ILJA A. SERŽANT

Johannes-Gutenberg-University of Mainz
ilja.serzants@uni-mainz.de

ILYA KHAIT

University of Jena
ekh.itd@gmail.com

Analyzing Latgalian vocalism: A usage-based statistical approach

This is a work-in-progress paper. Our aim is to compare possible vowel combinations on the level of wordforms. We draw our data from the TriMCo-dialectal corpus (https://www.trimco.uni-mainz.de/trimco-dialectal-corpus/) that encompases oral speech records from several subdialects of Latgalian. A special software developed for this study extracts all wordforms found in the TriMCo-corpus and analyzes it according to special criteria such as the quality of the stressed vowel, the palatalization/non-palatalization of the immediately preceding consonant, etc., excludes dublets, etc. In the next step, the extracted data are subjected to statistical analysis. We pursue two goals: (i) our aim is to identify factors that determine the phonological patterns of the variety of Latgalian documented in the corpus and (ii) try to describe the variation in the phonological systems of the Latgalian subdialects of the corpus (south-east Latgale).

The advantage of the method as opposed to methods applied in the traditional dialectology is that we provide for an „all-inclusive-approach" which means that we are able to accommodate variation within one subdialect and even within one speaker's speech. In contrast to the traditional dialectology, we do not invoke or imply such concepts as „ideal subdialect" or its „ideal speaker" that attest a particular isogloss with no exceptions (e.g. used to create the dialectological maps). It is well-known that dialectal variation represented by the dialectological maps is just a first approximation necessary to organize the empirical data. However, on a more fine-grained level, as has been often emphasized by dialectologists, there are a lot of „exceptions" or „mistakes" that have to be ignored on the large-scale methods of the traditional dialectology. Our aim is to overcome this problem and represent the subdialects „as they are" with all their exceptions and mistakes.

The authors of this paper have carried out different parts of the research:

Ilja Seržant: the concept of the paper, research and statistical analysis;

Ilya Khait: programming to provide for automatic parsing and annotation of all word forms into syllables, vocalizations, provenience, speaker, token frequency, etc.