

Michael Rießler, Mervi de Heer, Terhi Honkola, Unni-Päivä Leino, Kaj Syrjänen, Outi Vesakoski
University of Freiburg, University of Turku

Basic vocabulary and the phylogenetic approach to the study of Uralic language history

In this presentation we offer an overview a) of the basic vocabulary data with cognate (correlate) assessments we have collected for Uralic languages, and b) of the Bayesian phylogenetic methods we are using to analyse these data for studying the history of the Uralic language family.

We will start by introducing our data, which consists of cognate sets taken from basic vocabulary in 26 Uralic languages. Cognate assessments are taken from earlier literature on Uralic historical linguistics and show which meanings share a common origin. We also discuss – and illustrate with examples – how the more and less stable subsets in the vocabulary affect the analyses and how the benefits of the phylogenetic approach are lost if the amount of data is too small.

Since the type of data we are using resembles binary data used in evolutionary biology it can be analysed with the well-established phylogenetic computational tools. We will show, how the computational approach offers new tools to study historical linguistics. This is especially interesting for the objective handling of large data sets. However, it is important to understand that using of these methods as well inferring the outcomes are not completely transparent tasks. We will discuss these issues to be taken into account in planning the analyses and in inferring the language family trees using Bayesian algorithms.

Our presentation aims specifically at clarifying some of the common misunderstandings regarding our approach to historical linguistics, which were recently also raised in a monograph on “The Indo-European Controversy” by Pereltsvaig & Lewis (2015). Although their critics are not entirely unjustified, we will describe how our approach, called language phylogenetics”, takes the advantage of the vast knowledge resulting from earlier basic research in qualitatively oriented historical-comparative linguistics, and converts it into quantitatively-oriented “applied historical linguistics”.

References

- Lehtinen, J., Honkola, T., Korhonen, K., Syrjänen, K., Wahlberg, N. & Vesakoski, O. (2014) "Behind family trees: Secondary connections in Uralic language networks", *Language Dynamics and Change* 4: 189–221. DOI: 10.1163/22105832-00402007.
- Pereltsvaig, A. & Lewis, M. W. (2015) *The Indo-European Controversy: Facts and Fallacies in Historical Linguistics*. Cambridge.
- Syrjänen, K., Honkola, T., Korhonen, K., Lehtinen, J., Vesakoski, O., & Wahlberg, N. (2013) "Shedding more light on language classification using basic vocabularies and phylogenetic methods: a case study of Uralic", *Diachronica* 30:3, 323–352. DOI: 10.1075/dia.30.3.02syr.