LANGUAGE FAMILY TREES REFLECT GEOGRAPHY AND DEMOGRAPHY BEYOND NEUTRAL DRIFT

CHRISTIAN BENTZ[*1,2], DAN DEDIU[3], ANNEMARIE VERKERK[4], and GERHARD JÄGER[1,2]

[*]Corresponding Author: chris@christianbentz.de
[1]Department of General Linguistics, University of Tübingen, Tübingen, Germany
[2]DFG Center for Advanced Studies, University of Tübingen, Tübingen, Germany
[3]Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands
[4]Max Planck Institute for the Science of Human History, Jena, Germany

1. Introduction

Languages constantly evolve. Language families unfold in time and space, and span the globe. A core question of evolutionary linguistics is whether this remarkable diversity is the outcome of mere neutral drift, or if further pressures relating to geography, climate, population size, and human biology guide diversification. We apply phylogenetic signal analyses to measure the reflection of longitudes, latitudes, altitudes and population sizes on language family trees. In particular, we evaluate the evidence that language family tree structure and external factors have evolved completely independently (phylogenetic signal of zero), if they have evolved by neutral drift (phylogenetic signal of one), or whether there are further adaptive and non-adaptive pressures at play (phylogenetic signal between zero and one or bigger than one).

2. Methods

2.1. *Tree samples*

We collect phylogenetic trees from three major sources: 1) A database harnessing openly available tree topologies (Dediu, in press). 2) Bayesian posterior trees supplied by authors of recent phylogenetic studies for overall eight language families. 3) Trees derived via the maximum likelihood (ML) method applied to ASJP word lists (Wichmann et al., 2013). Overall, we arrive at a sample of 52588 trees representing 54 language families.

2.2. *Geographic and demographic information*

Latitude and longitude information per language is taken from Glottolog (Hammarström, Forkel, Haspelmath, & Bank, 2016). Altitudes are estimated using the

Google Maps Elevation API[1]. Population size data is taken from the Ethnologue (Lewis, Simons, & Fenning, 2013). We arrive at a sample of 6834 languages (unique ISO-639-3 and glottocodes) for which latitude, longitude, altitude, and population size is available.

2.3. *Phylogenetic signal metrics*

In a linguistic context, phylogenetic signal can be conceptualized as the extent to which languages close to each other on a phylogenetic tree also resemble each other in their population-level "external" traits. There is a range of methods to estimate phylogenetic signal (Münkemüller et al., 2012; Blomberg & Garland, 2002). We focus here on two metrics in particular: *Blomberg's K* (Blomberg, Garland Jr, Ives, & Crespi, 2003), and *Pagel's* λ (Pagel, 1997).

3. Results and Discussion

We investigate 4 external factors by both signal metrics, K and λ, for all three sources of trees. For these 24 combinations, geographic factors and population size have mean phylogenetic signals significantly higher than zero (Wilcoxon test: $p < 0.05$). The mean values across the 12 combinations per signal metric range from $\bar{\lambda} = 0.19$ ($SD = 0.26$) to $\bar{\lambda} = 1.001$ ($SD = 0.071$), and from $\bar{K} = 0.35$ ($SD = 0.15$) to $\bar{K} = 4.32$ ($SD = 4.54$) respectively. These are also generally significantly different from one ($p < 0.05$). Analyses for 128 combinations by signal metric, external factor and language family yield qualitatively similar results.

Hence, phylogenetic signals are generally stronger than expected under the null hypothesis of completely independent evolution (i.e. $\bar{K} = 0$ and $\bar{\lambda} = 0$). Moreover, the prediction of neutral drift (i.e. $\bar{K} = 1$ and $\bar{\lambda} = 1$) is mostly not confirmed. This suggests that there are further adaptive and non-adaptive pressures at play when language families evolve and diversify. Based on systematic differences between language families we argue that these include convergent evolution, niche occupancy, heterogeneous rate drift and lateral transfer of lexical and structural material. Finally, there is preliminary evidence that language families expand more along longitudes than latitudes, thus confirming the hypothesized east-west tendency of human migrations (Diamond, 1999; Güldemann & Hammarström, in print).

Acknowledgements

---

[1]https://developers.google.com/maps/documentation/elevation

References

Blomberg, S. P., & Garland, T. (2002). Tempo and mode in evolution: phylogenetic inertia, adaptation and comparative methods. *Journal of Evolutionary Biology*, *15*(6), 899–910.

Blomberg, S. P., Garland Jr, T., Ives, A. R., & Crespi, B. (2003). Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution*, *57*(4), 717–745.

Dediu, D. (in press). Making Genealogical Language Classifications Available for Phylogenetic Analysis: Newick Trees, Unified Identifiers, and Branch Length. *Language Dynamics and Change*.

Diamond, J. M. (1999). *Guns, germs and steel: the fates of human societies.* New York/ London: W. W. Norton.

Güldemann, T., & Hammarström, H. (in print). Geographical axis effects in large-scale linguistic distributions. In M. Crevels & P. Muysken (Eds.), *Language dispersal, diversification and contact.* Oxford: Oxford University Press.

Hammarström, H., Forkel, R., Haspelmath, M., & Bank, S. (Eds.). (2016). *Glottolog 2.7.* Jena: Max Planck Institute for the Science of Human History.

Lewis, M. P., Simons, G. F., & Fenning, C. D. (Eds.). (2013). *Ethnologue: Languages of the world* (17th ed.). Dallas, Texas: SIL International.

Münkemüller, T., Lavergne, S., Bzeznik, B., Dray, S., Jombart, T., Schiffers, K., & Thuiller, W. (2012). How to measure and test phylogenetic signal. *Methods in Ecology and Evolution*, *3*(4), 743–756.

Pagel, M. (1997). Inferring evolutionary processes from phylogenies. *Zoologica Scripta*, *26*(4), 331–348.

Wichmann, S., Müller, A., Wett, A., Velupillai, V., Bischoffberger, J., Brown, C. H., Holman, E. W., Sauppe, S., Molochieva, Z., Brown, P., Hammarström, H., Belyaev, O., List, J.-M., Bakker, D., Egorov, D., Urban, M., Mailhammer, R., Carrizo, A., Dryer, M. S., Korovina, E., Beck, D., Geyer, H., Epps, P., Grant, A., & Valenzuela, P. (2013). *The ASJP database (version 16).*