Journal of Linguistic Geography (2022), **10**, 11–22
doi:10.1017/jlg.2021.12

CAMBRIDGE UNIVERSITY PRESS

Article

Linguistic traits as heritable units? Spatial Bayesian clustering reveals Swiss German dialect regions

Noemi Romano[1], Peter Ranacher[2,4], Sandro Bachmann[3] and Stéphane Joost[1]

[1]Laboratory of Geographic Information Systems (LASIG), School of Architecture, Civil and Environmental Engineering (ENAC), École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland, [2]Department of Geography, University of Zurich (UZH), Zurich, Switzerland, [3]German Department, University of Zurich (UZH), Zurich, Switzerland and [4]URPP Language and Space, University of Zurich (UZH), Zurich, Switzerland

Abstract

In the early 2000s, the SADS, an extensive linguistic atlas project, surveyed more than three thousand individuals across German-speaking Switzerland on over two hundred linguistic variants, capturing the morphosyntactic variation in Swiss German. In this paper, we applied TESS, a Bayesian clustering method from evolutionary biology to the SADS to infer population structure, building on parallels between biology and linguistics that have recently been illustrated theoretically and explored experimentally. We tested three clustering models with different spatial assumptions: a nonspatial model, a spatial trend model with a spatial gradient, and a spatial full-trend model with both a spatial gradient and spatial-autocorrelation. Results reveal five distinct morphosyntactic populations, four of which correspond to traditional Swiss German dialect regions and one of which corresponds to a base population. Moreover, the spatial trend model outperforms the nonspatial model, suggesting a gradual transition of morphosyntax and supporting the idea of a Swiss German *dialect continuum*.

Keywords: dialect continuum; spatial population genetics; geographical clines; morphosyntax; Bayesian clustering

1. Introduction

In dialectology, the neogrammarian tradition of describing a single local dialect with all the relevant features through an *Ortsgrammatik* (i.e., a grammar of a local dialect) was still prevalent in the first half of the twentieth century (Murray, 2010). Around the same time, however, the tradition of dialect geography evolved, where the spatial distribution of linguistic features or lexemes were mapped and interpreted (Schrambke, 2010). Originally, dialect geography focused on mapping single linguistic features or single lexemes one at a time (Girnth, 2010). In the late twentieth century, this approach was criticized and accused of missing general patterns of dialect variation (Wieling et al., 2015). At the same time, language atlases became available, and novel computational methods were developed, which sparked the development of dialectometry, a quantitative branch of dialect geography (Nerbonne & Kretzschmar, Jr., 2013). Dialectometry explores the general spatial variation of dialects using distance or similarity measures rather than single features (Goebl et al., 1984).

Dialectometry was first introduced by Séguy (1973) who analyzed the dialectal variation in the Gascogne in southwest France using the Hamming distance metric (HD) (Hamming, 1950), which captures the co-occurrence of linguistic features at different locations. Later, Goebl (1982) introduced the Weighted Identity Value (WIV), a similarity measure that also accounts for the rareness of a linguistic feature. Both HD and WIV take categorical or nominal features as inputs, which makes them especially well-suited for syntactic and morphological data. For example, HD was used to explore the syntactic variation in Dutch (Spruit et al., 2008) and French dialects (Goebl, 2010). By contrast, Levenshtein distance (LD) takes strings or words as input. LD computes the minimum number of single-character operations (insertions, deletions, or substitutions) required to change one string to another (Levenshtein, 1966). LD was first introduced to dialectometry by Kessler (1995) and was later used to measure phonetic and lexical distances between dialect variants, such as in Dutch (Heeringa, 2004), Norwegian (Gooskens et al., 2004), and Catalan (Valls et al., 2012).

HD, WIV, LD, and other distance and similarity measures gave rise to the use of cluster analysis to explore dialect variation. Cluster analysis partitions a set of objects into similar groups, such that distances within the group are minimized while distances between groups are maximized. Initially, researchers predominately applied hard-clustering methods to dialect data, such as Hierarchical Clustering (Goebl, 2008; Prokić et al., 2008; Scherrer et al., 2016; Szmrecsanyi, 2011) or k-means clustering (Lundberg, 2005). Hard- clustering assigns each object to a single group, generating clear-cut boundaries between groups. Thus, hard-clustering methods miss out on gradual variations in continuous dialect data.

In contrast, dimensionality reduction (DR) aims to reduce the number of variables in a dataset while preserving the variation as much as possible. DR allows for mapping complex patterns in high-dimensional data to low-dimensional space. Methods of dimensionality reduction, such as Multidimensional Scaling, Principal Component Analysis, and Factor Analysis have been applied to

Author for correspondence: Noemi Romano, E-mail: noemi.romano@epfl.ch
Cite this article: Romano N, Ranacher P, Bachmann S, and Joost S. (2022) Linguistic traits as heritable units? Spatial Bayesian clustering reveals Swiss German dialect regions. *Journal of Linguistic Geography* **10**: 11–22, https://doi.org/10.1017/jlg.2021.12

© The Author(s), 2022. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

https://doi.org/10.1017/jlg.2021.12 Published online by Cambridge University Press

explore patterns in dialect data (Embleton, 1993; W. J. Heeringa, 2004; Nerbonne, W. Heeringa, et al., 1999; Pröll et al., 2014; Shackleton Jr., 2005). While DR appropriately reflects dialect continua, the resulting patterns are not always informative, since the remaining dimensions might not fully capture the variation in the data (Wieling et al., 2015). More problematic still, there is no direct way of mapping visual patterns to explicit clusters.

Unlike DR, Bayesian clustering (BC) both captures the gradual variation in dialects and returns explicit clusters instead of mere visualizations. In BC, the assignment of objects to clusters is fuzzy: each object belongs to every cluster with a certain probability. Bayesian clustering yields core regions where objects predominantly belong to a single cluster and gradual boundaries where objects belong to multiple clusters with almost equal probabilities. BC was initially developed in evolutionary biology: the STRUCTURE algorithm aims to infer population structure from individual genotypes and gene frequencies (Pritchard et al., 2000). Genotypes are usually admixed. Due to genetic exchange, individuals carry genetic information from several different populations. Using Markov Chain Monte Carlo sampling (MCMC), STRUCTURE estimates both the unknown populations and the degree to which an individual belongs to each population.

Spatial Bayesian clustering (Durand, Jay, et al., 2009; Chen et al., 2007; Corander et al., 2008, Guillot et al., 2005) is largely based on the Structure algorithm, but, in addition, accounts for spatial trends—the systematic increase or decrease of a variable along a spatial axis, and spatial autocorrelation—the similarity of neighboring objects in space. While BC has been applied to both languages and dialects, including Tasmanian (Bowern, 2012), Sahul (Reesink et al., 2009), Melanesian (Dunn et al., 2008), and Finnish (Syrjänen et al., 2016), spatial Bayesian clustering has not yet been applied to linguistic data. This is surprising, especially given that geography has long been identified as a key factor of dialect variation (Nerbonne and Kleiweg, 2007).

In this article, we contribute to closing this research gap. We apply TESS (Chen et al., 2007; Durand, Jay, et al., 2009), a spatial Bayesian clustering algorithm, to Swiss German dialect data. The data are adopted from the *Syntaktischer Atlas der deutschen Schweiz* (SADS) (Bachmann et al., 2021; Glaser & Bart, 2021). The SADS data comprise syntactical and morphosyntactical constructions in Swiss German, seeking to complement the *Sprachatlas der deutschen Schweiz* (SDS) (Hotzenköcherle, Trüb, et al., 1962–2003), which largely focuses on lexical items as well as phonological and morphological features. Recently, several studies have used quantitative methods to explore both the SDS and SADS data (see Bart et al., 2013; Jeszenszky, Stoeckle, et al., 2017; Kellerhals, 2014; Leemann, 2016; Scherrer et al., 2016; Sibler et al., 2012).

TESS models admixture levels using both global trends and local spatial autocorrelation. We test three different models: (1) a *nonspatial* model, largely equivalent to the original Structure algorithm; (2) a *spatial trend* model, which accounts for global spatial trends in the admixture levels; and (3) a *spatial full-trend* model, which accounts for both global trends and local spatial autocorrelation. We apply all three models to the SADS data, evaluate their performance, and identify the most suitable model for inferring clusters.

2. Dialectal and biological populations

Parallels in biological and linguistic evolution were already noted by Darwin (1859), and many analogies between the two fields have since been made, especially between biology and dialectology. While unrelated or distantly related varieties ("languages") are investigated at a macrogeographic scale, the study of closely related varieties ("dialects") looks into microgeographic areas where speakers can understand each other without significant effort. Related dialect speakers are what is called mutually intelligible and, thus, likely to influence each other and interchange linguistic features (Hock et al., 2009; Reesink et al., 2009), leading to analogous processes as observed in evolutionary biology, as summarized in Figure 1.

Biological populations and dialect groups undergo similar evolutionary processes (Figure 1.C). Natural selection occurs when species adapt to a specific environment and their genetic information changes across generations. A similar process can be found in dialects in the form of social selection (Atkinson et al., 2005; Pagel, 2009), for instance caused by the presence of administrative boundaries or natural barriers. Gene flow occurs when two distinct biological populations come into contact and transfer genetic information between populations (Jobling et al., 2013). The counterpart of gene flow in linguistics is termed borrowing (Atkinson et al., 2005; Pagel, 2009), which can occur when languages or dialect speakers from different regions come in contact and influence each other. As was explained above, borrowing is more likely between related languages and dialects. In biology, gene flow opposes the divergent force of selection in the areas where populations come into contact, thus blurring interpopulation transitions. This leads to the formation of geographical clines, a gradual change in a biological trait across space. Similarly, a dialect continuum results from the mutual intelligibility between neighboring dialects; the linguistic difference between dialects increases with distance and vice versa, creating a continuum rather than discrete boundaries.

These parallels show that treating dialect data analogous to genetic evidence is far from arbitrary. From a data perspective, dialectal phenomena are expected to be continuous and dialect groups are likely admixed due to borrowing and contact. Bayesian clustering is tailored to finding populations in admixed traits, making it well-suited to cluster dialect data.

3. Materials and methods

3.1 Data

The *Syntaktischer Atlas der deutschen Schweiz* (SADS) (Bucheli et al., 2002; Glaser & Bart, 2015) was created to capture the morphosyntactic variation in Swiss German. Between 2000 and 2002, a total of three thousand participants in 383 Swiss German-speaking municipalities were surveyed, with three to twenty-six individuals per municipality (median value of 6.5). The SADS data are particular in that they have multiple samples per municipality, while comparable data campaigns only interviewed a single respondent per spatial unit (Jeszenszky & Weibel, 2015), leading to less robust conclusions about linguistic variation. For this study, only respondents with a known geographic origin were considered, resulting in 2,970 individuals (Map 1) for which 211 morphosyntactic variants were collected. From these we derived the variables for the analysis, of which ultimately eighty-eight moderately associated variables were used for clustering (Figure 2).

3.2 Population structure

We applied TESS 2.3 (Chen et al., 2007; Durand, Jay, et al., 2009) to cluster the dialect data. TESS is a spatial Bayesian clustering algorithm inspired by STRUCTURE, a widely used algorithm in population genetics to infer population structure from multilocus

https://doi.org/10.1017/jlg.2021.12 Published online by Cambridge University Press

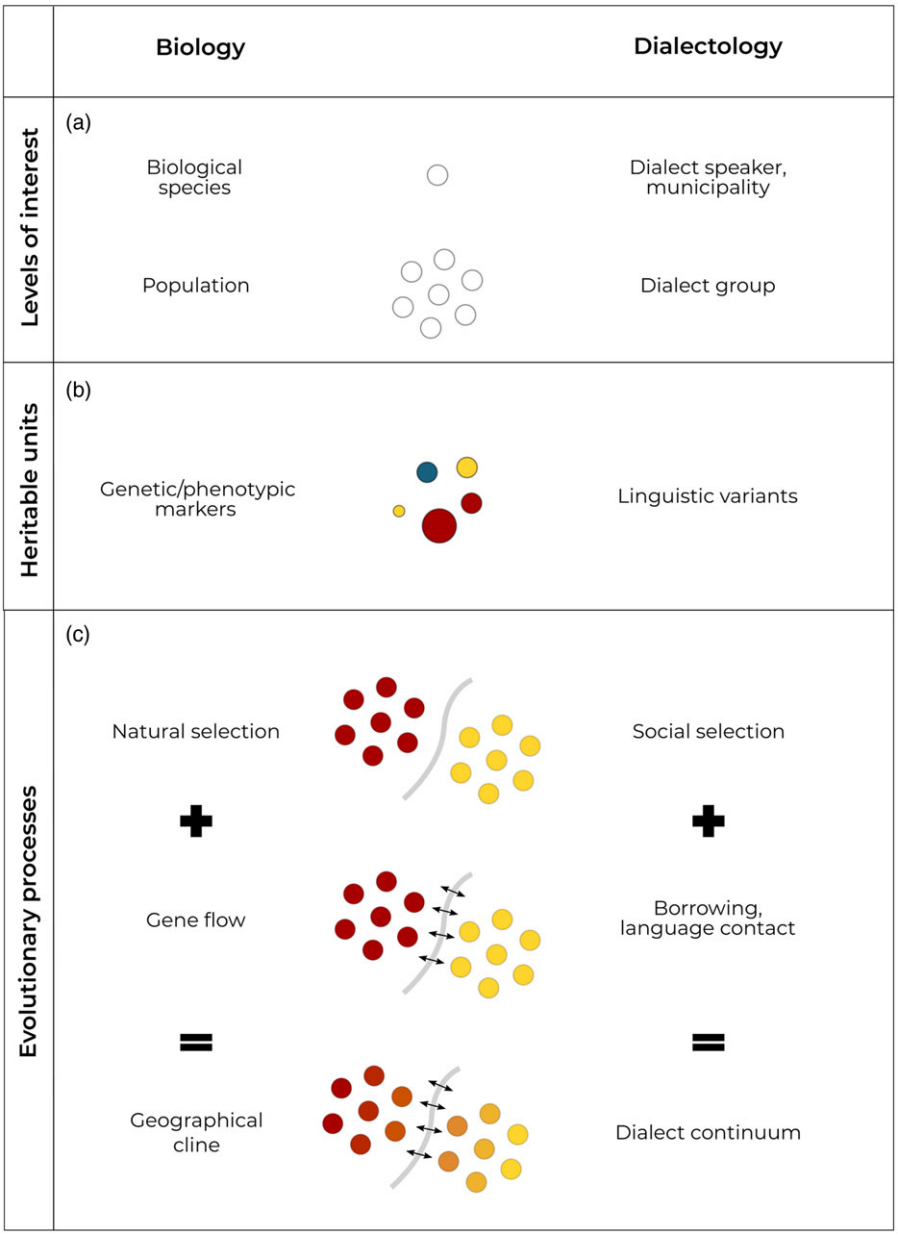


Figure 1. Parallels between biology and dialectology. Table adapted from Atkinson (2005).

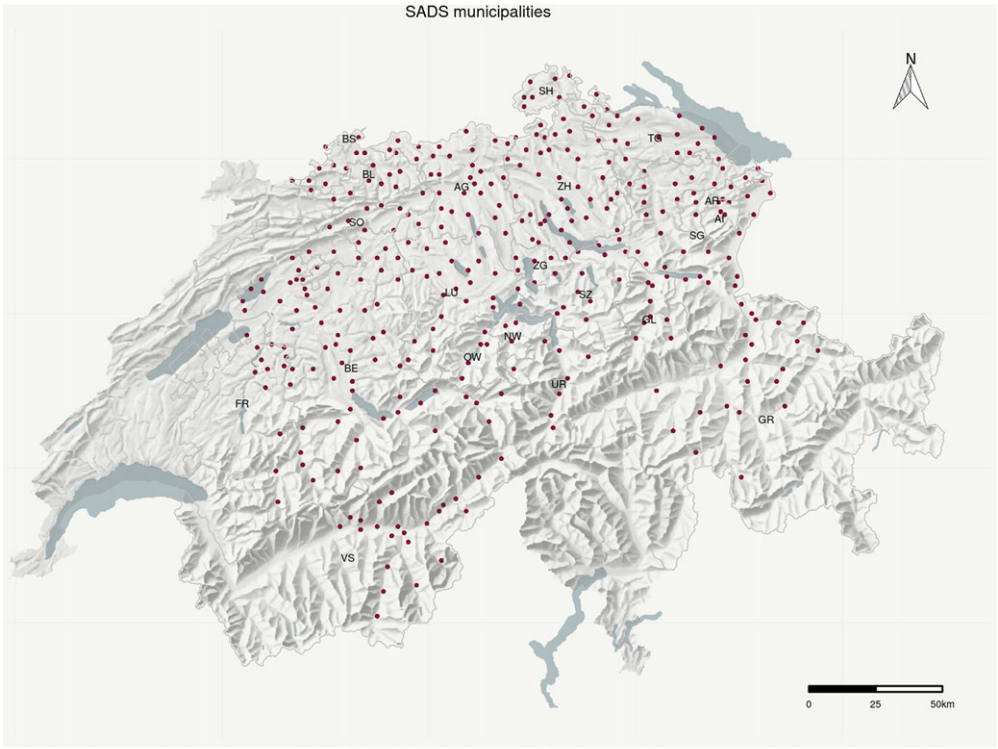
genotype data. TESS assumes that individuals have mixed ancestry, such that each individual has traits from at most K_{max} populations. A so-called admixture model infers the unknown ancestral populations, which—having been in continuous contact with each other—give rise to individuals with mixed ancestry. For Swiss German extensive interdialectal contact is expected, which is why we consider the admixture model well-suited for clustering the SADS data.

When inferring ancestral populations, TESS assumes alleles at different loci to be independent within populations, which is referred to as linkage equilibrium in evolutionary biology. This biological assumption cannot be fully paralleled in linguistics since linguistic features are not carried in a specific physical location such as loci in biology. However, the strong association between linguistic variants could be interpreted as a linkage disequilibrium state. We computed Cramer's V to measure the association between all morphosyntactic features in the SADS data and retained only those that were moderately associated ($V \leq 0.3$). While moderate correlations in the data improve the clustering, strong correlations could lead to biased, overly confident clusters and overestimate the number of inferred populations (Falush et al., 2003; Kaeuffer et al., 2007). For comparison, we also performed the cluster analysis on the full dataset, resulting in similar but overly clear-cut clusters (see supplementary materials).

TESS 2.3 incorporates spatial priors to account for global and local spatial effects. Global spatial priors assume that the influence of population K changes gradually along a spatial trend surface with gradients β_K. Local spatial priors assume that the probability of an individual to belong to population K depends on its spatial neighbors. We tested three models with different spatial priors (Durand, Jay, et al., 2009):

- the *nonspatial* model without a spatial prior (similar to STRUCTURE);
- the *spatial trend* model with global spatial priors;
- the *spatial full-trend* model, with both global and local spatial priors.

https://doi.org/10.1017/jlg.2021.12 Published online by Cambridge University Press



Map 1. All municipalities in the SADS data

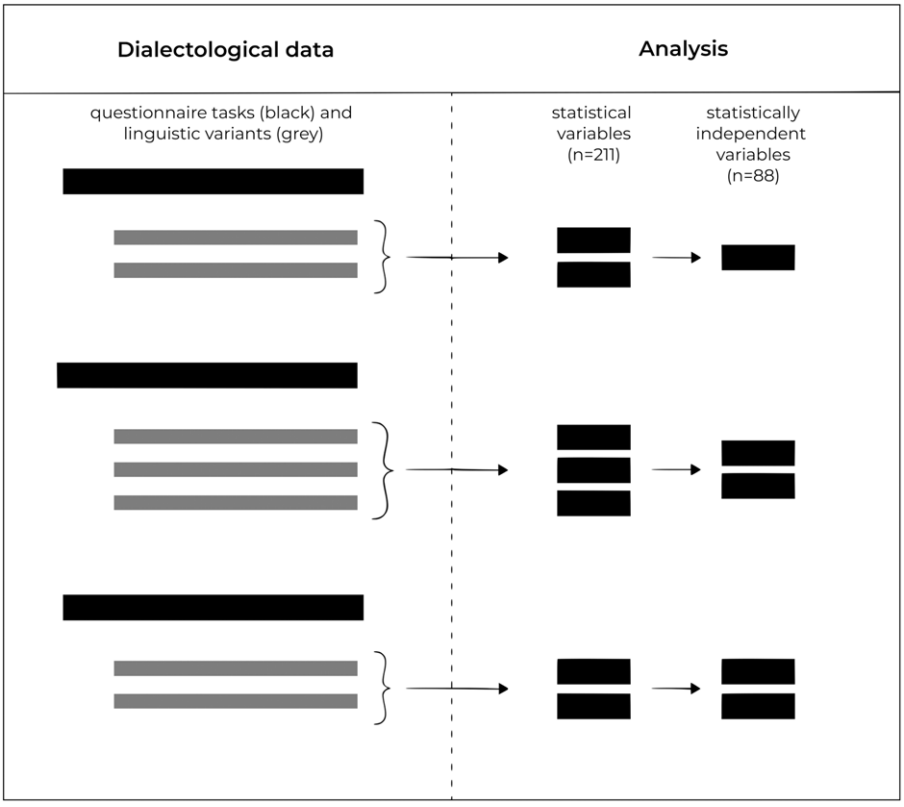


Figure 2. Pre-processing the dialectological data: the SADS derives linguistic variants from questionnaire tasks. These serve as statistical variables for the analysis, of which only the statistically independent ones were used for Bayesian clustering.

https://doi.org/10.1017/jlg.2021.12 Published online by Cambridge University Press

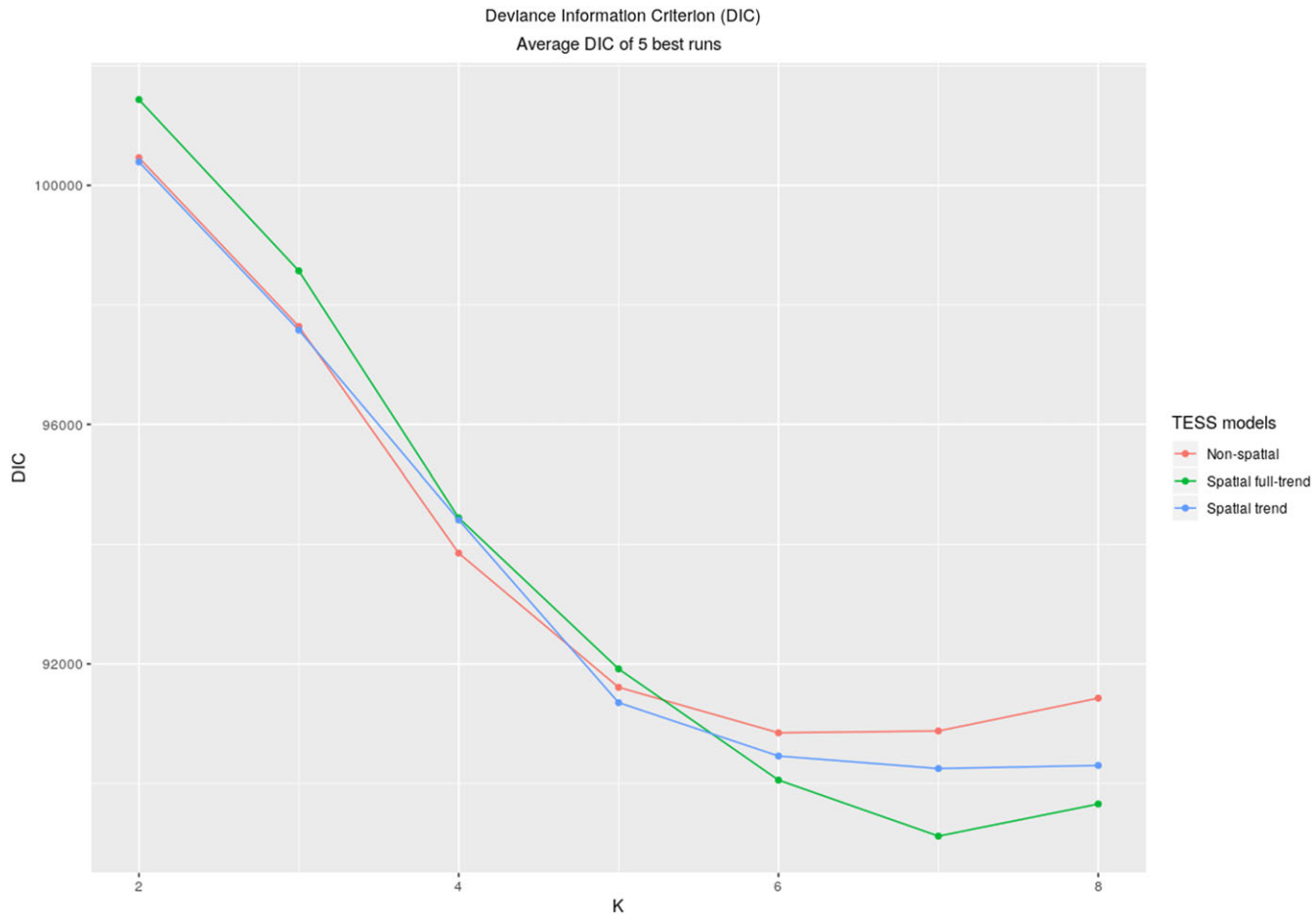


Figure 3. Average Deviance Information Criteria (DIC) of the five best MCMC runs for each model. The DIC levels off between five and seven populations. For $K = 5$ the *spatial trend* model outperforms both the *spatial full-trend* model and *the nonspatial* model.

TESS estimates population structure using Markov Chain Monte Carlo (MCMC) sampling. For the SADS data, 30,000 iterations were sufficient for the MCMC to converge to a stationary distribution. We discarded the first ten thousand iterations as burn-in, ensuring to sample from high probability regions. We tested different numbers of ancestral populations ($K \in 2, \ldots, 8$). For each K we ran fifteen independent runs. To have better estimates for the individual admixture proportions (Durand, Chen, et al., 2009), the results were averaged over the five best runs using the software CLUMPP (Jakobsson et al., 2007).

3.3 Model performance

TESS provides a statistical measure—the Deviance Information Criterion (DIC)—for testing which model and which number of ancestral populations best fit the data (Durand, Jay, et al., 2009; Spiegelhalter et al., 2002). The DIC comprises two terms: an estimate of model fit and an estimate of the effective number of parameters, which for TESS correspond to the number of ancestral populations K. There is a trade-off between model fit and the number of parameters: increasing K improves the fit, making it easier to capture the variation in the data but at the same time makes the model more likely to overfit and pick up random noise in the data. In general, the lower the DIC, the better the model. As a rule of thumb, the most suitable number of populations is where the DIC levels off, such that adding more populations decreases or only marginally improves performance.

4. Results

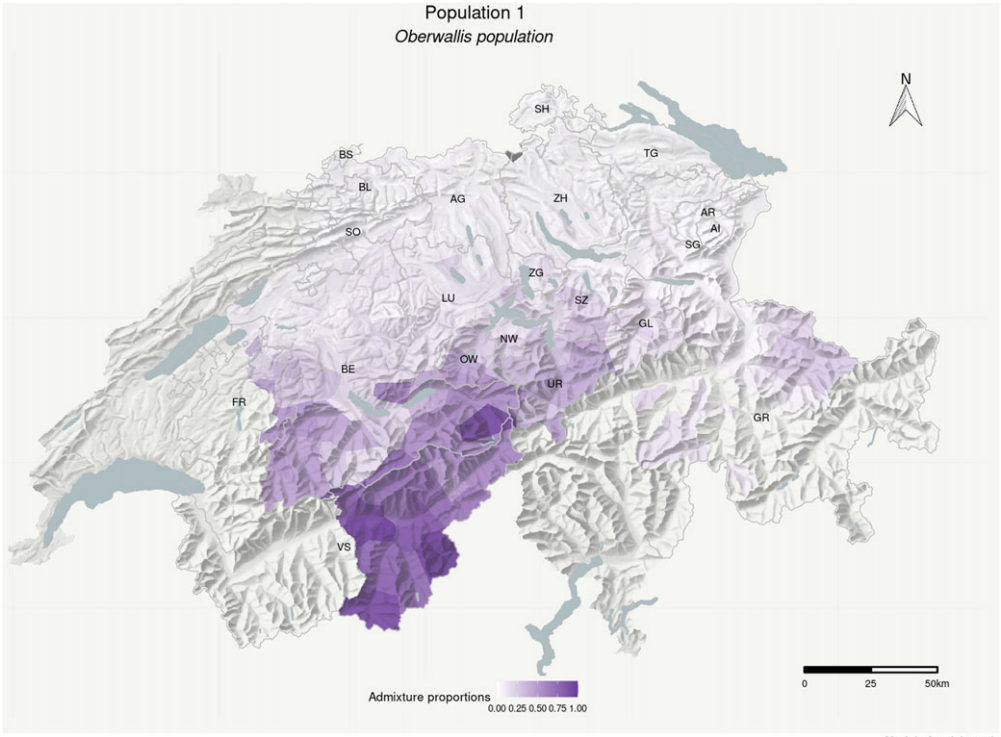
4.1 Number of ancestral populations

Figure 3 shows the DIC for $K \in 2, \ldots,$ eight for the nonspatial model, the spatial trend model and the spatial full-trend model averaged over the five best runs. For all three models, the DIC levels off between five and seven populations. For $K > 5$ the spatial trend model outperforms the nonspatial model. For $K > 6$ the spatial full-trend model outperforms all other models.

The purpose of cluster analysis is that of generalization: we apply TESS to reveal the main patterns of dialectal diversity in Swiss German morphosyntax.

If supported by the data, large clusters are preferred over small ones as they provide a higher level of generalization and, thus, deeper linguistic insights not apparent from the raw data. Considering that there is at best a moderate increase between $K = 5$ and $K = 7$ over all three models, we concluded that five populations best explain the structure in the data. For $K = 5$ the spatial trend model outperforms all other models. In the next section, we present both the spatial distribution of all five population clusters and their spatial trends. For comparison, we provide additional results for $K = 6$ in the supplementary materials.

https://doi.org/10.1017/jlg.2021.12 Published online by Cambridge University Press



Map 2. Admixture proportions of Oberwallis population

4.2 Spatial distribution of populations

Population 1 (Map 2) is predominantly present in canton VS and becomes gradually less present toward the border with BE in the north, the eastern part of FR, and some municipalities in the center of Switzerland (for Swiss canton codes see Table 1). The population continues eastward through OW, NW, SZ, and UR, eventually reaching the canton of GR. We refer to this population as *Oberwallis population*.

Population 2 (Map 3) covers most municipalities in the cantons of BE and FR. The population spreads upward to BS and BL in the north, the border to VS in the south, and to ZH and SZ in the east. As the core of the population is in BE, we refer to this population as the *Bern population*.

Population 3 (Map 4) has its core in northern Switzerland (canton SH) and gradually decreases in frequency to the cantons SO, LU, UR, OW, GL, SG, AI, and AR in the south. Accordingly, we label it as *Northern population*.

Population 4 (Map 5) does not show a distinctive spatial pattern. It lacks a clear core, and most municipalities seem to partly belong to it. Therefore, we refer to this population as the *Swiss base population*.

Population 5 (Map 6) is located in the southeastern part of Switzerland. Its core is in GR from where it continues to cantons UR, LU, AG, ZH, SH, and TG, gradually decreasing toward the west. Therefore, we label the population as the *Graubünden population*.

4.3 Spatial trends

The β coefficients quantify the magnitude of the trend surface in the spatial trend model. Table 2 shows the β coefficients and their relative Credible Intervals (CI) for all five populations.

The Oberwallis population has a significant southward-pointing latitudinal trend but no longitudinal trend (the CI of β_{long} includes 0). On the contrary, the Bern population shows a significant westward-pointing longitudinal trend but no latitudinal trend. The Northern population has a significant northward-pointing latitudinal trend with no longitudinal trend. The Swiss base population does not show any spatial trend (the CI of both β_{long} and β_{lat} trend center around zero). The Graubünden population has small but significant latitudinal and longitudinal trends.

5. Discussion

We identified five populations in the SADS data, for which the spatial trend model outperforms both the nonspatial and the spatial full-trend model. When comparing admixture levels of the municipalities to traditional dialect regions (Hotzenköcherle, Bigler, et al., 1984), we find four correspondences: the areas of dialectal populations 1, 2, 3, and 5 match the four traditional dialectal macroregions constituted by bundles of North-South and East-West isoglosses (Glaser, 2014:25–26; Christen et al., 2019:29–30). Population 4, on the contrary, can be interpreted as a "base" population, which has left traces in all Swiss German dialects. Population 1 (Oberwallis population) corresponds to the dialect region commonly classified as Highest Alemannic, whereas populations 2, 3, and 5 together are commonly classified as High Alemannic (Hotzenköcherle, Bigler, et al., 1984).[1]

In traditional dialectology and folk linguistics, dialect boundaries are usually discrete rather than gradual (Stoeckle, 2012), as they are constructed by few salient, overlapping isoglosses. This is also the case for Swiss German dialects. In contrast, the populations in our study do not exhibit clear-cut boundaries since grammar—and to a certain extent even single linguistic features (Seiler, 2005:330–33)—show gradual transitions (Glaser & Bart, 2015:90–92). Similar findings have also been made by other quantitative studies on Swiss German.

https://doi.org/10.1017/jlg.2021.12 Published online by Cambridge University Press

Table 1. Swiss canton codes

AG	Aargau	**AI**	Appenzell Innerrhoden
AR	Appenzell Ausserrhoden	**BE**	Bern
BL	Basel-Landschaft	**BS**	Basel-Stadt
FR	Fribourg	**GL**	Glarus
GR	Graubünden (Grisons)	**LU**	Luzern
NW	Nidwalden	**OW**	Obwalden
SG	St. Gallen	**SH**	Schaffhausen
SO	Solothurn	**SZ**	Schwyz
TG	Thurgau	**UR**	Uri
VS	Valais	**ZH**	Zürich

Jeszensky and Weibel (2016) show that syntactic features often show gradual transitions, even for individual features.

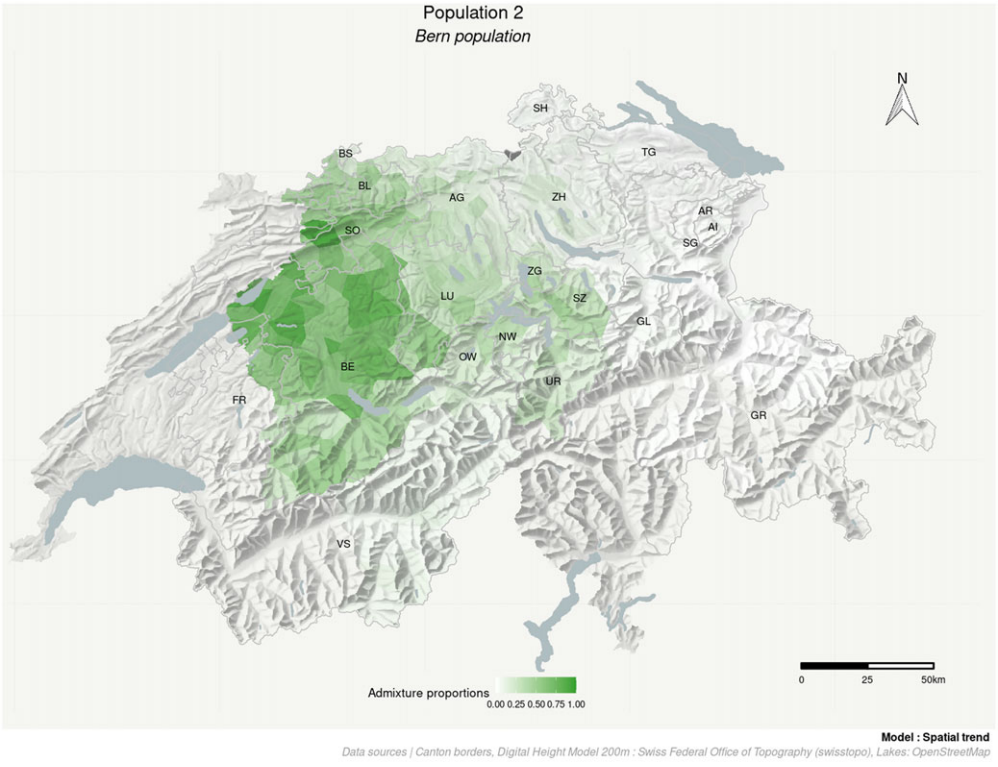
Depending on the linguistic feature and the way the data were collected, individual variables can show both crisp or gradual boundaries (Glaser, 2014; Jeszenszky, Stoeckle, et al., 2018; Scherrer et al., 2016). Bayesian clustering is flexible enough to adapt to the signal in the data and faithfully identifies both gradual transitions and clear-cut boundaries. On the contrary, discrete clustering will always yield discrete dialect regions, which in some cases "may not be an accurate representation of reality" (Grieve et al., 2011).

Concerning the spatial distribution of the populations, our results correspond to those by Kellerhals (2014), who identified similar dialect regions in the SADS data (depending on the method sometimes represented by more than a single area). Scherrer et al. (2016) compare the SADS data to phonological, morphological, and lexical data from the SDS and prove that all three follow a similar spatial distribution. The traditional dialect boundaries have also been confirmed recently by Lameli et al. (2020) using data from the Deutsche Sprachatlas (Wenker et al., 2013).

Four out of the five populations show an evident spatial structure with clear cores and a gradual increase of admixture toward the center. The β coefficients (Table 2)—which correspond to the spatial gradient of the admixture levels—are always oriented in the direction of the core: southward for the Oberwallis population, westward for the Bern population, northward for Northern population, and eastward for the Graubünden population. This is in line with the wave model of Schmidt (1872), which claims that language traits spread gradually from a central point.

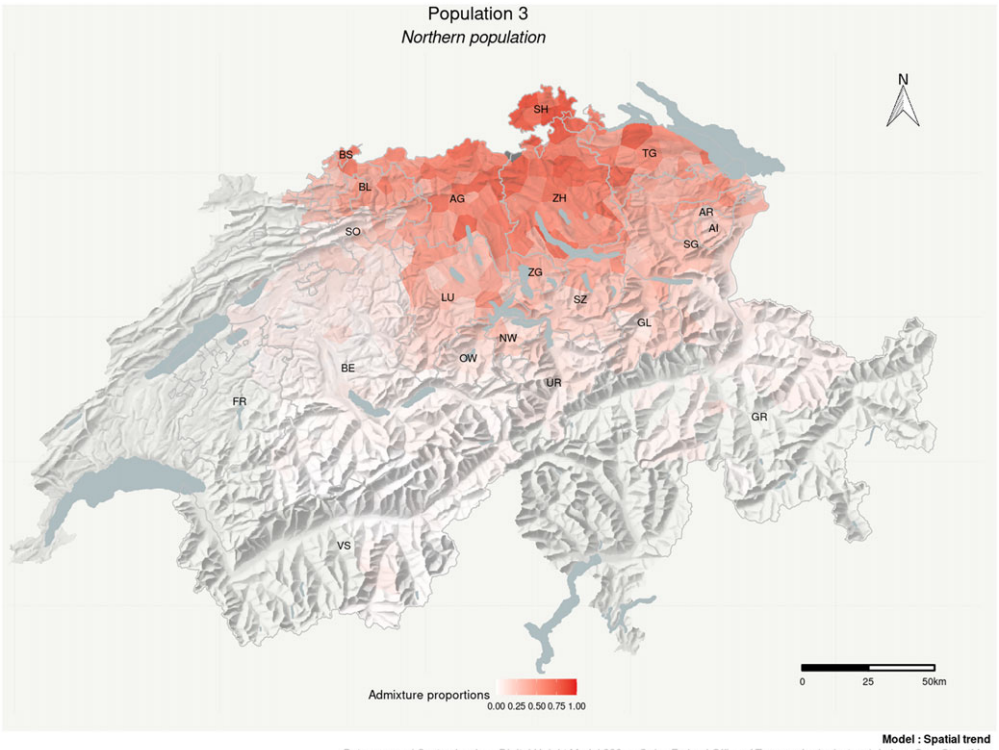
The cores of the clusters are located at the borders of the study area and admixture gradually increases toward the center. This suggests that there is more dialect diversity—with respect to the features in the SADS—in the center of the study area and less so at the borders. We compute the Shannon diversity index (H), a statistical measure of diversity (Shannon, 2001), to further test this. Map 7 shows that the least diverse municipalities with the lowest H are indeed located at the edges of the study area and overlap with the cores of the four populations, whereas the most diverse municipalities with the highest H are in the center.

In this study, we treat dialect features as heritable units (Syrjänen et al., 2016), analogous to genotypic and phenotypic markers (Figure 1). Our findings show that this analogy is indeed reasonable. The spatial pattern of the Oberwallis population (Map 2), for example, reflects one of the most relevant historical migration events in Swiss history known as the Walser migration. Due to particular sovereign rights in the area (Waibel, 2013), several groups of people emigrated in the 13–15th centuries from the

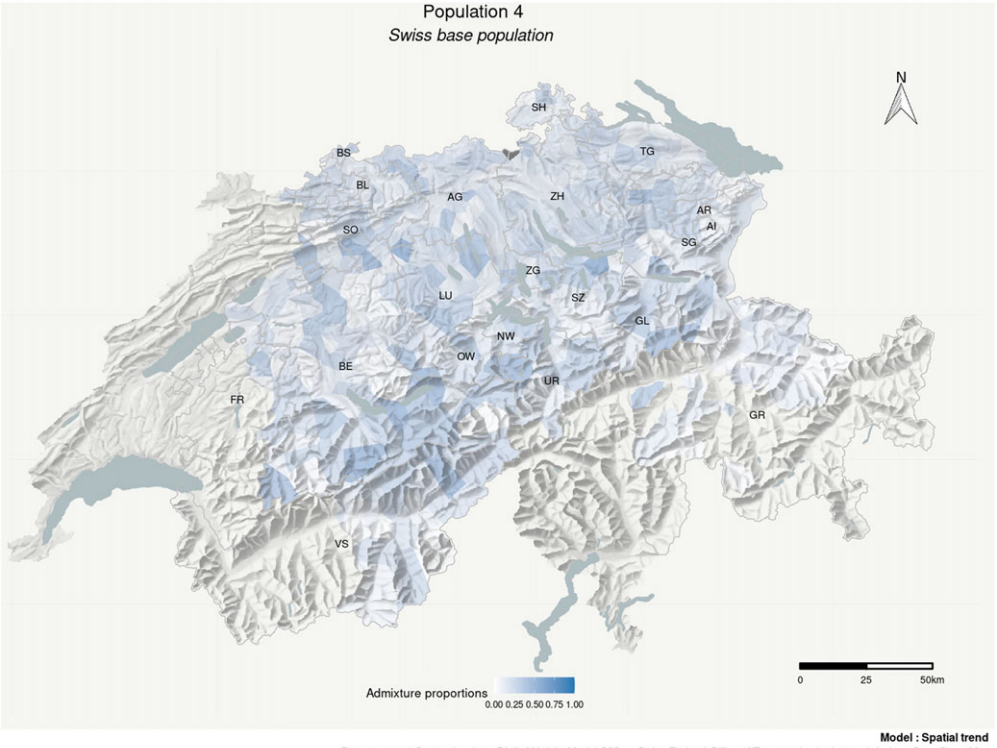


Map 3. Admixture proportions of *Bern* population

https://doi.org/10.1017/jlg.2021.12 Published online by Cambridge University Press



Map 4. Admixture proportions of *Northern* population



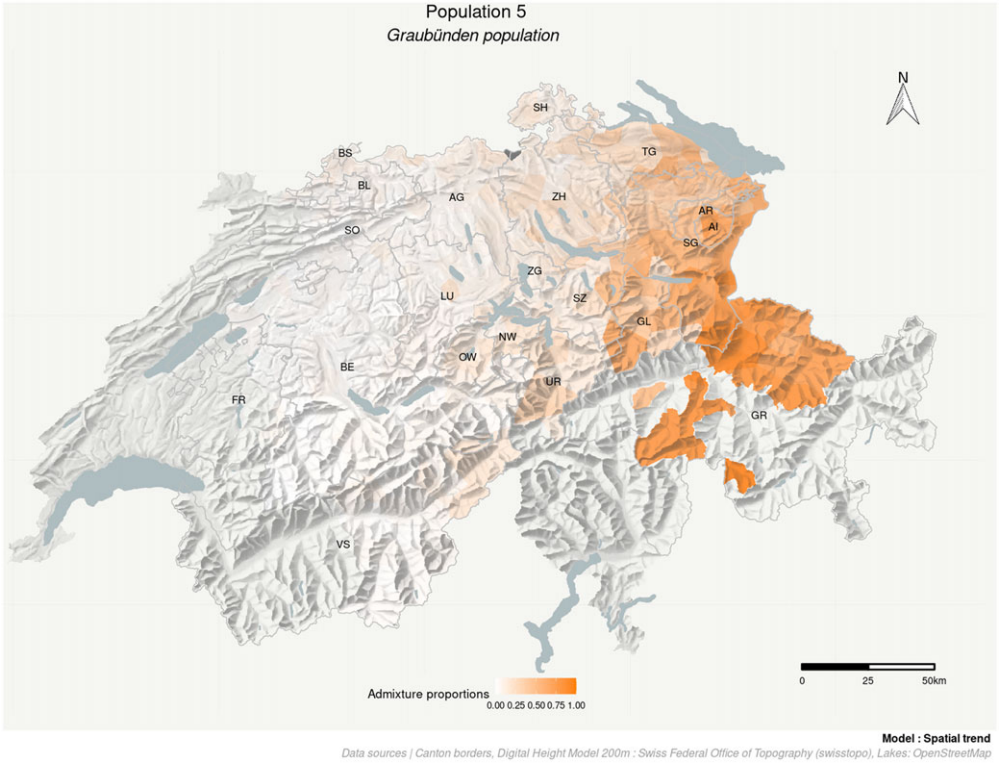
Map 5. Admixture proportions of *Swiss base* population

Oberwallis to the Berner Oberland, Uri, and Graubünden (Lessmann-Della Pietra, 2019; Waibel, 2013). The migration of the Walser population to Northern regions corresponds to a spread of dialects with Highest Alemannic features. The lighter areas in Graubünden in Map 6 would then be explained by the layering of the Oberwallis population with its Highest Alemannic features as a superstratum on top of the more prevalent Graubünden population. At the same time, the lighter areas

https://doi.org/10.1017/jlg.2021.12 Published online by Cambridge University Press

Table 2. β estimates and relative Credible Intervals (CI). Positive β_{long} represents west to east trends, meaning that the admixture proportions increase eastward. Positive β_{lat} represents south to north trends, meaning that the admixture proportions increase northward

	β_{long}	β_{lat}	β_{long} $CI_{95\%}$	β_{lat} $CI_{95\%}$
Population 1	−0.16	−5.39	[−0.83, 0.06]	[−5.84, −4.54]
Population 2	−2.83	0.27	[−3.05, −1.97]	[−0.05, 0.92]
Population 3	0.42	4.43	[−0.07, 0.66]	[3.1, 4.86]
Population 4	−0.21	0.12	[−0.41, 0.30]	[−0.53, 0.67]
Population 5	3.17	−1.20	[2.79, 3.38]	[−1.60, −0.58]



Map 6. Admixture proportions of *Graubünden* population

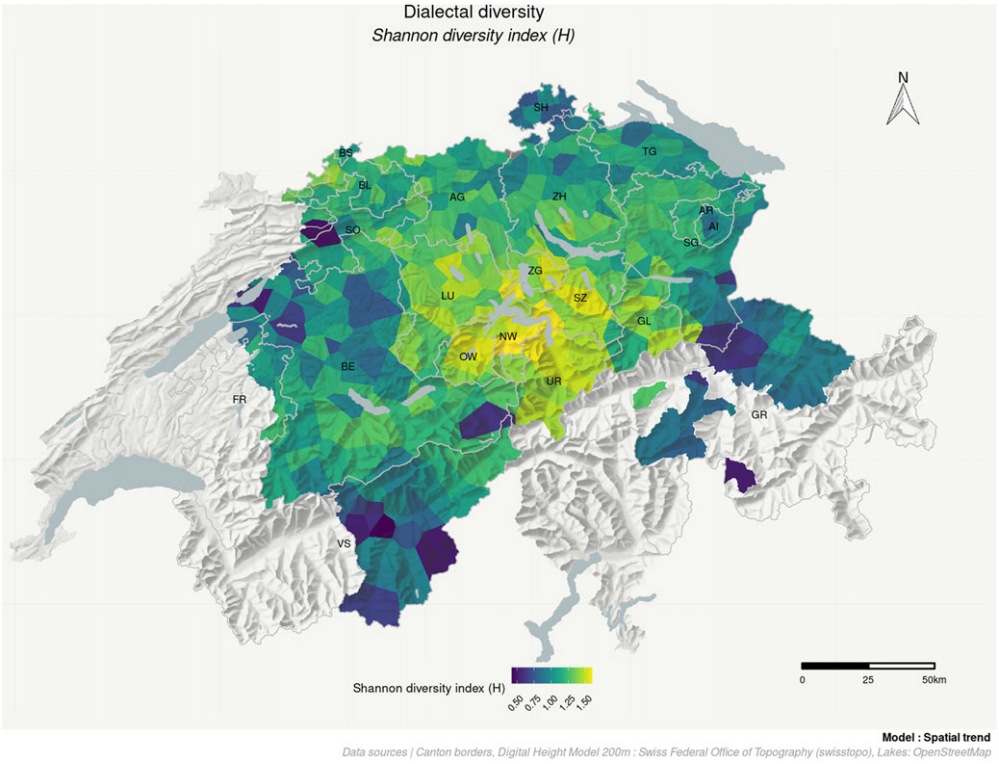
in Fribourg and Central Switzerland in Map 3 would be explained by a case of substratum layering, where the Oberwallis population's dialects are superimposed by Bern population dialects. This pattern is confirmed in Map 2: the core of the cluster is in VS, from where it continues toward the east along the migration corridor, reaching municipalities in the cantons UR and GR.

Bayesian clustering rests on several assumptions that must be met to yield meaningful results. We briefly summarize how our analysis satisfies these assumptions. First, the analysis is based on a sufficient number of linguistically independent features, which are likely to cause balanced and unbiased populations. The dialect sample is dense and evenly distributed over the study area, which is key for estimating both spatial trends and spatial autocorrelation. All dialects are related and are likely to belong to a single continuum. This in line with the objectives of the admixture model and the DIC, which aim to assign data points to as few meaningful clusters as possible. Reliable information about the historical background of the dialects is available and was used as an implicit prior to decide for an optimal number of clusters (Brooks et al., 2002). Finally, TESS infers populations that are as close as possible to Hardy-Weinberg equilibrium (HWE). HWE assumes that the allele frequencies in populations do not change between generations, which in the context of dialects implies that ancestral linguistic features do not change over time. This assumption seems contradictory for dialects in contact, which are likely to interchange features and are expected to evolve faster than biological populations (Boyd et al., 2005; Dawkins, 1976). If HWE does not hold, the resulting clusters might not adequately represent ancestral populations but rather reflect current population structure. This is why one must be careful when interpreting the five populations with regard to evolutionary history. All the same our results suggest that historical events, such as the Walser migration, are indeed picked up in the SADS data.

6. Conclusion

In this study, we applied TESS, a spatial Bayesian clustering algorithm from population genetics to Swiss German morphosyntactic dialect data. TESS reveals five populations, which largely correspond

https://doi.org/10.1017/jlg.2021.12 Published online by Cambridge University Press



Map 7. Dialectal diversity expressed in terms of the Shannon diversity index (H)

to traditional dialect regions. The model returns probabilistic clusters with fuzzy rather than discrete boundaries, appropriately reflecting the continuous nature of dialects.

From a linguistic point of view, Bayesian clustering revealed the structure of Swiss German dialectal strata (here: populations) and their diatopic/geographical distribution rather accurately, which is corroborated not only by earlier quantitative studies but also by traditional qualitative data and linguistic atlases. Thus, the results can be rated robust. As geographical patterns of linguistic features often reflect the diffusion paths of said features, there is reason to assume that inferring historical developments from populations may be possible to a certain degree (Nerbonne, 2010; Trudgill, 1974; Wolfram et al., 2003). Naturally, (especially more general) assumptions on diachronic developments become fuzzier with greater temporal depth. One example where Bayesian clustering reveals the connection between spatial patterns and linguistic change over time with almost absolute certainty is the Walser migration. At this point, most interpretations of other diachronic developments remain uncertain, since there is no historical data to which we could compare the results, and, hence, there is no conclusive answer as to which extent Bayesian clustering can truly reveal spatiotemporal aspects of linguistic change. Nevertheless, the validity of such an interpretation might be more likely in light of the fact that a temporal aspect of this kind is inherently attributed to the model. However, caution in terms of interpreting historical developments from Bayesian clustering applies for both population genetics and dialectology.

Acknowledgments. This research was supported by the University of Zurich Research Priority Program 'Language and Space'

Data Availability Statement. All TESS results (fifteen runs) may be found at https://github.com/romanoe/mapping-dialects/tree/master/TESS. To explore the interactive maps choosing six populations, please visit: Mapping Swiss German dialects: https://www.noemiromano.ch/mapping-dialects. For full dataset (correlated features not removed), please visit: https://github.com/romanoe/mapping-dialects/tree/master/Raw%20data

Supplementary material. To view supplementary material for this article, please visit https://doi.org/10.1017/jlg.2021.12

Note

1 For each population, there are in fact several linguistic variables (e.g., from the SADS), which, in their areal distribution, correspond to the boundaries of the individual populations. We want to mention one such linguistic variable per population for illustrative purposes:

- Population 1 (Oberwallis): depictive agreement; in the Oberwallis population area, in contrast to other dialectal areas, the (co)predicative adjective (*heiss* 'hot') takes a suffix corresponding to the number and case of the direct object (*die Milch* 'the milk')(e.g., *Du musst die Milch aber heisse trinken* 'You have to drink the milk hot.' (See SADS 4.1.1 Hauptkarte).
- Population 2 (Bern): case of the interrogative pronoun; in the Bern population area the interrogative pronoun is used in the accusative form *wen* 'whom' (cf., *Für wen/wer/wem sind denn die Blumen*? 'For who/whom are the flowers?' (See SADS 2.3.2 Hauptkarte).
- Population 3 (Northern): form of the comparative particle in predicative comparative constructions; only in the Northern population area the particle *wie* ('than') can be used in sentences like *Dann ist er ja älter, als/weder/wie/wan ich gedacht habe* 'So he is older than I thought.' (See SADS 5.3.3 Hauptkarte).
- Population 4 (Swiss base): order of pronouns in a pronoun cluster; both possible orders can be used all over Switzerland (cf., *Ich habe es ihr/ihr es gestern gegeben* 'I gave it to her yesterday'. (See SADS 2.6.3 Hauptkarte).
- Population 5 (Graubünden): verb order in two-verb clusters in subordinate clauses; in the Graubünden population area the order *heiraten will* 'wants to marry' can be used, whereas *will heiraten* 'want to marry' is used all over

German speaking Switzerland (cf., *Also ich weiss auch nicht, ober einmal heiraten will/will heiraten* 'I don't know, if he wants to get married one day.' (See SADS 3.1.1 Hauptkarte).

When inferring clusters for K = 6, five out of the six populations also correspond to some of the linguistic variables investigated (cf., the supplementary materials). Population 1 (pink) is different and rather serves as a transition zone for several variables (e.g., the final infinitival clause, see SADS 5.1.3 Hauptkarte; Jeszenszky & Weibel, 2016), with the exception of the distribution of the use of the prepositional dative marker (see SADS 1.3.1 Hauptkarte).

References

Atkinson, Quentin D. & Russell D. Gray. 2005. Curious parallels and curious connections: phylogenetic thinking in biology and historical linguistics. *Systematic Biology 54*(4). 513–26.

Bachmann, Sandro, Gabriela Bart, & Elvira Glaser. 2021. *Syntaktischer Atlas der deutschen Schweiz. Volume 2: Karten. Herausgegeben von Elvira Glaser, bearbeitet von Sandro Bachmann, Gabriela Bart und Elvira Glaser, sowie Claudia Bucheli Berger und Guido Seiler*. Tübingen: Francke.

Bart, Gabriela, Elvira Glaser, Pius Sibler, & Robert Weibel. 2013. Analysis of Swiss German syntactic variants using spatial statistics. In Xosé Afonso Álvarez Pérez, Ernestina Carrilho, & Catarina Magro, (eds.), *Current approaches to limits and areas in dialectology*, 143–69. Newcastle upon Tyne: Cambridge Scholars Publishing.

Bowern, Claire. 2012. The riddle of Tasmanian languages. *Proceedings of the Royal Society of London B: Biological Sciences 279*(1747). 4590–95.

Boyd, Robert & Peter J. Richerson. 2005. *The origin and evolution of cultures*. Oxford: Oxford University Press.

Brooks, S.P., Jim Smith, Aki Vehtari, Martyn Plummer, Mervyn Stone, Christian P. Robert, D.M. Titterington, J.A. Nelder, Anthony Atkinson, & A.P. Dawid, A. Lawson. 2002. Discussion on the paper by Spiegelhalter, Best, Carlin and van der Linde. *Journal of the Royal Statistical Society. Series B: Statistical Methodology 64*(4). 616–39.

Bucheli, Claudia & Elvira Glaser. 2002. The syntactic atlas of Swiss German dialects: empirical and methodological problems. *Syntactic Microvariation 2*. 41–73.

Chen, Chibiao, Eric Durand, Florence Forbes, & Olivier François. 2007. Bayesian clustering algorithms ascertaining spatial population structure: a new computer program and a comparison study. *Molecular Ecology Resources 7*(5). 747–56.

Christen, Helen, Elvira Glaser, & Matthias Friedli. 2019. *Kleiner Sprachatlas der deutschen Schweiz (KSDS)*, 8th edn. Frauenfeld: Huber.

Corander, Jukka, Jukka Sirén, & Elja Arjas. 2008. Bayesian spatial modeling of genetic population structure. *Computational Statistics 23*(1). 111–29.

Croft, William. 2000. *Explaining language change: An evolutionary approach*. London: Pearson Education.

Darwin, Charles. 1859. *On the origin of species by means of natural selection*. London: Murray.

Dawkins, Richard. 1976. *The selfish gene*. Oxford: Oxford University Press.

Dunn, Michael, Stephen C. Levinson, Eva Lindström, Ger Reesink, & Angela Terrill. 2008. Structural phylogeny in historical linguistics: methodological explorations applied in Island Melanesia. *Language 84*(4). 710–59.

Durand, Eric, Chibiao Chen, & Olivier François. 2009. *TESS version 2.3 reference manual*. http://membres-timc.imag.fr/Olivier.Francois/tess.html (4 April, 2022).

Durand, Eric, Flora Jay, Oscar E. Gaggiotti, & Olivier François. 2009. Spatial inference of admixture proportions and secondary contact zones. *Molecular Biology and Evolution 26*(9). 1963–73.

Embleton, Sheila. 1993. Multidimensional scaling as a dialectometrical technique: Outline of a research project. *Contributions to quantitative linguistics*, 76–267. New York: Springer.

Falush, Daniel, Matthew Stephens, & Jonathan K. Pritchard. 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics 164*(4). 1567–87.

Girnth, Heiko. 2010. Mapping language data. In Alfred Lameli, Roland Kehrein, & Stefan Rabanus (eds.), *Language and space: An international handbook of linguistic variation,* vol. *2*, 98–121. Berlin: De Gruyter Mouton.

Glaser, Elvira. 2014. Wandel und Variation in der Morphosyntax der schweizerdeutschen Dialekte. *Taal en Tongval 66*(1). 21–64.

Glaser, Elvira & Gabriela Bart. 2015. Dialektsyntax des Schweizerdeutschen. In Roland Kehrein, Alfred Lameli, & Stefan Rabanus (eds.), *Regionale Variation des Deutschen*, 81–108. Berlin: De Gruyter.

Glaser, Elvira & Gabriela Bart. 2021. *Syntaktischer Atlas der deutschen Schweiz. Volume 1: Einleitung und Kommentare. Herausgegeben von Elvira Glaser, bearbeitet von Elvira Glaser und Gabriela Bart, sowie Claudia Bucheli Berger, Guido Seiler, Sandro Bachmann und Anja Hasse, unter Mitarbeit von Matthias Friedli und Janine Richner-Steiner*. Tübingen: Francke.

Goebl, Hans. 1982. *Dialektometrie*, vol. 157. Vienna: Verlag der Osterreichischen Akademië der Wissenschaften.

Goebl, Hans. 2008. Le Laboratoire de dialectométrie de l'Université de Salzbourg. Un bref rapport de recherche. *Zeitschrift für französische Sprache und Literatur 118*(1). 35–55.

Goebl, Hans. 2010. Dialectometry and quantitative mapping. In Alfred Lameli, Roland Kehrein, & Stefan Rabanus (eds.), *Language and space: An international handbook of linguistic variation*, vol. *2*, 433–57. Berlin: De Gruyter Mouton.

Goebl, Hans, Karl Jaberg, & Jules Gilliéron. 1984. *Dialektometrische Studien: Anhand italoromanischer, rätoromanischer und galloromanischer Sprachmaterialien aus AIS und ALF*, Vol.3. Tübingen: Max Niemeyer.

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

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

Guillot, Gilles, Frédéric Mortier, & Arnaud Estoup. 2005. GENELAND: a computer package for landscape genetics. *Molecular Ecology Notes 5*(3). 712–15.

Hamming, Richard W. 1950. Error detecting and error correcting codes. *Bell System Technical Journal 29*(2). 147–60.

Heeringa, Wilbert Jan. 2004. *Measuring dialect pronunciation differences using Levenshtein distance*. Groningen: University of Groningen dissertation.

Hock, Hans Henrich & Brian D. Joseph. 2009. *Language history, language change, and language relationship. An introduction to historical and comparative linguistics*. Berlin: De Gruyter.

Hotzenköcherle, Rudolf, Niklaus Bigler, Robert Schläpfer, & Rolf Börlin. 1984. *Die Sprachlandschaften der deutschen Schweiz*, Vol. *1*. Aarau: Sauerländer.

Hotzenköcherle, Rudolf, Rudolf Trüb, & Heinrich et al. Baumgartner. 1962–2003. *Sprachatlas der deutschen Schweiz* (SDS). Francke.

Jakobsson, Mattias & Noah A. Rosenberg. 2007. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics 23*(14). 1801–06.

Jeszenszky, Péter, Philipp Stoeckle, Elvira Glaser, & Robert Weibel. 2017. Exploring global and local patterns in the correlation of geographic distances and morphosyntactic variation in Swiss German. *Journal of Linguistic Geography 5*(2). 86–108.

Jeszenszky, Péter, Philipp Stoeckle, Elvira Glaser, & Robert Weibel. 2018. A gradient perspective on modeling interdialectal transitions. *Journal of Linguistic Geography 6*(2). 78–99.

Jeszenszky, Péter & Robert Weibel. 2015. Measuring boundaries in the dialect continuum. Paper presented at the 18th AGILE International Conference on Geographic Information Science, Lisbon (Portugal), June 9–12, 2015.

Jeszenszky, Péter & Robert Weibel. 2016. Modeling transitions between syntactic variants in the dialect continuum. Paper presented at the 19th AGILE International Conference on Geographic Information Science, Helsinki (Finland), June 14–17, 2016.

Jobling, Mark, Matthew Hurles, & Chris Tyler-Smith. 2013. *Human evolutionary genetics: origins, peoples & disease*. New York: Garland Science.

Kaeuffer, R., D. Réale, D.W. Coltman, & D. Pontier. 2007. Detecting population structure using STRUCTURE software: effect of background linkage disequilibrium. *Heredity 99*(4). 374–80.

https://doi.org/10.1017/jlg.2021.12 Published online by Cambridge University Press

Kellerhals, Sandra. 2014. *Dialektometrische Analyse und Visualisierung von schweizerdeutschen Dialekten auf verschiedenen linguistischen Ebenen*. Zurich: Institute of Geography, University of Zurich dissertation.

Kessler, Brett. 1995. Computational dialectology in Irish Gaelic. In *Proceedings of the seventh conference on European chapter of the Association for Computational Linguistics*, 60–66. San Francisco: Morgan Kaufmann Publishers.

Lameli, Alfred, Elvira Glaser, & Philipp Stoeckle. 2020. Drawing areal information from a corpus of noisy dialect data. *Journal of Linguistic Geography 8*(1). 31–48.

Leemann, Adrian. 2016. Analyzing geospatial variation in articulation rate using crowdsourced speech data. *Journal of Linguistic Geography 4*(2). 76–96.

Lessmann-Della Pietra, Cristina. 2019. Geschichte von Bosco Gurin. In Sandro Bachmann & Elvira Glaser (eds.), *Bosco Gurin: Das Walserdorf im Tessin und seine Sprache(n)*, 21–35. Bern: Schweizerische Akademie der Geistes- und Sozialwissenschaften (SAGW).

Levenshtein, Vladimir I. 1966. Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Doklady 10*(8). 707–10.

Lundberg, Jan. 2005. Classifying dialects using cluster analysis. Gothenburg: Göteborg University master's thesis.

Murray, Robert W. 2010. Language and space: The neogrammarian tradition. In Peter Auer & Jürgen Erich Schmidt (eds.), *Language and space: theories and methods*, 70–87. Berlin: De Gruyter Mouton.

Nerbonne, John. 2010. Measuring the diffusion of linguistic change. *Philosophical Transactions of the Royal Society B: Biological Sciences 365*(1559). 3821–28.

Nerbonne, John, Wilbert Heeringa, & Peter Kleiweg. 1999. Edit distance and dialect proximity. In David Sankoff & Joseph Kruskal (eds.), *Time warps, string edits and macromolecules: the theory and practice of sequence comparison*, 5–15. Stanford: CSLI Press.

Nerbonne, John & Peter Kleiweg. 2007. Toward a dialectological yardstick. *Journal of Quantitative Linguistics 14*(2–3). 148–66.

Nerbonne, John & William A. Kretzschmar Jr. 2013. Dialectometry++. *Literary and Linguistic Computing 28*(1). 2–12.

Pagel, Mark. 2009. Human language as a culturally transmitted replicator. *Nature Reviews Genetics 10*(6). 405.

Pritchard, Jonathan K., Matthew Stephens, & Peter Donnelly. 2000. Inference of population structure using multilocus genotype data. *Genetics 155*(2). 945–59.

Prokić, Jelena & John Nerbonne. 2008. Recognising groups among dialects. *International Journal of Humanities and Arts Computing 2*(1–2). 153–72.

Pröll, Simon, Simon Pickl, & Aaron Spettl. 2014. Latente Strukturen in geolinguistischen Korpora. In Michael Elmentaler, Markus Hundt, & Jürgen Erich Schmidt (eds.), *Deutsche Dialekte. Konzepte, Probleme, Handlungsfelder. Akten des, 4.*, vol. 4, 247–58. Stuttgard: Steiner.

Reesink, Ger, Ruth Singer, & Michael Dunn. 2009. Explaining the linguistic diversity of Sahul using population models. *PLoS Biology 7*(11): e1000241.

Scherrer, Yves & Philipp Stoeckle. 2016. A quantitative approach to Swiss German: dialectometric analyses and comparisons of linguistic levels. *Dialectologia et Geolinguistica 24*(1). 92–125.

Schmidt, Johannes. 1872. *Die Verwantschaftsverhältnisse der Indogermanischen Sprachen*. Weimar: Böhlau.

Schrambke, Renate. 2010. Language and space: Traditional dialect geography. In Peter Auer & Jürgen Erich Schmidt (eds.), *Language and space: theories and methods*, 87–107. Berlin: De Gruyter Mouton.

Séguy, Jean. 1973. *La dialectométrie dans l'Atlas linguistique de la Gascogne*. France: Société de linguistique romane.

Seiler, Guido. 2005. Wie verlaufen syntaktische Isoglossen, und welche Konsequenzen sind daraus zu ziehen? In Eckhard Eggers, Jürgen Erich Schmidt, & Dieter Stellmacher (eds.), *Moderne Dialekte – Neue Dialektologie. Akten des 1. Kongresses der Internationalen Gesellschaft für Dialektologie des Deutschen (IGDD) am Forschungsinstitut für deutsche Sprache «Deutscher Sprachatlas» der Phillips-Universität Marburg vom 5.–8. März 2003*, 313–41. Stuttgart: Steiner.

Shackleton Jr, Robert G. 2005. English-American speech relationships: A quantitative approach. *Journal of English Linguistics 33*(2). 99–160.

Shannon, Claude Elwood. 2001. A mathematical theory of communication. *ACM SIGMOBILE Mobile Computing and Communications Review 5*(1). 3–55.

Sibler, Pius, Robert Weibel, Elvira Glaser, & Gabriela Bart. 2012. Methods for cartographic visualization of linguistic data in support of area formation in dialectology. Paper presented at AutoCarto 2021, Columbus (Ohio, USA), September 16–18, 2012.

Spiegelhalter, David J., Nicola G. Best, Bradley P. Carlin, & Angelika Van Der Linde. 2002. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 64*(4). 583–639.

Spruit, Marco René. 2008. *Quantitative perspectives on syntactic variation in Dutch dialects*. Thesis no. 174. Amsterdam: Netherland Graduate School of Linguistics thesis.

Stoeckle, Philipp. 2012. The folk linguistic construction of local dialect areas: linguistic and extra-linguistic factors. In Sandra Hansen, Christian Schwarz, Philipp Stoeckle, & Tobias Streck (eds.), *Dialectological and folk dialectological concepts of space. Current methods and perspectives in sociolinguistic research on dialect change*, 142–63. Berlin: De Gruyter.

Syrjänen, Kaj, Terhi Honkola, Jyri Lehtinen, Antti Leino, & Outi Vesakoski. 2016. Applying population genetic approaches within languages. *Language Dynamics and Change 6*(2). 235–83.

Szmrecsanyi, Benedikt. 2011. Corpus-based dialectometry: a methodological sketch. *Corpora 6*(1). 45–76.

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

Valls, Esteve, John Nerbonne, Jelena Prokić, Martijn Wieling, Esteve Clue, & Maria Rosa Lloret. 2012. Applying the Levenshtein Distance to Catalan dialects: A brief comparison of two dialectometric approaches. *Verba: anuario galego de filoloxıa 39*. 35–61.

Waibel, Max. 2013. Walser. *Historisches Lexikon der Schweiz (HLS)*. https://hls-dhs-dss.ch/de/articles/007950/2013-08-20/ (4 April, 2022).

Wenker, Georg, Alfred Lameli, Johanna Heil, & Constanze Wellendorf. 2013. *Schriften zum "Sprachatlas des Deutschen Reichs."* Hildesheim, Germany: Georg Olms Verlag.

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

Wolfram, Walt & Natalie Schilling-Estes. 2003. Dialectology and linguistic diffusion. In Brian D. Joseph & Richard D. Janda (eds.), *The handbook of historical linguistics*, 713–35. Malden, MA: Blackwell.

https://doi.org/10.1017/jlg.2021.12 Published online by Cambridge University Press