

A sociolinguistic analysis of linguistically sensitive dialectal word pronunciation distances

Martijn Wieling, University of Groningen, the Netherlands

Abstract

This presentation will focus on two important findings of my dissertation (Wieling, 2012). In the first part of the presentation I will show that linguistically sensible phonetic distances can be derived from atlases of alternative dialectal pronunciations (Wieling et al., 2012). The procedure uses alignments of the Levenshtein distance algorithm (Levenshtein, 1965), and applies an information-theoretic measure, pointwise mutual information (Church and Hanks, 1990), assigning smaller segment distances to segments which frequently correspond. I will illustrate that the obtained phonetic distances are linguistically sensible for six independent dialect data sets and can be used to yield more sensitive dialectal word pronunciation distances.

In the second part of the presentation I will use a novel approach to show that these sensitive dialectal word pronunciation distances (from the standard language) are determined by geography and several social and lexical factors (Wieling et al., 2011). The non-linear influence of geography is represented by a generalized additive model, while the word- and speaker-related variability is accounted for by using a mixed-effects regression analysis. This is a promising approach as it combines the merits of both dialectometry (i.e. a focus on geography and taking a large number of linguistic items into account simultaneously in order to obtain an objective view of dialect variation; Séguy, 1971) and dialectology (a focus on individual linguistic items and the inclusion of social factors; Chambers and Trudgill, 1998).

References

Church, Kenneth and Patrick Hanks (1990). Word association norms, mutual information, and lexicography. *Computational Linguistics*, 16(1):22–29.

Chambers, Jack and Peter Trudgill (1998). *Dialectology*. Cambridge University Press, Cambridge, 2nd edition.

Séguy, Jean (1971). La relation entre la distance spatiale et la distance lexicale. *Revue de Linguistique Romane*, 35(138):335–357.

Levenshtein, Vladimir (1965). Binary codes capable of correcting deletions, insertions and reversals. *Doklady Akademii Nauk SSSR*, 163:845–848. In Russian.

Martijn Wieling (2012). *A Quantitative Approach to Social and Geographical Dialect Variation*. PhD thesis, Rijksuniversiteit Groningen. Available at <http://www.martijnwieling.nl>.

Martijn Wieling, Eliza Margaretha and John Nerbonne (2012). Inducing a measure of phonetic similarity from pronunciation variation. *Journal of Phonetics*, 40(2), 307-314.

Martijn Wieling, John Nerbonne and R. Harald Baayen (2011). Quantitative social dialectology: Explaining linguistic variation geographically and socially. *PLoS ONE*, 6(9): e23613. doi:10.1371/journal.pone.0023613.