

Dynamics of differentiation of Russian dialects

Darja Jonjić, Matías Guzmán Naranjo & Søren Wichmann
(Kiel University, University of Freiburg & Kiel University)

Keywords: Russian, dialectometry, geography, population, gravity model

The Dialectological Atlas of the Russian Language (abbreviated DARJa based on its Russian title) represents more than 4 decades of data collection, from 1938 onwards, and was published in Moscow during 1986-2005. It contains 313 maps, each corresponding to a linguistic feature, and covers 4196 locations. In 2015-16, researchers at Kazan Federal University extracted linguistic features and their values directly from the physical maps and created Excel files giving the values of features across locations (Isaev et al. 2016). We have processed these materials further, georeferencing the map of locations covered and manually extracting the latitude and longitude location of every location. Thus, the DARJa data are now amenable to systematic, quantitative analyses.

Our initial research question concerned the extent to which administrative divisions pertaining to different epochs of Russian history fit and may help to account for dialectal divisions of the Russian language. For this, we interpolated all feature values using the Euclidean distance, performed hierarchical clustering on interpolated data and finally cut the classification tree into clusters. This quantitative analysis has revealed that the type of administrative unit that best fits the dialect clusters is the *ujezd*, which is the earliest unit, dating back to the 15th century. Now our interest lies in finding out which other factors play a role in the differentiation of Russian dialects.

Honkola et al.'s (2018) investigations of the diversification of Finnish dialects have shown that these dialects initially differentiate mainly conditioned by geographical factors and later seem to cluster and stabilize as a result of the influence of cultural factors, including variations in subsistence patterns. Nerbonne and Heeringa (2006: 22) also found that "geography influences linguistic variation massively", and, moreover, that population size correlates with linguistic distance. Contrary to the predictions of the gravity model of Trudgill (1974), their study revealed that the larger the population size, the greater the linguistic distance. The focus of this talk is to determine whether and to what extent there is an influence of geography and population size on the patterning of Russian dialects as represented in the DARJa data.

References

Honkola, Terhi, Kalle Ruokolainen, Kaj J. J. Syrjänen, Unni-Päivä Leino, Ilpo Tammi, Niklas Wahlberg & Outi Vesakoski (2018), Evolution within a Language: Environmental Differences Contribute to the Divergence of Dialect Groups, *BMC Evolutionary Biology* 18, 132.

Isaev, Igor', Valerii Solov'ev, Farid Salimov, Aleksandr Pilugin & Venera Bairacheva (2016), Creating a Database of Russian Dialects and Prospects for Dialectometric Studies, *Herald of the Russian Academy of Sciences* 86(6), 448-453.

Nerbonne, John & Wilbert Heeringa (2007), Geographic Distributions of Linguistic Variation Reflect Dynamics of Differentiation, in S. Featherston, and W. Sternefeld (eds), (2007), *Roots: Linguistics in Search of its Evidential Base*, Berlin/New York: Mouton de Gruyter, 267-297.

Trudgill, Peter (1974), Linguistic Change and Diffusion: Description and Explanation in Sociolinguistic Dialect Geography, *Language in Society* 3(2), 215-246.