

The Dialect Continuum Tree

Erik Elgh* & Harald Hammarström*

[This is a submission for the workshop 'Ancient languages and Algorithms: Demystifying new methods in historical linguistics']

The dialect continuum has been a perennial problem for the tree model of language relationship and subgrouping. The gradual spread of a language over an area with subsequent regional and overlapping innovations is in stark contrast to the 'clean splits' of sister groups suggested by the tree. Alternative models, highlighting the dialectal relations, have existed almost from the moment the tree truly entered historical linguistics: Already [Schmidt, 1872] proposed the so called *Wellentheorie*, or 'wave model', emphasizing the areal character of linguistic innovations. While showing the distribution of isoglosses, the wave model suffers from simply being a data display system and the graph in itself contains no inherent notion of diachronic process.

In the last 15 years there have been quite a few publications making use of Neighbor-Net [Bryant and Moulton, 2004] to represent language relationships with potential dialectal complications, e.g. [Hamed, 2005], [Holden and Gray, 2006] and [Lehtinen et al., 2014]. Just like the traditional wave model, these network graphs do not directly represent any process of divergence or contact. They are purely data display systems.

In contrast, the work of [Ross, 1988, 7-11] explicitly incorporates dialect continua ('linkages') into a tree based diachronic representation. It distinguishes clean, tree-like splits from dialect dispersion, both integrated in the same 'tree'. It does not indicate any substructure in extant linkages. Another recent approach which is not purely data display is *historical glottometry* [Kalyan et al., 2018, Kalyan and François, 2019]. This method emphasizes the overlapping pattern of shared innovations in a dialect area, and is explicit about the structure of the continuum.

While this presentation aims to outline the problems the dialect continuum poses in relation to the assumptions of the tree model, the main purpose is to introduce and define a new type of graph, the *Dialect Continuum Tree* (DCT). The DCT is built to represent language relationship through the history of small, more or less invariable dialects. These dialects diverge according to a strictly bifurcating tree pattern, but maintain contact with their mutually intelligible neighbours. When this contact eventually breaks down, one dialect continuum becomes two, effectively leading to a language split. These properties makes it so that, given a short amount of time between successive language splits, parts of a subgroup might share innovations with another subgroup that they do not share with all languages in their own. As noted by [Kalyan et al., 2018] this phenomenon is very similar to the concept of *incomplete lineage sorting* in population genetics. By introducing this type of dialectal substructure into the historical process of language diversification, the DCT can account for the wave like innovation patterns that come from dialect continua, while the resulting macro structure of language relationship is still a tree.

Keywords: phylogenetics, family tree, dialect continuum, wave model

References

- [Bryant and Moulton, 2004] Bryant, D. and Moulton, V. (2004). Neighbor-net: an agglomerative method for the construction of phylogenetic networks. *Molecular biology and evolution*, 21(2):255–265.
- [Hamed, 2005] Hamed, M. B. (2005). Neighbour-nets portray the chinese dialect continuum and the linguistic legacy of china's demic history. *Proceedings of the Royal Society B: Biological Sciences*, 272(1567):1015–1022.
- [Holden and Gray, 2006] Holden, C. J. and Gray, R. D. (2006). Rapid radiation, borrowing and dialect continua in the bantu languages. *Phylogenetic methods and the prehistory of languages*, 19:31.
- [Kalyan and François, 2019] Kalyan, S. and François, A. (2019). When the waves meet the trees: A response to jacques and list. *Journal of Historical Linguistics*, 9(1):168–177.
- [Kalyan et al., 2018] Kalyan, S., François, A., et al. (2018). 5. freeing the comparative method from the tree model: A framework for historical glottometry. *Senri Ethnological Studies*, 98:59–89.
- [Lehtinen et al., 2014] Lehtinen, J., Honkola, T., Korhonen, K., Syrjänen, K., Wahlberg, N., and Vesakoski, O. (2014). Behind family trees: secondary connections in uralic language networks. *Language Dynamics and Change*, 4(2):189–221.
- [Ross, 1988] Ross, M. (1988). *Proto Oceanic and the Austronesian languages of western Melanesia*. Dept. of Linguistics, Research School of Pacific Studies, The Australian National University.

*Department of Linguistics and Philology, Uppsala University

[Schmidt, 1872] Schmidt, J. (1872). *Die verwantschaftsverhältnisse der indogermanischen sprachen*. H. Böhlau.