Horizontal transfer and phylogenetic calibration in linguistics: a Bayesian approach

Luke Maurits 1  Mervi de Heer 2  Michael Dunn 2  Outi Vesakoski 1
1University of Turku, Turku, Finland  2University of Uppsala, Uppsala, Sweden

Introduction & Summary

• Dating language divergence poses a challenge in linguistics
• Bayesian methods and results of historical-comparative linguistics together help answer the challenge.
• Absolute calibrations are usually obtained to obtain divergence times of a linguistic subfamilies in calendar years. However, they are hard to establish[1], especially for families without written records.
• Most linguistic datasets contain results of horizontal transfer, e.g. loanwords, due to language contacts.
• Although often thought of as “noise”, loanwords carry valuable timing information as contact happens between contemporaneous languages.
• Uralic has long-standing contacts with Indo-European, a family which has been extensively researched and whose timing is better understood.
• We test “relative calibration” of linguistic phylogenies using Indo-European loanwords acquired into the Uralic family and the effects on dating Proto-Uralic and intermediate protolanguages.

Materials

Uralex 1.0. and its tagged borrowings
• Basic vocabulary dataset of 26 Uralic languages [3]
• Known borrowings tagged using etymological literature [4]
• Certainty estimate given using evaluative literature

Example: borrowings in Standard Estonian

Results & Conclusion

• Our relative calibration method estimates Proto-Uralic divergence at approx 5,250 YBP (95% HPD 4,000–6,600).
• Analysing same data with absolute calibrations based on loanwords yields a much older date of approx 6,700 YBP[6].
• Both estimates older than a recent popular hypothesis in Uralic linguistics of ca. 4,000 YBP[7, 8].

Methods

• Simultaneous sampling of Uralic and IE trees in a single MCMC chain using BEAST 2.
• Uralic tree includes UraLex data, whose evolution is modelled with a relaxed lognormal clock.
• The Indo-European tree has no data, but interior node times are constrained to match published posterior.
• Sampling is constrained so that trees are only accepted if there is some non-zero overlap in the lifespan of the ancestors of each pair of subfamilies with identified borrowings.

References


Acknowledgements


Contact Information

• Web: https://bedlan.net
• Email: luke@maurits.id.au, mervi.deheer@moderna.uu.se

Luke Maurits 1  Mervi de Heer 2  Michael Dunn 2  Outi Vesakoski 1
1University of Turku, Turku, Finland  2University of Uppsala, Uppsala, Sweden