Zeitschrift für
SPRACHVARIATION
& SOZIOLINGUISTIK
JOURNAL OF LANGUAGE
VARIATION AND SOCIOLINGUISTICS

https://doi.org/10.5282/jlvs/7

Geographical patterns in the Bavarian dialects of Austria and South Tyrol. A real-time comparison using dialectometric methods

Philipp Stöckle[1], Philip C. Vergeiner[2]

[1] Austrian Centre for Digital Humanities and Cultural Heritage, Austrian Academy of Sciences, Vienna, Austria
[2] Institute of German Philology, LMU Munich, Munich, Germany

E-mail: philipp.stoeckle@oeaw.ac.at

Received 28.10.2024
Accepted for publication 14.02.2025
Published 30.03.2025

Abstract

This paper presents a comparative real-time study of phonological variation in Bavarian dialects in Austria and South Tyrol, using a dialectometric approach. Drawing on historical data from the 'Dictionary of Bavarian Dialects in Austria' (*Wörterbuch der bairischen Mundarten in Österreich*) and contemporary data from the project 'Variation and Change of Dialect Varieties in Austria (in Real and Apparent Time)', supplemented by data from the project 'German Dialects in South Tyrol' (*Deutsche Dialekte in Südtirol*), we investigate geographic patterns of dialectal variation over time. The study addresses three central research questions: (1) What geolinguistic patterns emerge from historical and contemporary datasets? (2) How do these patterns align with traditional qualitative dialect classifications? (3) Can language change be identified through diachronic comparison? By analyzing a consistent set of 31 phonological variables using techniques such as multidimensional scaling and cluster analysis, we find a strong overall correspondence between historical and contemporary dialect classifications, which also align well with traditional qualitative dialect classifications. Our results highlight the persistence of major dialect divisions, particularly between (South) Central Bavarian and South Bavarian. At the same time, we identify more localized changes, such as the retreat of South Bavarian in southeastern Austria, increasing regionalization, and shifts in east-west variation patterns. In sum, the study demonstrates that real-time comparisons using dialectometric methods are feasible, despite certain methodological challenges.

Keywords: dialect classification, dialectometry, real-time change

1 Introduction

It has been noted that "German [...] is the only language for which the dialect dynamics of large parts of the language area can be traced and analyzed over an entire century based on empirical data" (Schmidt & Herrgen 2011: 89; own translation). This is particularly true for the dialects of Austria and South Tyrol, where the history of dialectological research extends over more than a century. Early studies, such as those by Schatz (1903) and Lessiak (1963 [1903]), concentrated on local dialects in selected research sites. By the first half of the 20th century, however, several large-scale projects for the collection of dialect data were initiated, enabling more comprehensive studies of geolinguistic patterns. For instance, a large-scale indirect dialect survey was launched after 1926, modeled on Georg Wenker's 'Language Atlas of the German Empire' (*Sprachatlas des Deutschen Reichs*), resulting in over 3,500 completed questionnaires (see Fleischer 2017 for details). An even more extensive data collection was conducted for the 'Dictionary of Bavarian Dialects in Austria' (*Wörterbuch der bairischen Mundarten in Österreich*), which amassed over three million dialect records in the decades following its inception in 1912 (see Stöckle 2021).

Remarkably, however, despite the focus on dialect change in recent dialectological research (see e.g., Moosmüller & Scheutz 2013; Vergeiner et al. 2021a; Vergeiner 2022), this historical data is seldom used to investigate dialect change. Most existing studies rely on apparent-time designs and/or compare newly collected data with older research, often using historical dialect descriptions and overviews as a basis for comparison (e.g., Kranzmayer 1956). Only few studies directly compare historical and contemporary data to investigate language change in real-time (e.g., Stöckle & Wittibschlager 2022; Vergeiner & Bülow 2024; for other German speaking regions, cf. e.g., Steiner et al. 2023; Streck 2012; Schwarz 2015). Since these studies tend to focus on only one linguistic variable at a time, drawing comprehensive conclusions about real-time change in Austria remains challenging.

To address this desideratum, the present study aims to investigate real-time change in the Bavarian dialects of Austria and South Tyrol across multiple features of dialect phonology. In doing so, we analyze and compare the geolinguistic structures of dialect phonology in the same set of variables in both historical and contemporary data. To abstract away from the individual features and to reveal more general tendencies we employ aggregative dialectometric methods (e.g., Vergeiner 2023). Our historical data are sourced from the 'Dictionary of Bavarian Dialects in Austria' (*Wörterbuch der bairischen Mundarten in Österreich,* WBÖ), while the contemporary data come from the project 'Variation and Change of Dialect Varieties in Austria (in Real and Apparent Time)', with additional data taken from the project 'German Dialects in South Tyrol' (*Deutsche Dialekte in Südtirol*; see Scheutz 2016).

Based on these datasets, our study will address the following research questions:

1. Can geographical patterns in dialect phonology for the Bavarian dialects of Austria and South Tyrol be identified in both historical and contemporary data? What is the linguistic basis for these patterns?
2. How do these patterns relate to traditional (qualitative) dialect classifications (e.g., Wiesinger 1983), and how do the geographical patterns in the historical data compare to those in the contemporary data?
3. Can differences be identified which are indicative of language change over the past 100 years? If language change is observed, how can it be explained?

In what follows, we will provide a brief overview of dialect variation and change in Austria and South Tyrol (Section 2). Subsequently, we introduce our data and methods in more detail (Section 3). In Section 4, our results will be reported. Finally, we discuss and conclude our main findings in Section 5.

2 Dialect variation and change in Austria and South Tyrol

Figure 1 presents the traditional dialect classification of Bavarian in Austria and South Tyrol according to Wiesinger (1983). Note that a very similar classification can be found in earlier research, such as Kranzmayer (1956) (cf. 'Hilfskarte 1').

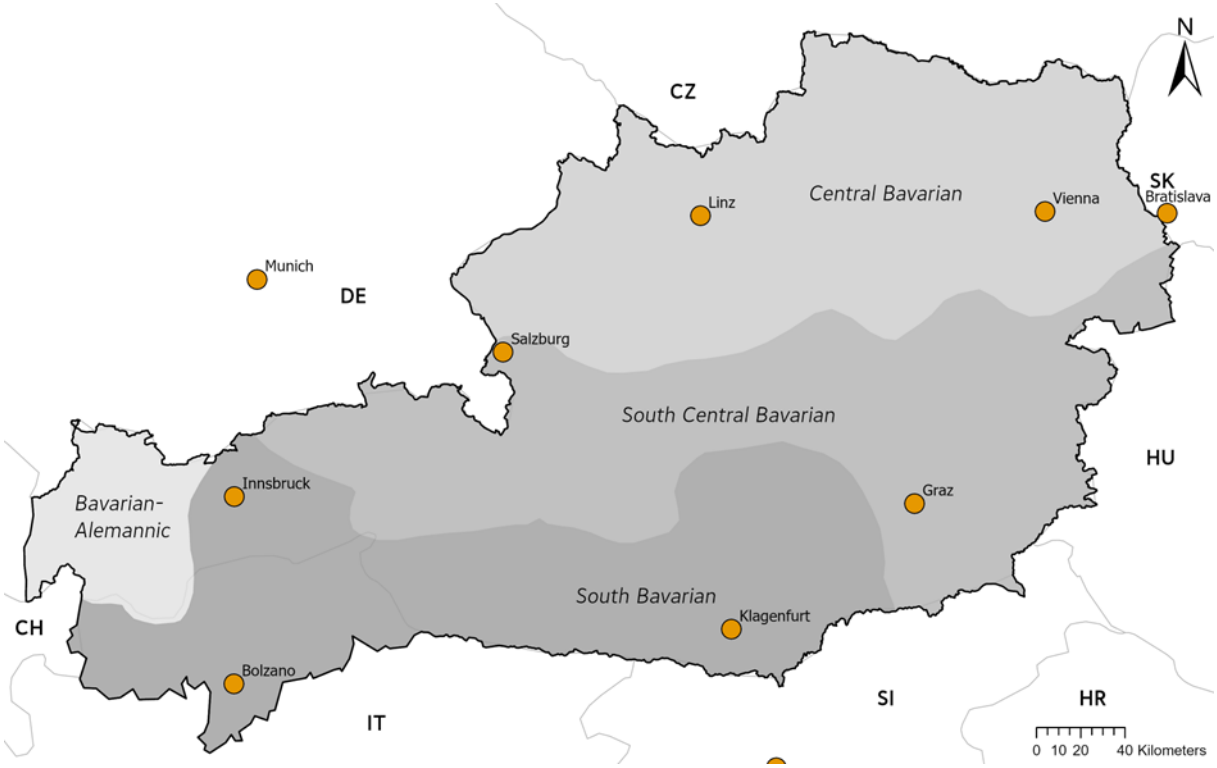

Figure 1: Bavarian dialects of Austria and South Tyrol based on Wiesinger (1983).

The majority of Austria and South Tyrol belong to the Bavarian dialect area, with only the westernmost Austrian state of Vorarlberg and a small part of North Tyrol falling within the Alemannic dialect group. Traditionally, the Bavarian dialects of Austria and South Tyrol are subdivided into three major parts: Central Bavarian (in northern Austria), South Bavarian (in southern and western Austria as well as in South Tyrol), and a broad South Central Bavarian transition zone in-between. Additionally, there is a transition zone between Bavarian and Alemannic in the western parts of Tyrol. These areas can be further divided into even smaller regions, such as East Central Bavarian and West Central Bavarian (see Wiesinger 1990).

The traditional dialect classification is based on a structuralist (qualitative) approach, focusing primarily on features of phonology and inflectional morphology. Although previous dialectometric research on dialect phonology – using both historical (see Stöckle 2024) and contemporary data (see Vergeiner 2025) – indicates a fairly good correspondence with this classification, there are some notable differences, e.g. regarding the South Central Bavarian area. Such differences are also evident when comparing the dialectometric studies themselves. However, direct comparison is challenging due to differences in regional scope (for example, Vergeiner 2025 includes the Alemannic regions of Austria, while Stöckle 2024 includes South Tyrol), the linguistic variables investigated, and the statistical methods used.[1] To overcome these obstacles, the present study will investigate the same areas and variables, employing the same statistical methods to ensure that any potential differences can be more easily interpreted as results of real-time change.

It is very plausible that such changes have indeed occurred. Although previous research suggests that some phonological features have remained remarkably stable over the past century (cf. e.g., Vergeiner et al. 2021b; Vergeiner & Wallner 2022), several studies demonstrate significant changes in individual variables (cf. e.g., Scheutz 1985, 2016; Scheuringer 1990; Moosmüller & Scheutz 2013; Vergeiner et al. 2021a; Vergeiner 2022). These studies reveal patterns of change, which are also observed in other German-speaking areas (see also Lenz 2019: 338–349; for other German-speaking regions e.g., Schmidt & Herrgen 2011):

While there is no general loss of dialects, significant dialect leveling has occurred due to dialect-to-standard

convergence, a phenomenon typical of diaglossic situations, such as those in most of Austria (see Auer 2005). Dialect change appears to be particularly pronounced for small-scale dialect features (Vergeiner et al. 2021b), which are being replaced by either standard German variants or more widespread dialect features. This results not only in increased regionalization of dialects but also in the movement of dialect isoglosses.[2] Several studies (e.g., Bülow et al. 2019; Vergeiner et al. 2021a; Vergeiner 2022) indicate that this process is leading to the diffusion of Central Bavarian features and a reduction in South Bavarian features, particularly in southeastern Austria. Generally, dialects in the west tend to be more stable, while those in the east appear to change more rapidly, possibly due to the influence of large urban centers, especially Vienna (e.g., Lenz 2019).

Our study aims to investigate whether these trends observed in individual dialect features have led to significant changes in the broader geolinguistic patterns of dialect phonology in Austria.

3 Data and methods

In this section, we present our data in more detail, starting with the historical data (Section 3.1) and then proceeding to the contemporary data (Section 3.2). In Section 3.3, we describe our variable set, and in Section 3.4 we outline our dialectometric approach.

3.1 Historical data

Our historical data set is based on the corpus of 'Dictionary of Bavarian Dialects in Austria' (*Wörterbuch der bairischen Mundarten in Österreich,* WBÖ). The WBÖ is a lexicographical long-term project at the Austrian Academy of Sciences that aims to research and document the Bavarian dialects that were part of the Habsburg Empire in 1912, when the project was founded.[3] The following decades were characterized by an intensive and comprehensive collection of material, in which mostly voluntary collectors gathered language data using questionnaires, noted it down on slips of paper and finally sent them to the dictionary chancelleries. The questionnaire-based collection was later supplemented by surveys, which were carried out by trained explorers in the form of field trips. In addition, excerpts from dialectological literature and other written sources were compiled, which made it possible to close gaps in the material.[4] In total, a collection of around 3 million paper slips was compiled in this way, with most material dating from the first half of the 20th century. To speed up the lexicographical work, the digitization of the paper slips was initiated in the 1990s (cf. Barabas et al. 2010).[5] Eventually, a database of around 2.4 million data entries was converted into XML/TEI format (cf. Bowers & Stöckle 2018) and is now publicly accessible via the "Lexical Information System Austria" (LIÖ).[6]

As the WBÖ material is comparatively heterogeneous, a number of things must be considered for its use in quantitative research. In terms of their composition and origin, the questionnaire-based data correspond most closely to a uniform corpus (such as the contemporary data used in this study). One challenge, however, lies in the transcription conventions, some of which were handled differently by the individual collectors, which can have an impact on quality, particularly in the case of phonological phenomena. For the present study, therefore, only those phenomena were selected for which transcripts were available in sufficiently reliable quality. Lameli et al. (2020) were able to show in their dialectometric studies on Swiss German Wenker data that, despite the underlying lay spelling, robust results can be achieved that largely correspond to other traditional and dialectometric dialect classifications in German-speaking Switzerland (for similar results see Lameli 2013; Vergeiner & Bülow 2023). In the case of literary excerpts, a distinction must also be made between academic texts (dissertations, dialect dictionaries) and poetic texts written in dialect (so-called "literarische Denkmäler der Mundart", 'literary monuments of dialect'; Geyer 2019: 486). For reasons of uniformity – both in terms of the survey methods and the time period in which it was collected – only data from the questionnaire surveys, the field trips and academic texts were used for this study, whereas the excerpts from dialect literature were not taken into account.

Although pronunciation is a central aspect of dialect lexicography, it was not systematically recorded in the dataset used for this analysis. As a result, a number of records lack information on pronunciation, limiting the possibilities of phonological comparisons. Additionally, the response rate during the data collection process was not uniform across regions, leading to an uneven geographical distribution of data. In areas with lower

data density, it was sometimes necessary to supplement the analysis by using multiple lexemes for a single phonological variable to fill these gaps (see Table 1). A further challenge arose from considerable variation in phonetic transcriptions. This variation is largely due to the involvement of a high number of collectors, each of whom may have had differing transcription practices or phonetic interpretation standards, resulting in inconsistencies within the dataset. Despite these challenges, linguistic types could be summarized based on the available transcriptions. For a detailed discussion of the methodological approaches used to account for transcription variation and the categorization of the data, see Stöckle (2024).

3.2 Contemporary data

Our contemporary data are sourced from two datasets: For Austria, we utilize the comprehensive dialect corpus of the project 'Variation and Change of Dialect Varieties in Austria (in Real and Apparent Time)'.[7] For South Tyrol, we draw on data from the project 'German Dialects in South Tyrol' (*Deutsche Dialekte in Südtirol*; see Scheutz 2016).

Both datasets were collected using the same methodology: The data consist of recordings of direct dialect interviews conducted by trained fieldworkers. The interviews were based on a traditional dialect questionnaire, which was very similar in both projects, as the questionnaire for the Austrian survey was modeled after the one used for South Tyrol. In both projects, speakers from rural locations were interviewed, with at least one older and one younger speaker selected per location. Apart from this, the selection process in both surveys adhered to traditional dialectological sampling criteria (Chambers & Trudgill 1998: 29–30). In sum, our contemporary data consist of 293 recordings from 109 locations: 269 recordings from Austria (97 locations) and 24 recordings from South Tyrol (12 locations).

One of the main challenges in the comparison of the two datasets was the establishment of a uniform network of locations or regions onto which the linguistic data could be projected. The contemporary data were collected from 109 survey locations, whereas the data for the WBÖ were gathered by collectors from over 2,000 municipalities across the research area. For the WBÖ, these municipalities were organized into regions

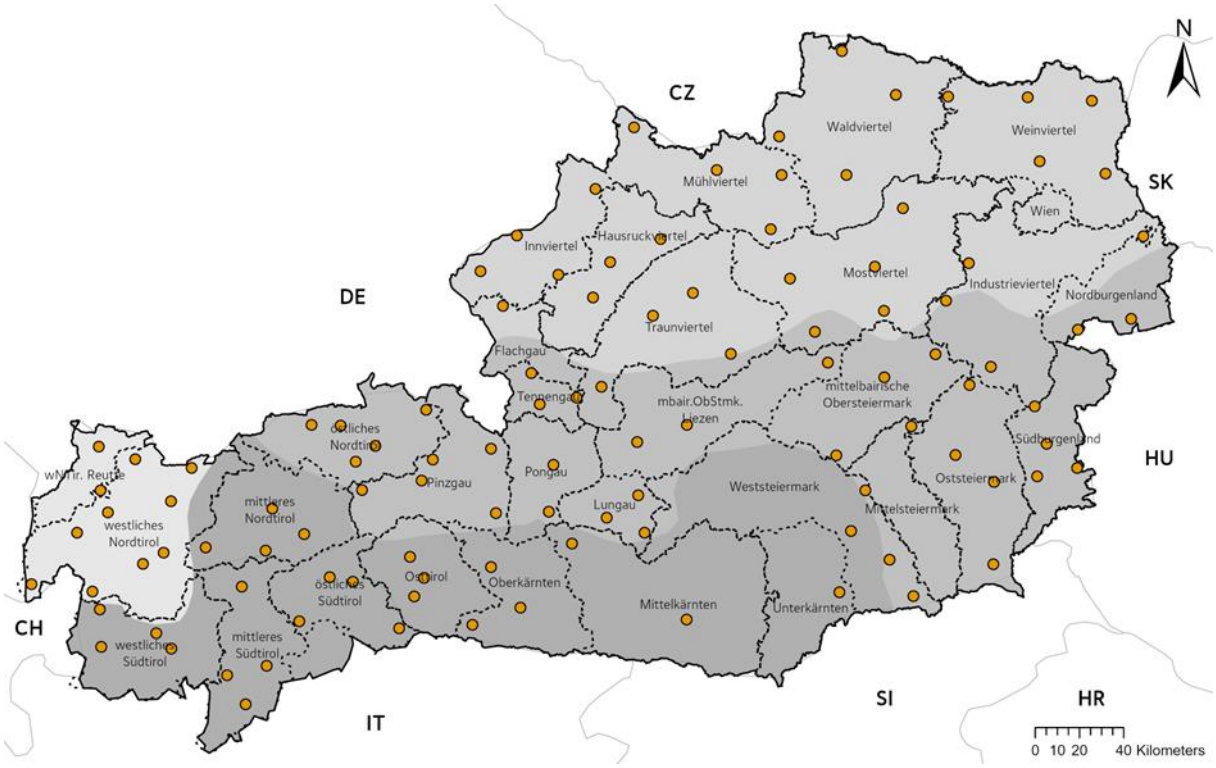

Figure 2: Research locations in the contemporary data and so-called "major regions" according to the WBÖ classification.

following a hierarchical system: at the lowest level, municipalities were grouped into small regions, which were then combined into major regions. At the highest level, the regions correspond to federal states.

To achieve the greatest possible commonality between the two datasets, while still preserving a sufficient level of regional differentiation, we decided to use the WBÖ major regions as geographical framework. Data from both datasets were projected onto these major regions with slight modifications to optimize their alignment with the distribution of the contemporary data.[8] One notable difference remains: While Vienna is not represented in the contemporary data, it constitutes a major region in the historical WBÖ data. Given the substantial volume of data originating from Vienna in the WBÖ, we decided to retain this region for the historical analyses. Therefore, the historical dataset consists of 32 regions, while the contemporary dataset consists of 31 regions.

Figure 2 displays the 109 contemporary survey locations along with the major regions according to the WBÖ classification. The results of the dialectometric analyses will be projected onto the centroids of these regions, as shown in Figure 5 and subsequent figures.

While data collection for the contemporary datasets from Austria and South Tyrol was largely similar, it is important to highlight key differences when comparing them with the historical data. The contemporary data were collected by a small, highly trained team of researchers, ensuring a consistent and controlled data set that accurately represents the (intended) base dialect. In contrast, the historical data were mostly collected by voluntary contributors, which results in less standardization and reduced control over the quality and consistency of the data.[9] This variability is particularly pronounced in the phonetic transcriptions. Additionally, the absence of audio recordings for the historical data makes it impossible to verify the transcriptions directly. As a result, phonological classification must be based solely on the written dialect transcriptions. This limitation necessitates careful interpretation of the data, particularly in the analysis of phonological features. Notably, however, prior research at various linguistic levels (e.g., Stöckle et al. 2021, Stöckle & Wittibschlager 2022) has demonstrated that, with the necessary caution, the historical data can be effectively analyzed and successfully compared to contemporary datasets, therefore providing a solid basis for comparison on a quantitative level.

3.3 Variables

To investigate the geolinguistic patterns in dialect phonology, we selected a set of 31 variables that can be compared between the historical and contemporary data. Admittedly, this is a relatively small data sample for a dialectometric analysis – especially considering that these 31 variables classify approximately the same number of geographic datapoints (31 in the contemporary data and 32 in the historical data).[10] Nevertheless, this dataset is sufficient to reveal clear geographical patterns. We restricted the dataset for two primary reasons: First, unlike most classical dialectometric analyses that rely on preexisting atlas data, our study is based on corpus data that required specific annotation. Second, we needed to select only those variables that were available and comparable in both datasets.

In addition to data availability and comparability, variable selection was based on previous research (for an overview, cf. e.g., Kranzmayer 1956; Wiesinger 1983, 1990; Lenz 2019) and aimed to capture the main phonological differences among the Bavarian dialects of Austria and South Tyrol. The variables cover various phonological levels, including vocalic phenomena in both stressed (#1 – #17) and unstressed positions (#18 – #25), as well as consonantal phenomena (#26 – #31). Table 1 presents the set of variables, which are defined according to the Middle High German (MHG) proto-system, a standard approach in German dialectology. Additionally, Table 1 also shows which lexemes were annotated from the different datasets. Whenever possible, we selected the same lexemes or, if not feasible, lexemes with similar phonological contexts to ensure comparability.

To enhance comparability, we coded the variables using a similar set of variants across all three datasets. To this end, the spoken data from Austria and South Tyrol were annotated using a broad phonological transcription based on traditional dialectological research, which intentionally avoided phonetic details to better match the layperson transcriptions in the WBÖ data. Subsequently, different spelling forms in the WBÖ data that appeared to represent the same phonological

Table 1: Variable set and lexemes annotated in the different datasets (WBÖ = historical data; DiÖ = contemporary data for Austria; ST = contemporary data for South Tyrol).

#	**variable**	**lexemes: WBÖ**	**lexemes: DiÖ**	**lexemes: ST**
1	MHG *â*	*Schaf, schlafen*	*Schaf, schlafen*	*Waage, schlafen*
2	MHG *ô*	*stoßen*	*hoch, rot, tot*	*hoch, rot, tot*
3	MHG *ê*	*Schnee, See*	*Schnee, See*	*Schnee, gehe*
4	MHG *œ* before *n*	*schön*	*schön*	*Schön*
5	MHG *ei*	*heiß, Seife*	*heiß, Meister*	*heiß, zwei*
6	MHG *ei* before *n*	*klein*	*Stein*	*Stein*
7	MHG *iu~ui*	*Feuer*	*Feuer, heuer*	*Feuer, heuer*
8	MHG *uo*	*Hut, gut*	*Fuß, gut*	*Fuß, Schuh*
9	MHG *ou* before *f*	*kaufen*	*kaufen, raufen*	*kaufen, raufen*
10	MHG o (lengthened)	*Hose*	*Ofen*	*Ofen*
11	MHG *o/ë* in *kommen*	*kommen*	*kommen*	*Kommen*
12	MHG *ë* + *r*	*Kerze, Herz*	*Kerze*	*Kerze*
13	MHG *o* + *r*	*morgen, Dorf*	*morgen*	*Morgen*
14	MHG *a* + *r* before *t(s)*	*hart, Garten*	*Bart, schwarz*	*Bart, schwarz*
15	MHG *a* + *l*	*Salz*	*kalt, alt*	*kalt, Stall*
16	MHG *i* + *l*	*spielen, viel*	*Bild, hilfst*	*Bild, hilf*
17	MHG *o* + *l*	*Holz*	*Holz, Wolf*	*Holz, Wolke*
18	prefix MHG *ge-* before plosive	*gekauft, getan, gekonnt, getrocknet*	*gebraucht, gebacken*	*gebissen, gebraten*
19	wordfinal MHG *-en* after nasal	*kommen, schwimmen*	*bringen, brennen*	*bringen, kommen*
20	wordfinal MHG *-en* after *ch*	*machen*	*machen, brauchen*	*machen, brauchen*
21	wordfinal MHG *-en* after *f*	*kaufen*	*helfen, raufen*	*helfen, raufen*
22	wordfinal MHG *-e*	*Hase, Henne*	*Schuhe, Tage*	*Schuhe, Tage*
23	wordfinal MHG *-r*	*Feuer*	*Feuer, Bauer*	*Feuer, Bauer*
24	wordfinal MHG *-el*	*Löffel, Apfel*	*Löffel, Nebel*	*Löffel, Nebel*
25	epenthetic vowel after MHG *l*	*Milch*	*Milch*	*Milch*
26	initial MHG *k(ch)* before *n*, *l*	*Knödel*	*Knödel, Klee*	*Knecht, klein*
27	postvocalic MHG *-s-* before *t*	*gestern, Gast*	*Ast, Meister*	*Mist, gestern*
28	intervocalic MHG *-b-*	*Gabel*	*lieber, gröber*	*lieber, gröber*
29	intervocalic MHG *-t-*	*Wetter*	*Wetter, füttern*	*Wetter, Butter*
30	wordfinal MHG *-g*	*Weg*	*Tag, Weg*	*Tag, Weg*
31	wordfinal MHG *-n*	*Mann*	*Wein, Stein*	*Wein, Stein*

types as the (main) variants identified in the spoken data were grouped together.[11]

Since both the historical and contemporary data contain different responses per variable and location, relative frequency distributions could be calculated for all variants. These relative frequencies were then used to aggregate the data, as described in the next section.

3.4 Statistical methods

To aggregate our data and reveal more general geolinguistic patterns in our data, we employ aggregative dialectometric methods (see, e.g., Wieling & Nerbonne 2015). These methods have proven effective in various studies across different languages and varieties (e.g., Heeringa 2004; Heeringa et al. 2009; Szmrecsanyi 2013; Lameli 2013; Pröll 2015; Scherrer & Stöckle 2016). As noted above (see Section 2), this also includes previous research conducted with the current data corpora (see e.g., Stöckle 2024 for the WBÖ data; Vergeiner 2023, 2025 for the DiÖ data).

The first step in conducting the dialectometric analyses is data aggregation. This is achieved by calculating the differences in dialect features between every pair of locations in the dataset, generating a site-by-site

distance matrix. Since this study uses numerical data (relative frequency values), distances can be computed using Euclidean distance, a standard measure for this type of data (e.g., Szmrecsanyi 2013: 28). The simplest way to obtain a distance matrix would be to calculate the Euclidean distance directly from the relative frequencies of all variants. However, since each variant belongs to a variable and the number of variants per variable varies, this approach would dis-proportionately weight variables with more variants. To avoid this, we adopted a more refined method: first computing separate distance matrices for each variable and then summing the distances to obtain a final distance matrix for all 31 variables.[12] This method produced two final distance matrices – one for contemporary data and one for historical data. Based on these distance matrices, different statistical analyses can be applied. The present study employs multidimensional scaling (MDS) and cluster analysis (CA), two of the most frequently used methods in dialectometry.

MDS visualizes relationships between objects based on their pairwise similarities or dissimilarities, aiming to reduce high-dimensional data to a lower-dimensional space while preserving the original distances as much as possible (e.g., Borg et al. 2018). In dialectometry, MDS is particularly well-suited for displaying dialect continua. To achieve this, a three-dimensional solution is typically used, with each dimension mapped to a color in the RGB (red, green, blue) color space. This allows locations to be represented as blends of these colors according to their coordinates in the MDS space (Nerbonne et al. 1999). For our analysis, we employ interval MDS with stress minimization via majorization, using the *mds* function from the *R* package *smacof* (Mair et al. 2022).

CA aims to form clusters with high intra-group similarity and significant inter-group differences (Backhaus et al. 2021: 452–454). To group locations into (distinct) clusters and to reveal the hierarchical relationships among these clusters, this study uses hierarchical agglomerative CA. As a clustering algorithm, we employ WPGMA (Weighted Pair Group Method with Arithmetic Mean).[13] A common limitation of CA is that it does not reveal the linguistic basis of cluster solutions. To overcome this, we use the cluster determinant method described by Prokić et al. (2012). This method identifies items that are most uniform within a cluster (i.e., minimal within-cluster difference) and most distinct when compared with locations outside the cluster (i.e., maximal between-cluster difference). A normalized score is then calculated to reveal which features might be regarded as "shibboleths". To compute the CA, we use the *hclust* function from the *R* package *stats* (R Core Team 2022). Cluster determinants are calculated using the *Gabmap* web application (Leinonen et al. 2016). All maps in this article were created using the *ArcGIS Pro* software.

4 Results

In this section, we report our results, starting with a comparison of the linguistic distances in the historical and contemporary data (Section 4.1). We then explore the geolinguistic patterns in these data using both MDS and CA (Section 4.2).

4.1 Linguistic distances in the historical and contemporary data

Table 2 provides key statistics for the aggregate linguistic distances in both the historical and contemporary data, while the histograms in Figure 3 display the distribution of these distances across both data sets.

As shown in Table 2, the mean and median values of both datasets are quite similar. However, the linguistic distances in the historical data are slightly higher, reflecting greater overall dissimilarity. In contrast, the

Table 2: Descriptive statistics of the aggregate linguistic distances in the historical and contemporary data.

	median	mean	sd	min	max	skew	kurtosis
historical data	18.8	18.8	5.2	5.7	32.7	0.01	-0.63
contemporary data	18.0	17.8	6.9	2.5	32.9	-0.04	-0.99

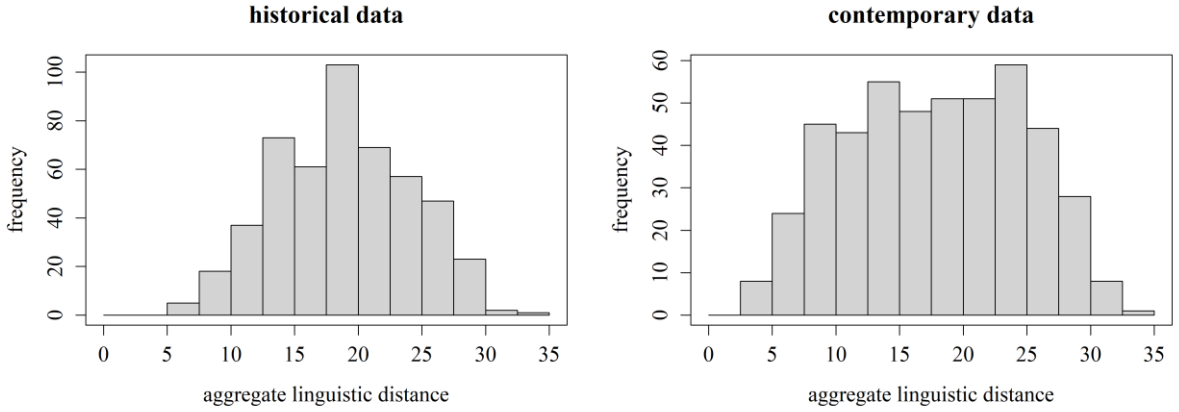

Figure 3: Distribution of the aggregate linguistic distances in the historical and the contemporary data.

larger standard deviation in the contemporary data indicates greater variability in linguistic distances. Both datasets exhibit skewness values close to zero, suggesting near symmetry, while the negative kurtosis values indicate flatter distributions with fewer extreme values (outliers) than a normal distribution. This flattening effect is more pronounced in the contemporary data. As illustrated in Figure 3, the contemporary dataset displays a broader and more nuanced variation within a specific range, whereas the historical dataset shows more tightly clustered values. This highlights the greater heterogeneity of linguistic distances in the contemporary data compared to the historical dataset.

It is also possible to map the historical[14] and contemporary distances directly onto each other to illustrate their relationship (see Figure 4).

Figure 4 shows a clear positive correlation[15] between the two distance matrices, indicating that locations with greater dissimilarities in the historical data also exhibit greater dissimilarities in the contemporary data. However, the correlation is far from perfect, which may

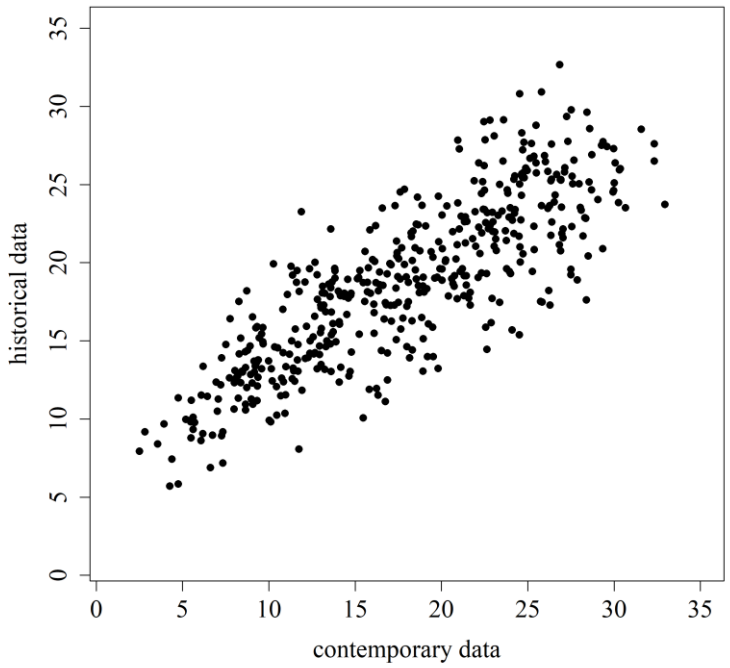

Figure 4: Correlation between the aggregate linguistic distances in the historical and contemporary data.

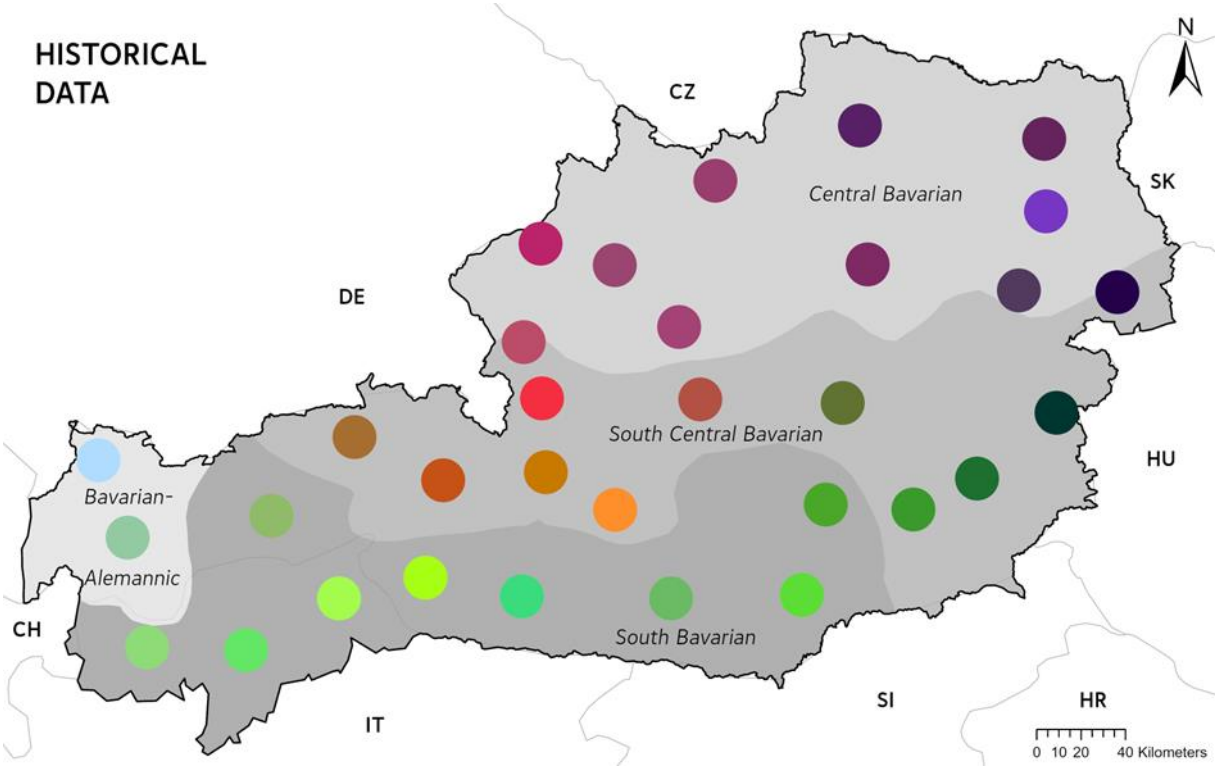

Figure 5: MDS map of the historical data.

be attributed to language change in real-time. In the next section, we will explore these differences in greater detail using CA and MDS to detect geolinguistic patterns in the data.

4.2 Geolinguistic patterns in the historical and contemporary data

Multidimensional Scaling

As noted above (Section 3.4), MDS allows for the representation of complex, high-dimensional data in a more interpretable, lower-dimensional format. For both the historical and contemporary data, a three-dimensional MDS has been computed and mapped onto the RGB color space (dimension 1 = green, dimension 2 = red, dimension 3 = blue). Figure 5 displays the results for the historical data (stress = 0.09), with each location (centroid of each respective region) colored according to their position in the MDS space. It is important to note that the colors themselves do not carry meaning; rather, they serve to display similarities, with locations that are linguistically similar being colored similarly.

Figure 5 reveals certain geographical structures in the data: A large continuous area in the south, displayed in green, covers most of Tyrol, Carinthia, and large parts of Styria. This area corresponds well with the South Bavarian region in the traditional dialect classification of Austria (Wiesinger 1983; see also Section 2). To the north, within Central and South Central Bavarian, there appears to be a broad continuum between a more western (reddish) pole in the greater Salzburg region[16] and western Upper Austria, and a more eastern (bluish) pole in Lower Austria and Burgenland. A similar east-west divide is also recognized by traditional dialectology (cf. Wiesinger 1990: 463). Within the western area, the Salzburgian Lungau region is evident due to its orange coloring (for the linguistic particularities of the region, cf. Mauser 2021), while in the east, the city of Vienna is distinguished by its purple coloring (for Vienna, see Wiesinger 1990: 465). Also notably distinct is the Reutte region in the far west, which stands out from its neighboring areas due to its light blue coloring. This is not surprising, as the region is heavily influenced by Alemannic (Swabian) dialects from the north (see Wiesinger 1990: 482–485).

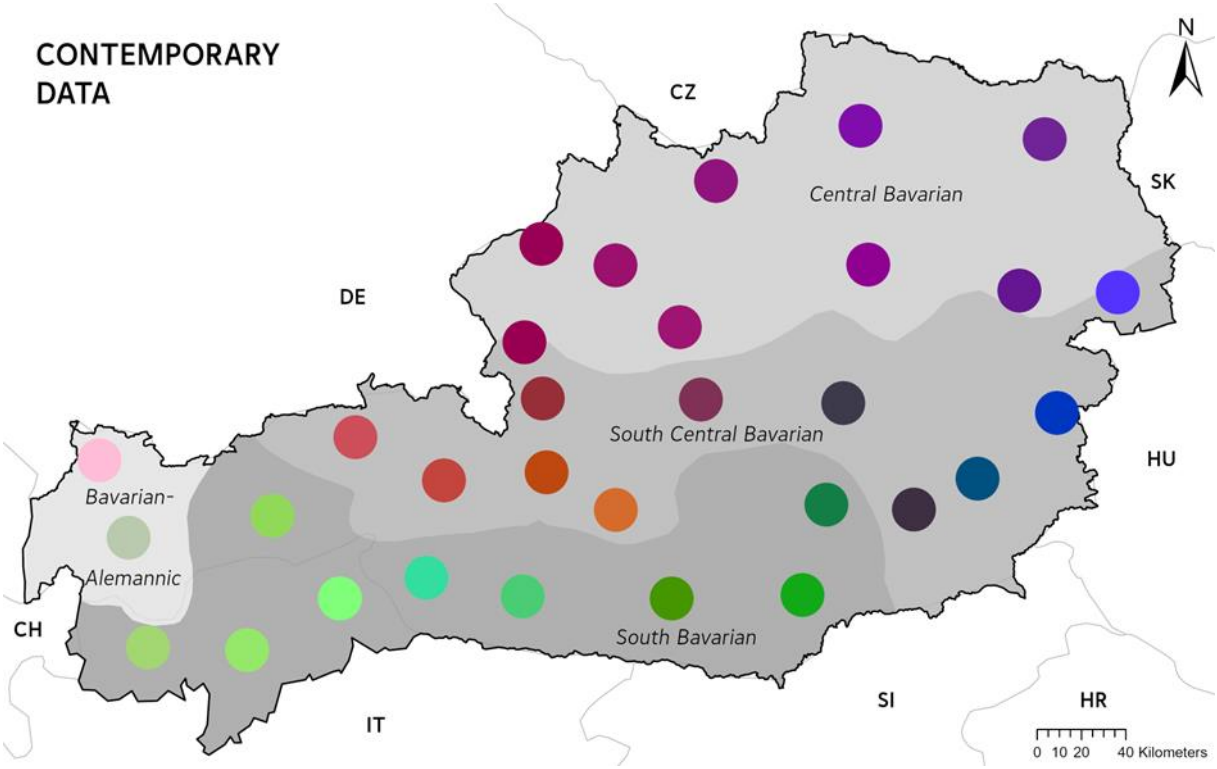

Figure 6: MDS map of the contemporary data.

The results for the historical data can be compared with those for the contemporary data, shown in Figure 6 (stress-1 = 0.06).

The results for the contemporary data closely resemble those for the historical data.[17] Again, there is a southern area in green that can be associated with South Bavarian. However, this region appears somewhat diminished, particularly in the east, where the Styrian locations are shown in greenish-black. To the north, three areas are more distinctly identifiable: Burgenland in blue, Upper and Lower Austria in purple, and the greater Salzburg region in various shades of orange. In the far west, Reutte is clearly distinguished by its light pink coloring. Overall, these differences support findings from other studies that, while the South Bavarian region is gradually diminished in the east, there are also tendencies toward greater regionalization and less continuous patterns outside the Central Bavarian area (see e.g., Vergeiner et al. 2021a; Vergeiner 2022). In the next sections, we will expand on these findings by applying an additional exploratory statistical method (CA) and investigating the linguistic basis for these differences.

Cluster Analysis

As mentioned above (see Section 3.4), we used the WGPMA algorithm to perform a hierarchical cluster analysis (CA). The clustering is done step by step and leads to a tree-like structure, i.e. a dendrogram. The clusters (or "dendremes", cf. Goebl 2006: 421) obtained in this way can ultimately be mapped and interpreted as dialect areas. In contrast to non-hierarchical methods such as the *k*-means method (cf. Bortz & Schuster 2010: 461), the number of groups is not fixed in advance in this approach. The decision for a particular number of clusters may be guided by theoretical considerations (e.g., comparisons with existing dialect classifications) or by specific statistical methods for determining the optimal number of clusters (cf. Everitt et al. 2011: 95). However, to understand the gradually emerging dialect-geographical differentiation of the research area, we will not discuss one cluster solution in what follows but compare different groupings with each other. Specifically, we compare visualizations for two (Figure 7), three (Figure 8), four (Figure 9) and five (Figure 10) clusters for the historical and contemporary data.

If the data sets are divided into two regions (see Figure 7), a comparable picture emerges for both the historical and the contemporary data, which also corresponds largely to the traditional dialect classification (interestingly, with a better fit in the contemporary data): a division into a southern region (green) and a larger northern region (purple), with the southern region matching the South Bavarian dialect area (including the Bavarian Alemannic transition zone in the west).

The division into three cluster reveals substantial differences between the historical and contemporary data (see Figure 8). In the historical data, the large (South) Central Bavarian region splits into two distinct sub-regions, following a clear west-east divide (purple vs. blue). In the north, this division follows the border between Upper and Lower Austria; moving to the south, the eastern region includes Burgenland and almost all of Styria, while the western cluster encompasses the

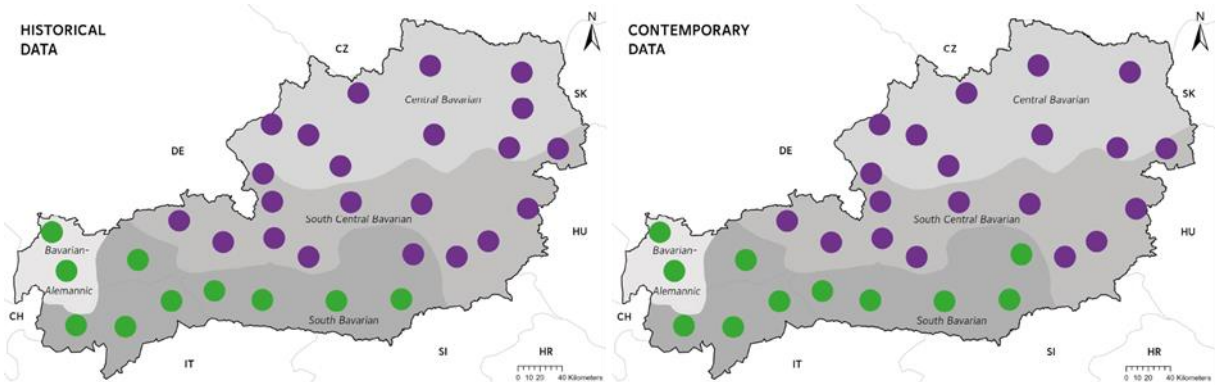

Figure 7: Two-cluster solution for the historical and contemporary data.

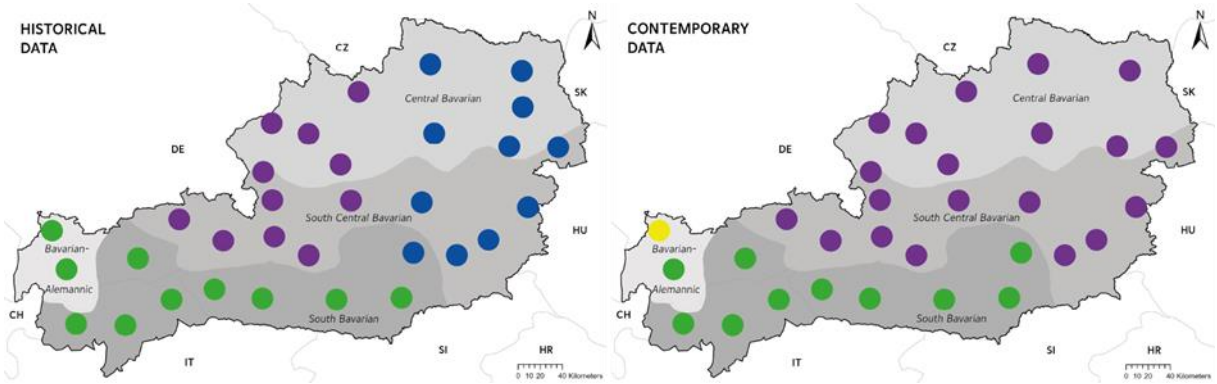

Figure 8: Three-cluster solution for the historical and contemporary data.

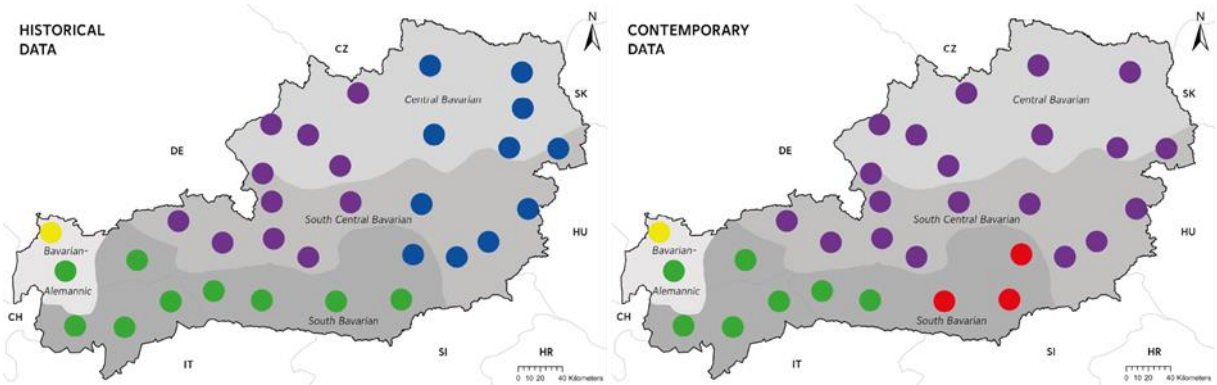

Figure 9: Four-cluster solution for the historical and contemporary data.

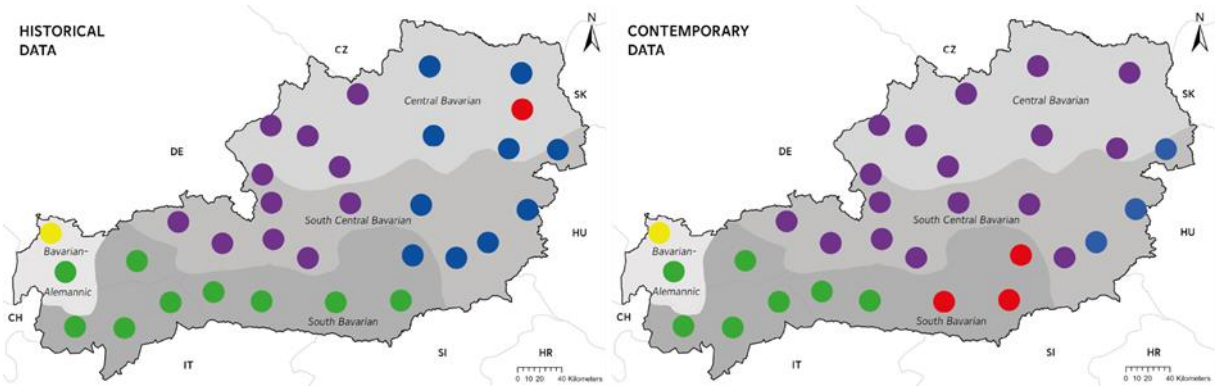

Figure 10: Five-cluster solution for the historical and contemporary data.

greater Salzburg region. In contrast, in the contemporary data, a cluster splits off in the far west from the South Bavarian area (green), which consists of only one region, the district of Reutte (yellow). As already explained in the context of the MDS results, the influence of Alemannic is clearly visible here.

Dividing the area into four clusters (see Figure 9), we observe a similar pattern in the historical data as in the three-cluster solution for the contemporary data (see Figure 8), with the district of Reutte forming a separate cluster (yellow). In comparison, in the contemporary data, the large northern area (purple) remains stable, while a west-east division is visible in South Bavarian. This partition, which was already clear in the MDS, runs through the federal state of Carinthia, separating Upper Carinthia and Tyrol in the west (green) from Central Carinthia and south-western Styria in the east (red).

Our final division into five clusters reveals some additional differences between the historical and contemporary data (see Figure 10). In the historical data, Vienna separates into its own cluster from the rest of the eastern region (red). In the contemporary data, the northern area is subdivided (purple vs. blue) with the eastern parts of South Central Bavarian (i.e. Burgenland and the eastern regions of Styria) forming its own cluster (cf. Moser et al. 2022 for the particularities of this region).

For a linguistic explanation and interpretation of these results, the linguistic variants most strongly associated with the clusters (i.e., the cluster determinants) will be examined in more detail in the next section.

Cluster determinants

In this section, we focus on the characteristic features of the dialect regions under investigation. Rather than analyzing all possible cluster solutions, we will concentrate on two solutions: the first, which presents the broadest division into two clusters, and the most differentiated solution, which consists of five clusters. This approach allows us to observe both the overarching patterns and the finer distinctions between the dialectal groups.

As outlined in section 3.4, for each cluster, we identified the variants that "differ the least within the given group and still differ a great deal with respect to the sites outside the group" (Prokic et al. 2012: 76). Notably, since this method compares the within-cluster differences with the between-cluster differences, the absence of a particular feature can also become a defining characteristic of a cluster. In the following tables we therefore included a "Freq" column, where an upward arrow indicates a notably high frequency of a variant, and a downward arrow denotes a particularly low frequency.[18] Another consequence of this procedure is that clusters consisting of only one site, such as Reutte and Vienna, cannot have cluster determinants calculated, as variance within a single-site cluster cannot be computed. However, between-cluster differences can be specified for these clusters, allowing for conclusions to be drawn about the most important characteristics of these regions, too. For each cluster, we will discuss the three most prominent variants in what follows, starting with the historical data before turning to the contemporary data.

Table 3: Cluster determinants for the historical data (for Vienna and Reutte only between-scores).

	Region	Variable	Variant	Example[19]	Freq	Score
2 clusters	north	MHG *i* + *l*	/iːl/	/ʃpiːln/ (*spielen*)	↓	1.61
		postvocalic MHG *-s-* before *t*	/ʃt/	/gɛʃtɐn/ (*gestern*)	↓	1.52
		prefix MHG *ge-* before plosive	deletion	/kaːft/ (*gekauft*)	↑	1.44
	south	wordfinal MHG *-el*	/ə, ɪ/	/lɛfɪ/ (*Löffel*)	↓	1.86
		MHG *i* + *l*	/iːl/	/ʃpiːln/ (*spielen*)	↑	1.79
		MHG *a* + *l*	/ɔɻ/	/sɔɻts/ (*Salz*)	↓	1.68
5 clusters	north-central	MHG *ei* before *n*	/ɔɐ̯/	/klɔɐ̯n/ (*klein*)	↑	1.23
		wordfinal MHG *-en* after nasal	/ɐ/	/kɔmɐ/ (*kommen*)	↑	1.16
		wordfinal MHG *-el*	/ə, ɪ/	/lɛfɪ/ (*Löffel*)	↑	1.07
	north-east	MHG *o* + *r*	/ɔr/	/mɔrgŋ/ (*morgen*)	↓	1.62
		MHG *iu~ui*	/ɔɪ/	/fɔɪɐ/ (*Feuer*)	↓	1.04
		initial MHG *k(ch)* before *n, l*	/kx/	/kxneːdl/ (*Knödel*)	↓	0.94
	south	wordfinal MHG *-el*	/ə, ɪ/	/lɛfɪ/ (*Löffel*)	↓	1.78
		MHG *i* + *l*	/iːl/	/ʃpiːln/ (*spielen*)	↑	1.64
		MHG *a* + *l*	/ɔɻ/	/sɔɻts/ (*Salz*)	↓	1.60
	Vienna	MHG *o/ë* in *kommen*	/ʊ/	/kʊmɐ/ (*kommen*)	↑	(3.18)
		MHG *ei* before *n*	/aː/	/klaːn/ (*klein*)	↑	(2.80)
		wordfinal MHG *-g*	/ç/	/veːç/ (*Weg*)	↑	(2.25)
	Reutte	MHG *ou* before *f*	/ɔʊ/	/kɔʊfə/ (*kaufen*)	↑	(3.65)
		MHG *ei* before *n*	/uɐ/	/kluɐn/ (*klein*)	↑	(2.50)
		MHG *o/ë* in *kommen*	/ɛɪ/	/kɛɪmə/ (*kommen*)	↑	(2.40)

First, we examine the key linguistic features that distinguish the northern and southern clusters identified in the 2-cluster classification for the historical data (see Table 3).

The northern cluster is primarily defined by the absence of the /iːl/ realization for MHG *i* + *l*, a key feature of the southern cluster. Instead, the northern cluster exhibits variants with mostly vocalized vowels (e.g., /ʃpyːn/). Another characteristic is postvocalic MHG *-s-* before *t*, resulting either in a palatalized form (e.g., /gɛʃtɐn/), common in Alemannic and South Bavarian dialects, or an alveolar fricative (e.g., /gɛstɐn/), prevalent in most other Bavarian dialects (see e.g., Lenz 2019: 330; Vergeiner & Wallner 2022). A third characteristic of the northern cluster is the deletion of the prefix MHG *ge-* before a plosive, as in *kauft* for *gekauft* ('bought'), marking a distinctive (morpho)phonological trait in northern dialects (see e.g., Wiesinger 1989: 63–67; Vergeiner et al. 2021b).

The southern cluster, in contrast, is defined by distinct features with higher overall scores. This indicates a greater uniformity, a pattern already evident in the results of the multidimensional scaling (MDS) and cluster analysis (CA). The most prominent feature is the realization of word-final MHG *-el*, appearing as either a retained lateral (/lɛfl/) or a vocalized form (/lɛfɪ/), with the lateral dominating in the south. Another key feature is the strong preference for /iːl/ as a realization of MHG *i* + *l*. A third defining trait is the realization of MHG *a* + *l*, as in *Salz* ('salt'): while vocalized forms appear in the north, the south retains *l* (e.g., /sɔlts/).

Altogether, the cluster determinants point to the importance of the *l*-vocalization in the north vs. the preservation of *l* in the south as an overall distinctive phenomenon for the dialect classification. This result aligns well with traditional dialectology where postvocalic *l* is viewed as the key feature for differentiating Bavarian dialects (see Wiesinger 1983: 840).

For the 5-cluster classification, the following results can be observed: The north-central cluster covers parts of Central and South Central Bavarian. Its most prominent characteristic is the realization of MHG *ei* before *n* as /ɔɐ/, a common Bavarian feature. However, Vienna and some South Bavarian dialects in Carinthia and Tyrol use /aː/, while /ʊɐ/ appears on the eastern and western edges of South Bavarian, and /iː/ in Alemannic transition areas (see e.g., Gabriel 1994: 97). Another key feature is the reduction of word-final MHG *-en* after nasals to schwa (/ɐ/), in contrast to most southern dialects, which retain the nasal (see Wiesinger 1989: 13–25; Vergeiner & Wallner 2022). Lastly, MHG *-el* is vocalized to /ə/ or /ɪ/, a feature also relevant to the two-cluster classification (see above).

The north-eastern cluster covers the eastern parts of Central and South Central Bavarian, as well as western Styria in South Bavarian. The highest scoring feature in this area is the realization of MHG *o* + *r*, where *r* is vocalized and appears as /ɔɐ/ or /ʊɐ/. Unlike in other regions, the /r/ sound is not preserved here (see Kranzmayer 1956: 38; Vergeiner 2022). The second feature is the realization of MHG *iu/ui* in words like *Feuer* ('fire'). In this cluster, we find the variant /aɪ/, while in other regions the more traditional diphthongs /ʊɪ/ or /ɔɪ/ are retained (see Vergeiner in press). A third significant feature is the realization of initial MHG *k(ch)* before *n* or *l*. In contrast to other regions where it is realized as an affricate, here it appears as a fortis plosive /k/, or in some cases within Central Bavarian, it is lenited to /g/ (see also Wiesinger 1990: 457–458, 477).

The southern cluster largely corresponds to the South Bavarian dialect region. The most noteworthy feature here is the consistent realization of *l* in wordfinal MHG *-el*, and following MHG *i* and *a*. This contrasts with the vocalization observed in other clusters.

The Vienna cluster stands out, as it comprises only one geographical location. The most significant characteristic is the realization of the main vowel in *kommen* as /ʊ/, a variant that is occasionally found in the east of South (Central) Bavarian, but is otherwise less common, with /ɛ/ being more widespread in other regions. The second feature, shared with the north-central cluster, is the realization of MHG *ei* as a /aː/, which has become a defining feature of Viennese speech and is widespread in eastern Austria (see Wiesinger 1990: 465, Lenz 2019: 329). The third feature is the realization of word-final MHG *-g*, which frequently appears in Vienna as the fricative /ç/, while in all other regions the plosive /k/ is more common.

The Reutte cluster, consisting of a single location at the transition zone to the Alemannic dialect area, exhibits several distinct features aligned with its proximity to Alemannic-speaking regions. The most prominent feature here is the realization of MHG *ou* before *f*, which takes the form of the diphthong /ɔʊ/. In most other areas, this is realized as /aʊ/ or /aː/ (see Lenz 2019: 329; Vergeiner et al. 2021b). The second feature concerns MHG *ei* before *n* (cf. also the north-central cluster and Vienna), which is being realized almost exclusively as /ʊɐ/. Lastly, the third key feature is the realization of MHG *o/ë* in *kommen* (cf. the Vienna cluster), which is realized as /ɛɪ/ in Reutte.

Next, we will discuss the cluster determinants for the contemporary data, as displayed in table 4. In the 2-cluster classification, MHG *o* + *l* emerges as the most defining feature for both clusters. In the northern cluster, the absence of /ɔl/ is most characteristic, while in the southern cluster, the absence of /ɔɪ/ is particularly distinctive. This pattern can be explained by the fact that, although /ɔɪ/ is the primary variant of MHG *o* + *l* in the north, other vocalized variants such as /ʊɪ/ are also frequently found in this area. In the south, however, /ɔl/ dominates alongside other non-vocalized variants like /ʊl/. Thus, each cluster is defined by the absence of the main variant that predominates in the other cluster. The other variants most distinctive for the northern cluster include the deletion of schwa in the MHG prefix *ge-* before plosives and the deletion of word-final MHG *-n*. Both variants also appear in some southern regions, but with less frequency. For the southern cluster, the absence of *l*-vocalization in the context of MHG *a* + *l* and word-final MHG *-el* are the second and third most characteristic features.

In summary, the cluster determinants for the 2-cluster solution resemble those found in the historical data. Once again, the results confirm the importance of the *l*-vocalization in classifying Austrian dialects. The other two features – deletion of MHG *ge-* before plosives and word-final MHG *-n* – are also recognized as key distinctions between Central and South Bavarian

Table 4: Cluster determinants in the contemporary data (for Reutte only between-scores).

	Region	Variable	Variant	Example	Freq	Score
2 clusters	north	MHG *o* + *l*	/ɔl/	/hɔlts/ (*Holz*)	↓	1.61
		MHG *ge-* before plosive	deletion	/bɔxŋ/ (*gebacken*)	↑	1.59
		wordfinal MHG *-n*	deletion	/mɔ̃ː/ (*Mann*)	↑	1.54
	south	MHG *o* + *l*	/ɔɪ/	/hɔɪts/ (*Holz*)	↓	1.87
		MHG *a* + *l*	/ɔɪ/	/kɔɪt/ (*kalt*)	↓	1.82
		wordfinal MHG *-el*	/ə, ɪ/	/lɛfə/ (*Löffel*)	↓	1.70
5 clusters	north	MHG *a* + *l*	/ɔɻ/	/kɔɪt/ (*kalt*)	↑	2.01
		MHG *o* + *l*	/ɔɪ/	/hɔɪts/ (*Holz*)	↑	1.90
		wordfinal MHG *-el*	/ə, ɪ/	/lɛfə/ (*Löffel*)	↑	1.56
	south-east	wordfinal MHG *-g*	deletion	/vɛɻ/ (*Weg*)	↑	2.23
		MHG *æ* before *n*	/ɛɪ/	/ʃɛɪn/ (*schön*)	↑	1.74
		MHG *o* (lengthened)	/ɔʊ/	/ɔʊfm/ (*Ofen*)	↑	1.19
	south	initial MHG *k*(*ch*) before *n*, *l*	/kʰ/	/kʰnɛdl̩/ (*Knödel*)	↑	1.89
		MHG *o* + *l*	/ɔɪ/	/hɔɪts/ (*Holz*)	↓	1.33
		wordfinal MHG *-en* after nasal	/ɐn/	/brɛnɐn/ (*brennen*)	↑	1.24
	south-west	MHG *o* + *l*	/ɔl/	/hɔlts/ (*Holz*)	↑	2.07
		MHG *i* + *l*	/ɪl/	/bɪlt/ (*Bild*)	↑	1.96
		postvocalic MHG *-s-* before *t*	/ʃt/	/ɔʃt/ (*Ast*)	↑	1.95
	Reutte	MHG *ou* before *f*	/ɔʊ/	/kɔʊfə/ (*kaufen*)	↑	(3.31)*
		MHG *o*/*ë* in *kommen*	/ɛɪ/	/kɛɪmə/	↑	(2.07)*
		intervocalic MHG -b-	/b/	/lɪɐbr/	↑	(1.96)*

dialects in traditional dialectology (see, e.g., Wiesinger 1990: 459, 476). Strikingly, the southern cluster displays again higher scores, indicating the greater uniformity of the southern dialects.

For the five-cluster solution, the following results are observed: The northern cluster encompasses what is traditionally considered Central Bavarian, but also great parts of South Central Bavarian (i.e., Upper and Lower Austria, Salzburg, and some parts of Styria). In this cluster, all three determinants pertain to *l*-vocalization, resulting in /ɔɪ/ as a reflex of MHG *a* + *l* and MHG *o* + *l*, and /ə, ɪ/ for word-final MHG *-el*.

The south-eastern cluster comprises large parts of the area traditionally regarded as eastern South Central Bavarian (for the particularities of this region, see e.g., Moser et al. 2022). This cluster is primarily distinguished by the deletion of word final *-g* in words like *Weg* ('way'). Although this feature is not unique to this area, it has been well preserved due to the region's linguistic conservatism. Another key feature is the realization of the diphthong /ɛɪ/ for MHG *æ* before *n*. This feature sets these dialects apart from those in the north, where the monophthong /ɛː/ prevails, and from the dialects in the west, which use centralizing diphthongs (/ɛɐ/ or /ɪɐ/; see Vergeiner et al. 2021a). A third significant feature is the realization of /ɔʊ/ for MHG *o* when lengthened; this variant is a typical feature of eastern South Central Bavarian (Wiesinger 1990: 471), and might even be diffusing to other phonological contexts in this region (see Vergeiner et al. 2021a).

The southern cluster includes eastern South Bavarian, i.e. most parts of Carinthia and western Styria. For initial MHG *k*(*ch*) before *n* and *l*, these dialects are characterized by the presence of an aspirated plosive (/kʰ/) rather than forms with lenition (/g/) or an affricate (/kx/ or /kç/) (see above for the historical data). Other characteristics of this region are the absence of vocalized /ɔɪ/ for MHG *o* + *l* and the realization of word-final MHG *-en* after a nasal as /ɐn/, whereas most northern dialects have /ɐ/ (as also found for the historical data).

The western cluster encompasses western Carinthia and most of Tyrol, i.e. the western half of the South Bavarian region. The most distinctive features of this region are the non-vocalized variants /ɔl/ and /ɪl/ for MHG *o* + *l* and MHG *i* + *l*, respectively. Another defining characteristic of this region is the palatalization of MHG *-s-* before *t* (see above for the historical data).

In the northwestern-most part of Tyrol, the Reutte region forms its own distinct cluster. Just as in the

historical data, the most distinct features of the Reutte dialect include the diphthong /ɔʊ/ in place of MHG *ou* before *f*, and the use of /ɛɪ/ in the word *kommen* ('come'). Another characteristic is the retention of the intervocalic plosive MHG -*b*-, as opposed to the lenition to /β/, which is widespread in Bavarian dialects (Wiesinger 1990: 453).

In sum, the cluster determinants show a strong consistency between the historical and contemporary data, with a north-south distribution pattern that remains stable across both datasets. Key features, especially the vocalization of /l/ and the prefix *ge-*, exhibit a high degree of alignment. Also, within the five-cluster solution, the vocalization of /l/ plays an important role, distinguishing the (South) Central Bavarian dialect clusters from South Bavarian dialects. Further characteristic features in both data sets include the realization of ending *-en* in infinitive forms after nasals, the *s*-palatalization, and initial *k* before *n*, *l*. The differentiation of Reutte as a separate cluster, marked by Alemannic features such as /ɔʊ/ in place of MHG *ou* before *f*, is another consistent finding across both datasets. Furthermore, there is substantial agreement with existing literature on dialectal features, though one must bear in mind that our selection of features was based on this literature. Interestingly, however, some features typically used to differentiate Central and South Bavarian, such as the diphthongisation of MHG *ê* or *ô*, the lenition of MHG -*t*- or the vocalisation of postvocalic *r* (see Wiesinger 1990: 457–459), do not appear central to this cluster classification. A more detailed examination of individual features and their developments might further clarify the clustering patterns and their potential changes. However, we defer this to future research and conclude our analysis with a broader discussion of the overarching findings from our comparative analysis.

5 Discussion and conclusion

In recent decades, quantitative dialectology has experienced a notable surge. However, despite numerous dialectometric studies examining geolinguistic patterns across various regions, studies on language change using these methods – especially real-time studies comparing older and newer data sets – remain a significant desideratum (but see, e.g., Pröll et al. 2022 working with historical data; for dialectometric studies using the apparent time construct, see e.g., Wieling & Nerbonne 2015: 251–253). One important reason for this is the scarcity of suitable data for such analyses. However, in the German-speaking area, including Austria and South Tyrol, the long-standing dialectological tradition offers opportunities for such empirical comparisons (Schmidt & Herrgen 2011: 89). It was the aim of this paper to address this issue and to conduct the first-ever real-time dialectometric study for the Bavarian dialects of Austria and South Tyrol. In doing so, we compared the geolinguistic patterns and their linguistic foundations across two data sets: (1) historical data from the first half of the 20th century, drawn from the 'Dictionary of Bavarian Dialects in Austria' (WBÖ), and (2) contemporary data from the projects 'Variation and Change of Dialect Varieties in Austria (in Real and Apparent Time)' and 'German Dialects in South Tyrol'. By aggregating over a set of 31 phonological variables, we applied the same statistical methods to both data sets for detecting geolinguistic patterns, namely multidimensional scaling (MDS) and cluster analysis (CA). To account for the linguistic foundations of these patterns, we also computed cluster determinants (Prokić et al. 2012).

Although there are notable differences between the datasets in terms of data collection and representation (see Section 3), our comparison of the aggregated historical and contemporary data yielded fruitful results. Overall, we found that the main geolinguistic patterns are remarkably similar in both datasets. The aggregated linguistic distances between the research sites in the historical and contemporary data exhibit a strong correlation. Additionally, the MDS results and key aspects of the CA analyses demonstrate a high degree of similarity. Furthermore, our general findings align well with traditional dialect classifications, such as those proposed by Wiesinger (1983).

In both datasets, the most prominent difference within our research area lies between the dialects of the south (most of Tyrol, Carinthia, and parts of Styria) and those in the rest of the region. According to the cluster determinants, the most significant phonological difference – one could call it a "shibboleth" – between these regions is the behavior of postvocalic *l*, which is preserved as a consonant in the south but vocalized in

the north. This finding is, of course, not unexpected, as postvocalic *l* is recognized as a key feature for subclassifying Bavarian dialects in traditional dialectology (see, e.g., Wiesinger 1983: 840). Our southern area corresponds closely to what is traditionally labeled as South Bavarian.

Another major correspondence between our datasets and traditional dialectology is the exceptional behavior of the dialects in the northwestern edge of Tyrol, specifically in the Reutte region. In both the MDS and CA analyses, this region is clearly distinguished from the rest of South Bavarian, aligning with the traditional view that it may even be classified as part of the Alemannic (Swabian) dialects rather than Bavarian. This also corresponds to our finding that features such as /ɔʊ̯/ for MHG *ou* and /b/ for intervocalic MHG *-b-* are characteristic of this region.[20]

However, our findings also reveal some key differences with the traditional dialect classification. We were unable to identify a distinct South-Central transition zone between the southern and northern dialects. In traditional dialectology, this zone is typically explained by the "various rates at which innovations from Central Bavarian have entered the south" (Wiesinger 1983: 471). However, few features are consistently shared across the dialects in this area and not found elsewhere, which may explain why this region did not emerge clearly in our analyses. Instead of distinguishing between Central and South Central Bavarian, our analyses emphasized east-west differences. While traditional dialectology also recognizes these differences, it typically considers them of secondary importance. To understand this, however, it is important to note that these approaches generally account for the Bavarian dialects across Germany, Austria, and South Tyrol. In this context, the tripartite division into North, Central, and South Bavarian, with intermediate transition zones, still appears plausible.

The most significant difference between our data sets lies in the specific manifestation of these east-west differences. In the historical data, a broad continuum of eastern and western dialects cuts across the northern and central parts of the research area (with the city of Vienna as a kind of 'dialect island' in the eastern half). In contrast, the contemporary data show greater uniformity in the north (Central Bavarian), while the east-west distinctions are more pronounced in the southern dialects. Accordingly, the MDS analysis suggests a gradual decline of the South Bavarian region in its eastern half in the contemporary data. Finally, our results also indicate greater fragmentation and less continuous patterns in the contemporary data, particularly for the southern and western dialects, compared to the historical data.

These differences may – at least partially – be attributed to language change. Our results suggest a retreat of South Bavarian in the south-east, some dialect leveling in northern Austria, and increased regionalization in the south and west. Similar patterns have been identified in studies on individual dialect features (see e.g., Vergeiner et al. 2021a; Vergeiner 2022). However, one should be cautious not to overestimate these hypothetical patterns of change. It is important to consider that the data sets were collected differently, with the contemporary survey being much more controlled in terms of participant selection and data transcription. Moreover, the regional coverage of the contemporary data is less dense and limited to rural locations. Ultimately, these differences could explain why the contemporary data appear less gradual and show clearer distinctions.

Bearing this in mind, the main finding of our real-time analysis is the considerable stability over the past century in the regional structuring of Bavarian dialects in Austria and South Tyrol. Although this finding is supported by other studies focusing on individual dialect features (e.g., Vergeiner et al. 2021b; Vergeiner & Wallner 2022), it is often assumed that significant dialect leveling has occurred due to dialect-to-standard convergence, a phenomenon typical in diglossic situations such as those in most of Austria (see Auer 2005). While we cannot dismiss this possibility, it has not fundamentally altered the overall geographical structure of Bavarian dialects in Austria and South Tyrol, at least in our data sets (see also Vergeiner & Bülow 2023 for the whole Alpine region). It is important to emphasize, however, that our contemporary data only represent the most traditional, rural dialects. Moreover, due to the methods used for data collection, it cannot be ruled out that the informants may have reported forms during the interviews that they no longer use in their everyday lives (cf. e.g., Streck 2012

and Schwarz 2015, who identified considerable differences between data from direct dialect surveys and actual corpus data), potentially obscuring language change. A comparison with more modern, innovative forms of dialect would be valuable, which remains a topic for future research.

In sum, our study demonstrates that real-time comparisons using dialectometric methods are indeed feasible. Despite various methodological challenges, such studies can validate and extend findings from studies on individual features by incorporating larger datasets and employing more objective analytical methods.

Author Statement

Both authors contributed equally to the study's design, data preparation, and analysis. They were also equally involved in drafting, revising, and finalizing the manuscript for submission.

Data availability

Relevant data and supporting materials for this study are made available on OSF under the following URL:

https://osf.io/6d7f4/?view_only=135fd6d914f64bed88af41d999df9077

Endnotes

1 In this regard, it is important to note that the statistical methods used in dialectometry are highly dependent on the data, and even small differences in the datasets (e.g., other regions or variables) can significantly affect the results.

2 One might question whether this leads to the replacement of dialects by regiolects, a trend observed in other German-speaking regions as well (cf. Schmidt & Herrgen 2011). However, this issue is beyond the scope of our article, and we follow the tradition of Austrian dialectology in referring to these varieties as dialects as well (cf. also Wiesinger 1990, who distinguishes between *Verkehrsdialekt* 'contact dialect' and *Basisdialekt* 'base dialect').

3 The project was originally founded together with the dictionary chancellery in Munich with the aim of creating a joint "Bavarian-Austrian dictionary". However, the two dictionary projects separated in the mid-1950s and since then the "Bavarian Dictionary" in Munich and the WBÖ in Vienna have been edited separately.

4 For more details on the history of the WBÖ, cf. Geyer (2019) and Stöckle (2021).

5 As the lemmas for the initial letters *A*, *B*/*P* and *C* had already been edited in the first WBÖ volumes, the digitization started with the letter *D*.

6 See https://lioe.dioe.at [accessed 21.10.2024]

7 The project is funded by the Austrian Science Fund (F 6002-G23) as part of the Special Research Program 'German in Austria' (F 60).

8 Specifically, the major regions "mittelbairische Obersteiermark" ('Central Bavarian Upper Styria') and "westliches Nordtirol" ('Western North Tyrol') were each subdivided into two smaller regions. Additionally, "südbairische Obersteiermark" ('South Bavarian Upper Styria') was merged with "Weststeiermark" ('Western Styria'), and "Mittelburgenland" ('Central Burgenland') was combined with "Südburgenland" ('Southern Burgenland').

9 Naturally, the overall concept of the WBÖ was also designed to document the rural base dialects. For further details on this topic, see Geyer (2019: 485–488).

10 However, cf. e.g., Lameli (2013), who demonstrated that it is possible to classify the varieties in Germany using just 66 morpho-phonological variables, while Szmrecsanyi's (2013) dialectometric analysis of English in Great Britain is based on only 57 variants (some of which belong to the same variables).

11 For some variants, however, this was not feasible, resulting in a larger set of variants for the historical data (n/variants = 132) compared to the contemporary data (n/variants = 104). This difference, however, involves only minor variants with a low overall frequency.

12 We would like to thank Johann-Mattis List for pointing this out to us.

13 We opted for this algorithm over others (e.g., the frequently employed Ward algorithm) because it showed a better fit with our original distance matrices, as indicated by the higher cophenetic correlation.

14 Since the contemporary data does not include Vienna, the distance values for this location are excluded from the historical data in this comparison.

15 Using a Mantel test, a strong correlation of $r = 0.8$ was found ($p < 0.000$***).

16 We use this term to refer to the province of Salzburg, the Liezen district in Styria, and north-eastern Tyrol, which were historically closely connected to Salzburg, leading to many linguistic similarities (see Wiesinger 1990: 473).

17 Using a Mantel test, a strong correlation of r = 0.82 is found (p < 0.000***) between the distances in the MDS space for both data sets.

18 Note that for binary variables – i.e., those with only two possible realizations – both variants yield identical cluster determinant values. We have chosen to present only the variant that deviates from the standard as the reference point in the tables. This provides a more informative picture of the dialect-specific features that distinguish each cluster.

19 As explained in section 3.1, the dialect pronunciation in the historical data was noted on paper slips by the collectors, who used the transcription system Teuthonista, which was very common in traditional German dialectology. Due to the resulting range of different transcriptions (cf. Stöckle 2024: 306) and for better comparability with the contemporary data, all examples represent categorized variants and are reproduced in IPA.

20 It may be surprising that the distinction between Bavarian and Alemannic did not emerge as the most significant. However, this can be explained by two factors: First, western Tyrol can be considered a transition zone between Bavarian and Alemannic, with even the dialects in Reutte sharing many features with other Tyrolean dialects. Second, our feature selection focused on internal variation within Bavarian rather than differences between Bavarian and Alemannic.

References

Auer, Peter. 2005. Europe's sociolinguistic unity, or: a typology of European dialect/standard constellations. In Nicole Delbecque, Johan van der Auwera, & Dirk Geeraerts (eds.), *Perspectives on variation. Sociolinguistic, historical, comparative*, 7–42. Berlin: de Gruyter.

Backhaus, Klaus, Bernd Erichson, Sonja Gensler, Rolf Weiber & Thomas Weiber. 2021. *Multivariate Analysis. An Application-Oriented Introduction*. Wiesbaden: Springer.

Barabas, Bettina, Claudia Hareter-Kroiss, Birgit Hofstetter, Lana Mayer, Barbara Piringer & Sonja Schwaiger. 2010. Digitalisierung handschriftlicher Mundartbelege. Herausforderungen einer Datenbank. *Germanistische Linguistik* 199–201. 47–64.

Borg, Ingwer, Patrick J. F. Groenen & Patrick Mair. 2018. *Applied Multidimensional Scaling and Unfolding*. 2nd ed. Cham: Springer.

Bortz, Jürgen & Christof Schuster. 2010. *Statistik für Human- und Sozialwissenschaftler*. 7th ed. Berlin: Springer.

Bowers, Jack & Philipp Stöckle. 2018. TEI and Bavarian dialect resources in Austria: updates from the DBÖ and WBÖ. In Andrew U. Frank, Christine Ivanovic, Francesco Mambrini, Marco Passarotti, & Caroline Sporleder (eds.), *Proceedings of the Second Workshop on Corpus-Based Research in the Humanities (CRH-2)*, 45–54. Vienna: Gerastree Proceedings.

Bülow, Lars, Philip C. Vergeiner, Hannes Scheutz, Dominik Wallner & Stephan Elspass. 2019. Wie regelhaft ist Lautwandel? Grundsätzliche Überlegungen anhand einer Fallstudie zu mhd. *ô* in den bairischen Dialekten Österreichs und Südtirols. *Zeitschrift für Dialektologie und Linguistik* 86(1). 5–50.

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

Everitt, Brian S., Sabine Landau, Morven Leese & Daniel Stahl. 2011. *Cluster Analysis*. 5th ed. London: Wiley

Fleischer, Jürg. 2017. *Geschichte, Anlage und Durchführung der Fragebogen-Erhebungen von Georg Wenkers 40 Sätzen: Dokumentation, Entdeckungen und Neubewertungen*. Hildesheim: Olms.

Gabriel, Eugen (ed.). 1994. *Vorarlberger Sprachatlas mit Einschluss des Fürstentums Liechtenstein, Westtirols und des Allgäus (VALTS). Vol. II. Lautgeographie II. Langvokale und Diphthonge, Silbendehnungen und -kürzungen*. Bregenz: Vorarlberger Landesregierung.

Geyer, Ingeborg. 2019. Wörterbuch der bairischen Mundarten in Österreich: Rückblick auf 105 Jahre Erheben, Aufbereiten und Auswerten im institutionellen Rahmen der ÖAW. In Sebastian Kürschner, Mechthild Habermann, & Peter O. Müller (eds.), *Methodik moderner Dialektforschung. Erhebung und Auswertung von Daten am Beispiel des Oberdeutschen*, 471–488. Hildesheim: Olms.

Goebl, Hans. 2006. Recent Advances in Salzburg Dialectometry. *Literary and Linguistic Computing* 21(4). 411–435.

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

Heeringa, Wilbert, Martijn Wieling, Boudewijn van den Berg & John Nerbonne. 2009. A quantitative examination of variation in Dutch Low Saxon morphology. In Alexandra N. Lenz, Charlotte Gooskens & Siemon Reker (eds.), *Low Saxon Dialects across Borders*, 195–216. Stuttgart: Steiner.

Kranzmayer, Eberhard. 1956. *Historische Lautgeographie des gesamtbairischen Dialektraumes*. Vienna: Böhlau.

Lameli, Alfred. 2013. *Strukturen im Sprachraum. Analysen zur arealtypologischen Komplexität der Dialekte in Deutschland.* Berlin: de Gruyter.

Lameli, Alfred, Elvira Glaser & Philipp Stöckle. 2020. Drawing areal information from a corpus of noisy dialect data. *Journal of Linguistic Geography* 8(1). 1–18.

Leinonen, Therese, Çağrı Çöltekin & John Nerbonne. 2016. Using Gabmap. *Lingua* 178. 71–83.

Lenz, Alexandra N. 2019. Bairisch und Alemannisch in Österreich. In Joachim Herrgen & Jürgen Erich Schmidt (eds.), *Language and Space. An International Handbook of Linguistic Variation*, Volume 4: *Deutsch*, 318–363. Berlin: de Gruyter.

Lessiak, Primus. 1963 [1903]. *Die Mundart von Pernegg in Kärnten.* Marburg: Elwert.

Mair, Patrick, Patrick J. F. Groenen & Jan de Leeuw. 2022. More on multidimensional scaling in R: smacof version 2. *Journal of Statistical Software* 102(10). 1–47.

Mauser, Peter. 2021. *Wiarach ba ins ret: Das Lungauer Sprachbuch.* Tamsweg: Pfeifenberger.

Moosmüller, Sylvia & Hannes Scheutz. 2013. Chain shifts revisited. The case of monophthongisation and E-merger in the city dialects of Salzburg and Vienna. In Peter Auer, Javier Caro Reina & Göz Kaufmann (eds.), *Language variation – European perspectives IV: Selected papers from the Sixth International Conference on Language Variation in Europe (ICLaVE 6), Freiburg, June 2011*, 173–186. Amsterdam: Benjamins.

Moser, Ann-Marie, Philip C. Vergeiner, Nina Weihs & Jan Luttenberger. 2022. Das „Hianzische“ aus empirischer Sicht: Zu Sprachvariation und -wandel im Burgenland. *Wiener Linguistische Gazette* 91. 68–93.

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*, v–xv. Stanford: CSLI.

Prokić, Jelena, Çagri Çöltekin & John Nerbonne. 2012. Detecting Shibboleths. In *Proceedings of the 2012 Joint Workshop of LINGVIS and UNCLH*. Stroudsburgh, PA: Association of Computational Linguistics. 72–80.

Pröll, Simon. 2015. *Raumvariation zwischen Muster und Zufall. Geostatistische Analysen am Beispiel des Sprachatlas von Bayerisch-Schwaben.* Stuttgart: Steiner.

Pröll, Simon, Daniel Hrbek, Johanna Schirmer, Victor Christensen Franziska Haug & Irena Rietzler. 2022. Dialektometrische Analysen graphematischer Variation im Althochdeutschen und Altsächsischen. *Sprachwissenschaft* 47(3). 293–328.

R Core Team. 2022. *R: A language and environment for statistical computing.* R Foundation for Statistical Computing: Vienna. https://www.R-project.org/

Schatz, Josef. 1903. Die tirolische Mundart. *Zeitschrift des Ferdinandeums für Tirol und Vorarlberg* 47. 1–94.

Scherrer, Yves & Philipp Stöckle. 2016. A quantitative approach to Swiss German – Dialectometric analyses and comparisons of linguistic levels. *Dialectologia et Geolinguistica* 24. 92–125.

Scheuringer, Hermann. 1990. *Sprachentwicklung in Bayern und Österreich. Eine Analyse des Substandardverhaltens der Städte Braunau am Inn (Österreich) und Simbach am Inn (Bayern) und ihres Umlandes.* Hamburg: Buske.

Scheutz, Hannes. 1985. *Strukturen der Lautveränderung. Variationslinguistische Studien zur Theorie und Empirie sprachlicher Wandlungsprozesse am Beispiel des Mittelbairischen von Ulrichsberg/Oberösterreich.* Vienna: Braumüller.

Scheutz, Hannes. 2016. *Insre Sproch. Deutsche Dialekte in Südtirol.* Bozen: Athesia.

Schmidt, Jürgen Erich & Joachim Herrgen. 2011. *Sprachdynamik. Eine Einführung in die moderne Regionalsprachenforschung.* Berlin: Schmidt.

Schwarz, Christian. 2015. *Phonologischer Dialektwandel in den alemannischen Basisdialekten Südwestdeutschlands im 20. Jahrhundert. Eine empirische Untersuchung zum Vokalismus.* Stuttgart: Steiner.

Steiner, Carina, Péter Jeszenszky & Adrian Leemann. 2023. Variation and change in Swiss German agreement morphology: Spatial, social, and attitudinal effects. *Journal of Linguistic Geography* 11. 8–24.

Stöckle, Philipp. 2021. Wörterbuch der bairischen Mundarten in Österreich (WBÖ). In Alexandra N. Lenz, & Philipp Stöckle (eds.), *Germanistische Dialektlexikographie zu Beginn des 21. Jahrhunderts*, 11–46. Stuttgart: Steiner.

Stöckle, Philipp. 2024. Historische bairische Dialekte in Österreich und Südtirol – ein dialektometrischer Zugang. *Jahrbuch für Germanistische Sprachgeschichte* 15. 301–316.

Stöckle, Philipp, Christina Hemetsberger & Manuela Stütz. 2021. Die WBÖ-Belegdatenbank als Quelle für syntaktische Analysen – Möglichkeiten, Grenzen, Perspektiven. *Wiener Linguistische Gazette* 89. 579–626.

Stöckle, Philipp & Anja Wittibschlager. 2022. Zur Sprachdynamik des Konjunktivs im Bairischen in Österreich. *Linguistik online* 114(2). 43–65.

Streck, Tobias. 2012. *Phonologischer Wandel im Konsonantismus der alemannischen Dialekte Baden-Württembergs. Sprachatlasvergleich, Spontansprache und dialektometrische Studien.* Stuttgart: Steiner.

Szmrecsanyi, Benedikt. 2013. *Grammatical variation in British English dialects. A study in corpus-based dialectometry.* Cambridge: Cambridge University Press.

Vergeiner, Philip C. 2022. Variation und Wandel des postvokalischen r. Zum intra- und interdialektalen Levelling in Österreich. *Taal en Tongval* 74(2). 215–246.

Vergeiner, Philip C. 2023. Quantitative Perspectives on the Geolinguistic Structures of Dialect Morphology in Austria. *Dialectologia et Geolinguistica* 31. 87–115.

Vergeiner, Philip C. 2025. Dialect Classification and Everyday Culture. A Case Study from Austria. *Languages* 10(2). 17.

Vergeiner, Philip C. In press. Diasystematik und Dialektwandel: Zu den Reflexen von germ. /eu/ in den österreichischen Dialekten. *Linguistik Online*.

Vergeiner, Philip C., Dominik Wallner & Lars Bülow. 2021a. Reihenschritte im rezenten Lautwandel? Ergebnisse einer real- und apparent-time-Untersuchung zur Entwicklung der Vokalreihe mhd. *ê – œ – ô* im Süd- und Südmittelbairischen. *Zeitschrift für Sprachwissenschaft* 40(1). 31–64.

Vergeiner, Philip C., Stephan Elspaß & Dominik Wallner. 2021b. Zur Stabilität dialektaler Formen: Eine real- und apparent-time Analyse remanenter Merkmale in den ruralen Basisdialekten Österreichs. *Linguistik Online* 110(5). 9–32.

Vergeiner, Philip C. & Dominik Wallner. 2022. On explaining stable dialect features: A real- and apparent-time study on the word-final syllable *-en* in Austrian base dialects. *Open Linguistics* 8(1). 65–86.

Vergeiner, Philip C. & Lars Bülow. 2023. Geolinguistic structures of dialect phonology in the German-Speaking Alpine region. A dialectometric approach using crowdsourcing data. *Open Linguistics* 9(1). 20220252.

Vergeiner, Philip C. & Lars Bülow. 2024. Zum Schwund des Präteritumschwunds bei sein und wollen. Eine Fallstudie zur regionalen Sprachgeschichte der österreichischen Dialekte. *Jahrbuch für germanistische Sprachgeschichte* 15. 285–300.

Wieling, Martijn & John Nerbonne. 2015. Advances in Dialectometry. *Annual Review of Linguistics* 1(1). 243–264.

Wiesinger, Peter. 1983. Die Einteilung der deutschen Dialekte. In Werner Besch, Ulrich Knoop, Wolfgang Putschke & Herbert Wiegand (eds.), *Dialektologie*, Zweiter Halbband, 807–900. Berlin: de Gruyter.

Wiesinger, Peter. 1989. *Die Flexionsmorphologie des Verbums im Bairischen*. Vienna: Verlag der Österreichischen Akademie der Wissenschaften.

Wiesinger, Peter. 1990. The Central and Southern Bavarian Dialects in Bavaria and Austria. In Charles V. Russ (ed.), *The Dialects of modern German. A linguistic survey*, 438–519. Stanford: Stanford University Press.