Persistence and Change of Colexifications in Indo-European

Languages differ in how they organize their lexicon: two given concepts can be expressed by two different word forms in one language, but by the same form in another language. This encoding of different concepts with the same word form is known as colexification (François 2008). Studies have shown that highly associated concepts are colexified more frequently across languages (Xu et al. 2020), and that colexification patterns are less genetically stable than the phonological form of vocabulary, at least in Europe (Gast & Koptjevskaja-Tamm 2022). Colexification patterns can also be borrowed, as various examples of areally common colexifications show (e.g., Schapper 2022 on the colexification of BONE and STRENGTH in Melanesia or Segerer & Vanhove 2022 on the colexification of color terms in Africa).

The present study is concerned with colexification in the Indo-European language family, and its relationship to lexical material. It aims to answer the question of how strongly a colexification is tied to the phonological material its concepts are expressed with, and if it is likely to persist when phonological material is replaced. To investigate this, we use vocabulary lists from the CLICS database (Rzymski et al. 2021), and etymology data from IELex (Chang et al. 2015) in order to split the vocabulary items into cognate sets.

When relating colexifications to etymological data, there are two main options:

(i)   Colexifications are independent of etymology and are equally found among unrelated lexemes denoting the same concept.
(ii)   Colexifications are dependent on etymology and are more frequently found with lexemes of a certain etymology.

(i) suggests that colexifications arise, persist, and spread independently of phonological substance, and that, more generally, the development of patterns is unrelated to the development of matter. (ii) on the other hand, would suggest that colexifications arise rarely and are then bound to a certain phonological form. In this scenario colexifications mostly spread and persist together with phonological matter.

As it has been shown that certain colexifications are more common cross-linguistically even among unrelated languages in different areas, we expect that colexifications in the Indo-European family are largely (but not completely) independent from the etymology of their word forms.

In a further stage of this study, we plan to use phylogenetic methods in order to investigate the history of certain colexifications in the Indo-European language family. We also plan to conduct a similar study on Austronesian languages in order to compare language families that differ strongly in terms of culture and geographical environment.

Literature

Chang, Will, David Hall, Chundra Cathcart, and Andrew Garrett. "Ancestry-constrained phylogenetic analysis supports the Indo-European steppe hypothesis." *Language* (2015): 194-244.

François, Alexandre. "Semantic maps and the typology of colexification." *From polysemy to semantic change: Towards a typology of lexical semantic associations* 106 (2008): 163.

Gast, Volker, and Maria Koptjevskaja-Tamm. "Patterns of persistence and diffusibility in the European lexicon." *Linguistic Typology* 26, no. 2 (2022): 403-438.

Rzymski, Christoph, Tiago Tresoldi, Simon J. Greenhill, Mei-Shin Wu, Nathanael E. Schweikhard, Maria Koptjevskaja-Tamm, Volker Gast et al. "The Database of Cross-Linguistic Colexifications, reproducible analysis of cross-linguistic polysemies." *Scientific data* 7, no. 1 (2020): 13.

Schapper, Antoinette. "Baring the bones: The lexico-semantic association of bone with strength in Melanesia and the study of colexification." *Linguistic Typology* 26, no. 2 (2022): 313-347.

Segerer, Guillaume, and Martine Vanhove. "Areal patterns and colexifications of colour terms in the languages of Africa." *Linguistic Typology* 26, no. 2 (2022): 247-281.

Xu, Yang, Khang Duong, Barbara C. Malt, Serena Jiang, and Mahesh Srinivasan. "Conceptual relations predict colexification across languages." *Cognition* 201 (2020): 104280.