The impact of phonetic narrowness on the dialectometry of Hungarian

FRUZSINA SÁRA VARGHA
Eötvös Loránd University Budapest

The aim of this paper is to illustrate how the digitized Hungarian dialect data and a dedicated data management tool can be used for dialectometric studies, with particular regard to the effects of the quantity of phonetic detail contained in the transcribed data.

In the last fifteen years more than one million data items from The Atlas of Hungarian Dialects and several regional Hungarian linguistic atlases have been appropriately digitized in a series of interrelated, computational dialectology projects. The software tool, developed according to the evolution of the research needs, fully supports the Hungarian standard transcription system (a special narrow phonetic transcription). It is also possible to automatically convert the data in a broader transcription, with fewer or no diacritics (and other simplifications as required). As every piece of dialect data is geographically localized, we can generate integrated interactive maps with data coming from different atlases. The digitized atlases constitute a large corpus suitable to quantitative analyses.

The research presented here, as part of a larger project financed by the Hungarian Scientific Research Fund, compares different similarity matrices generated by an automatic string edit distance analysis (based on the Levenshtein algorithm) from data in narrow vs. broader transcription. In order to validate the method, the comparison is extended to a matrix based on a manual classification of data items in the same regional atlas.

Preliminary results suggest that the matrices generated from narrower vs. broader transcripts exhibit appreciable differences especially in the case of dialect enclaves and speaker communities that are distant in time or space from the originally related dialect area. Broader transcripts tend to accentuate the effects of linguistic phenomena less resistant to change due to recent influences from neighboring dialects or languages.