Transitivizing/detransitivizing typology and language family history

Johanna Nichols, University of California, Berkeley johanna@berkeley.edu
Riho Grünthal, University of Helsinki rgruntha@mappi.helsinki.fi

The transitivizing/detransitivizing typology of Nichols, Peterson, Barnes 2004 (NPB) also proves useful to historical linguistics. We focus on families of northern Eurasia, chiefly the three oldest families (Indo-European [IE], Uralic, Nakh-Daghestanian), daughter branches aged about 2000-3000 years, and one younger family for which we have data on enough daughters to support a family phylogeny (Tungusic). We use neighbor net (SplitsTree) because it can reveal not only phylogeny but also typological and geographical affinities. The NPB 36-item verb list reveals either phylogeny or historical geography, under identifiable conditions of datasets and coding.

Whole-pair gross typological classification (each pair classified simply as causativizing, decausativizing, etc.) accurately reveals family-wide geographical trends but is less good on phylogeny. E.g. Slavic: a major north-south split and cline revealing the spread of reflexivization as detransitivizing device from south to north. Uralic: the young [[Saami-Finnic]-Mordvin] branch is evident, but more evident is a division into steppe-influenced vs. northern languages; when an outgroup is added it attaches to a typological, not genealogical, outlier. The IE tree is strongly geographical or typological, an elongate shape with detransitivization (reflexivization; an innovation) at one end and transitivization (causative or factitive derivation; forms inherited, distribution innovative) at the other. Slavic is a clear clade (though with incorrect internal structure); both Germanic and Indo-Iranian correctly cluster together, but each only in a branch that also incudes an extraneous language. Tungusic: the expected division of northern (Even, Evenki) vs. southern (Manchu, Nanai, Udehe) is very clear. Nakh-Daghestanian: the major branches are clear (their geography is entirely consistent with subgrouping), as is the pure geography of north vs. south Caucasus.

Considering just the 9 pairs with prototypically animate S/O yields much the same groupings as the entire 18 pairs, while the 9 pairs with prototypically inanimate S/O behave differently, supporting the NPB claim that the inanimate set are lower-frequency, less stable verbs reflecting a mix of universals and quirks.

Coding not for gross whole-pair structural type but for three different processes (derivation; stem alternation; extensions, etc.) per form brings trees into closer approximation to actual phylogeny while also producing a superior, more informative typology with finer discrimination of closely related languages and stronger cross-categorial correlations. Importantly, this coding greatly improves potential integration with other databases and with word-and-paradigm morphological models, and we recommend that it replace the gross NPB type.

Adding information about cognate status of the root morpheme of each pair member yields a very good family tree, in line with known results of comparative-historical work, for all families and branches tested. (Uralic and Tungusic analysis are just beginning but they appear to improve phylogeny in the same way.)

Thus very small closed datasets, collected originally for typology, yield rich information about language family history — strikingly, a mere 18 items (9 pairs), coded for morphological type and cognacy, yield a very good genealogical tree — while historical methods have also improved the typology.

Nichols, Johanna, David A. Peterson and Jonathan Barnes. 2004. Transitivizing and detransitivizing languages. *Linguistic Typology* 8:2: 149-211.

498 words