

AUTOMATICALLY DETECTING BOUNDARIES AND TRANSITION ZONES IN SWISS GERMAN MORPHO-SYNTACTIC VARIATION

Péter Jeszenszky

University of Zurich, Department of Geography

Curdin Derungs

University of Zurich, Department of Geography

Robert Weibel

University of Zurich, Department of Geography

It has long been proposed (Trudgill, 1974) that in order to quantify the impact of extralinguistic factors on dialect variation in a realistic way, the distribution of individual dialectal variants has to be modelled quantitatively along with the patterns in which the usage of these variants transitions into each other in space.

This project aims at the quantitative modelling of such transitions using methods familiar in geographic information science. The SADS dataset (Syntactic Atlas of German-speaking Switzerland), used in this study, contains data for 383 survey sites (~25% of the municipalities) in the German-speaking part of Switzerland (Bucheli & Glaser, 2002). Multiple respondents per survey site allow for investigating the variation in individual phenomena with an elevated spatial granularity. A methodology to automatically detect crisp boundaries and transition zones between variants is established.

Such transitions can be abrupt (characterised by crisp isoglosses) or more gradual (Seiler, 2005), but their spatial patterns are rarely regular. Transition zones are often described and used in dialectology (Scherrer, 2012; Pickl, 2013; Scholz et al., 2016), but the concept lacks a clear definition. We account for this by, first, deriving transition zones in a bottom-up, data-driven way and, second, modelling these zones as fuzzy objects. Thus, two processing steps are iterated – over multiple threshold values. First, a given threshold value is applied to variant usage proportions at all survey sites in order to select candidate transition zone locations (*cl*) - i.e. sites with no clear dominant variant. Second, *cl* are fed to a modified DBSCAN algorithm (Ester et al., 1996) in order to distinguish *cl* that geometrically resemble transition zones from those considered geometric outliers. Finally, *cl* resulting from iterations over multiple threshold values are aggregated, to constitute what we above called fuzzy transition zones.

Results are also validated using multiple spatial statistic measurements. Additionally, boundaries found are compared to physical and social borders (Hotzenköcherle, 1986) in a quantitative way, providing empirical evidence for longstanding (dialect contact) claims in geolinguistics about Swiss German syntax.