Bottom-up dialectometry using the *GeoLing* package

SIMON PRÖLL
Univerisity of Augsburg

SIMON PICKL
University of Salzburg

AARON SPETTL
Ulm University

STEPHAN ELSPAß
University of Salzburg

WERNER KÖNIG
Univerisity of Augsburg

VOLKER SCHMIDT
Ulm University

In this talk, we will present new methods for examining aspects of dialect data that have (for the most part) been disregarded in previous quantitative research: the gradient between idiosyncratic and recurring patterns in individual variants' distributions and the complex relations between these distributions and whole dialects.

To this end, we introduce *GeoLing*, a novel tool for analyzing large sets of dialect data. In contrast to other dialectometrical software, *GeoLing* allows the relation between individual variants and whole varieties to be kept transparent throughout the analyses - the focus lies on identifying recurring patterns in variants' distributions and detecting latent structures even below the threshold of overall, large-scale patterns in the aggregate.

This is achieved by adopting techniques from geostatistics, image analysis and data mining: Density estimation is used to estimate continuous fields of variants' usage probabilities from geographically scattered dialect data; cluster analysis is applied to obtain groups of variables that share spatial similarities; and with factor analysis, similar distributions of variants can be condensed to fuzzy co-occurrence areas of varying size and strength. Thus, the deeper structure of spatial variation becomes accessible from different angles.

In this talk we will showcase the software package in action, using the example of dialect data from Southern Germany. Both software and results are original work from a third-party funded research project concerned with the development of novel methods and tools. This paper will be the first presentation of the overall outcome of this project.