The position and internal connections of Finno-Saami:
A computational perspective

Computational methods of language phylogeny, originating in the study of biological evolution, have seen ever wider use in historical linguistics in recent years (for Uralic languages, see Honkola et al. 2013; Syrjänen et al. 2013). This has been motivated by the will to develop a quantifiable way of analyzing historical-linguistic data with more comprehensive and flexible models than earlier quantitative methods in historical linguistics, e.g. glottochronology. Here it is shown how a more diverse use of different computational methods and linguistic data types can make our picture of the diversification of Finnic and Saami languages more accurate.

The phylogenetic approach does not aim to replace the traditional historical-comparative method. Instead, it depends on traditional research while providing a straightforwardly data-driven and quantifiable way of analysing linguistic material and assessing the results of the analyses. This dependence is true especially of analyses on lexical, phonological and morphological innovations, in which prior in-depth linguistic research is central.

An approach that is alternative to using linguistic innovations determined in traditional research is to model the typological variation between languages. This has been performed e.g. in Melanesian languages (Dunn et al. 2008) as well as those of the wider Sahul region (Reesink et al. 2009) in whose largely unknown genealogical relations are claimed in these studies to be recoverable with patterns in large structural datasets. However, in long-standing linguistic areas large amounts of structural influence can serve to bring the typological profiles of languages closer than their genealogical relatedness would entail (see e.g. Thomason 2001: 125–126).

This presentation brings new insights into the study of Uralic language history, concentrating on the connections of the Finnic and Saami subgroups. Results obtained with both phylogenetic tree (Bayesian MCMC analysis) and network (NeighborNet) methods are compared with analyses done with population clustering methods such as Structure. Using methods designed for the study of the development of both between-species (macroevolutionary; Bayesian MCMC and NeighborNet) and within-species variation (microevolutionary; Structure), it is determined which of these methods can best model the different stages in the diversification of these languages.

The linguistic macroevolutionary results from vocabulary data show the robustness of the layer of Proto-Finno-Saami lexical innovations, as well as later divergent and convergent patterns between

the languages. For examining recent dialect continua in both the Finnic and Saami subgroups, the Bayesian tree inference algorithms are seen to be non-optimal. With networks, factors in recent diversification within closely related languages are pointed out. Different kinds of lexicon, analysed in distance networks, show the effect of wave-like divergence and language contact.

Finally, it is briefly explored how typological, structural data could be used in diversifying the picture of Uralic language connections. Computational analysis of typological data is argued to be useful in inferring factors of areal developments and language contact both between closely and more distantly related languages.

References

Dunn, M., Levinson, S. C., Lindström, E., Reesink, G. & Terrill, A. (2008). Structural phylogeny in historical linguistics: Methodological explorations applied in Island Melanesia. *Language* 84, 710–59.

Honkola, T., Vesakoski, O., Korhonen, K., Lehtinen, J., Syrjänen, K. & Wahlberg, N. (2013). Cultural and climatic changes shape the evolutionary history of the Uralic languages. *Journal of Evolutionary Biology* 26, 1245–53.

Reesink, G., Singer, R. & Dunn, M. (2009). Explaining the linguistic diversity of Sahul using population models. *PLoS Biology* 7, e1000241.

Syrjänen, K., Honkola, T., Korhonen, K., Lehtinen, J., Vesakoski, O. & Wahlberg, N. (2013). Shedding more light on language classification using basic vocabularies and phylogeneti methods: A case study of Uralic. *Diachronica* 30, 323–52.

Thomason, S. G. (2001). *Language contact*. Edinburgh University Press.