

Lexical variation in American English: A web-based dialect survey

JACK GRIEVE
Aston University

COSTANZA ASNAGHI
Università Cattolica del Sacro Cuore & KU Leuven

RUETTE TOM
Humboldt-Universität zu Berlin & KU Leuven

This presentation describes the results of a survey of lexical variation in modern American English. This survey is based on a unique approach to data collection, which involves searching for the variants of lexical alternation variables in hundreds of newspaper websites from across a region of interest using site-restricted web searches. The hit counts returned by these web searches are then used to estimate the proportion of each variant in each newspaper, which are mapped across the cities in which these newspapers are published and subjected to statistical analysis to identify patterns of regional variation. The method is first introduced and validated by mapping lexical alternation variables with known distributions in American English (e.g. bag/sac, frosting/icing, casket/coffin). The method is then used to map hundreds of lexical alternation variables in American English, many of which have never been mapped before (e.g. cemetery/graveyard, sofa/couch, cinema/movie theater). Based on a spatial and multivariate statistical analysis of this dataset, individual and common patterns of regional lexical variation are identified in American English, as well as modern American dialect regions. In particular, this analysis identifies four major American dialect regions, which correspond to the four major modern cultural regions of the United States: the Northeast, the Southeast, the Midwest, and the West. The identification of Northeastern and Midwestern dialect regions in particular contrasts with the results of previous American dialect surveys, which have generally identified Northern and Midland dialect regions in their place. It is argued that these results show that American dialect regions are currently undergoing change, no longer reflecting historical settlement patterns, but rather reflecting modern American cultural regions. Linguistic explanations for these patterns are also proposed, based on the sets of lexical alternation variables that are identified as exhibiting similar patterns of regional variation.