

Benedikt Szmrecsanyi. *Grammatical variation in British English dialects.* Cambridge: Cambridge University Press. 2013. 211 pp. ISBN 978-1-107-00345-3. Reviewed by **James A. Walker**, York University, Toronto.

Methods in dialectology have come a long way since the first large-scale dialect studies conducted in the late 19th century: thanks to mechanical (and now digital) recording techniques and exponential increases in data storage and processing, the quality and amount of data available to 21st-century dialectologists would have been unimaginable even 50 years ago. Yet studies in dialectology still tend to use analytical methods developed in the 19th century, focusing largely on lexical and phonetic features obtained through elicitation or questionnaires and relying on impressionistic interpretation of dialect maps to infer regional patterns of variation. In this book, Szmrecsanyi pushes the analysis of British English dialects in new methodological and analytical directions, taking on the less salient but more problematic question of grammatical variation between dialects, as well as expanding the statistical and cartographic toolbox of dialectology.

Chapter 1 ‘Introduction’ (pp. 1–14) reviews previous ‘big-picture’ accounts of British dialects and introduces the rationale, method and objectives of the book. While the Survey of English Dialects (SED; Orton and Dieth 1962–1971) is generally considered to be the first large-scale British dialect study, Ellis’s (1889) survey of differences in accent serves as a 19th-century precursor. Large-scale studies of British dialects have made use of the SED materials (e.g. Trudgill 1990; Shackleton 2007) as well as adopting the perspectives of social-network theory (McMahon *et al.* 2007) and perceptual dialectology (Inoue 1996). In contrast, the study reported in this book uses the framework of Corpus-Based Dialectometry (CBDM), which differs from traditional dialectology in using corpus data rather than elicitation and in aggregating features rather than analyzing features on an individual basis.

The corpora on which the study is based are detailed in Chapter 2, ‘Data and methods’ (pp. 15–32), which also outlines the steps involved in conducting CBDM. Data were drawn from the Freiburg Corpus of English Dialects

(FRED), oral history interviews recorded in the 1970s and 1980s with 427 elderly working-class people in 158 locales throughout Britain, as well as from two ‘reference’ corpora of standard British and American English, based on subsamples of the International Corpus of English (ICE-GB) and the Santa Barbara Corpus, respectively. Conducting CBDM involves first establishing a ‘feature catalogue’, the set of linguistic features that serve as dependent variables in the analysis. Occurrences of these features are then extracted from the corpus and used to construct a $n \times m$ frequency matrix (where n = the set of dialects sampled and m = the feature catalogue). Distances between dialects are then calculated by taking the Euclidean difference between each pair of dialects (i.e. the square root of the sum of the squared differences between the frequency values for each feature), which yields a correlation matrix of distances that can then be compared to different cartographic measurements. Chapter 3 ‘The feature catalogue’ (pp. 32–70) provides more detail on the selection and extraction of the 57 grammatical features that constitute the feature catalogue for this study. While acknowledging a certain degree of subjectivity in the selection of features (p. 35), Szmrecsanyi’s aim was to include not only features that are normally considered regional or dialectal but also those that occur across all dialects. Given the quantitative nature of the analysis, a feature’s inclusion in the catalogue was also determined by a minimum frequency in FRED and the ability to extract occurrences using automatic or semi-automatic procedures. The remainder of this chapter is devoted to a detailed discussion of the nominal, verbal and clause-/discourse-level features included in the catalogue.

Chapter 4 ‘Surveying the forest: On aggregate morphosyntactic distances and similarities’ (pp. 71–88) provides a first look at the aggregated data and introduces a number of different types of cartographic representation: network maps to illustrate similarities and dissimilarities among dialects, and beam maps and honeycomb maps to represent interpoint relationships. Two regionally selective analyses explore distances and similarities within England and Scotland. Skewness maps and kernel maps, which represent similarities and differences by means of different shades of colour, are used to identify *Sprachausgleiche*, “linguistic compromise and exchange” areas (p. 81), where there is an accumulation of intermixing between dialects, and dialect ‘kernels’, which constitute nuclei within the dialect landscape. A comparison of dialectal similarities and distances with the reference corpora reveals greater differences within Britain than between Standard British and Standard American English (p. 84).

The question of whether the British dialect landscape is characterized as a continuum (smooth transitions) or as dialect areas (abrupt breaks) is addressed in Chapter 5 ‘Is morphosyntactic variability gradient? Exploring dialect con-

tinua' (pp. 89–110) and Chapter 6 'Classification: The dialect area scenario' (pp. 111–127). Plots of multidimensional scaling and continuum maps that take geographical proximity as a "proxy for the likelihood of social contact and communicative interaction" (p. 91) show that Britain is characterized by a mixture of continua and dialect areas. Of four measures to quantify geographical distance ('as the crow flies', walking, least-cost travel and Trudgill's (1974) 'gravity' model), the gravity model (which considers not only distance between population centres but also their relative sizes) makes predictions that best match the study's findings. Hierarchical cluster analysis, expressed as dendrograms and cluster maps, reveals clustering that tends to agree with customary dialect divisions and perceptual dialectology studies (pp. 124, 126).

Chapter 7 'Back to the features' (pp. 128–150) looks for bundles of features through Principal Component Analysis (PCA), which identifies four principal components, expressed in component score maps: a nonstandard component involving a number of nonstandard forms led by past-tense *come*, which creates a North-South continuum; a component involving *do* and *have*; a component involving *be*; and a component involving *would*. Outlier dialects turn out to be due to poor sampling (that is, small numbers of speakers or small frequencies of some features).

Chapter 8 'Summary and discussion' (pp. 151–163) summarizes the findings of the study of morphosyntactic variability in British English dialects. The concern with geography has tended to bias the selection of features in traditional dialect atlases, which may have led to an exaggeration of the importance of the role of geography in dialectal variation. Szmrecsanyi's conclusion is that "geography is important, but it is not all-important" (p. 158). Chapter 9 'Outlook and concluding remarks' (pp. 164–165) addresses remaining issues having to do with the role of frequency, language ideology/attitudes and speaker identity in conditioning the variation, as well as the problem with dialectology's traditional focus on NORMs (Non-mobile Older Rural Males). There are three appendices that provide the summary statistics of the feature catalogue, the matrix of PCA component loadings and the colour maps referred to throughout the book.

The study presented in this book brings an impressive array of statistical techniques to bear on our understanding of the grammatical component of British dialect variation, and the use of colour provides an extra dimension to the detail of dialect maps. However, the study is limited in certain ways by decisions made about the selection and extraction of features. As Szmrecsanyi acknowledges (p. 38), the reliance on (semi-)automatic extraction means that certain 'hard-to-retrieve' null features had to be purposely excluded. Unfortunately, such features are among those that contribute to important distinctions between

varieties of English (cf. Walker and Meyerhoff 2006; Torres Cacoullous and Walker 2009). Automatic extraction also leads to a reliance on what Szmrecsanyi calls ‘absolute’ frequencies (i.e. number of occurrences, whether log-transformed or ‘normalized’ per 100,000 words) (p. 26) as input to the analysis. While this approach may be appealing because it cuts down on the seemingly tedious task of extraction and coding, the principle of accountability (Labov 1972) requires that we extract not only tokens of the feature we are interested in but also instances where that feature could have occurred but did not. Szmrecsanyi acknowledges in several places (pp. 38, 101, 165) the importance of examining not only the frequency with which a feature occurs but also the elements of the linguistic context that condition its occurrence; yet that examination is not undertaken as part of the study. Despite these limitations, this book should serve as a model for future work in the study of dialect variation, especially when not only grammatical features but also phonetics and the lexicon are considered.

References

- Ellis, Alexander John. 1889. *On early English pronunciation*. Vol. V: *The existing phonology of English dialects*. London: N. Trübner. [Republished (1968) New York: Greenwood Press.]
- Inoue, Fumio. 1996. Subjective dialect division in Great Britain. *American Speech* 71 (2): 142–161.
- Labov, William. 1972. *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.
- McMahon, April, Paul Heggarty, Robert McMahon and Warren Maguire. 2007. The sound patterns of Englishes: Representing phonetic similarity. *English Language and Linguistics* 11 (1): 113–142.
- Orton, Harold and Eugen Dieth. 1962–1971. *Survey of English dialects: Basic materials*. Leeds: E.J. Arnold & Son.
- Shackleton, Robert. 2007. Phonetic variation in the traditional English dialects: A computational analysis. *Journal of English Linguistics* 35 (1): 30–102.
- Torres Cacoullous, Rena and James A. Walker. 2009. On the persistence of grammar in discourse formulas: A variationist study of *that*. *Journal of Linguistics* 47 (1): 1–43.
- Trudgill, Peter. 1974. Linguistic change and diffusion: Description and explanation in sociolinguistic dialect geography. *Language in Society* 3 (2): 215–246.

- Trudgill, Peter. 1990. *The dialects of England*. Oxford and Cambridge, MA: Blackwell.
- Walker, James A. and Miriam Meyerhoff. 2006. Zero copula in the eastern Caribbean: Evidence from Bequia. *American Speech* 81 (2): 146–163.