Benedikt Szmrecsanyi (Freiburg, Germany)

Corpus-based dialectometry: the grammar of traditional British English dialects

Abstract

Dialectometry, a branch of geolinguistics founded by Jean Séguy (1971) and refined by Hans Goebl (e.g. 1982) and John Nerbonne (e.g. 2006), is concerned with aggregate dialect similarities where the crucial explanatory factor is geographical distance. By contrast to dialectology, dialectometrical analysis focuses not on individual dialect features but on aggregate distances or similarities between dialects, drawing on a large number of dialect features and utilizing a variety of cartographic visualization techniques to project aggregate linguistic distances and similarities to geography. There is also a strong emphasis on exploratory data analysis to infer patterns from feature aggregates.

The research reported in this paper departs from most previous work in dialectometry in several ways. Empirically, it draws on frequency vectors derived from naturalistic corpus data (specifically, from the Freiburg Corpus of English Dialects [FRED]; cf. Szmrecsanyi & Hernández 2007) and not on discrete atlas classifications, as is customary in orthodox dialectometry. Substantially, it is concerned with morphosyntactic and grammatical (as opposed to lexical or pronunciational) variability. Methodologically, it marries the careful philological analysis of dialect phenomena in authentic, naturalistic texts to aggregational-dialectometrical techniques.

The aggregate analysis is based on frequency information concerning 57 typically non-standard grammatical phenomena (for instance, multiple negation or non-standard weak verb forms), drawn from eleven major domains of English morphosyntax (for instance, negation and verb morphology). Three questions guide the paper: First, on methodological grounds, what are the major methodical challenges in corpus-based dialectometry? Second, to what extent is grammatical variation in non-standard British dialects patterned geographically? Third, can a data-driven line of analysis like the one presented in this paper objectively establish dialect areas?

References

Goebl, H. (1982). Dialektometrie: Prinzipien und Methoden des Einsatzes der Numerischen Taxonomie im Bereich der Dialektgeographie. Wien: Österreichische Akademie der Wissenschaften.

Nerbonne, J. (2006). Identifying linguistic structure in aggregate comparison. Literary and Linguistic Computing 21(4), 463–475.

Séguy, J. (1971). La relation entre la distance spatiale et la distance lexicale. Revue de Linguistique Romane 35, 335–357.

Szmrecsanyi, B. and N. Hernández (2007). Manual of Information to accompany the Freiburg Corpus of English Dialects Sampler ("FRED-S"). http://www.freidok.uni-freiburg.de/volltexte/2859/. Freiburg: English Dialects Research Group.