

## Reconstructing prehistoric sociolinguistics from modern grammatical evidence

Johanna Nichols, UC Berkeley

Though a good deal is known to prehistorians about early centers of population growth and dispersal, and linguistics can identify some grammatical symptoms of sociolinguistic dominance and language shift, four problems remain unsolved. (1) Typology can now identify favored targets of selection in sociolinguistically asymmetrical language contact (e.g. canonical typology, Trudgill 2011), but these have not been applied to determining which language families descend from backwater refugees and which from expanding and sociolinguistically dominant ones. (2) Nor have they been applied to diachronic studies of head-marking, especially polysynthetic, languages with templatic morphosyntax. (3) Where past expansions can be identified, it is usually not known whether that involved spread into (near-)uninhabited land, dominance and absorption of a prior population, or sidelining a previous population with minimal substratal effects. (4) The effects of dense vs. sparse networks and short vs. long connections can now be modeled (Fagyal et al. 2010), but it is also known that very long travels in pre-Neolithic societies were routine (Graeber & Wengrow 2021:173); should ancient mobile groups (and selection in them) be modeled as nodes in sparse populations? as separate small populations (Bickel 2022)? as individuals in large, densely connected populations?

This paper uses case studies of four known or likely centers of expansion to propose answers by pushing back the temporal reach of sociolinguistic reconstruction. Additional theoretical considerations are the patch-and-pump model of first and early settlements (Author in press); staging areas and cost-path modeling (Anderson & Gillam 2001, Anderson et al. 2013) identifying centers and trajectories of spread; relational complexity (Author in press) to identify targets of selection in polysynthetic languages; self-similarity at different levels as an effect of selection (Nichols 2018); improved typological descriptions of features subject to selective pressure (e.g. Authors 2022); and isolation-by-distance modeling to identify centers and peripheries of spreads (cf. Grünthal et al. 2022).

Six variables are traced here across four case studies: (1) The early Columbia Plateau, for which the set of "Penutian" families is shown to be a likely early frontier population preserving a Eurasian-like typology as subsequent immigrants brought or developed a very different typology. (2) The later Columbia Plateau, where postglacial desiccation triggered the various "Penutian" spreads south and west, argued here to have begun in spreads along existing networks with minimal contact effects; (3) the Lower Mississippi Valley, a long-standing staging area (Kaufman 2014). (4) The Altai region (upper Irtysh and Yenisei, Minusinsk Basin, northern Kazakhstan and Mongolia), from which Pre-Uralic, Pre-Turkic, and Pre-Mongolic may have dispersed; the very self-similar Ural-Altai typology is barely emergent in reconstructed ProtoUralic (c. 4500 BP) and highly developed in Proto-Turkic (c. 2000 BP) and Proto-Mongolic (c.1000 BP). This evolution points to long-term selection without sociolinguistic dominance. Variables: Harmonic pronoun consonantism; self-similar morphology and syntax (morpheme/word order, head/dependent marking); fixed base lexical valence; high/low causativization (base intransitivity); consistency in finiteness across different clause types; configurational/templatic. All are relatively stable in families, and high frequencies of either polar value (e.g. head-final vs. head-initial) are favored in selection.

## **Keywords**

Centers and peripheries, language spreads, linguistic networks, linguistic selection, sociolinguistic typology.

## **References**

- Anderson, David G., Thaddeus G. Bissett, and Stephen J. Yerka. 2013. The Late-Pleistocene human settlement of interior North America: The role of physiography and sea-level change. *Paleoamerican Odyssey*, ed. K. E. Graf, C.V. Ketron, and M. R. Waters, 183–203. College Station, TX: Center for the Study of the First Americans, Texas A & M University.
- Graeber, David G. and J. Christopher Gillam. 2001. Paleoindian interaction and mating networks: Reply to Moore and Moseley. *American Antiquity* 66:3.530–535.
- Bickel, Balthasar. 2022. Large and ancient linguistic areas. *Language dispersal, diversification, and contact*, ed. Mily Crevels and Pieter Muysken, 78–99. Oxford: Oxford University Press.
- Fagyal, Zsuzsanna, Samarth Swarup, Anna María Escobar, Les Gasser, and Kiran Lakkaraju. 2010. Centers and peripheries: Network role in language change. *Lingua* 120:8.2061–2079.
- Graeber, David and David Wengrow. 2021. *The dawn of everything: A new history of humanity*. New York: Farrar, Straus, and Giroux.
- Grünthal, Riho, Volker Heyd, Sampsa Holopainen, Juha Janhunen, Olesya Khanina, Matti Miestamo, Johanna Nichols, Janne Saarikivi, and Kaius Sinnemäki. 2022. Drastic demographic events triggered the Uralic spread. *Diachronica* 39:4.490–524.
- Kaufman, David V. 2014. The Lower Mississippi Valley as a language area. PhD dissertation, University of Kansas.
- Nichols, Johanna. 2018. Non-linguistic conditions for causativization as a linguistic attractor. *Frontiers in Psychology* 8.2356.
- Trudgill, Peter. 2011. *Sociolinguistic Typology: Social determinants of linguistic structure and complexity*. Oxford: Oxford University Press.