5 Dialectology for Computational Linguists

John Nerbonne, Wilbert Heeringa, Jelena Prokić and Martijn Wieling

1 Introduction

This chapter provides an overview of computational work in dialectology. We have published similar surveys in the not-too-distant past (Wieling and Nerbonne, 2015; Heeringa and Prokić, 2018), but these were aimed at dialectologists and general linguists, respectively. This chapter is written for computational linguists, so that we will focus less on the technical details of exploiting the computer in research on dialects (which is documented in the articles we cite) but rather more on background assumptions and emerging issues and opportunities.

2 Dialectology

James Walker's and Miriam Meyerhoff and Steffen Klaere's introductory chapters clarify that one normally reserves the term *dialect* for what they also call *diatopic variation*, in particular, a geographically restricted language variety, and we focus our attention here on the study of this sort of variety, ignoring the other dimensions of variation, including social, situational, and temporal. We hasten to add both that these are interesting (see other chapters) and also that many of the same issues arise in the (computational) analysis of nongeographical variation.

2.1 Research Questions and Research Perspectives

Many researchers and laypeople alike are fascinated by dialects, what makes them up, and where they are spoken. This means that there is a substantial intrinsic motivation for the study. The two questions – what is special about a given dialect and where is it spoken – are the two main research questions that are asked in dialectology, the scientific study of dialects.

Lay interest seldom extends to explanations of dialectal facts – why certain language elements are different and why the geographical distribution takes the form it does, but dialectologists are passionately interested in these questions, too, and we will include attempts at explanation arising from computational work as well. Naturally we strive as scientists to provide not only accurate answers to these questions but also insightful ones, i.e., compact characterizations with defensible assumptions.

Computational linguists interested in developing classifiers for dialectal material, i.e., procedures to assign dialectal material to one of a given set of varieties, may note that this is indeed a natural way to construe the first research question. It will be sufficient if the classifiers to be developed are discriminative rather than generative in the sense of Ng and Jordan (2002). See also chapter 9 in this volume.

2.2   Comparative, Nongenerative Perspective

Implicit in the way we have sketched the research questions is that we are seldom interested in a complete description of a dialect in the sense of generative grammar, i.e., a description that would allow one to decide for any putative utterance, whether it belongs to the dialect or not. This goal has been abandoned in Chomskyan generative linguistics in any case (Nerbonne, 2018). Instead we are primarily interested in what is different in the dialect under study with respect to other dialects, and we largely ignore what is common. In any case, there have been few attempts at developing comprehensive grammars of a range of dialects.[1] The focus on comparison means that computational techniques that proceed by processing comparable material from a range of dialects will be most useful, as we shall see in this chapter.

This does not mean, however, that there has been no interest in dialectology on the part of generative grammar; on the contrary, many generativists have been interested. Barbiers (2009) explores *syntactic microvariation*, i.e., the sort of detailed variation that one finds across a range of dialects, in an attempt to characterize that range in generative terms. See Hinskens (2018) and references therein for a recent survey of generative work in phonology. But their interest is also primarily comparative, e.g., in exploring how a syntactic phenomenon such as inflected complementizers or a phonological phenomenon such as final devoicing varies from one dialect to another. They add to this an interest in what's common with respect to a particular phenomenon under study, but their perspective is likewise discriminative.

2.3   Data Collection

Dialectologists have experimented with myriad data collection techniques, attending carefully to sampling (Macauley, 2018), which must go beyond random sampling of the usual sort in order to ensure coverage of the entire area where a variety is spoken. Dialectologists have experimented not only with questionnaires (Llamas, 2018) and other written surveys (Chambers, 2018), which better guarantee that responses are comparable, but also with personal interviews (Bailey, 2018), which, in the hands of a good interviewer, provide a better chance at obtaining natural speech. Increasingly, dialectologists are turning to the World Wide Web as a means of conducting surveys, e.g., the *Atlas zur deutschen Alltagssprache* at www.atlas-alltagssprache.de (Möller and Elspaß, 2008), as well as

---

[1]Wagner and O'Baoill's *Linguistic Atlas and Survey of Irish Dialects* (1958) is perhaps a counterexample, but well outside the generative tradition (see discussion in text).

to smart phone apps and crowdsourcing (Scherrer et al., 2012; Goldman et al., 2014).

Dialectology has always been a discipline that has cherished data, collecting notably it in the form of *dialect atlases* (Kretzschmar Jr et al., 1993), which in modern times, are normally accompanied by a database of dialect speech material (Van den Berg, 2003; Elspaß and König, 2008). Computational linguists interested in pursuing research in dialectology should definitely consider collaboration with one of the many dialect atlases across the world.

3   Dialectometry

As the opening chapters of this handbook note, one of the primary data analysis steps in traditional dialectology involved decorating a map with codes corresponding to the different ways of expressing the same thing in the language area under study. This might be [ɑr] vs. [a] (the latter represents the "r-less" "ah" in the pronunciation of *car* in eastern New England and in most of Great Britain). In a following step the researcher attempted to draw an *isogloss* separating the area with the one code ([ɑr]) from the area with the other ([a]). In fact this would run north–south in New England (see Labov et al. (2006a, Map 11.12) for a larger geographical view). If several linguistic features share their geographical distributions, one might speak of *isogloss bundles*, groups of isoglosses with roughly the same geographical course. But the problems with this rough procedure have long been appreciated. Even in simple cases it might be impossible to draw a single line, and since one can draw tens of thousands of isoglosses, the choice of which to regard as definitive was vexing (Nerbonne, 2009). Areas were normally drawn with sharp borders, even while dialectologists conceded that they often witnessed gradual, unsharp boundaries.

Traditional dialectology often relied on extremely knowledgeable practitioners – both field workers and project managers. They drew dialect maps without resorting to counting the number of differing features or to recording the reactions of dialect speakers. As Bill Kretzschmar (2009) once said of Kurath, the dean of American dialectology in his time "Kurath knew that he wanted to subdivide the survey region, and he only needed to find diagnostic isoglosses to match his perceptions, guided by his long experience." In the absence of more replicable methods, the authority of researchers such as Kurath were accepted.

Jean Séguy (1971, 1973) initiated the work in *dialectometry*, i.e., the "measure of dialect". He worked on the dialect atlas of Gascony in southwest France, and sought a way to avoid choosing the basis for his dialect characterizations and his dialect maps in an arbitrary way. His solution was simple: he examined the differences per item between the data collection sites, counting different items as contributing one to a difference and identical ones as contributing zero. He derived from this an aggregate difference, which in fact is a distance: zero for identical sites, symmetrical, and conforming to the triangle inequality. Séguy published only the two papers cited on his ideas before he met an unexpected death in an automobile accident.

Hans Goebl (1982) was in correspondence with Séguy and adopted his ideas enthusiastically even while elaborating a great deal on them. Goebl examined each point (data collection site) and the distribution of distances (actually, Goebl always preferred to analyze similarities, i.e., inverse distances) to all the other sites, e.g., in an effort to characterize transitional zones.

Although Goebl emphasizes the care that he has always taken in collecting and codifying data (a step he refers to as "taxation"), a clear disadvantage in his work from a computational point of view is that care *must* be taken. Only categorical data are analyzed by the later step in Goebl's procedures. This is perhaps okay as the basis to analyze lexical differences, since lexemes are either the same at two sites or different.[2] Pronunciation (a sound shift) or even syntax might have been part of the data collection in early dialectometry, but then manual procedures were used to extract the categorical differences that are then subject to further analysis. There are enormous opportunities for computational linguistic work in extracting the interesting differences automatically.

Perhaps Goebl's most lasting contribution to dialectometry was the introduction of *clustering* as a means of detecting dialect areas. The number of features that two varieties share may be regarded as the inverse of a distance, as we just noted, and that distance gives rise to a half matrix of distances when the entire set of data collection sites is analyzed. Goebl noted further that dialect areas are often hierarchically organized – e.g., Badisch, Swabian and Bavarian are all southern German dialects, meaning that they share features that distinguish them from the northern varieties. This made *hierarchical clustering* the method of choice (as opposed to *k-means*, for example, which simply partitions the input without attention to potential further internal structure), and Goebl was the first to champion this. When a half matrix of aggregate differences (or, conversely, similarities, which Goebl always preferred to work with) were clustered – with no reference to their locations, the results normally projected nicely to convex geographical regions. This was in itself a confirmation of the approach. Goebl's work on clustering pushed the discussion concerning the determination of dialect boundaries to a new, more replicable level.

Clustering algorithms are, however, notoriously unstable: small differences in the input can be magnified to quite large differences in results. For this reason further work has been conducted to develop more stable versions. One technique, the *bootstrap* (and note that the term we took from biological work is used differently from computational linguistics), re-clusters several times (typically 100 or 1.000) using random subsets of the original data, e.g., the distances among a subset of the entire set of words. Alternatively, one may recluster using variable small amounts of noise. These procedures indeed solve the stability problem when parameters are chosen judiciously (Nerbonne et al., 2008). A further problem arises in the choice of algorithm. Further computational linguistic work has examined several of the various hierarchical algorithms, concluding that those where distances between non-atomic clusters were determined using

---

[2]But this identity can be difficult to ascertain. See Heeringa and Nerbonne (2006) for an example of a set of five word forms that is difficult to partition into lexeme subsets.

nearest neighbor regimens or schemes using the (weighted) means of distances between cluster elements (WPGMA amd UPGMA) yield acceptable results, while centroid-based elements and complete-link (aka "farthest neighbor") did not. Evaluation using standard measures of cluster quality, such as the modified Rand index, cophenetic correlation coefficient, entropy, and purity (Prokić and Nerbonne, 2008) as well as a comparison to the more stable multi-dimensional scaling procedure (MDS, see later).

As computational linguists are well aware, the interest in clustering and in developing novel algorithms and techniques is large and is not likely to diminish, given the impossibility of designing perfect clustering algorithms (Kleinberg, 2004).

4   Edit distance on phonetic transcriptions

In the past three decades, the computational analysis of dialects has been performed at various linguistic levels, including phonetics, morphology, lexicology and syntax, with pronunciation differences being the most studied. In this section we discuss *edit distance* – a very simple, yet efficient way commonly used in dialectometry to infer pronunciation differences between language varieties. It has successfully been applied to many languages in aggregate analysis of dialect varieties: Dutch (Nerbonne et al., 1996; Heeringa, 2004; Wieling et al., 2007), Sardinian (Bolognesi and Heeringa, 2002), Norwegian (Gooskens and Heeringa, 2004), German (Nerbonne and Siedle, 2005), American English (Nerbonne, 2015b), and Bulgarian (Osenova et al., 2009).

Edit distance, also known as *Levenshtein distance* (Levenshtein, 1965), is a measure of how distant (or dissimilar) two strings are by computing the minimal number of substitutions, insertions, and deletions needed to transform one string into the other. At the same time, it can be used to align two strings. For example, given two pronunciations of the Dutch word *hart* – [hɑrt] and [ærtə] – the algorithm produces the following alignment:

h	ɑ	r	t	
	æ	r	t	ə
1	1			1

In order to transform [hɑrt] into [ærtə] 3 edit operations are needed: the [h] has to be deleted, the [ɑ] replaced by [æ], and [ə] has to be inserted. If all operations are assigned cost 1, the total distance between these two strings is 3.

Kessler (1995) introduced the use of edit distance in dialectometry by successfully utilizing it to compare phonetic transcriptions of Gaelic Irish and compute linguistic distance between each pair of sites under investigation. In his approach, the strings are compared by simply counting the number of mismatching segments in phonetic transcriptions, i.e., assigning all operations the cost of 1 as in the example. This simple technique is called *phone string comparison*,

and two segments are always counted either as the same or as different, never as partially similar or similar to a certain degree. For example, in this simple approach [e] is considered equally different from [ε] and [s]. In a following step, Kessler applied clustering to the calculated distances and automatically obtained dialect boundaries that corresponded fairly well to those in traditional scholarship.

The most comprehensive study of the application of edit distance in dialectometry can be found in Heeringa (2004). In order to take into account the varying dissimilarity between different segments, Heeringa examines feature representations of segments, where each segment is represented as a bundle of phonetic features, but also acoustic segment representations, using canonical spectograms. The results showed that the more discriminating feature and acoustic representations, although linguistically more intuitive, do not lead to better results in the task of dialect classification. When compared to perceptual studies of dialects, the use of binary costs (0/1) (in the phone string comparison) actually outperforms distances obtained using gradual costs of segments.

In his work on automatic sequence alignment and cognate detection, List (2012a) groups sounds into predefined classes in order to reduce the number of phonetic segments and achieve more efficient and accurate string comparison. The concept of sound classes was introduced to historical linguistics by A. B. Dolgopolsky (1986), who grouped sounds into different classes based on observations of sound frequency correspondences. The main idea is that sounds within one sound class correspond more often with each other than with sounds from other sound classes. In List's approach transition probabilities between sound classes are automatically computed from the data and then employed by the alignment algorithm. Sound classes have been successfully applied in a couple of approaches, including phonetic alignment (List, 2012b), and automatic cognate detection (List, 2012a). The method is implemented as a part of the LingPy library, an open-source software package for quantitative historical linguistics (List et al., 2017).

4.1 Automatically Induced Segment Distances

Wieling et al. (2009) proposed a data-driven solution to segment comparison in edit distance algorithm. They collect the frequencies of the segments aligned by the phone string comparison in a large contingency table. They then weighted the edit operation costs using (an inverse of) *pointwise mutual information* (PMI) with edit distance algorithm in order to let operation costs reflect the likelihood of two segments being involved in a substituion. PMI is an information-theoretic association measure that estimates the amount of information one event tells us about the other. Applied to the phonetic transcriptions of words, PMI can gauge how similar or distant two phones are; the more often they are aligned, the bigger the similarity between them and vice versa. The procedure of calculating the association strength between the phones and improving the alignments at the same time is iterative: a) all word transcriptions are initially aligned using the binary (0/1) Levenshtein algorithm; b) from the

obtained alignments, for all pairs of phone segments PMI association values are calculated; c) all word transcriptions are aligned once more using Levenshstein algorithm, but based on the phone distances generated in the previous step; and d) steps b and c are repeated until there are no further changes in phone distances and alignments.

Wieling et al. (2009) showed that Levenshtein PMI produces more correct alignments when compared to the simple Levenshtein algorithm and to two slightly modified versions of this algorithm. Wieling et al. (2012) found that PMI induced segment distances correlate well with acoustic distances in formant space ($0.61 < r < 0.76$) measured on six dialect data sets.

A similar procedure has been used in computational historical linguistics by Gerhard Jäger (2013, 2015) to infer language phylogenies from word lists using PMI-weighted alignments. Naturally, the concept of string similarity in historical linguistics is to be understood as relative to the sound changes that related languages have undergone. This need not be the same as the global similarity that one sometimes perceives among dialects. By discounting the substitution of sounds related by regular sound change, PMI alignments reflect the results of those changes. Jäger (2013) has shown that applied to 5644 word lists taken from the Automated Similarity Judgment Project database (Wichmann et al., 2012), the PMI-weighted alignment improves the accuracy of phylogenetic inference in comparison to plain Levenshtein-based alignments by 1 to 3 percent, depending on the evaluation method used. Applied to a collection of approximately 1000 Eurasian languages and dialects, PMI-weighted alignments, combined with phylogenetic inference, produced a classification in excellent agreement with established findings of historical linguistics as well as strong statistical support for several putative macrofamilies (Jäger, 2015).

4.2 Some Criticisms

Although successful in detecting dialect groups at the aggregate level, edit distance has often been criticized as linguistically naive and too simple. The criticism arises because linguists are interested not only in determining main dialect areas, but also in discovering linguistic details behind those divisions (see Sec. 5.5). An aggregate approach based on simple edit distance fails to provide such details and thus fails to shed light on the linguistic processes that lead to the observed dialect divisions. In order to overcome the simplicity of the edit distance approach, several advances in dialectometry have been put forward in the past decade. These include a technique to automatically induce a more sensitive measure between individual sound segments in phonetic transcriptions (see Sec. 5.4.1), and a method that enables researchers to identify characteristic features of dialect areas.

Another point of criticism of edit distance as applied in dialectometry is that it is based on manually prepared phonetic transcriptions, often extracted from traditional dialect atlases. These manually prepared data are on the one hand expensive to acquire and on the other hand do not represent natural speech but data already selected and transcribed by linguists. Further research in acoustic

phonetics but also in articulography promises to overcome the problem of relying too heavily on phonetic transcriptions (see Section 5.6.3).

4.3 Distinctive Elements

In order to identify distinctive characteristics of dialect areas, Prokić et al. (2012) suggested a method inspired by Fisher's linear discriminant that seeks features that differ little within the group in question and a great deal outside that group. The method starts from a number of sites already grouped into two or more dialects and examines one candidate group at a time, seeking features that characterize it best in distinction to elements outside the group. It does not assume that the groups were obtained via clustering, only that candidate groups have somehow been identified. This method, in comparison to others proposed earlier (see remarks in Section 5.5 on multidimensional scaling), is general and it can be used with any type of dialect charateristic – word, segment, or syntactic construction – as long as one can define a numerical difference metric between the features. Prokić et al. (2012) tested the method on German and Dutch dialect sets that were first analyzed by means of the Levenshtein algorithm to calculate distances between each pair of sites and Ward's clustering algorithm to detect main dialect areas. The results show that the words identified successfully distinguish the group of dialects in focus from the rest.

Even if we are able to detect which individual variables are most characteristic for a given variety, the scientific challenge of characterizing these remains. We do not suppose that dialect speakers monitor all the thousands of potential variables individually. In keeping with the quantitative spirit of computational dialectology, we focus here on dimension reduction techniques that are used in the attempt to detect commonalities among these variables. We refrain from an explanation of alternative dimension reduction techniques, referring the reader to Tabachnick and Fidell (2012) for detailed information at a practical level.

Shackleton (2010) applies principal component analysis (PCA) to the a site × variable matrix, where his matrix contained not aggregates of differences but rather, variable values.[3] This allowed him to show, e.g., that certain vowel shifts patterned similarly across English dialects and, in particular, that vowel shifts involved several vowels. This is, of course, scientifically more satisfying than simply noting where patters seem to overlap. The commonalities (the PCs) may be regarded as latent variables that are shared across varieties.

Nerbonne (2006) applied factor analysis, a technique similar to PCA, to site × vowel matrices, where the vowels were characterized by vectors of feature values. The results were similar to Shackleton's, but Nerbonne's analysis really succeeded only in showing that some vowel tokens (for example, the reduced vowel in the the English plural morpheme {əz}) behaved in a common fashion, suggesting that a single sound (phoneme) varied from one set of varieties to the next. This might be seen as an indication that the phoneme was playing a role in variation, which is encouraging, but less insightful than linguists would like.

---

[3]Labov et al. (2006b) had used PCA on the acoustic values of twenty-one vowels in American English.

The search for compact and insightful characterizations of dialect differences is essential for the scientific study of dialect and worthy of much more attention.

5   Geography of Distributions

Let's return to the research question on the location of dialects, which, as we noted, was answered wisely in traditional dialectology, but in a way that left the justification for the answer often unclear. Séguy's step to examine large aggregates of dialectal material brought clarity into the basis of dialectal division, which Goebl's clustering elaborated on algorithmically to extract relatively similar groups of collection sites. Subsequent work examined alternative clustering algorithms and various means of making results more reliable (see Section 5.3 above).

From there it was a small step to project those collections of sites onto maps, with the gratifying result that they often projected to (nearly) coherent dialect regions, which, moreover, normally differed little from those that experts had outlined earlier. There was a difference, naturally, in the explicitness of the methods, and in the efficiency with which analyses could be conducted.

An alternative to clustering arose when Embleton (1993) suggested the use of *multi-dimensional scaling* (MDS) as a means of analyzing the site × site matrix of dialect distances (in fact, since distances are symmetric, half matrices suffice). MDS (Kruskal and Wish, 1978) was originally developed for psychometric applications, and it inputs a half matrix of distances and outputs coordinates of the sites in a low-dimensional space from which the original distances can be derived (approximately). The number of dimensions need not be specified in advance, and existing software normally produces solutions in one, two, and more dimensions. The quality of the solution is specified via a loss function known as *strain*, or alternatively, as the correlation coefficient of the scaled distances with those in the input. As might be expected, strain falls as the correlation rises.

Nerbonne et al. (1999) took advantage of the fact that three-dimensional solutions accounted for 80-90 percent of the variance in the dialect material they studied, in order to take the the further step of interpreting coordinates chromatically, mapping the three coordinates to red, green and blue (RGB). Each site was then colored using each color to the degree of its coordinate in the three-dimensional MDS solution. Figure 1 shows the results of applying MDS to the results of an analysis of Dutch pronunciations (see Heeringa (2004) for details on the data set).

With respect to our second research question, how are varieties geographically distributed, the MDS analysis is noteworthy for two reasons. First, unlike cluster analysis, MDS does not assume that varieties are aligned with discrete sets of data collection sites that then project to convex (and exhaustive) dialect areas. Instead, *dialect continua* can likewise be detected. Leinonen (2010) analyzes Swedish varieties as a continuum, where areas play no role. Second, when we examine the MDS coordinates of data collection sites, these are distributed continuously, not categorically, so that "borders" between areas of relative con-
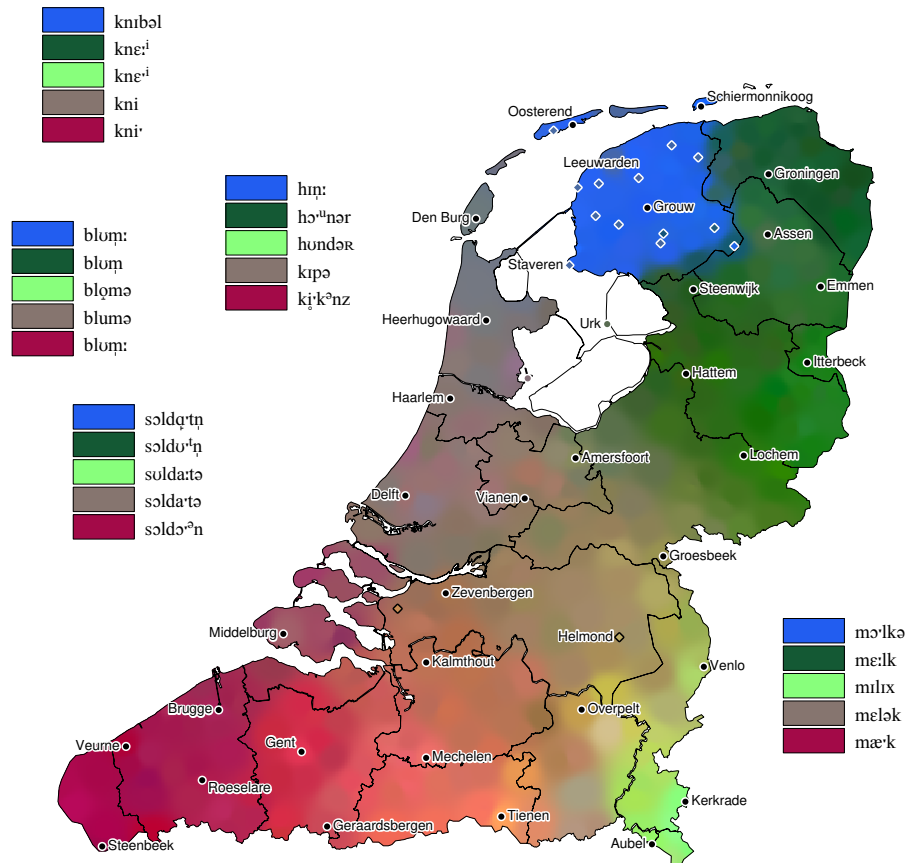
Figure 1: Heeringa (2004) first compared the pronunciations of 125 words at each pair of 360 sites, aggregated the site × site differences, and analyzed those results via MDS. The first three MDS dimensions were then interpreted chromatically (see text). The legends (added later) indicate the pronunciations of some indicative words at sites with the given color.

centration are typically gradual. MDS analyses thus do away with dialect maps showing areas, which often unfortunately suggest that a traveller might move from one dialect area to another in a single step. In general this is found congenial among dialectologists.

It is important to attempt to interpret the MDS dimensions. Naturally, the maps described constitute one dimension of interpretation, the geographic one, but linguists are likewise passionately interested in the linguistic basis of the differences. The legends in Figure 1 were obtained by examining which words correlate most strongly with each MDS dimension, yielding indications of the linguistic basis. But words – taken as sequences of sounds (phonemes) – are rougher indications than most linguists want. In the analysis behind Figure 1 entire word pronunciations served as the input to the (edit distance) analysis, so finer indications are not possible.

Pickl (2013) suggests using *factor analysis* (FA) rather than MDS (see Tabachnick and Fidell (2012) for detailed information at a practical level on the different dimension-reducing techniques), and distinguishes twenty factors, whose geographic projections he also examines. Most of these are nearly negligible (≤ 1 percent) in explained variance, but the attempt to go beyond the usual aggregate analysis in dialectometry should not be the last. See Nerbonne (2015a) for more discussion of Pickl's work.

5.1 Geostatistics and Topography

Jack Grieve has pioneered the effort to include a geostatistical perspective on dialect distributions (Grieve et al., 2011; Grieve, 2018). *Geostatistics* is a well-developed field motivated not only by geography proper, but also by the many other fields where geography plays an important role, e.g. epidemiology, geology, ecology and environmental protection. It has studied various sorts of geographic sampling regimes as well as interpolations, and it is perhaps surprising that dialectology was pursued for so long without attending to this major potential source of further insight. Grieve's work is therefore important for bridging the gap.

Burridge (2017) is intrigued by the parallelism between dialect distributions on the one hand and borders induced by *surface tension* on the other. Burridge shows, e.g., how border (coastal) features such as estuaries tend to engender perpendicular boundaries due to the relative likelihood of social contact (and thus dialectal experience) on the one side of the boundary as opposed to the other. He thus derives predictions about dialect boundaries from physical properties of the space where they are spoken. In more elaborate models the influence of population centers is also studied.

5.2 Other Relations

By providing numeric measures of the overall differences between varieties, dialectometry has stimulated various studies on the relations between language variety and other communal properties of human beings.

Manni et al. (2006) investigate whether there is link between genetics and language variation. The authors naturally do not suppose that language users inherit linguistic tendencies genetically, but were interested instead in the question of whether dialects might not persist in families due to the predominance of language learning in the family. It turned out that language variation and genetic make-up – both measured at the level of the municipality – correlated moderately ($r = 0.42$), but that the relationship disappeared when one controlled for the effect of geography, with which both correlate independently. See Manni et al. for further references on this topic.

Falck et al. (2012) investigate the potential influence of dialect (or perhaps cultural identification) on mobility. Their study focuses on mobility, where it was known that people tend to move to places close by – presumably to remain near friends, family, and well-known surroundings. After controlling for this factor, the authors were able to show that, when people move house, they also tend to remain within their dialect area, where the patterns of speech were also familiar.

Given the numeric turn throughout the humanities and social sciences, many further relations could be promising foci of investigation.

6 Validation

Computational linguistics has emphasized the need to validate putative means of measuring linguistic features and linguistic differences, so it is perhaps not surprising that early work by computational linguists was the first dialectological work to seek validation for its efforts (Heeringa et al., 2002). As it was early on in computational dialectology, Heeringa et al. (2002) measured the degree to which their clustering results jibed with earlier scholarship in the well-studied case of Dutch dialectology.

Wieling et al. (2009) evaluated several versions of edit distance on the basis of their ability to align phonetic transcriptions (as well as human experts). This is a sensitive task that allowed authors to tease apart the performance of rather similar approaches.

6.1 Human Judgments of Similarity

As we noted, clustering can be unstable, making it a poor partner in validation efforts. And while comparing one's results to earlier scholarship is indispensable, one does not wish to canonize the earlier work to the degree that computational work can never improve on it. Gooskens and Heeringa (2004) meet both of these objections by comparing edit distance measures of pronunciation differences to dialect speakers' judgments of how similar dialect speech sounded to their own. It was sensible to ask the naive subjects only how similar the speech samples sound to their *own* speech since they had the most experience with this, and since one assumes that dialects function primarily as emblems of geographic identity. Gooskens and Heeringa (2004) also judiciously chose Norway for their

experiment, which recommended itself for Norwegians' very accepting attitudes toward dialectal speech. While dialects and accents are frowned upon in many places, Norwegians normally retain their accents even in public speech. The authors were able to show that the computational measure correlated well with speakers' judgments of dissimilarity ($r = 0.67$).

Wieling et al. (2014b) used the more sensitive PMI-based edit distance to measure the strength of foreign accents in American English in the Speech Accent Archive hosted at George Mason University (Weinberger and Kunath, 2011). Although the degree of difference in foreign accents is perhaps not perceived in exactly the same way as the degree of difference in dialect pronunciations, it is likely to be comparable. The subjects in Wieling et al.'s experiment were predominantly linguistically informed listeners, as they had responded to an appeal in the Language Log (http://languagelog.ldc.upenn.edu) to participate as subjects. So these subjects were more discriminating than the untrained participants in Gooskens and Heeringa's (2004) work. Wieling et al. applied a logarithmic transformation to the PMI edit distance in keeping with psychometric custom and obtained a high correlation ($r = 0.81$), which was nearly as high as human performance, measured through the average correlation of individual raters' judgments ($r = 0.84$).

6.2 Acoustic Characterizations

A great deal of the work reported on thus far in this chapter involved the automatic analysis of phonetic transcriptions recorded in large dialect atlas projects. In many cases this is the only option available, as sensitive acoustic recordings of the material were simply never made. Tape recorders were unwieldy until the second half of the twentieth century and were for that reason not used on the extensive field trips necessary for data collection.

But many dialectologists are skeptical about the reliance on phonetic transcription, which can vary from transcriber to transcriber. For this reason, but also in order to pursue the computational lines sketched here, it has been important to follow up the transcription-based work with work that does not rely on transcription. Leinonen (2010) therefore applied dialectometric techniques to the SweDia database of dialect recordings (Eriksson, 2004a,b). She first extracted the formants of all the twenty Swedish vowels of the more than thousand speakers using the mean of those formant distances as measures of pronunciation distance. The results were most satisfying, both in confirming the Swedish dialect continuum as well as in demonstrating how dialects have been losing distinctions between speakers forty years apart in age (Leinonen, 2010, Figure 7.8, p 140).

It would clearly be desirable to experiment with more acoustic analyses, but the proper acoustic analysis of consonants is still very difficult (Thomas, 2018).

6.3 Articulography

Instead of proceeding from transcriptions (see Section 4), or an acoustic characterization of dialects (see Section 5.6.2), a new direction of research in dialectology is focusing on the underlying movements of the tongue and lips. While the articulation obviously results in the acoustic speech signal, it is not straightforward to infer the precise articulatory trajectories of tongue and lips on the basis of an acoustic analysis. For example, while the first and second formant are frequently assumed to model height and backness of the tongue, this relationship is far from perfect (Rosner and Pickering, 1994; Wieling et al., 2016). Consequently, direct articulatory measurements, such as those that can be made through, e.g., electropalatography, ultrasound tongue imaging, and electromagnetic articulography are sometimes used to study dialect or sociolinguistic variation.

Electropalatography uses a custom-made artificial palate with electrodes, which is able to detect when and where the tongue touches the palate during speech. The drawback of this approach is that each participant requires their own custom-made artificial palate, therefore making this type of research relatively slow and expensive. In addition the position of the tongue when it is not touching the palate is unknown. Ultrasound tongue imaging, by contrast, uses an ultrasound probe positioned below the chin by which a series of images over time are obtained visualizing the shape of the tongue surface. While it is relatively easy to collect data using this approach, the analysis is more complex as image analysis is necessary to process the resulting (grainy) images. Electromagnetic articulography, as a third technique, tracks the movement over time of small sensors attached to the tongue and lips. The advantage of this approach compared to ultrasound tongue imaging is that the analysis is more straightforward (i.e., the positional information is available for each sensor over time), but the disadvantage is that it is more invasive and labor intensive to collect these types of data.

Although these articulatory studies require more effort than simply recording the acoustic signal, there have been some studies – mostly including only a limited number of speakers – investigating dialect or sociolinguistic variation. For example, Recasens and Espinosa (2007) studied articulatory differences using electropalatography between two Catalan dialects in a sample of ten speakers and found clear articulatory differences regarding the fricatives. Lawson et al. (2011) used ultrasound tongue imaging in a sample of fifteen speakers to show that the /r/ pronunciation in Scottish English was socially stratified. Middle-class speakers generally using bunched articulations, while working-class speakers more frequently used tongue-tip raised variants. More recently, Wieling et al. (2016) used electromagnetic articulography in a sample of forty speakers to show that across the pronunciation of about 100 words, speakers from a dialect in the north of the Netherlands had a more posterior tongue position than those from a more southern dialect.

Particularly, the study of Wieling et al. (2016) continues themes from dialectometry. First by providing aggregate results over a large set of words, rather

than focusing on individual sound contrasts. But also due to using a very large data set (due to collecting data over time at a high sampling rate, usually more than 100 Hz) and by using replicable analysis techniques that have also been employed in dialectometry. For example, Wieling et al. (2016) use a non-linear regression technique, generalized additive modeling, which has been employed in dialectometry to model the non-linear geographical patterns of dialect variation (Wieling et al., 2011, 2014a).

7 Emerging Opportunities and Issues

We have tried to indicate emerging opportunities and issues in this chapter as we proceeded, but we shall summarize them here and also suggest other areas that we have not treated in depth.

First, we have concentrated on the areas of dialectology that have attracted the most attention from computational linguists. We have therefore said little about morphology and syntax, even though these have attracted some work (Heeringa and Prokić, 2018) and could definitely profit from more.

Second, although we described Séguy's and Goebl's efforts at handling linguistic data at a categorical level, we have not elaborated on the opportunities for extracting such data (e.g., signals of lexical relationships) at length, even though some have been shown to be promising (Nerbonne and Kleiweg, 2007; Szmrecsanyi, 2008) and deserve further attention.

Third, the role of linguistic frequency in detecting affinities is worth more attention. Goebl (1982) has long championed an inverse frequency weighting, regarding similar (or identical) linguistic items that occur infrequently as especially strong evidence of dialect affinity – just as historical linguists also tend to emphasize the significance of unusual shared innovations, albeit in a non-quantitative fashion. Others recommend discounting infrequent linguistic items entirely, regarding them as potential "noise" (Carver, 1987). Nerbonne and Kleiweg (2007) provide evidence for Goebl's position, and Wieling and Montemagni (2015) show that removing infrequent data from analyses worsens their quality, but much too little work has been devoted to this question.

Fourth, the excellent initiatives of Grieve and Burridge presented in Section 5.1 clearly deserve further attention. In particular, Burridge promises an explanation of why dialect divisions have the geographic contours they do. It would be interesting to know how much dialect variation can be explained along these lines.

Fifth, as we hinted in Section 2, geographic variation is part of a palette of variation types. In spite of some efforts to apply dialectometric techniques to the sociolinguistic question of dialect change (Nerbonne et al., 2013; Valls et al., 2013), there has been little resonance. While this may be related to sociolinguists' focus on individual changes in progress, some work has been at pains to indicate how the aggregate perspective may enrich the usual sociolinguistic one (Wieling et al., 2018). We suspect that computational efforts will likely be fruitful in the study of diastratic variation as well.

Sixth, but most interesting to linguists, would be improvements in the linguistic characterization of the dialect differences that dialectometry so reliably unearths. To what extent can these be subsumed under general characterizations of the sort favored in dialect handbooks? Are there perhaps machine learning techniques that might be brought to bear?

8 Conclusions

After identifying the essential research questions in dialectology – namely what's varying and where – this article reviewed computational work to date in dialectology, indicating where progress has been made vis-à-vis those questions. Measures of linguistic proximity have served as the basis for dialectolometry, where notably one was derived from computer science – namely edit distance. Dialectometry has added a number of tools for the investigation of the geographic distribution of linguistic variation, in particular, clustering and multidimensional scaling, and more advanced work is probing geostatistics as well as models inspired by the mathematics of surface tension. Dialectometry has played an important role in enabling studies of the relations between language variation and both genetics and mobility.

We also reviewed efforts at characterizing the linguistic differences among varieties. Computational work has certainly provided means of cataloging such differences exhaustively, and in identifying characteristic elements. We would prefer to have ways of characterizing the differences more insightfully and more concisely, however, and the success in this direction has been modest. We also reviewed efforts at validating the more popular computational measures, a consideration that was certainly inspired by work in computational linguistics. This section closed with remarks on how textually focused work is now branching into the acoustics and the articulation of dialect speech, a further potentially important sort of validation.

We closed with a list of six issues where further progress would be valuable, and where prospects seem promising.

References

Bailey, Guy. 2018. Field interviews in dialectology. Pages 284–299 of: Boberg, Charles, Nerbonne, John, and Watt, Dominic (eds.), *The Handbook of Dialectology.* Hoboken, NJ: Wiley-Blackwell.

Barbiers, Sjef. 2009. Locus and limits of syntactic microvariation. *Lingua*, **119**(11), 1607–1623.

Bolognesi, Roberto, and Heeringa, Wilbert. 2002. De invloed van dominante talen op het lexicon en de fonologie van Sardische dialecten. Pages 45–84 of: Bakker, Dik, Sanders, Ted, Schoonen, Rob, and van der Wijst, Per (eds.),

Gramma/TTT: Tijdschrift voor taalwetenschap. Nijmegen: Nijmegen University Press.

Burridge, James. 2017. Spatial evolution of human dialects. *Physical Review X*, **7**(3), 031008.

Carver, Craig M. 1987. *American Regional Dialects: A Word Geography.* Ann Arbor: University of Michigan Press.

Chambers, J. K. 2018. Written dialect surveys. Pages 268–283 of: Boberg, Charles, Nerbonne, John, and Watt, Dominic (eds.), *The Handbook of Dialectology.* Hoboken, NJ: Wiley-Blackwell.

Dolgopolsky, A. B. 1986. A probabilistic hypothesis concerning the oldest relationships among the language families of northern Eurasia. Pages 27–50 of: Shevoroshkin, Vitalij V. (ed), *Typology, Relationship and Time.* Ann Arbor: Karoma Publisher.

Elspaß, Stephan, and König, Werner (eds.), 2008. *Sprachgeographie digital. Die neue Generation der Sprachatlanten.* Berlin: Walter de Gruyter GmbH & Co. KG.

Embleton, Sheila. 1993. Multidimensional scaling as a dialectometrical technique: Outline of a research project. Pages 267–276 of: Köhler, Reinhard, and Rieger, Burghard (eds.), *Contributions to Quantitative Linguistics.* Dordrecht: Kluwer.

Eriksson, Anders. 2004a. *SweDia 2000: A Swedish Dialect Database.* Department of Linguistics/Institutionen för lingvistik.

Eriksson, Anders. 2004b. SweDia-projektet: dialektforskning i ett jämförande perspektiv. *Folkmålsstudier*, **43**, 11–32.

Falck, Oliver, Heblich, Stephan, Lameli, Alfred, and Südekum, Jens. 2012. Dialects, cultural identity, and economic exchange. *Journal of Urban Economics*, **72**(2-3), 225–239.

Goebl, Hans. 1982. *Dialektometrie: Prinzipien und Methoden des Einsatzes der numerischen Taxonomie im Bereich der Dialektgeographie.* Wien: Österreichische Akademie der Wissenschaften.

Goldman, Jean-Philippe, Leemann, Adrian, Kolly, Marie-José, Hove, Ingrid, Almajai, Ibrahim, Dellwo, Volker, and Moran, Steven. 2014. A crowd-sourcing smartphone application for Swiss German: Putting language documentation in the hands of the users. Pages 3444–3447 of: Calzolari, Nicoletta et al. (ed), *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14).* Reykjavik, Iceland: European Language Resources Association (ELRA).

Gooskens, Charlotte, and Heeringa, Wilbert. 2004. Perceptive evaluation of Levenshtein dialect distance measurements using Norwegian dialect data. *Language Variation and Change*, **16**(3), 189–207.

Grieve, Jack. 2018. Spatial statistics for dialectology. Pages 415–434 of: Boberg, Charles, Nerbonne, John, and Watt, Dominic (eds.), *The Handbook of Dialectology*. Hoboken, NJ: John Wiley & Sons.

Grieve, Jack, Speelman, Dirk, and Geeraerts, Dirk. 2011. A statistical method for the identification and aggregation of regional linguistic variation. *Language Variation and Change*, **23**(2), 193–221.

Heeringa, Wilbert. 2004. Measuring Dialect Pronunciation Differences Using Levenshtein Distance. PhD thesis, University of Groningen, Groningen.

Heeringa, Wilbert, and Nerbonne, John. 2006. De analyse van taalvariatie in het Nederlandse dialectgebied: Methoden en resultaten op basis van lexicon en uitspraak. *Nederlandse Taalkunde*, **11**(3), 218–25.

Heeringa, Wilbert, and Prokić, Jelena. 2018. Computational dialectology. Pages 330–347 of: Boberg, Charles, Nerbonne, John, and Watts, Dominic (eds.), *The Handbook of Dialectology*. Hoboken, NJ: Wiley-Blackwell.

Heeringa, Wilbert, Nerbonne, John, and Kleiweg, Peter. 2002. Validating dialect comparison methods. Pages 445–452 of: Gaul, Wolgang, and Ritter, Gerd (eds.), *Classification, automation, and new media. Proceedings of the 24th conf. of the Gesellschaft für Klassifikation*. Heidelberg: Springer.

Hinskens, Frans. 2018. Dialectology and formal linguistic theory: The blind man and the lame. Pages 88–105 of: Boberg, Charles, Nerbonne, John, and Watt, Dominic (eds.), *The Handbook of Dialectology*. Hoboken, NJ: Wiley-Blackwell.

Jäger, Gerhard. 2013. Phylogenetic inference from word lists using weighted alignment with empirical determined weights. *Language Dynamics and Change*, **3**(2), 245–291.

Jäger, Gerhard. 2015. Support for linguistic macrofamilies from weighted alignment. *Proceedings of the National Academy of Sciences of the United States of America*, **112**(41), 12752–12757.

Kessler, Brett. 1995. Computational dialectology in Irish Gaelic. Pages 60–66 of: *Proceedings of the 7th Conference of the European Chapter of the ACL*. San Francisco: Morgan Kaufmann.

Kleinberg, Jon M. 2004. An impossibility theorem for clustering. Pages 463–470 of: Thrun, Sebastian, and Saul, Lawrence K.and Schölkop, Bernhard (eds.), *Advances in Neural Information Processing Systems 16: Proceedings of the 2003 Conference (NIPS 16)*. Cambridge, MA: MIT Press.

Kretzschmar, William A. 2009. *The Linguistics of Speech*. New York: Cambridge University Press.

Kretzschmar Jr, William A, McDavid, Virginia G, Lerud, Theodore K, and Johnson, Ellen. 1993. *Handbook of the Linguistic Atlas of the Middle and South Atlantic States*. Chicago: University of Chicago Press.

Kruskal, Joseph B., and Wish, Myron. 1978. *Multidimensional Scaling*. Newbury Park, London, New Delhi: Sage Publications.

Labov, William, Ash, Sharon, and Boberg, Charles. 2006a. *The Atlas of North American English: Phonetics, Phonology, and Sound Change*. Berlin, New York: Mouton de Gruyter.

Labov, William, Ash, Sharon, and Boberg, Charles. 2006b. *Atlas of North American English: Phonetics Phonology and Sound Change*. Berlin, New York: Mouton de Gruyter.

Lawson, E., Scobbie, J. M., and Stuart-Smith, J. 2011. The social stratification of tongue shape for postvocalic /r/ in Scottish English. *Journal of Sociolinguistics*, **15**(2), 256–268.

Leinonen, Therese N. 2010. *An Acoustic Analysis of Vowel Pronunciation in Swedish Dialects*. PhD thesis, Rijksuniversiteit Groningen, Groningen.

Levenshtein, V. I. 1965. Dvoičnye kody s ispravleniem vypadenij, vstavok i zameščenij simbolov. *Doklady Akademij Nauk SSSR*, **163**(4), 845–848.

List, Johann-Mattis. 2012a. LexStat: Automatic detection of cognates in multilingual wordlists. Pages 117–125 of: *Proceedings of the EACL 2012 Joint Workshop of LINGVIS & UNCLH*. Shroudburg, PA: Association for Computational Linguistics.

List, Johann-Mattis. 2012b. SCA: Phonetic alignment based on sound classes. Pages 32–51 of: Slavkovik, Marija, and Lassiter, Dan (eds.), *New Directions in Logic, Language, and Computation*. Berlin, Heidelberg: Springer.

List, Johann-Mattis, Greenhill, Simon, and Forkel, Robert. 2017. *LingPy. A Python library for quantitative tasks in historical linguistics*. Software library maintained at https://doi.org/10.5281/zenodo.1065403.

Llamas, Carmen. 2018. The dialect questionnaire. Pages 253–267 of: Boberg, Charles, Nerbonne, John, and Watt, Dominic (eds.), *The Handbook of Dialectology*. Hoboken, NJ: Wiley-Blackwell.

Macauley, Ronald. 2018. Dialect sampling methods. Pages 241–252 of: Boberg, Charles, Nerbonne, John, and Watt, Dominic (eds.), *The Handbook of Dialectology*. Hoboken, NJ: Wiley-Blackwell.

Manni, Franz, Heeringa, Wilbert, and Nerbonne, John. 2006. To what extent are surnames words? Comparing geographic patterns of surnames and dialect variation in the Netherlands. *Literary and Linguistic Computing*, **21**(4), 507–528.

Möller, Robert, and Elspaß, Stephan. 2008. Erhebung dialektgeographischer Daten per Internet: ein Atlasprojekt zur deutschen Alltagssprache. Pages 115–132 of: Elspaß, Stephan, and König, Werner (eds.), *Sprachgeographie digital. Die neue Generation der Sprachatlanten (mit 80 Karten)*. Hildesheim, Zürich, New York: Olms.

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

Nerbonne, John. 2009. Data-driven dialectology. *Language and Linguistics Compass*, **3**(1), 175–198.

Nerbonne, John. 2015a. Simon Pickl. 2013. Probabilistische Geolinguistik. Geostatistische Analysen lexikalischer Variation in Bayerisch-Schwaben. *Zeitschrift für Rezensionen zur germanistischen Sprachwissenschaft*, **7**(1-2), 124–129.

Nerbonne, John. 2015b. Various variation aggregates in the LAMSAS south. Pages 369–382 of: Picone, Michael, and Davis, Catherine (eds.), *New Perspectives on Language Variety in the South: Historical and Contemporary Approaches. (LAVIS III)*. Tuscaloosa: University of Alabama Press.

Nerbonne, John. 2018. Vaulting ambition. Pages 445–468 of: *Grammar and corpora*. Heidelberg: Heidelberg University Publishing.

Nerbonne, John, Heeringa, Wilbert, and Kleiweg, Peter. 1999. Edit distance and dialect proximity. Pages v–xv of: Sankoff, David, and Kruskal, Joseph (eds.), *Time Warps, String Edits and Macromolecules: The Theory and Practice of Sequence Comparison*. Stanford, CA: CSLI Press.

Nerbonne, John, Heeringa, Wilbert, van den Hout, Eric, van de Kooi, Peter, Otten, Simone, and van de Vis, Willem. 1996. Phonetic distances between Dutch dialects. Pages 185–202 of: Durieux, Gert, Daelemans, Walter, and Gillis, Steven (eds.), *CLIN VI: Proceedings of the Sixth CLIN Meeting*. Antwerp: Centre for Dutch Language and Speech (UIA).

Nerbonne, John, and Kleiweg, Peter. 2007. Toward a Dialectological Yardstick. *Journal of Quantitative Linguistics*, **14**(2), 148–166.

Nerbonne, John, Kleiweg, Peter, Manni, Franz, and Heeringa, Wilbert. 2008. Projecting dialect differences to geography: Bootstrapping clustering vs. clustering with noise. Pages 647–654 of: Preisach, Christine, Schmidt-Thieme, Lars, Burkhardt, Hans, and Decker, Reinhold (eds.), *Data Analysis, Machine Learning, and Applications. 31st Annual Meeting of the German Classification Society*. Berlin: Springer.

Nerbonne, John, and Siedle, Christine. 2005. Dialektklassifikation auf der Grundlage aggregierter Ausspracheunterschiede. *Zeitschrift für Dialektologie und Linguistik*, **72**(2), 129–147.

Nerbonne, John, van Ommen, Sandrien, Gooskens, Charlotte, and Wieling, Martijn. 2013. Measuring socially motivated pronunciation differences. Pages 107–40 of: Borin, Lars, and Saxena, Anju (eds.), *Approaches to Measuring Linguistic Differences*. Berlin: De Gruyter.

Ng, Andrew Y, and Jordan, Michael I. 2002. On discriminative vs. generative classifiers: A comparison of logistic regression and naive Bayes. Pages 841–848 of: Becker, Susanna, Thrun Sebastian, and Obermayer, Klaus (eds.), *Advances in Neural Information Processing Systems: Proceedings of the 2002 Conference (Vol. 15)*, vol. 14. Cambridge, MA: MIT Press.

Osenova, Petya, Heeringa, Wilbert, and Nerbonne, John. 2009. A quantitative analysis of Bulgarian dialect pronunciation. *Zeitschrift für Slavische Philologie*, **66**(2), 425–458.

Pickl, Simon. 2013. *Probabilistische Geolinguistik: Geostatistische Analysen lexikalischer Variation in Bayerisch-Schwaben*. Stuttgart: Steiner.

Prokić, Jelena, Çöltekin, Çağri, and Nerbonne, John. 2012. Detecting shibboleths. Pages 72–80 of: *Proceedings of the EACL 2012 Joint Workshop of LINGVIS & UNCLH*. Shroudsburg, PA: Association for Computational Linguistics.

Prokić, Jelena, and Nerbonne, John. 2008. Recognizing groups among dialects. *International Journal of Humanities and Arts Computing*, **2**, 153–171.

Recasens, Daniel, and Espinosa, Aina. 2007. An electropalatographic and acoustic study of affricates and fricatives in two Catalan dialects. *Journal of the International Phonetic Association*, **37**(2), 143–172.

Rosner, Burton S, and Pickering, John B. 1994. *Vowel Perception and Production*. Oxford University Press.

Scherrer, Yves, Leemann, Adrian, Kolly, Marie-José, and Werlen, Iwar. 2012. *Dialäkt Äpp – A smartphone application for Swiss German dialects with great scientific potential*. Lecture at 7th Congress of the International Society for Dialectology and Geolinguistics (SIDG Congress-Dialect 2.0). Available at https://archive-ouverte.unige.ch/unige:22817.

Shackleton, Robert. 2010. Quantitative assessment of English-American speech relationships. PhD thesis, Rijksuniversiteit of Groningen.

Szmrecsanyi, Benedikt. 2008. Corpus-based dialectometry: Aggregate morphosyntactic variability in British English dialects. *International Journal of Humanities and Arts Computing*, **2**(1–2), 279–296.

Tabachnick, Barbara G, and Fidell, Linda S. 2012. *Using Multivariate Statistics*. Boston/Munich: Pearson.

Thomas, Erik R. 2018. Acoustic phonetic dialectlogy. Pages 314–329 of: Boberg, Charles, Nerbonne, John, and Watt, Dominic (eds.), *The Handbook of Dialectology*. Hoboken, NJ: Wiley-Blackwell.

Valls, Esteve, Wieling, Martijn, and Nerbonne, John. 2013. Linguistic advergence and divergence in north-western Catalan: A dialectometric investigation of dialect leveling and border effects. *Literary and Linguistic Computing*, **28**(1), 119–146.

Van den Berg, Boudewijn L. 2003. Phonology & morphology of Dutch & Frisian dialects in 1.1 million transcriptions, Goeman-Taeldeman-Van Reenen project 1980–1995, Meertens Instituut Electronic Publications in Linguistics 3. Meertens Instituut (CD-ROM), Amsterdam.

Weinberger, Steven H, and Kunath, Stephen A. 2011. The speech accent archive: Towards a typology of English accents. *Language and Computers*, **73**(1), 265–281.

Wichmann, Søren, Müller, André, Velupillai, Viveka, Wett, Annkathrin, Brown, Cecil H., Molochieva, Zarina, Bishoffberger, Julia, Holman, Eric W., Sauppe, Sebastian, Brown, Pamela, Bakker, Dik, List, Johann-Mattis, Egorov, Dmitry, Belyaev, Oleg, Urban, Matthias, Hammarström, Harald, Carrizo, Agustina, Mailhammer, Robert, Geyer, Helen, Beck, David, Korovina, Evgenia, Epps, Pattie, Valenzuela, Pilar, and Grant, Anthony. 2012. *The ASJP Database*. Accessible at https://asjp.clld.org/.

Wieling, Martijn, Bloem, Jelke, Mignella, Kaitlin, Timmermeister, Mona, and Nerbonne, John. 2014b. Measuring foreign accent strength in English. Validating Levenshtein distance as a measure. *Language Dynamics and Change*, **4**(2), 253–269.

Wieling, Martijn, Heeringa, Wilbert, and Nerbonne, John. 2007. An aggregate analysis of pronunciation in the Goeman-Taeldeman-van Reenen-Project data. *Taal en Tongval*, **59**(1), 84–116.

Wieling, Martijn, Margaretha, Eliza, and Nerbonne, John. 2012. Inducing a measure of phonetic similarity from pronunciation variation. *Journal of Phonetics*, **40**(2), 307–314.

Wieling, Martijn, and Montemagni, Simonetta. 2015. Infrequent forms: Noise or not? Pages 215–224 of: Côté, Marie-Hélèen, Knooihuizen, Remco, and Nerbonne, John (eds.), *The Future of Dialects*. Berlin: Language Science Press.

Wieling, Martijn, Montemagni, Simonetta, Nerbonne, John, and Baayen, R. Harald. 2014a. Lexical differences between Tuscan dialects and standard

Italian: Accounting for geographic and sociodemographic variation using generalized additive mixed modeling. *Language*, **90**(3), 669–692.

Wieling, Martijn, and Nerbonne, John. 2015. Advances in dialectometry. *Annual Review of Linguistics*, **1**, 243–264.

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

Wieling, Martijn, Prokić, Jelena, and Nerbonne, John. 2009. Evaluating the pairwise string alignment of pronunciations. Pages 26–34 of: *Proceedings of the EACL 2009 Workshop on Language Technology and Resources for Cultural Heritage, Social Sciences, Humanities, and Education (LaTeCH – SHELT&R 2009)*. Shroudsburg, PA: Association for Computational Linguistics.

Wieling, Martijn, Tomaschek, Fabian, Arnold, Denis, Tiede, Mark, Bröker, Franziska, Thiele, Samuel, Wood, Simon N, and Baayen, R Harald. 2016. Investigating dialectal differences using articulography. *Journal of Phonetics*, **59**, 122–143.

Wieling, Martijn, Valls, Esteve, Baayen, R Harald, and Nerbonne, John. 2018. Border effects among Catalan dialects. Pages 71–97 of: Speelman, Dirk, Heylen, Kris, and Geeraerts, Dirk (eds.), *Mixed-Effects Regression Models in Linguistics*. Berlin: Springer.