An Alternative to Trees in Bayesian Language Histories

T. Mark Ellison[1] & Luisa Miceli[2]

[1]Universität zu Köln, Germany, [2]University of Western Australia

The tree model of the history of language families is credited to Schleicher (1853) - although the common origins of a number of languages had long been suspected (see Campbell & Poser 2008 for details). The tree model sees vertical - so-called "normal" - transmission from parents to children as the fundamental cause of shared structure between languages. Schleicher's student Schmidt (1872) saw limitations in the tree model, and proposed an alternative account of language relatedness, the wave model. In this account, horizontal - between-peer - transmission allows different forms or structures to spread through a linguistic area from distinct starting points to overlapping extents.

These two approaches have continued to the present day, with historical linguists tending to favour tree models, and those looking at synchronic variation recognising models of transmission through social-networks. Recently, the tree model has been reinvigorated by the introduction of Bayesian phylogenetic methods, originally developed to reconstruct biological lineages (e.g. Gray & Atkinson 2003, see Bowern 2018 for an overview).

In this talk, we address one of the assumptions made in the context of recent phylogenetic analyses: that convergence of the tree-construction algorithm is evidence that the historical origins of the languages was a tree-like sequence of three stages - unity, split, then independent divergence. We argue instead that horizontal transmission between language communities can give rise to distributions of characters that cause Bayesian phylogenetic analyses to converge on a tree, whether or not their actual history was tree-like. This develops a similar argument to that proposed about dialects by Murawaki (2015).

This argument is developed in computational simulations, in which linguistic communities start with unrelated languages, but develop with limited horizontal transmission subject to geographic impediments to communication. We show that Bayesian tree-building methods can and do interpret simulated results in such scenarios as having tree-like phylogenetic signal, leading them to converge. Furthermore, current homoplasy measures do not necessarily flag data generated in this way as not tree-like.

In this talk we also explore an alternative model of language histories which can, like phylogenies, be tested with Bayesian reasoning. In these models, rates of convergence and divergence of characters are determined by parameters defining the influence of contact between communities - parameters that can change over time.

In conclusion, care needs to be taken in imputing historical verisimilitude to tree models just because they are able to account for a set of data. We show that non-tree models offer equally plausible explanations of some data sets.

Bowern, C. 2018. **Computational Phylogenetics.** *Annual Review of Linguistics*, 4:281-296.

Campbell, L. and Poser, W.J., 2008. *Language classification: history and method*. Cambridge University Press.

Gray, Russell D., and Quentin D. Atkinson. 2003. **Language-tree divergence times support the Anatolian theory of Indo-European origins.** *Nature* 426.435–9.

Murawaki, Y. 2015. **Spatial Structure of Evolutionary Models of Dialects in Contact.** *PLOS ONE*, 10(7), e0134335. https://doi.org/10.1371/journal.pone.0134335

Schleicher A. 1853. **Die ersten Spaltungen des indogermanischen Urvolkes.** *Allgemeine Monatsschrift für Wissenschaft und Literatur*. 3:786–787.

Schmidt, Johannes. 1872. *Die Verwandtschaftverhältnisse der indogermanischen Sprachen*. Weimar: Böhlau.