HORIZONTAL TRANSFER OF LINGUISTIC FEATURES AS AN INTEGRAL PART OF THE EVOLUTIONARY HISTORY OF LANGUAGE FAMILIES – PHYLOGENETIC RECONSTRUCTION OF LINGUISTIC TREE–BASED NETWORKS

Lina Herbst[*1,2,3], Volker Gast[1], Denise Kühnert[2,3], and Mary Walworth[2,3]

[*]Corresponding Author: lina.herbst@uni-jena.de
[1]Department of English and American Studies, Friedrich Schiller University Jena, Jena, Germany
[2]Max Planck Institute for the Science of Human History, Jena, Germany
[3]Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany

More than 7000 different languages from more than 200 language families are spoken worldwide today. The evolutionary relationships within a language family are often illustrated and studied using a linguistic phylogeny, also called a language family tree. Based on linguistic features this allows inference of the internal structure and the time-depth of a language family. However, linguistic features are not always "inherited", they can also be acquired horizontally through language contact.

Non-treelike evolution is also common in evolutionary biology. Scholars have represented these horizontal relationships within phylogenetic networks, a tree with horizontal edges. By allowing linguistic features to move horizontally, a more accurate representation of linguistic evolution is possible. However, in historical linguistics network reconstructions are still in their infancy and no standard has been established so far.

Pioneering work by Nelson-Sathi et al. uncovered hidden lexical borrowing in Indo-European (IE) languages, showing that borrowing is more widespread than previously thought (Nelson-Sathi et al., 2011). A software library based on this approach was provided by List and Moran (List & Moran, 2013). In addition, minimal lateral networks (MLNs) were applied to 40 IE languages by investigating presence-absence patterns of cognate sets (List, Nelson-Sathi, Geisler, & Martin, 2014). The MLN approach uses weighted parsimony and thus allows for a certain proportion of parallel evolution (List et al., 2014). List et al. have also investigated the history of 40 Chinese dialects using the MLN approach (List, Nelson-Sathi, Martin, & Geisler, 2014). In addition, Willems et al. used distance data from IE languages to infer hybridization networks (Willems et al., 2016). Cathcart et al. introduced a novel method for investigating areal dispersal of linguistic features, and they applied this approach to 117 IE languages (Cathcart, Carling, Larsson,

Johansson, & Round, 2018). In 2019, a mixture-model approach was applied to the Austronesian, Sinitic, IE and Japonic language families to detect non-tree-based signals (Verkerk, 2019), and there are even more examples for the detection of linguistic transfer events and the reconstruction of linguistic networks.

In this work, we develop a novel process that generates rooted and timed linguistic networks in which horizontal edges account for transfer events between languages. These networks will be based on a birth-death process, which will be adapted to account for horizontal transfers to detect the timing and the location of linguistic transfer. The novel process will be integrated in an open-source Bayesian phylogenetic framework, such that knowledge about the past can be included e.g. on the depth of a sub-family. We will evaluate and quantify the effects horizontal transfer has on different aspects of language from the grammar to the lexicon. The process will be applied to the Polynesian language family, which has been suggested to have a highly reticulate history with many linguistic transfers. This will shed light on the timing and phylogenetic placement of linguistic transfers among the Polynesian languages and reveal which linguistic features were transferred.

Acknowledgements

Lina Herbst was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – 461279234.

References

Cathcart, C., Carling, G., Larsson, F., Johansson, N., & Round, E. (2018). Areal pressure in grammatical evolution. *Diachronica. International Journal for Historical Linguistics*, *35*(1), 1–34.

List, J.-M., & Moran, S. (2013). An open source toolkit for quantitative historical linguistics. *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, 13–18.

List, J.-M., Nelson-Sathi, S., Geisler, H., & Martin, W. (2014). Networks of lexical borrowing and lateral gene transfer in language and genome evolution. *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology*, *36*(2), 141–150.

List, J.-M., Nelson-Sathi, S., Martin, W., & Geisler, H. (2014). Using phylogenetic networks to model chinese dialect history. *Language Dynamics and Change*, *4*(2), 222–252.

Nelson-Sathi, S., List, J.-M., Geisler, H., Fangerau, H., Gray, R. D., Martin, W., & Dagan, T. (2011). Networks uncover hidden lexical borrowing in indo-european language evolution. *Proceedings. Biological Sciences*, *278*(1712), 1794–1803.

Verkerk, A. (2019). Detecting non–tree–like signal using multiple tree topologies. *Journal of Historical Linguistics*, *9*(1), 9–69.

Willems, M., Lord, E., Laforest, L., Labelle, G., Lapointe, F.-J., Sciullo, A. M. D., & Makarenkov, V. (2016). Using hybridization networks to retrace the evolution of indo-european languages. *BMC Evolutionary Biology*, *16*(180).