Why do we need loans? A comparative-contrastive study on Eurasian lexical borrowability

Languages borrow words when there is a need for it, all languages contain loanwords, and no part of the lexicon is entirely “loan-proof”. These are statements about lexical borrowing that are typically found in linguistic textbooks (Hock and Joseph 1996). Further, we know that there are large discrepancies in the borrowability of different lexical concepts, where core vocabulary domains (sense perception, spatial relations, the body, kinship, and motion) in general are more resistant to borrowing, whereas culture-dependent domains (religion and belief, clothing and grooming, the house, law, social and political relations, agriculture and vegetation) belong to a more loan-intense part of the lexicon. We are also aware that there are large differences in borrowability between languages, something that has multiple connotations, including language history, populations size, language contact, grammatical structure, and so forth (Haspelmath and Tadmor 2009).

Our study aims at investigating borrowability more carefully from a historical perspective, using quantitative and statistical methods, focusing on the families of Indo-European, Nakh-Dagestani, Northwest Caucasian, Kartvelian, Uralic, and Turkic. We use a lexical dataset, which contains 100 lexical meanings each from the domains of basic vocabulary (Swadesh) and culture vocabulary (domains of agriculture, vegetation, food, warfare, hunting, animals, technology), compiled from around 250 languages of the previously mentioned families (around 25,000 lexemes in total). The lexemes of the dataset have been coded (manually, from dictionaries) for cognacy according to a tree model and are distinguished by borrowability versus inheritance at every historical stage in the tree structure. The dataset also systematically includes ancient language forms (including reconstructed ones), and traces continued development of cognates, involving further semantic change. Connected to the dataset, there is language metadata that includes geographic extension, time period, relative population size and family tree topology. Taken together, this makes the dataset a unique and yet unexplored source for investigating large-scale borrowability statistically.

In our study, we are specifically aiming the following research questions:

- What is the general level of borrowability in our vocabulary (culture words), compared to the average borrowability of the same lexical meaning concluded by the cross-linguistic study of (Haspelmath and Tadmor 2009)?
- In our data, are there any internal differences between the borrowability of lexical concepts, depending on semantic domain?
- Is there a general connection between loanword directionality and population size of languages?
- Is the amount of borrowing generally equal over time (gradually increasing with increasing amounts of documentation), or are some periods and geographic areas more intense in borrowing?
- What happens to lexemes upon borrowing? Do they continue to change with the language or are they more likely to be frozen, semantically and/or morphologically?
