Exploring global and local patterns of the correlation of geographic distances with morpho-syntactic variation in Swiss German dialects

PÉTER JESZENSZKY
PHILIPP STOECKLE
ROBERT WEIBEL
University of Zurich

This project aims at the investigation of linguistic variational difference between dialects using methods familiar in geographic information science (GIScience). The SADS dataset (Syntactic Atlas of Swiss German Dialects) is used in this study (Bucheli & Glaser, 2002). It contains data for 383 survey sites in the German-speaking part of Switzerland, with about 25% of the municipalities surveyed. The study presented here is aimed at establishing the degree of correlation between dialectal variation, expressed by a linguistic distance measure, and geographic distance. A linguistic distance matrix is built from binary differences between survey sites (Nerbonne & Kleiweg, 2007; Spruit, 2008; Scherrer, 2012). The linguistic distances are then compared to a travel time matrix (Axhausen et al., 2006) and the crow-fly (Euclidean) distance matrix, respectively, and correlation is computed to test the hypothesis that geographically proximate dialects tend to be more similar than distant ones and linguistic distance correlates more with travel distance than with Euclidean distance (Gooskens 2005).

Apart from global patterns we also investigated local patterns correlation. If distances are normalized, differences between linguistic and geographic distance values per site can be analysed and visualized, showing where in geographic space the correlation between the distances is higher or lower than expected, respectively. Thus, breaks in correlation can also be compared to geographical borders (political, cultural, historical; Hotzenköcherle, 1986). The full survey (118 questions about 50 syntactic phenomena) is taken into consideration for the linguistic distance while travel times for various years from 1850 to 2014 by car *and* public transport are used, enabling to investigate spatial and temporal trends of variation. Because SADS used multiple informants per site (median 5-6), we compare whether calculating correlations when only an ”ideal speaker” is considered gives us different trends to the scenario taking into account all informants.