order to ease qualitative investigations. In this context, we will discuss the CLICS² in detail, pointing not only to its strong and weak points, but also illustrating how its design principles can be used by the participants themselves as starting point for individual investigations. In addition, we will discuss basic techniques of data exploration, which go beyond the search of patterns in web-based interfaces, enabling the participants to pursue their own computer-assisted exploratory data analysis, and discuss how scholars have analysed datasets such as CLICS² in the past (Gast and Koptjevskaja-Tamm 2018).

References

Gast, Volker, and Maria Koptjevskaja-Tamm. 2018. "The Areal Factor in Lexical Typology. Some Evidence from Lexical Databases." In *Aspects of Linguistic Variation*, edited by Daniël Olmen, Tanja Mortelmans, and Frank Brisard, 43–81. Berlin; New York: de Gruyter.

List, Johann-Mattis, Simon J. Greenhill, Cormac Anderson, Thomas Mayer, Tiago Tresoldi, and Robert Forkel. 2018. "CLICS². An Improved Database of Cross-Linguistic Colexifications Assembling Lexical Data with Help of Cross-Linguistic Data Formats." *Linguistic Typology* 22 (2): 277–306.