Wright, Susan. 1997. Speaker innovation, textual revision and the case of Joseph Addison. In Terttu Nevalainen & Leena Kahlas-Tarkka (eds.), *To explain the Present: Studies in the changing English language in honour of Matti Rissanen*, 483–503. Helsinki: Société Néophilologique.

(Received 9 November 2016)

doi:10.1017/S1360674316000587

Benedikt Szmrecsanyi, *Grammatical variation in British English dialects: A study in corpus-based dialectometry*. Cambridge: Cambridge University Press. Pp. xvii+211. 2013 hardback ISBN 9781107003453 and 2015 paperback ISBN 9781107515772.

Reviewed by James N. Stanford, Dartmouth College

In this innovative and thoughtful study, Benedikt Szmrecsanyi uses dialectometry to explore morphosyntactic patterns in traditional dialects across Great Britain. The book breaks new ground by (1) choosing morphosyntax as the topic of dialectometry research, and (2) taking a corpus-based approach to dialectometry. Szmrecsanyi's book is a welcome contribution to a field that more commonly focuses on phonological variables, usually based on data from reading passages or word lists.

Many linguists and members of the general public are fascinated by dialect differences and the role of geographic distance between communities. Despite the increasing digital connectedness of the modern world, the great majority of spoken (or signed) communication still involves human beings in close physical proximity to each other. Physical distance therefore continues to have a crucial role in language variation and change, following Bloomfield's principle of density and face-to-face interaction (1933: 476) and Nerbonne & Kleiweg's (2007) Fundamental Dialectological Postulate. Dialectometry is one way to empirically explore the effects of physical distance and take another step toward satisfying the age-old human curiosity about language variation and geography.

Correlations between physical distance and language variation can be studied in multiple ways. The approach in this book, aggregation and dialectometry, is relatively uncommon on the western side of the Atlantic (Grieve 2016), but numerous studies in Europe, East Asia and elsewhere have contributed to a long history of dialectometric aggregration (Séguy 1971, 1973). Examples include Goebl (1993); Heeringa (2004); Nerbonne & Kretzschmar (2003); Nerbonne (2006, 2009, 2010); Nerbonne & Kleiweg (2007); Heeringa, Johnson & Gooskens (2009); Yang (2010); Wieling & Nerbonne (2015); Pelkey (2015), inter alia. Such studies have primarily targeted phonological features, and Szmrecsanyi's study of British English morphosyntax is the next logical step.

Szmrecsanyi's dialectometric analyses of British English are precise and thorough, including a wide range of metrics, schools of thought, and mapping methods. The author has an excellent ability to clearly describe complex concepts, such as multidimensional scaling (pp. 91–9),[1] including appropriate discussions of strengths and weaknesses of different approaches. In this respect, the book would be appropriate for all graduate levels of linguistics and scholarly use, as well as advanced undergraduate classes.

The book has three main goals, quoted below from page 6:

(i) Does a frequency-derived measure of morphosyntactic variability in traditional British English dialects exhibit a geographic signal?
(ii) If there is a geographic signal, exactly how are morphosyntactic distances and similarities distributed? Specifically: are we dealing with a dialect continuum scenario or with a dialect area scenario?
(iii) Do feature subsets make a difference, and what is the extent to which individual features gang up to create areal (sub)patterns?

Szmrecsanyi's primary data set is the *Freiburg Corpus of English Dialects* (FRED). Most of the speakers in this corpus were born around the turn of the twentieth century and recorded in the 1970s and 1980s when they were senior citizens. With the FRED data, Szmrecsanyi is able to probe geographic distinctions from an earlier era and also draw comparisons with modern English corpora (the *International Corpus of English – Great Britain* and the *Santa Barbara Corpus of Spoken American English*). FRED consists of 368 interviews involving 427 speakers, for a total of 2,437,000 words of transcribed text (pp. 15–16). One of the advantages of FRED is that it enables researchers to examine older 'traditional dialects…which are dying out fast' (p. 5). There are also some well-known drawbacks to this type of data set, as the author recognizes. In particular, FRED primarily involved 'elderly speakers with a working-class background – so-called NORMs (*non-mobile old rural males*)' (p. 4, citing Chambers & Trudgill 1998: 29). Data sets based on NORMs have well-known problems with gender imbalance, lack of demographic diversity, and a narrow focus on rural communities. At the same time, as the traditional dialectologists pointed out during their heyday (e.g. Kurath 1939), the NORM approach is one way to explore some of the oldest and most regionalized forms, albeit not a very inclusive approach from a modern viewpoint. Regardless of the ideology of its era, the FRED data set exists and it is worthy of detailed analysis.

Szmrecsanyi's dialectometric study of FRED examines 57 morphosyntactic features (listed on p. 24) in 158 locations representing 34 counties in Great Britain (pp. 137–50). Most of book's analyses use these 34 counties as the geographic data points; an exception is the section on Principal Components Analysis (section 7.2) where calculations are based on the 158 individual speaker locations. The author created a

[1] Page numbers refer to the 2015 paperback edition.

https://doi.org/10.1017/S1360674316000587 Published online by Cambridge University Press

34x57 frequency matrix, i.e. each of the 34 counties has a single value for each of the 57 features. In the matrix, speaker means were computed for each feature and coded as gradient frequency counts averaged over all speakers in the county, regardless of the number of speakers in the county. As discussed below, this approach appropriately follows standard dialectometry methods, although it raises some questions as well. The author also confirmed the overall consistency of this feature matrix by computing a Cronbach's alpha of 0.77, which is in the conventionally acceptable range (p. 26).

Overview

The introduction (chapter 1) provides a concise literature review of prior dialectology and dialectometry research on British English, including the *Survey of English Dialects* (SED; Orton & Dieth 1962), the foundational role of Trudgill (1990) in establishing dialect divisions, and a description of major contemporary schools of thought in dialectometry and related fields.

Chapter 2 introduces the methods and data sources of the present study, including background on the FRED data. The chapter also introduces software tools: Peter Kleiweg's RuG/L04 dialectometry software (website given on p. 30) and *Visual DialectoMetry* software (Haimerl 2006).

Chapter 3 provides a detailed discussion of the 34 morphosyntactic features in terms of their individual distributions and relative frequencies in FRED. The chapter reports on each feature's geographic distribution across Great Britain, and provides statistical results for the geographic significance of each one. Considered as individual features, 14 (25 percent) of the 57 morphosyntactic features have significant geographic distributions in various parts of Great Britain (see p. 152 for a summary list of the 14 features and their geographic distributions). These results help lay a foundation for the aggregate analysis in the remainder of the book, i.e. although only 14 of 57 features are geographically significant as individual features, many other patterns and relationships appear when the data are considered in aggregate.

Chapter 4 computes the aggregate morphosyntactic distances and similarities of the FRED corpus according to the 34 British English counties, along with comparisons to Standard American English and Standard British English. The aggregation involves correlations of pairwise geographic distances against pairwise dialect differences that are calculated as 'the square root of the sum of all fifty-seven pairwise feature frequency differentials' (p. 153). This chapter also refers the reader to interesting color maps in the appendix that highlight geographic contrasts and similarities, network analyses, histograms comparing skewness and other properties, along with a variety of other cartographic and analytical approaches. The chapter also discusses dialect kernels and patterns of compromise/exchange between the North of England and newer dialects in Wales and the Scottish Highlands. For example, the 'relatively young dialects in the Scottish highlands and on the Hebrides are overall quite close [to Standard British English]' (p. 85) since these are areas which historically had mostly Scottish Gaelic speakers. In these regions, English is relatively new, and so

https://doi.org/10.1017/S1360674316000587 Published online by Cambridge University Press

traditional dialect features are not present. Likewise, the results suggest that there is greater morphosyntactic similarity between British and American dialects than across different British regions (p. 84).

Chapters 5 and 6 tackle the theoretical and empirical question of dialect continua versus dialect areas (Heeringa & Nerbonne 2001). That is, Szmrecsanyi uses the FRED data to determine whether British English morphosyntactic variation is best interpreted in a dialect continua model (chapter 5) or in terms of dialect areas (chapter 6). These chapters highlight the value of dialectometry in supplying aggregated empirical data to address questions of this type. For FRED morphosyntax, Szmrecsanyi finds that some regions are more like a dialect continuum and others are more like a dialect area. On the whole, Great Britain's morphosyntax is 'not very continuum-like' (p. 110). Some individual regions, such as the Southwest of England and the Central Scottish Lowlands, are better suited to a dialect continuum model. Overall, the statistical predictors of dialect continua proved more relevant for Scotland than England, whose sharp discontinuities are best analyzed in terms of dialect areas (pp. 154–6). These sections of the book use multidimensional scaling, cluster analysis, as well as a variety of different measures of physical distance: simple 'as-the-crow-flies' distance (pp. 101–2); Google Maps' walking distance (p. 103), car travel distances (pp. 104–5); and gravity-based equations (pp. 105–7, following Trudgill 1974). The gravity approach produced the best fit to the data ($R^2 = 24.1\%$). Citing Johnstone (2009: 6), Szmrecsanyi notes that 'all of these measures can be seen as proxies of the likelihood of social contact and communicative interaction, as the underlying variable that is supposed to shape the diffusion of linguistic features in geographic space' (p. 100).

Chapter 7 examines some outlier locations (Banffshire, Denbighshire, Dumfriesshire, Middlesex, Leicestershire, Warwickshire) which behaved somewhat differently than other nearby data points, and possible reasons for these discrepancies are proposed. The chapter then explores how Principal Components Analysis (PCA) can be used to analyze feature bundles and co-occurrence patterns in the FRED data. Unlike the preceding chapters, this section uses the locations of all 158 speakers as data points, not grouping them into 34 counties, because the larger-sized matrix is better for PCA computations. The results suggest that four 'layers' (components) are most significant: (1) the 'non-standard *come* component; (2) a component involving variants of *do* and *have*; (3) a component involving *be*; and (4) a component involving *would.* Each of these principal components is discussed in detail, and numerous example sentences are provided from the FRED corpus.

The book concludes with two well-organized concluding chapters (chapters 8 and 9) that provide summaries and an outlook for future research.

Variationist perspectives

Linguistics is a diverse field with diverse viewpoints and methodologies, all of which have their own strengths and weaknesses. In this section, I will discuss the view from

https://doi.org/10.1017/S1360674316000587 Published online by Cambridge University Press

the world of variationist sociolinguistics. Variationist studies in the model of Labov (1966) and subsequent work in that paradigm sometimes lack a robust geographic component, focusing instead on social and linguistic factors. Dialectometry's in-depth statistical analyses of geographic factors are therefore appealing on many levels. Many variationists would appreciate the geographic patterns uncovered in this study, but also raise questions and different perspectives about some of the methods.

First, variationists reading this book may note the lack of emphasis on variability across individual speakers, including issues of gender, age, social class, mobility, identity, speech style and other factors. Dialectology and dialectometry traditionally average all the responses into a single value for each location: 'All the results from a given locality are presented together, making it impossible to distinguish the responses of individual informants' (Shackleton 2007: 33). Naturally, any researcher would be glad to find a data set that has *both* a vast range of geographic locations *and* a vast number of individual speakers from each site, perfectly balanced for age, gender, social class and all other factors, as well as multiple speech styles and social contexts for each speaker. Obviously, no such perfect data set exists for Great Britain or anywhere else. The author is using the available FRED data as precisely and thoroughly as the data set will allow. Even so, factors like gender are well established as potential contributors to language variation and change (e.g. Labov 1972/1991, 1990, 2001; Trudgill 1974), and such potential factors could be discussed in more detail (e.g. Wieling, Nerbonne & Baayen 2011; see also the sociolectometry of Ruette, Geeraerts, Peirsman & Speelman 2014).

For example, all of Banffshire county is represented by a single (female) speaker, and the author rightly considers this thin amount of sampling to be a likely reason for the outlier behavior of Banffshire in the analysis (p. 130). Likewise, Denbighshire county is represented by 4 male speakers and 2 female speakers, and so on, whereas other counties are much more solidly represented. The FRED data set for the Hebrides has 53 speakers and 19 locations; Shropshire has 38 speakers and 14 locations, etc. (pp. 18–22). Yet most of the dialectometric analyses in the book treat each of these counties as a single point. The author recognizes that this puts considerable weight on a few speakers in certain locations versus a much larger sample for averaging in other locations, such as Shropshire (most of the outlier data points discussed in section 7.1 appear to be locations with few speakers). It would also be helpful to see a map of the individual 158 locations, rather than just the 34 counties (see Shackleton 2007: 34).

In sum, because the study is encompassing all of Great Britain with just 34 county-level data points (except for the PCA work which uses 158 locations) and varying numbers of samples representing each data point, the results will always be a bit less than fully satisfying in terms of granularity. Compare Shackleton's (2007) phonetic study of England alone, which enjoys a more robust geographic data set (313 locations from the *Survey of English Dialects*). Such a data set improves the likelihood of geographic patterns that are not dependent on a few speakers. That said, morphosyntactic data with robust geographic coverage is much harder to find than phonological data. FRED is still an excellent choice for corpus data on morphosyntax

https://doi.org/10.1017/S1360674316000587 Published online by Cambridge University Press

with geographic patterns. Szmrecsanyi should be applauded for cleverly extracting so many interesting patterns and facts from this data set.

Second, variationists may voice concerns about aggregation of dialect features. To be fair, at least one type of aggregation – isogloss bundles – has been in the linguistics toolbox since Bloomfield and earlier. Bloomfield states that bundling 'offers a more suitable basis of classification than does a single isogloss that represents, perhaps, some unimportant feature' (1933/1984: 342, quoted by Szmrecsanyi on p. 2). This tradition of feature-bundling was found throughout the era of traditional dialectology. It continues in Labov, Ash & Boberg's (2006) geographic analysis of US dialect areas, including bundles that define the US Inland North (Northern Cities Vowel Shift), the US South and many others. Szmrecsanyi also argues that dialects are multidimensional objects that naturally lend themselves to an aggregate approach:

> The point is that so-called 'single feature-based studies' Nerbonne (2009: 176), with their atomistic focus on typically just one feature, are fine when it is the features themselves that are of analytic interest. They are woefully inadequate, however, when it comes to characterizing multidimensional objects… The problem with single-feature-based studies – in linguistics as well as everywhere else – is that feature selection is ultimately arbitrary (see Viereck 1985: 94), and that the next feature down the road may or may not contradict the characterization suggested by the previous feature. (pp. 2–3)

At the same time, aggregation of dialect features can potentially miss some interesting patterns in individual features (Szmrecsanyi checks for such patterns in individual features in chapter 3), and the aggregation approach also risks overlooking features that may be rare in frequency yet have an important part in an enregistered set of dialect features (Agha 2003; Johnstone 2011). The author recognizes this potential issue: 'we acknowledge explicitly that low-frequency features, despite their absence in speech corpora, may be structurally interesting and perceptually in fact salient' (p. 37).

Perhaps we can take advantage of the complementary strengths of both dialectometry and variationist sociolinguistics. It seems likely that a combination of the two approaches could be advantageous in some situations. In my own variationist work in rural China, I have found that aggregate methods can be a valuable complement. In particular, dialectometry has helped me explore time-depths beyond the typical horizon of Labovian methods. The Sui minority of Guizhou province is a clan-based society, and my synchronic data of the Sui language showed that clan dialect boundaries are relatively 'airtight': most clan dialect features remain stable across all ages, despite exogamous mobility and contact. The effects of dialect accommodation and contact between clans are limited because speakers have strong clan identities, and because the community strongly proscribes acquisition of other clans' dialect features. Moreover, a real-time comparison across fifty years also showed long-term stability: little or no change in the dialect boundaries over half a century. But a dialectometric analysis (using word-list pronunciation data) showed that simple geography was in fact a significant factor in Sui variation, despite the social strength of clan-based boundaries (Stanford 2012). The Sui dialectometry evidence suggests

https://doi.org/10.1017/S1360674316000587 Published online by Cambridge University Press

that long-term geographic diffusion has been occurring in the region at a much greater time-depth than I could have imagined using only sociolinguistic methods. In other words, clan dialect boundaries are a crucial part of Sui social organization and dialect patterns, but simple physical distance has an important long-term effect as well. Dialectometry helped bring this perspective to light.

Like other communities of scholars, linguists will probably always be engaged in debates about the merits of different methodologies, data sources and theoretical approaches. *Grammatical Variation in British English Dialects* is a welcome addition to these discussions. With this book, Szmrecsanyi has conclusively established that dialectometric methods can provide empirical perspectives on the role of physical distance in morphosyntactic variation extracted from corpus-based data. Future scholars can build on this sturdy foundation, and perhaps complement it with other methodologies as well.

Reviewer's address:
Program in Linguistics
Dartmouth College
6220 Reed Hall
Hanover NH 03755-3562
USA
James.N.Stanford@Dartmouth.edu

References

Agha, Asif. 2003. The social life of a cultural value. *Language and Communication* 23, 231–73.

Bloomfield, Leonard. 1933/1984. *Language*. Chicago: University of Chicago Press.

Chambers, J. K. & Peter Trudgill. 1998. *Dialectology*, 2nd edition. Cambridge: Cambridge University Press.

Goebl, Hans. 1993. Dialectometry: A short overview of the principles and practice of quantitative classification of linguistic atlas data. In Reinhard Kohler & Burghard Rieger (eds.), *Contributions to quantitative linguistics*, 277–315. Dordrecht: Kluwer.

Grieve, Jack. 2016. *Regional variation in written American English.* Cambridge: Cambridge University Press.

Haimerl, Edgar. 2006. Database design and technical solutions for the management, calculation, and visualization of dialect mass data. *Literary and Linguistic Computing* 21(4), 437–44.

Heeringa, Wilbert. 2004. Measuring dialect pronunciation differences using Levenshtein distance. PhD thesis, University of Groningen.

Heeringa, Wilbert, Keith Johnson & Charlotte Gooskens. 2009. Measuring Norwegian dialect distances using acoustic features. *Speech Communication* 51(2), 167–83.

Heeringa, Wilbert & John Nerbonne. 2001. Dialect areas and dialect continua. *Language Variation and Change* 13(3), 375–400.

Johnstone, Barbara. 2009. Language and geographical space. In Peter Auer & Jürgen Erich Schmidt (eds.), *Language and space: An international handbook of linguistic variation*, vol. 1: *Theories and methods*, 1–18. Berlin: Mouton de Gruyter.

https://doi.org/10.1017/S1360674316000587 Published online by Cambridge University Press

Johnstone, Barbara. 2011. Dialect enregisterment in performance. *Journal of Sociolinguistics* 15(5), 657–79.

Kurath, Hans. 1939. *Handbook of the linguistic geography of New England*. Providence, RI: Brown University Press.

Labov, William. 1966. *The social stratification of English in New York City*. Washington, DC: Center for Applied Linguistics; Cambridge: Cambridge University Press.

Labov, William. 1972/1991. *Sociolinguistic patterns*. Philadelphia, PA: University of Pennsylvania Press.

Labov, William. 1990. The intersection of sex and social class in the course of linguistic change. *Language Variation and Change* 2, 205–54.

Labov, William. 2001. *Principles of linguistic change*, vol. 2: *Social factors*. Malden, MA: Blackwell.

Labov, William, Sharon Ash & Charles Boberg. 2006. *The atlas of North American English*. Berlin: Mouton de Gruyter.

Nerbonne, John. 2006. Identifying linguistic structure in aggregate comparison. *Literary and Linguistic Computing* 21(4), 463–75.

Nerbonne, John. 2009. Data-driven dialectology. *Language and Linguistics Compass* 3(1), 175–98.

Nerbonne, John. 2010. Measuring the diffusion of linguistic change. *Philosophical Transactions of the Royal Society B* 365, 3821–8.

Nerbonne, John & Peter Kleiweg. 2007. Toward a dialectological yardstick. *Journal of Quantitative Linguistics* 14(2), 148–66.

Nerbonne, John & William Kretzschmar Jr. 2003. Introducing computational methods in dialectometry. *Computers and the Humanities* 37(3), 245–55.

Orton, Harold & Eugen Dieth. 1962. *Survey of English dialects*. Leeds: E. J. Arnold.

Pelkey, Jamin. 2015. Reconstructing phylogeny from linkage diffusion: Evidence for cladistic hinge variation. *Diachronica* 32(3), 397–433.

Ruette, Tom, Dirk Geeraerts, Yves Peirsman & Dirk Speelman. 2014. Semantic weighting mechanisms in scalable lexical sociolectometry. In Benedikt Szmrecsanyi & Bernhard Wälchli (eds.), *Aggregating dialectology, typology, and register analysis*, 205–303. Berlin: Walter de Gruyter.

Séguy, Jean. 1971. La relation entre la distance spatiale et la distance lexicale. *Revue de Linguistique Romane* 35, 335–57.

Séguy, Jean. 1973. La dialectometrie dans l'atlas linguistique de la Gascogne. *Revue de Linguistique Romane* 37, 1–24.

Shackleton, Robert G., Jr. 2007. Phonetic variation in the traditional English dialects: A computational analysis. *Journal of English Linguistics* 35(1), 30–102.

Stanford, James N. 2012. One size fits all? Dialectometry in a small clan-based indigenous society. *Language Variation and Change* 24(2), 247–78.

Trudgill, Peter. 1974. Linguistic change and diffusion: Description and explanation in sociolinguistic dialect geography. *Language in Society* 3(2), 215–46.

Trudgill, Peter. 1990. *The dialects of England*. Cambridge: Blackwell.

Viereck, Wolfgang. 1985. Linguistic atlases and dialectometry: The survey of English dialects. In John M. Kirk, Stewart Sanderson & J. D. A. Widdowson (eds.), *Studies in linguistic geography: The dialects of English in Britain and Ireland*, 94–112. London: Croom Helm.

Wieling, Martijn & John Nerbonne. 2015. Advances in dialectometry. *Annual Review of Linguistics* 1, 243–64.

Wieling, Martijn, John Nerbonne & R. Harald Baayen. 2011. Quantitative social dialectology: Explaining linguistic variation geographically and socially. *PloS ONE* 6(9), e23613.

https://doi.org/10.1017/S1360674316000587 Published online by Cambridge University Press

Yang, Cathryn. 2010. Lalo regional varieties: Phylogeny, dialectometry, and sociolinguistics. PhD thesis, La Trobe University.

(Received 16 December 2016)

doi:10.1017/S1360674317000041

Douglas Biber and **Bethany Gray**, *Grammatical complexity in academic English* (Studies in English Language). Cambridge: Cambridge University Press, 2016. Pp. xiv + 277. ISBN 9781107009264.

Reviewed by Viviana Cortes, Georgia State University

In true tradition of the School of Flagstaff of corpus-based linguistic analysis, Biber & Gray bring us a very informative volume in which they analyze a variety of aspects of grammatical complexity in English academic writing. The book is arranged in seven chapters topically organized around different features of grammatical complexity that are strategically related to highlight the primary focus of the book: to discuss 'phrasal complexity features and the associated phrasal discourse style that is typical of present day science research writing' (p. 39).

Chapter 1 is much more than a simple introduction. In this chapter, the authors identify in previous language research studies major stereotypes about grammatical complexity, language change processes and academic writing. They set out to challenge these stereotypes in the rest of the book, showing empirically based descriptions of academic writing patterns of use studied over time and across different language registers and subregisters.

In chapter 2, Biber & Gray introduce the different corpora used to conduct the studies presented in this volume and they provide a clear and detailed description of the various procedures they implemented for the grammatical analyses conducted. A wide range of corpora were analyzed, including corpora specially collected for this book as well as more established corpora previously used in other research studies. The *20th Century Research Article Corpus* was specially compiled by the authors and contains published research articles from journals in science, social science and the humanities from three twenty-year intervals (1965, 1985 and 2005), comprising 570 texts and about 3.6 million words. Other corpora and subcorpora used for comparison were ARCHER (*A Representative Corpus of Historical English Registers*), CETA (*Corpus of English Texts in Astronomy*), samples from the *Philosophical Transactions of the Royal Society*, the LSWE (*The Longman Spoken and Written English*) Corpus, subcorpora from the T2KSWAL (*TOEFL 2000 Spoken and Written Academic Language*) Corpus and a group of texts from Project Gutenberg, among others. The corpora were grammatically annotated using the Biber tagger, which relies on