Bern, 01.04.2021

UNIVERSITÄT
BERN

Bern Data Science Day 2021 - April 23

Contribution - Reduction of survey sites in dialectology: a new methodology based on clustering

Digital Humanities and Economics

Péter Jeszenszky, Carina Steiner and Adrian Leemann

Center for the Study of Language and Society, Faculty of Humanities, University of Bern

peter.jeszenszky@csls.unibe.ch

Abstract

Many language change studies aim for a partial revisitation, i.e., selecting a subset of survey sites from previous dialect studies. This survey site reduction has often been addressed only qualitatively in linguistics. Cluster analysis, however, offers an innovative means of identifying the most representative survey sites among a set of original survey sites. We present a general methodology for finding representative sites for an intended study, potentially applicable to any digital collection of data about dialects or linguistic variation. The methodology consists of the quantitative steps of finding clusters of original survey sites based on an association measure (linguistic distance), appointing a central site in these clusters, and the qualitative step of revising this set of candidate survey sites considering sociodemographic and linguistic changes that potentially occurred since the original data was recorded. We demonstrate the quantitative steps of the proposed methodology in the context of the 'Linguistic Atlas of Japan' (LAJ). Further, we present the full application of the methodology on the 'Linguistic Atlas of German-speaking Switzerland' (SDS), with the explicit aim of selecting survey sites corresponding to the aims of the current project 'Swiss German Dialects Across Time and Space' (SDATS), which revisits SDS 70 years later. We find that the proposed methodology allows for a greater objectivity in comparison to traditional approaches, depending on the circumstances and requirements of a study. We suggest, however, that the suitability of any set of candidate survey sites resulting from the proposed methodology should be rigorously revised by experts due to potential incongruences such as the overlap of objectives and variables across the original and intended studies, and ongoing dialect change.