

Modeling and measuring language change in the former Soviet Union

Kyle L. Marquardt*

November 14, 2019

Language has been a consistently important element of post-Soviet politics. Indeed, the tension between programs of linguistic revitalization in post-Soviet regions and speakers of other languages (especially Russian speakers) has been at the forefront of vital research on post-Soviet identity and politics (e.g. Laitin 1998). Understanding patterns of language change is therefore of clear importance to the study of post-Soviet societies and politics.

However, cross-regionally and cross-temporally consistent data on language change in the former Soviet Union do not exist. While Soviet censuses generally included questions regarding self-identification with relevant languages (e.g. the language a citizen considers her “native” language and a second Soviet language which she spoke fluently), only a few post-Soviet censuses continued with these formulations. Instead, some censuses only asked about native language, while others introduced their own idiosyncratic measures of linguistic capabilities; one country (Uzbekistan) has yet to conduct an official census. As a result, assessing the degree to which linguistic demographics have changed in post-Soviet states and ethnofederal regions is difficult: even in those countries that have had multiple post-Soviet censuses, alterations in the format of language questions (or the absence of such questions) makes comparing change in linguistic demographics over time and space in the former Soviet Union a highly tentative endeavor.

In this paper I integrate insights from the fields of linguistics, sociology and political science to develop a statistical model which provides cross-regionally and cross-temporally consistent measures of the titular language proficiency across ethnofederal regions of the former Soviet Union. The model combines a structural model of linguistic change with Bayesian factor analysis, allowing for the iterative estimation of the relationship between different measures of linguistic demographics and the underlying concept (titular language fluency). The model also parameterizes language status, demographic changes and regime changes as factors that could influence linguistic trajectories.

The new measures of titular language proficiency will allow researchers to quantitatively analyze outcomes such as language policy success, as well as the relationship between linguistic demographics and outcomes of interest such as democratization and civil conflict. The model’s language status parameters are also themselves of substantive interest, providing insight into the political and cultural factors that influence linguistic trajectories.

*School of Politics and Governance and International Institute for the Study of Institutions and Development, National Research University Higher School of Economics; kmarquardt@hse.ru