05: Cross-Linguistic Strategies of Denotation: Advancing data and methods in lexical typology

Lead supervisor:

Johann-Mattis List
Max Planck Institute for the Science of Human History
Independent Research Group Leader: CALC
Kahlaische Str. 10
07745 Jena, Germany
Email: list@shh.mpg.de
http://www.shh.mpg.de/375796/calc

Co-supervisor:

Prof. Dr. Volker Gast
Friedrich-Schiller-Universität Jena
Institut für Anglistik und Indogermanistik
07743 Jena, Germany
Ernst Abbe Platz 8
E-Mail: volker.gast@uni-jena.de
https://www.iaa.uni-jena.de/

Involved subjects: historical linguistics, language evolution, cognition

Number of positions requested: 1

Abstract:

Although on first sight there are many different ways in which the languages classify the world around us, many of the techniques used for naming are surprisingly similar across very different languages. Researchers of the Department of Linguistic and Cultural Evolution have already illustrated this in ongoing projects, such as the Database of Cross-Linguistic Colexifications (https://clics.clld.org), which make use of advanced techniques for data curation and analysis, involving reference catalogs that link concepts to specific identifiers (https://concepticon.clld.org), and data formats that allow for a quick access and reuse to available dataset (https://cldf.clld.org). Despite these achievements, both the data basis underlying these projects and the methods that are currently used to curate and analyze cross-linguistic language data to uncover global and areal naming patterns, are – however – still in need of further improvement.

To learn more about language evolution, for example, the Concepticon reference catalog needs to be further annotated, especially by adding additional meta-data from external sources established in psycholinguistic and neurolinguistic studies, including norm data on age-of-acquisition, word frequencies across different languages, and speaker ratings on word concreteness and familiarity. Furthermore, the CLICS database needs to be expanded further by adding languages from specific areas that are not yet sufficiently covered. For the expansion of the data, reliable workflows developed by members of the DLCE are available, as represented specifically by the Lexibank initiative that seeks to serve as a repository for cross-linguistic wordlist data, similar to the role that GenBank plays in biology.

In addition to the expansion of the data using our well-established frameworks, new methods need to be developed, allowing for a more fine-grained analyses. Given the great potential of cross-linguistic language data to help in answering not only concrete questions about human prehistory, but also abstract questions dealing with the evolution of human perception and cognition, a richer data basis of cross-linguistic lexical associations will need new and innovative methods to answer important questions on human prehistory. In concrete, we hope that improved clustering techniques for

community detection, advanced graph-based methods for fuzzy clustering, as well as various techniques developed for machine learning applications may help to gain more insights into the power of quantitative approaches to lexical typology.

In addition to addressing purely linguistic questions of language, perception, and cognition, the project can also integrate across disciplines, given that a study of naming patterns, like, for example, the words that are used by different languages to denote artifacts, domesticated plants or animals, or animal products, like milk, also provide hints on the concrete role these artifacts played in different cultures.

The ideal applicant should have a background in corpus linguistics / computational linguistics, or related topics. Alternatively, a historical linguist with experience in data managment or computing may also provide the qualifications we need. Applicants should have a good command on Python or be willing to learn the programming language quickly. A knowledge of modern visualization techniques (such as JavaScript) is an absolute plus, but no general requirement. Strong team player qualities will be needed, as candidates will work in an interdisciplinary and international team where communication and team work among the members will be crucial for the success of the project.