CORPORA, VECTORS, MODELS: COMPUTATIONAL APPROACHES TO LINGUISTIC TYPOLOGY - ABSTRACT

Aleksandrs Berdicevskis

University of Gothenburg, Gothenburg, Sweden

Abstract

In this paper, I will make several strong claims about language change and typology. The claims should be perceived not as reliably established truths, but rather as my views of how we can productively approach certain research questions. The claims are as follows:

1. When studying language change, we should try to explain it, but be aware that it may not always be possible (see Claims 2 and 3). Explanations can involve both language-internal and language-external (such as social, cultural, cognitive or even geographical) factors.

2. Sometimes the change cannot be explained since there is no «explanation» as such, just a complex interaction of random processes. Even in these cases, however, it can sometimes be established whether certain kinds of change are more likely than others.

3. Sometimes the change cannot explained simply because there is not enough data. Explanation typically involves causal claims, and establishing causal links is difficult. Even if change A in language X occurs after event B, this fact per se is usually not enough to claim that B causes A. This fact, however, can be enough to hypothesize about a causal link, and the hypothesis should considered true if:

a) it is supported by nomothetic (large-scale) quantitative cross-linguistic studies;

b) the causal mechanism (how B can cause A) is convincingly described, and it has been shown that this mechanism really works (for instance, via psycholinguistic experiments);

c) there is converging evidence from various other sources (corpus studies; diachronic studies; computational modelling etc).

4. The distribution of properties across the world's languages is partly shaped by weak universal pressures on language change generated by extra-linguistic factors (cf. Claims 1 and 2). Thus, understanding language change is the key to answering many questions traditionally viewed as belonging to language typology (and vice versa). To an extent, sociolinguistics, diachronic linguistics and typology should be viewed as facets of the same research direction, sometimes labeled language evolution.

5. One productive way of approaching the evolutionary questions is using the opportunities of modern natural language processing: corpus linguistics and beyond.

I will illustrate how this research agenda can be pursued using three examples from my own work.