John Nerbonne en Wilbert Heeringa

Computationele vergelijking en classificatie van dialecten

Summary

In traditional dialectology, maps are divided into dialect areas on the basis of isoglosses or with the use of the arrow method (Daan & Blok 1969). However, the choice of isoglosses is subjective. A second shortcoming for both methods is that there is no way to characterize relations between entire varieties. Comparison is inevitable atomistic. Third, existing methods records differences in varieties without distinguishing degrees of differences. The view that some differences vary along a continuum has no analytical foundation. This paper focuses on the Levenshtein method. Using this method, pronunciation differences can be measured for corresponding pronunciations in two varieties. We can apply the method to a large sample, providing an objective foundation for further analysis. The differences can be added, which allows one to relate entire varieties, aggregating the atomic differences. After comparing dialects on the basis of their distances the dialects can be classified by clustering or multidimensional scaling. Using clustering we get a sharp classification in the form of a tree, where the dialects are the leaves. Using multidimensional scaling we get a plot on which like dialects are plotted nearby and unlike dialects are distant. When scaling to three dimensions, a map can be colored, where the dimensions represent respectively the intensity of red, green and blue, while the areas between the dialects are colored by interpolating from the dimensions of the dialects. In that way, dialect variation is visualized as a continuum.