

Correlational causality with phylogeny

Annemarie Verkerk (University of Reading)

oral

Comparative phylogenetic methods are a set of statistical analyses that can be used to study the evolution and distribution of typological characteristics of related languages while accounting for the statistical independence introduced by their genealogical relations (Harvey and Pagel 1991). They can be used to study the correlation between two or more typological characteristics, as they include various types of correlation and regression analyses. In addition, they also include a small set of methods that can be used to assess the relative timing of evolutionary change of two or more features. This small set includes the evolutionary lag analysis proposed by Deaner and Nunn (1999), who study change in brain size as 'lagging behind' (or: caused by) change in body size. Given that the application of these methods is quite new in typology, this paper aims to answer questions about the types of data that they require, as well as the language sample size and sample structure that they call for.

This paper is mostly methodological in nature and discusses a range of case-studies in order to demonstrate the relevance of comparative phylogenetic methods for studying correlational causality. The data for these case-studies is taken from WALS (Dryer and Haspelmath 2013) and other sources, with the WALS data supplemented by additional data taken from grammars for the Austronesian language family. The two major data types discussed are discrete data and continuous data, because each needs to be analysed with different evolutionary models. Discrete data takes either two states (binary) or three or more distinct states (categorical), while continuous data are real number measurements on an interval or ratio scale. Case-studies are given of correlation analyses using both data types. Most WALS chapters arrange responses in three or more categorical classes. I discuss why it is difficult to model correlations between such multistate classes, but show that it is possible to make most WALS chapters binary in an insightful way. Lastly, I discuss sample structure and size, and I consider how different samples and sample sizes may effect the outcome, as well as how big a language family needs to be in order to use comparative phylogenetic methods in a meaningful way.

References

- Deaner, Robert O. & Charles L. Nunn. 1999. How quickly do brains catch up with bodies? A comparative method for detecting evolutionary lag. *Proceedings of the Royal Society of London. Series B: Biological Sciences* 266: 687-94.
- Dryer, Matthew S. & Martin Haspelmath (eds.) 2013. *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at <http://wals.info>)
- Harvey, Paul H. & Mark D. Pagel. 1991. *The comparative method in evolutionary biology*. Oxford: Oxford University Press.