Research Article

Philip C. Vergeiner* and Lars Bülow

Geolinguistic structures of dialect phonology in the German-speaking Alpine region: A dialectometric approach using crowdsourcing data

https://doi.org/10.1515/opli-2022-0252
received June 27, 2023; accepted October 25, 2023

Abstract: The Alpine region stands out in the German-speaking world for its well-preserved traditional dialects, which continue to play a significant role in daily life. However, the vast geographical range of the Alpine region and the limitations imposed by national and regional borders have hindered comprehensive investigations of the entire Alpine area. To overcome these obstacles, this study utilizes crowdsourcing data from the VerbaAlpina project to investigate phonological and morpho-phonological variation in the German-speaking Alpine region. Although the data were collected in a written form and with a focus on lexical variation, it is shown that the data contain detailed phonological information. By using multivariate dialectometric measurements (i.e. factor analysis [FA]) based on 19 (morpho-)phonological variables (comprising 8,582 tokens), the study explores the geolinguistic structures of the German dialects in the Alpine region. The results confirm, on a general level, the validity of the traditionally established dialect classification. However, the findings also reveal previously underrated border effects, in particular for Bavarian dialects along the German-Austrian border, which are argued to be the outcome of divergence processes. Hence, the study highlights the importance of cross-dialectal and cross-national perspectives in understanding dialect variation and emphasizes the value of crowdsourcing data for dialectological research.

Keywords: phonological variation, dialect geography, dialectometry, German, crowdsourcing

1 Introduction

The Alpine region is a unique area within the German-speaking world because traditional dialects have been well preserved and continue to play a significant role in people's daily lives. Among other things, this has led to extensive research on regional variation within the Alpine region, making it one of the most thoroughly researched dialect areas in the German-speaking world (for a research overview on the different alpine regions, see e.g. Krefeld and Lücke 2014, Eller-Wildfeuer et al. 2018, Streck 2019, Koch 2019, Lenz 2019, Christen 2019, Rabanus et al. 2019, Gaeta and Seiler 2021). Large-scale research projects such as the *Sprachatlas der deutschen Schweiz* ['Language Atlas of German Switzerland'] (SDS), the *Bayerischer Sprachatlas* ['Bavarian language atlas'] (BSA), and the special research program *Deutsch in Österreich* ['German in Austria'] (DiÖ) have been devoted to studying the German dialects in the Alpine area (often encompassing neighbouring non-Alpine regions as well). However, national and regional borders have limited the scope of these projects due to the vast geographical range of the Alpine region, spanning over five countries with German-speaking

*** Corresponding author: Philip C. Vergeiner**, Institute of German Philology, LMU Munich, Germany, e-mail: philip.vergeiner@lmu.de
Lars Bülow: Institute of German Philology, LMU Munich, Germany

populations (Austria, Germany, Switzerland, Liechtenstein, and Italy with the autonomous province of Bozen – South Tyrol) and two major Upper German dialect areas (Bavarian and Alemannic).

In addition to administrative factors, the lack of a comprehensive perspective on the Alpine region can be attributed to a shortage of appropriate data to conduct investigations of the entire area. While German dialectology – starting with Georg Wenker's *Sprachatlas des Deutschen Reichs* ['Language Atlas of the German Empire'] – initially employed indirect (written) dialect surveys to collect a large amount of data across a wide area and in a relatively short period, later research has primarily relied on direct (oral) dialect surveys. Although direct dialect surveys are considered to provide more reliable data, they are significantly more expensive and time consuming, resulting in limitations on regional coverage (for an overview, see e.g. Kunzmann and Mutter 2021, Seiler 2010). Only recently has there been a renewed interest in large-scale indirect dialect surveys due to the emergence of new data collection techniques through the internet and smartphone devices (e.g. Elspaß and Möller 2015; see also the contributions in the edited volume by Leemann and Hilton 2021). One of the first endeavours in this regard within the German-speaking world was the long-term research project *VerbaAlpina* (Krefeld and Lücke 2014, Krefeld and Lücke 2017; www.lmu.de/verbaalpina). As a large-scale digital humanities project, VerbaAlpina aims to document the languages and dialects of the entire Alpine region, including data collection through an online crowdsourcing platform (Kunzmann and Mutter 2020). These data allow for a holistic view of the entire German-speaking Alpine region – and even other Alpine dialects (e.g. Italian, French, and Slovenian dialects), regardless of their language family. Thus, it enables investigation of previously under-researched topics such as the large-scale geolinguistic structuring of the Alpine region, the influence of national borders on dialect structures, and the status of transition zones.

This article aims to utilize the VerbaAlpina crowdsourcing data to explore these issues. Before that, we discuss the applicability of the data for dialectological research. Our study focuses on phonological variation – also including morpho-phonological variation. Given that the VerbaAlpina data were collected in a written form with a particular emphasis on vocabulary, our first objective is to evaluate whether the data are suitable for such an investigation. We will show that the data contain remarkably detailed phonological information and can therefore be transformed into a form that can answer questions about phonological variation and change.

Our study wants to examine the geolinguistic structures of the German-speaking Alpine region. To investigate multivariate patterns, we will use non-aggregative dialectometric analyses (i.e. factor analyses, see Grieve 2014, Pickl and Pröll 2019), abstracting away from individual linguistic variables and revealing more general geographical structures. On the basis of this approach, we will investigate whether the geolinguistic structures in the Alpine region correspond to traditionally assumed dialect areas such as those established by Wiesinger (1983). Notably, these traditionally assumed dialect areas do not correspond to the national borders in the Alpine region. However, research on individual dialect features has shown that different processes of dialect change have taken place since the beginning of the early twentieth century. This includes not only processes of dialect-to-dialect convergence (causing increased regionalisation heavily influenced by urban centres) but – at least in Austria and Germany – also processes of dialect-to-standard convergence (e.g. Bülow et al. 2019, Auer and Schwarz 2015, Vergeiner et al. 2021). At the same time, this may have led to divergence along national – and maybe also regional (e.g. with regards to federal states) – borders. In particular, for Central Bavarian along the Austrian-German border, such divergence processes have previously been reported (cf. e.g. Scheuringer 1990, Scheutz 2007). However, the wider effects of these processes on the geolinguistic structures of the Alpine region remain unexplored and are not taken into account in dialect classifications.

Consequently, the following three research questions are targeted in our study:

- RQ1: Can geographical patterns for dialect phonology in the Alpine region be identified based on indirectly collected crowdsourcing data?
- RQ2: How do the geographical patterns revealed relate to the traditionally assumed dialect areas? If there are any differences – can they be attributed to language change or methodological factors?
- RQ3: What is the role of national and regional borders in the Alpine region, and are they reflected in the geolinguistic structures?

In what follows, we will first provide an overview of dialectological research on the German-speaking Alpine region (Section 2), before we present the VerbaAlpina project and its corpus in more detail (Section 3). In Section 4, we discuss the applicability of our crowdsourcing data for dialectological research and provide more information on how the data were processed and analysed. The results of our dialectometric analyses are presented in Section 5. The findings are discussed and summarised in Section 6.

2 Dialectological research on the German-speaking Alpine region

The Alpine region is characterised by great linguistic heterogeneity, including different official languages (e.g. German, French, Italian, Slovenian, Romansh, Ladin, Friulian) and language varieties (e.g. Bavarian in Germany, Austria and Italy, Lombard in Italy, or Littoral in Slovene and Italy) (for an overview, see Krefeld and Lücke 2014, Gaeta and Seiler 2021). This heterogeneity is particularly evident in the German-speaking area, which is home not only to different standard varieties of German (Austrian Standard German, Swiss Standard German, Federal German Standard German) but also to different dialects (e.g. Bavarian, Alemannic), which will be the focus of our study. Figure 1 provides an overview of the German dialects in the Alpine region (excluding language islands) according to Wiesinger's (1983) classification based on (mostly) phonological characteristics.

As Figure 1 shows, Alemannic and Bavarian dialects are spoken in the German Alpine region, both of which belong to the Upper German dialect group. In the west, the Alemannic dialects are located, which – viewed from north to south – are divided into Low Alemannic, Swabian, Central Alemannic, High Alemannic, and Highest Alemannic, with several transition zones in between (cf. Christen 2019, Lenz 2019, Streck 2019). For the Alpine region, Central Alemannic, High Alemannic, and Highest Alemannic dialects are of central importance. These dialects are spoken in Switzerland, in Liechtenstein, in western Austria (mostly in Vorarlberg), and in south-western Germany (in Bavaria and Baden-Württemberg).

In the east of the Upper German region, the Bavarian dialect group extends into the Alpine region. Mostly on the basis of phonological features, a distinction is made – again viewed from north to south – between North Bavarian, Central Bavarian, and South Bavarian dialects (cf. Koch 2019, Lenz 2019), with Central Bavarian and South Bavarian being particularly relevant for the Alpine region. Again, in-between these dialect areas, there are transition zones, most notably the South-Central Bavarian transition zone. In addition, there is

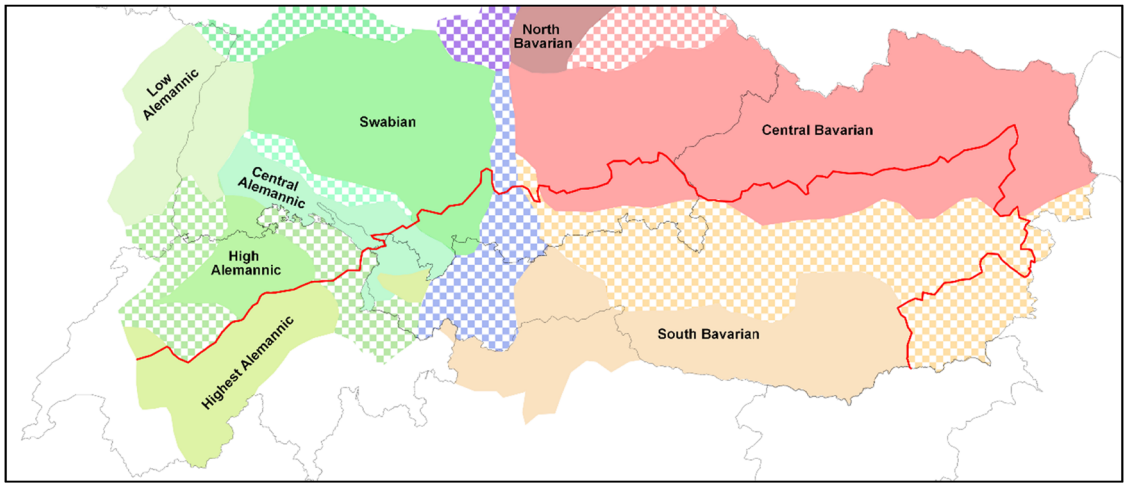

Figure 1: Bavarian and Alemannic dialects according to Wiesinger (1983) (the chequered areas are transition zones, the red line marks the borders of the Alpine region according to the Alpine Convention; for further information on the Alpine Convention, see https://www.alpconv.org/en/home/convention/).

a Bavarian-Alemannic transition zone. Bavarian is spoken in large parts of Bavaria in Germany, in most of Austria (except Vorarlberg), and in South Tyrol in Italy.

The German-speaking Alpine region is well researched from a dialectological point of view. Various direct and indirect surveys have been conducted in the nineteenth, twentieth, and twenty-first centuries, most of which focus on phonetics, phonology, and morphology. In addition to numerous local dialect grammars (e.g. Bauer 1967, Lessiak 1963) and the *Deutscher Sprachatlas* ['German Language Atlas'], the regional language atlas projects are particularly worth mentioning here. Among the latter, the *Sprachatlas der deutschen Schweiz* ['Language Atlas of German Switzerland'] (SDS) for Switzerland, the *Bayerischer Sprachatlas* ['Bavarian Language Atlas'] (BSA) for Germany, and the *Tirolischer Sprachatlas* ['Tyrolean Language Atlas'] (TSA), the *Vorarlberger Sprachatlas mit Einschluss des Fürstentums Liechtenstein, Westtirols und des Allgäus* ['Vorarlberg Language Atlas with Inclusion of the Principality of Liechtenstein, West Tyrol and the Allgäu'] (VALTS), and the *Sprachatlas von Oberösterreich* ['Language Atlas of Upper Austria'] (SAO) for Austria are worth to mention. An overview of the large-scale regional language atlas projects in the Alpine region is provided by Krefeld and Lücke (2014, 190). There are also dialect surveys in the realm of larger and smaller projects. For example, for Austria, the project part *Variation und Wandel dialektaler Varietäten in Österreich (in real und apparent time)* ['Variation and change of dialect varieties in Austria (in real and apparent time)'] of the SFB project (= Spezialforschungsbereich ['special research program']) *Deutsch in Österreich* ['German in Austria'] (DiÖ) can be mentioned, for South Tyrol the project *Deutsche Dialekte in Südtirol* ['German dialects in South Tyrol'] (Scheutz 2016) and the project *Varietäten im Kontakt* ['varieties in contact'] (VinKo).[1] Note that most of these projects only focus on one dialect, region, or country at a time (cf. Colcuc and Mutter 2021, 24).

Thus, although the dialectological situation is in principle well researched at the level of phonetics, phonology, and morphology, only few studies have yet been carried out that look at the various dialects at one point in time and across different types of borders (e.g. state borders, dialect borders), and there are hardly any studies covering the entire German-speaking Alpine region. There are, of course, pragmatic and methodological reasons for this. They can effectively be handled with the possibilities of digital surveys such as those used in the VerbaAlpina project, i.e. in particular the techniques of georeferencing and crowdsourcing (cf. Krefeld and Lücke 2014, 194). In the following, the project and its objectives are briefly outlined.

3 The VerbaAlpina project

VerbaAlpina is a transnational, cross-disciplinary digital humanities project dealing with the different languages and varieties of the Alpine region (cf. Colcuc and Mutter 2021, 25). The primary goal of the VerbaAlpina project is to study "the linguistic and cultural area of the entire Alpine region where dialects and languages of three huge language families (German, Romance and Slavonic) are spoken" (Colcuc and Mutter 2021, 24). To achieve this goal, VerbaAlpina created a cross-linguistic online platform that not only combines various linguistic data sets but also invites interested and dialect-competent Internet users from the Alpine region to systematically enrich the data via crowdsourcing (Krefeld and Lücke 2014, 202).

The focus of the VerbaAlpina project is on the lexis of different domains such as Alpine pasture and dairy farming, nature (e.g. terms for weather phenomena, landscape formations, fauna, and flora), and modern life (e.g. ecology and tourism). To better understand the dynamics of lexis in the Alpine region, VerbaAlpina draws on two data sources: linguistic atlases and dictionaries as well as crowdsourcing data. The basic idea behind VerbaAlpina is to overcome the closed nature of traditional printed publications and to create a multilingual online platform in which as much georeferenced linguistic data as possible are continuously collected, documented, classified, and published (cf. Krefeld and Lücke 2014, 195, 2019, 141). The use of data from linguistic

1 The VinKo-project shares some similarities with the VerbaAlpina project as it aims to document and analyse Germanic as well as Romanic language varieties in northeastern Italy. It employs crowdsourcing for data collection but focusses on spoken data to analyse phonological and morphosyntactic variation (Kruijt et al. 2023).

atlases, dictionaries, and a crowdsourcing tool "allows to gain knowledge not only about a certain historical period but also about the evolution of the lexis in the Alps" (Colcuc and Mutter 2021, 26).

In this article, we are focusing on the crowdsourcing data. These data are collected using a crowdsourcing tool,[2] which was developed specifically for the purposes of the VerbaAlpina platform. The tool was designed to invite dialect-competent Internet users from the Alpine region, i.e. the relevant 'crowd', to participate in small questionnaires in which they name concepts onomasiologically (e.g. 'Wie sagt man zu EIMER in INNSBRUCK?' ['How do you name BUCKET in INNSBRUCK?'], with an input field to type in variants such as *Kübel*, *Eimer*, or *Amper*). The answers are very often written in dialect, i.e. the graphemes used deviate from the orthographic norm in order to reflect dialect sounds. In most cases, this enables an analysis of the data regarding phonological structures (see Section 4.2). Note that the data are processed on the fly as "[d]ialect words that come directly from speakers via the crowdsourcing platform are not transcribed, but entered into the database immediately after speakers have sent them via crowdsourcing" (Colcuc and Mutter 2021, 30). This means that, after the speakers have provided their answers, the data are immediately accessible and visible on the platform. Crowdsourcing expands the network of data points, i.e. the number of speakers and locations per region. This is particularly important for those regions for which there are no regional language atlas projects.

The VerbaAlpina project places great emphasis on ensuring that all data comply with the FAIR principles.[3] Thus, the research data are 'findable', 'accessible', 'interoperable', and 'reusable'. Most importantly, VerbaAlpina produces stable data, i.e. the user interface of the platform and the database are 'frozen' or versioned every 6 months. These versions are then updated with new data, while older versions remain as such, with a DOI, and are stored in a data repository. Data can be exported via an API interface that is publicly available on the Internet.[4] Data can also be transferred to other formats and standards. Furthermore, in addition to the language data, VerbaAlpina also publishes metadata, software, and code (Krefeld and Lücke 2019, 141). In the following section, we briefly describe how we processed the data in order to answer our research questions.

4 Data processing

4.1 Regional aggregation

As of November 2022, the VerbaAlpina crowdsourcing data consisted of 11,214 responses from 896 German-speaking participants, with response frequencies (= number of items provided by the same participant) varying from 1 to 262. The mean and median response frequencies were calculated to be 12.5 and 4, respectively. Due to the significant variation in response frequencies among participants, it was necessary to aggregate the data at a regional level. We have selected the 'Nomenclature of Territorial Units for Statistics' (NUTS) as the framework for regional aggregation, utilizing the NUTS 3 level, which is the smallest administrative unit within this system (see https://ec.europa.eu/eurostat/web/nuts). In practice, the NUTS 3 system corresponds to the administrative divide of Swiss cantons, Italian provinces, and German counties (*Landkreise* or *kreisfreie Städte*), while for Austria, the NUTS 3 regions represent groups of several counties (*Bezirke*). Our analyses are based on the 30 regions delineated in Figure 2 (with corresponding numbers shown in Figure 3). Note that to increase the data basis per region, we merged some smaller NUTS 3 regions (primarily *kreisfreie Städte* in Germany were merged with surrounding *Landkreise*, and also some small Swiss cantons such as Glarus and Schwyz were merged together).

Although the NUTS 3 classification was primarily designed for administrative purposes, it aligns well with the commonly assumed dialect areas. Figure 4 shows to which dialect areas the 30 regions (primarily) belong according to traditional dialect classifications such as in Wiesinger (1983).

2 https://www.verba-alpina.gwi.uni-muenchen.de/crowdsourcing.

3 https://www.go-fair.org/fair-principles/.

4 https://www.verba-alpina.gwi.uni-muenchen.de/?page_id=8844&db=222.

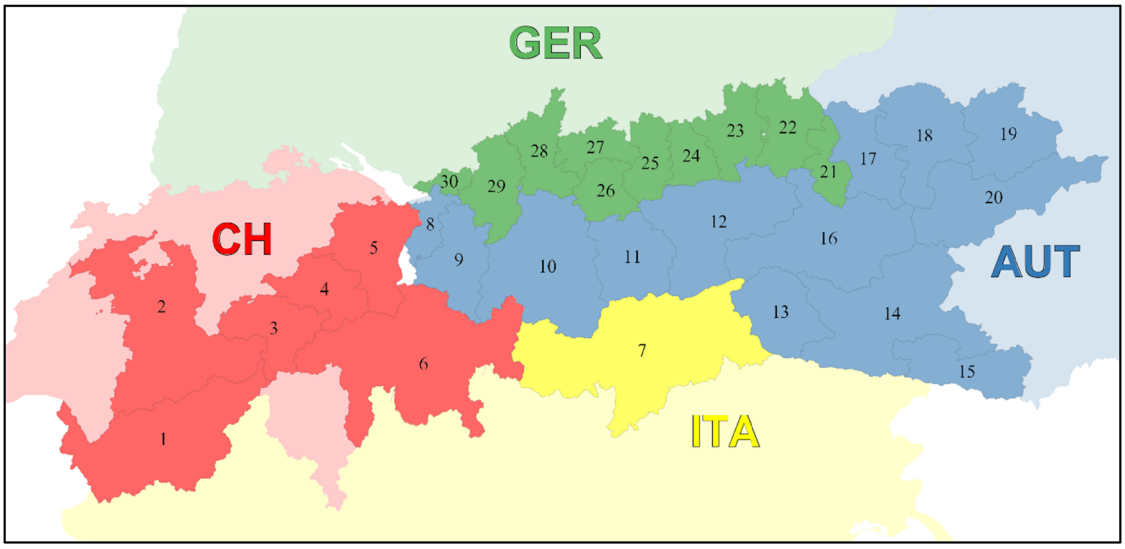

Figure 2: The 30 Alpine regions investigated (based on the NUTS 3 classification).

1	Valais	11	Innsbruck	21	Berchtesgarden
2	Bern	12	Tyrolean Unterland	22	Traunstein
3	Uri, Nidwalden, Obwalden	13	East Tyrol	23	Rosenheim Stadt, Rosenheim Landkreis
4	Glarus, Schwyz	14	Upper Carinthia, Lungau	24	Miesbach
5	St. Gallen, Appenzell IR, Appenzell AR	15	Klagenfurt-Villach	25	Bad Tölz Wolfratshausen
6	Graubünden	16	Pinzgau-Pongau	26	Garmisch-Partenkirchen
7	Bozen	17	Salzburg	27	Weilheim-Schongau
8	Rheintal-Bodenseegebiet	18	Traunviertel	28	Eastern Allgäu, Kaufbeuren
9	Bludenz-Bregenzer Wald	19	Steyr-Kirchdorf	29	Upper Allgäu, Kempten
10	Außerfern, Tyrolean Oberland	20	Liezen	30	Lindau

Figure 3: The 30 Alpine regions investigated (the running numbers refer to Figure 2).

4.2 Phonological information in the VerbaAlpina crowdsourcing data

Our primary goal in this study is to analyse the geolinguistic structures of dialect phonology in the Alpine area. However, the VerbaAlpina data represent phonology only indirectly since the survey was conducted in a written form. This poses a challenge, as non-expert transcription can hinder phonological interpretation. In addition, writing tasks may be vulnerable to interference with the standard variety, given that dialect speakers might not be accustomed to writing in their dialect (see, e.g., Seiler 2010, 516). Nevertheless, upon closer examination, we found detailed information on dialect phonology in the VerbaAlpina data. In what follows, two examples will be presented to illustrate this point: one for the lexeme *Käse* [‘cheese’] and another for the lexeme *Kuh* [‘cow’].

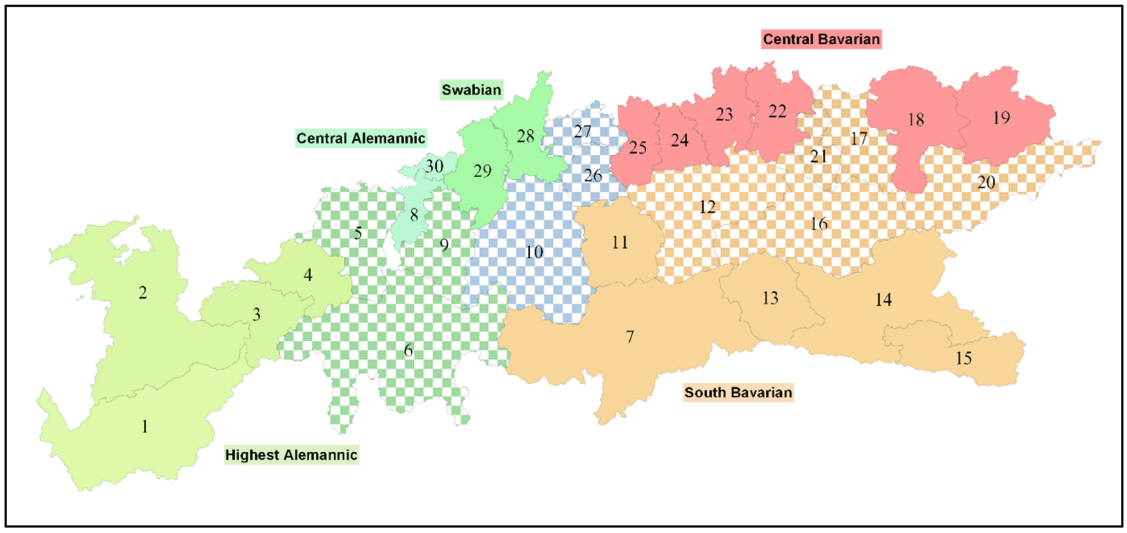

Figure 4: Traditionally assumed dialect areas (according to Wiesinger 1983) corresponding to the 30 Alpine regions.

Table 1: Spelling variants for *Käse* [‘cheese’] in the VerbaAlpina data

	<k>		**<ch>**		**<kch>**
<a(a)>	*kas* (115) *kaas* (64)	*kase* (5)			*kchaas* (2)
<ä(ä)>	*käs* (27) *kääs* (2) *käse* (1)		*chäs* (13) *chääs* (11) *chäse* (3)	*chääsa* (1) *chääse* (1)	*kchäs* (2)
<e(e)>	*kees* (1)		*chees* (6)	*ches* (1)	
Others	*käis* (1)		*cheis* (2)	*chies* (1)	

For the lexeme *Käse* [‘cheese’], the participants provided a total of 259 valid answers using 19 different spelling variants (capitalization will be ignored). Table 1 exhibits the variants along with their frequencies in parentheses, revealing notable differences in occurrence. *Kas* is the most frequent, with 115 instances, followed by *kaas* (64 instances), *käs* (27 instances), *chäs* (13 instances), and *chääs* (11 instances). Most other variants were only realized once or twice, including the variant adhering to standard German spelling (*käse*), which was only used once.

Certain recurring patterns are observable regarding the graphemes utilized by the participants. Specifically, the graphemes <k>, <kch>, and <ch> are frequently used at the beginning of the lexeme, while the first vowel is almost always represented as <a(a)>, <ä(ä)>, or <e(e)>. Prior dialectological research readily explains these patterns: The spellings <kch> and <ch> are most likely indicative of the initial consonant being realized as either an affricate (/kx/) or a fricative (/x/), respectively. Previous research shows that the former variant occurs in some of the South Bavarian and Central Alemannic dialects, while the latter variant is a characteristic feature of High and Highest Alemannic. Both variants result from the German consonant shift (Schirmunski 2010 [1962], 352–6). The varying representations of the vowel, on the other hand, can be attributed to a lowering and retraction of Middle High German (MHG) *æ* (/ɛ:/) in Bavarian, resulting in the Phoneme /a:/, whereas Alemannic dialects retained a front vowel (Wiesinger 1990, 450).

These interpretations are corroborated by plotting the individual variants on a map, see Figure 5. The left map indicates that spellings with <ch> are prevalent in the High and Highest Alemannic regions in German-speaking Switzerland, while spellings with <k> (and some spellings with <kch>) are used in the remaining

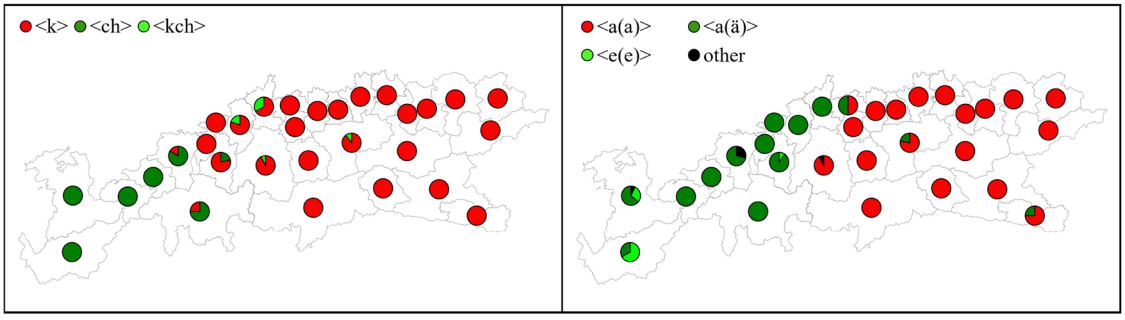

Figure 5: Regional distribution of the spelling variants for *Käse* [‘cheese’], on the left for the initial consonant, on the right for the first vowel.

Table 2: Spelling variants for *Kuh* [‘cow’] in the VerbaAlpina data

	<k>		**<ch>**		**<k(c)h>**	
<ua>	*kua* (104)	*kuah* (73)	*chua* (3)		*kchua* (3)	*khua* (2)
<ue>	*kue* (2)	*kueh* (1)	*chue* (18)	*chueh* (5)		
<uä>			*chuä* (13)	*chuäh* (6)		
<ui>	*kui* (9)	*kuih* (3)				
<üe>	*küeh* (5)		*chüe* (2)	*chüeh* (1)		
<u>	*kuh* (5)	*kuha* (4)				
<üi>	*küi* (3)					

Alemannic and Bavarian regions. The right map demonstrates that <a> spellings are restricted to the Bavarian area, whereas in the Alemannic regions in the west, spellings with <e(e)> or <ä(ä)> are present.

Consequently, the observed distribution of spelling variants for the lexeme *Käse* [‘cheese’] is indicative of phonological variation within our research area. This is not an isolated case, as evidenced by other lexemes such as *Kuh* [‘cow’], which yielded 261 valid responses from the participants, with a total of 19 distinct spelling variants. Table 2 illustrates the variants, along with their respective frequencies. The data again reveal a clear preference for certain spelling variants, with *kua* being the most frequently used variant with 104 instances, followed by *kuah* (73 instances), *chue* (18 instances), and *chuä* (13 instances). The remaining variants are minor, including the variant that corresponds to standard German spelling (*kuh*), which occurred five times.

Once more, there are certain patterns visible: For the beginning of the lexeme, the graphemes <k>, <k(c)h>, and <ch> are used, as discussed earlier with regards to the lexeme *Käse* [‘cheese’]. As for the vowel, there are seven distinct variants, with <ua> being the most frequent, followed by <ue>, <uä>, and <ui>. These variants align with prior research in dialectology, pointing towards the retention of a diphthong for MHG *uo*, which is a characteristic of Upper German dialects. This contrasts with the standard German long monophthong /u:/ in *Kuh* [‘cow’], which resulted from (Early) New High German monophthongisation, primarily found in the Central German dialects further north.

Despite Upper German dialects generally preserving a diphthong, there are differences among them in the quality of the diphthong offset. While the realization as /uɐ/ is predominant in most Central and South Bavarian dialects (e.g. Wiesinger 1990, 447), some Alemannic dialects tend to show a more centralized offset (= /uə/, e.g. Russ 1990, 370). The most frequent spelling variants (<ua> versus <ue, uä>) reflect these differences, while minor variants such as <ui> or <üe> indicate specific developments in smaller dialect regions. For instance, in the South Tyrolean Pustertal, MHG *uo* evolved into /uɪ/ (Scheutz 2016, 48–9), whereas in the Valais and some parts of Tyrol, MHG *uo* was palatalized to /yə/ or /yɐ/ (cf. Kranzmayer 1956, 57, Christen 2019, 252).

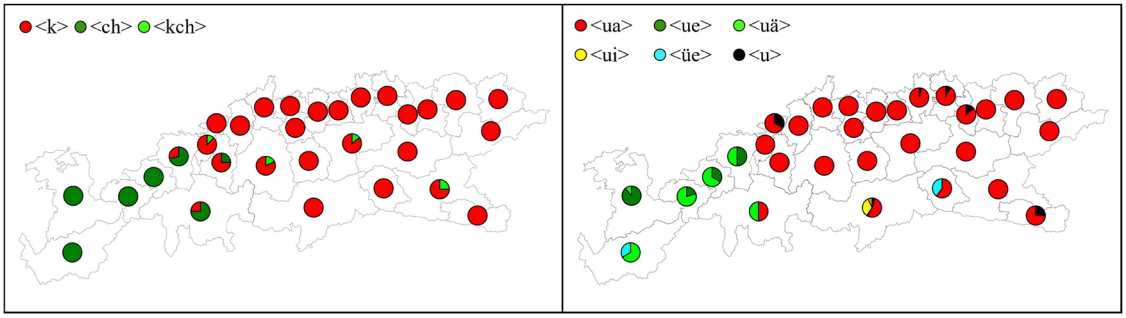

Figure 6: Regional distribution of the spelling variants for *Kuh* ['cow'], on the left for the initial consonant and on the right for the first vowel.

Again, the individual variants have been plotted on a map, as shown in Figure 6, providing further evidence for these interpretations. It is worth noting that the geographical distribution of <k> versus <ch> for *Kuh* ['cow'] is strikingly similar to that of *Käse* ['cheese'], depicted in Figure 5. Remarkably, the primary vowel variants (<ua> versus <ue, uä>) exhibit a similar regional distribution as the main variants for the initial consonant. This indicates the presence of underlying geolinguistic patterns that are the main focus of our study.

4.3 Variables

In order to uncover broader geographical patterns, we apply a multivariate dialectometric approach that involves abstraction from individual variables. Our dialectometric analyses are based on a selection of 19

Table 3: Variables and variants for the dialectometric analyses

	Variable	Items (base forms only)	Var 1	Var 2	Var 3	*n*
V1	MHG *a*	*Acker, Anke, backen, Dach, Gabel, hacken, laden, machen, Schlag, Stadel, Wagen, Wasser*	<a>	<o>		585
V2	MHG *æ*	*Käse*	<ä(ä), e (e)>	<a(a)>		547
V3	MHG *û*	*aus, Haus, Haut, sauer*	<au>	<u(u)>		175
V4	MHG *ü*	*Hütte, Kübel*	<ü>	<i>		473
V5	MHG *uo*	*Bub, Futter, Grube, Kuh*	<u>	<uo, ua>	<ue, uä>	415
V6	MHG *ei*	*breiten, Ei, Geiß, heim, klein, Laib, Stein, Teig*	<ei>	<oa>	<oi>	227
V7	MHG *a* before *l*	*Alm/Alp, Kalb, Schmalz, Stall, Salz*	<al>		<oi>	900
V8	MHG *e/ë* before *l*	*Feld, Fels, gelb, melken, Pelz, Schelle*	<el, äl>	<ai, äi, ei>	<ö>	328
V9	MHG *i* before *l*	*Milch*	<il(l)>	<ü(ü)>	<ui>	720
V10	Final *-er*	*Acker, Butter, Euter, Futter, Wetter*	<(e)r, (a)r>	<a, o>		974
V11	Final *-en* after velars/palatals	*anken, hacken, kochen, machen, melken, Rechen, stocken*	<(e)n>	<a, e, ä>		332
V12	Final *-en* after alveolars/dentals	*Boden, hüten, käsen, lassen, misten, sieden*	<(e)n>	<a, e, ä>		289
V13	Diminutive MHG *-lîn*	*Beillein, Blättlein, Fässlein, Glöcklein, Häuslein, Kälblein, Kännlein Kätzlein, Lämmlein, Männlein*	<(a)l, (e)l>	<li/le>	<e, ö, i>	217
V14	MHG *-e* in nominative feminine singular	*Hütte*	<(e)n>	<a, e, ä>		290
V15	Postvocalic *st*	*Ast, Biest, Gast, Mist, Post*	<st>	<scht>		298
V16	Lenition of *b*	*Gabel, Kübel, Nebel, Zwiebel*	<b(b)>	<v, w>		253
V17	Lenition of *t*	*Butter, Futter, Wetter*	<t(t)>	<d(d)>		465
V18	Fricativisation of *k*	*Käse, Kuh*	<k>	<ch>		907
V19	Vocalization of *r*	*Berg, Birke, Bürste, Germ, Hirte, Werk*	<r>	<a>		187

variables, which are presented in Table 3. All variables relate to spelling differences within single graphematic segments of either one or several lexemes in the VerbaAlpina dataset. In order to bolster the number of tokens per variable, we have included multiple lexemes for most variables and also taken into account morphologically complex words like compounds or derivatives. For instance, in variable 18 (V18), which concerns the realization of initial *k* before a vowel, we have not only included *Käse* ['cheese'] and *Kuh* ['cow'] but also morphologically complex words such as *käsen* ['make cheese'] or *Geißkäse* ['goat cheese'].

As illustrated earlier, spelling variants in the VerbaAlpina data often reflect differences in dialect phonology. Consequently, the selection of variables was made with the intention of capturing main differences in dialect phonology[5] in the Alpine region, based on earlier dialectological descriptions (e.g. Kranzmayer 1956, Schirmunski 2010 [1962], Christen 2019, Koch 2019, Lenz 2019, Streck 2019). The variables selected correspond to different phonological levels, including stressed (V1–V9) and unstressed (V9–12) vocalic phenomena, as well as consonantal phenomena (V15–19). Furthermore, two variables account for morpho-phonological features (V13 and V14).

Table 3 also shows the main variants. To facilitate the statistical analysis, we grouped together different spelling forms that apparently represent similar phonological features. To ensure that each variable carries similar weight in the analysis, only the two or three most frequent variants per variable are considered. Our subsequent analyses are based on a total of 8,582 tokens.

4.4 Statistical analyses

Dialectometric studies typically rely on aggregation to create a distance matrix that represents differences (or similarities) between dialect varieties, which is then analysed using statistical techniques such as multidimensional scaling or cluster analysis (e.g. Wieling and Nerbonne 2015). Although this approach has led to valuable insights, it obscures the distribution of individual linguistic features, limiting the ability to trace the source of aggregate dialect differences. Non-aggregative methods such as FA or principal component analysis avoid this problem by directly analysing the underlying linguistic features. In this study, FA is used as it has been shown to be particularly effective at identifying regional patterns (e.g. Leino and Hyvönen 2008, Grieve 2014, Pickl and Pröll 2019, Pickl et al. 2019).

FA is a statistical technique that aims to uncover underlying patterns of variation within a multivariate dataset. The method is based on a correlation matrix, where FA identifies and combines variables that are correlated with each other but largely independent of other variables into a smaller set of underlying constructs, so-called factors. This reduces the complexity of the original data and restructures it by identifying latent patterns that explain most of the variation in the data. Therefore, FA is a valuable tool for data reduction while preserving as much information as possible from the original dataset.

The current study utilizes FA to examine the correlations among the research locations in terms of the occurrence of the variants detailed in Table 3. It should be noted that FA does not assume any prior knowledge of the geographical position of the research locations and therefore only reveals regional patterns if there is a clear geographic signal present in the linguistic data itself.

The following two parameters are crucial for interpreting the factor solution:

- Factor loadings, which measure the relationship between a factor and the research locations. A factor loading close to 1 indicates a strong positive correlation between the factor and a location, meaning that the factor explains most of the variation in that location. Conversely, a loading close to 0 suggests no correlation, and a loading less than 0 indicates a negative correlation.
- Factor scores, which indicate how a specific variant ranks on a given factor. A high positive factor score indicates a positive association between a variant and a factor, while a score close to 0 indicates no association. Negative associations are reflected in negative factor scores.

5 In this article, we generally speak of "phonological variation" because the different variants have been described as part of the phonological systems of the dialects in previous studies. Moreover, there are "diasystematic contrasts" between different variants and the standard German forms.

In this study, principal axis factoring with varimax rotation was utilized to conduct the FA using the IBM SPSS Statistics software. The regression method was used to estimate the factor scores. Both the Kaiser–Meyer–Olkin Measure of Sampling Adequacy (= 0.68) and Bartlett's test of sphericity (X^2 = 1874.76, p < 0.000***) suggest that the data are suitable for the analysis.

5 Geolinguistic analyses

In this section, we report the outcomes of a five-factor solution based on the Kaiser–Guttman Criterion (eigenvalues <1). Collectively, the five factors explain 80.4% of the variability in the data. Factor 1 (F1) captures approximately 40.2% of the variance, followed by F2 with 25.4%, F3 with 7.9%, F4 with 4%, and finally, F5 with 3%.

5.1 Dominant factor loadings

Figure 7 offers an overview of the geographical patterns. The colours in the map represent the dominant factors in each area, i.e. the factors which have the highest factor loadings. For instance, a red colour implies that F1 is the prevailing factor in that particular region, purple corresponds to F2, green represents F3, and blue represents F4. It is noteworthy that F5 does not exhibit dominance in any region, hence its exclusion from Figure 7 (see Section 5.2 for the geographical patterns of F5). The shadings depict the relative strength of the dominant factor loadings, with darker shading indicating higher factor loadings. Therefore, the darker-shaded areas correspond to core dialect regions, whereas the lighter-shaded regions represent peripheral areas.

Figure 7 provides clear evidence of a robust geographical signal in our data. On the basis of the dominant factor loadings, we observe four coherent areas:

- F1 (red) is the primary factor in the western parts of the Alpine region, encompassing Switzerland, Vorarlberg in Austria, and the district of Bavarian Swabia (*Bayerisch-Schwaben*) in Germany.
- F2 (purple) is dominant in the northeast, specifically in the district of Upper Bavaria (*Oberbayern*) in Germany.
- F3 (green) has the highest loadings in the southeast, including Carinthia and Tyrol in both Austria and Italy.
- F4 (blue) dominates the north-eastern half of Austria, covering the Tyrolean Unterland, north-western Styria, most of Salzburg, and Upper Austria.

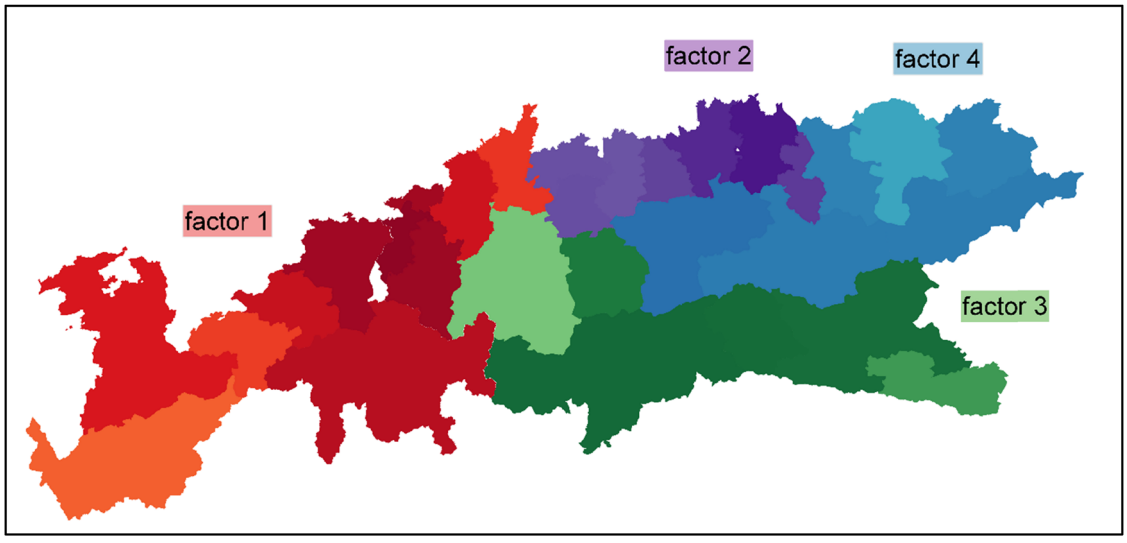

Figure 7: Dominant factor loadings (red = factor 1, purple = factor 2, green = factor 3, blue = factor 4).

The patterns observed in Figure 7 closely resemble the traditional dialect classification of the Alpine area described in Section 2. The regions where F1 dominates include the Alemannic dialects found in Switzerland, Germany, and Austria. F3 outweighs all other factor loadings in the South Bavarian regions in Austria and Italy. However, the usual distinction between Central Bavarian and the South-Central Bavarian transition zone is not visible in the same manner. Although a north–south division is apparent in that area, its boundary runs more towards the west and north. It thus runs through the commonly assumed Central and South-Central Bavarian regions but aligns well with the border between Austria and Germany. Therefore, F2 groups the German and F4 groups the Austrian Central and South-Central Bavarian dialects. This finding shows the relevance of the divergence processes along the Austrian-German border (Scheuringer 1990, Scheutz 2007) for the geolinguistic structures in the Alpine region (see Section 6 for a discussion).

5.2 Individual factor loadings

Our analysis of the dominant factor loadings reveals that the observed geographical patterns are generally consistent with the traditional dialect classification outlined in Section 2, with the exception of a potential border effect noted for (South-)Central Bavarian. Yet, considering only the dominant factors neglects the variation that exists below the threshold of dominance, that is, the regionally non-dominant parts of the dominant factors as well as factors that are entirely non-dominant (Pickl 2016, 91). To account for this variability, we need to concentrate on one factor at a time, as illustrated by the five separate heatmaps shown in Figure 8.

The data displayed in Figure 8 reveal a gradual decrease in factor loadings for all five factors when moving away from the core dialect areas, rather than a sharp drop. This observation provides evidence for the complex and gradual nature of dialect boundaries and the existence of dialect continua, and also sheds light on the relationships between various dialect areas. Specifically:

- The highest loadings for F1 are found in Alemannic regions, but they are also relatively high in the Bavarian-Alemannic transition zones in Austria and Germany, and become negative when moving eastwards.
- F2 not only has the highest loadings in the (South-)Central Bavarian regions in Bavaria but also exhibits high loadings in the (South-)Central Bavarian regions in Austria. F2 has even positive loadings in South Bavarian regions, while showing negative values only in the Alemannic west.
- Similarly, F3 not only shows the highest loadings in the South Bavarian regions but also has positive loadings throughout the Bavarian area, and, to a minor degree, in parts of Highest Alemannic (in particular in the Valais).
- F4 not only displays the highest loadings in the (South-)Central Bavarian regions of Austria but also has rather high loadings in the eastern South Bavarian dialects, with again positive loadings throughout the Bavarian area.
- Finally, F5 (which is not dominant in any region) has the highest loadings in the west of the Alemannic area and rather high values in all Swiss regions. Interestingly, in other Alemannic regions, the loadings are generally lower or even negative. As a result, F5 delineates the Swiss Alemannic dialects which possibly reveals another border effect. However, F5 might also separate the Highest Alemannic dialects from the rest of the Alemannic dialect area.

5.3 Factor scores

As elaborated in Section 4.4, non-aggregative dialectometric measures, like factor analyses, have a significant advantage as they reveal the linguistic basis for the geographical patterns they detect. This aspect can be

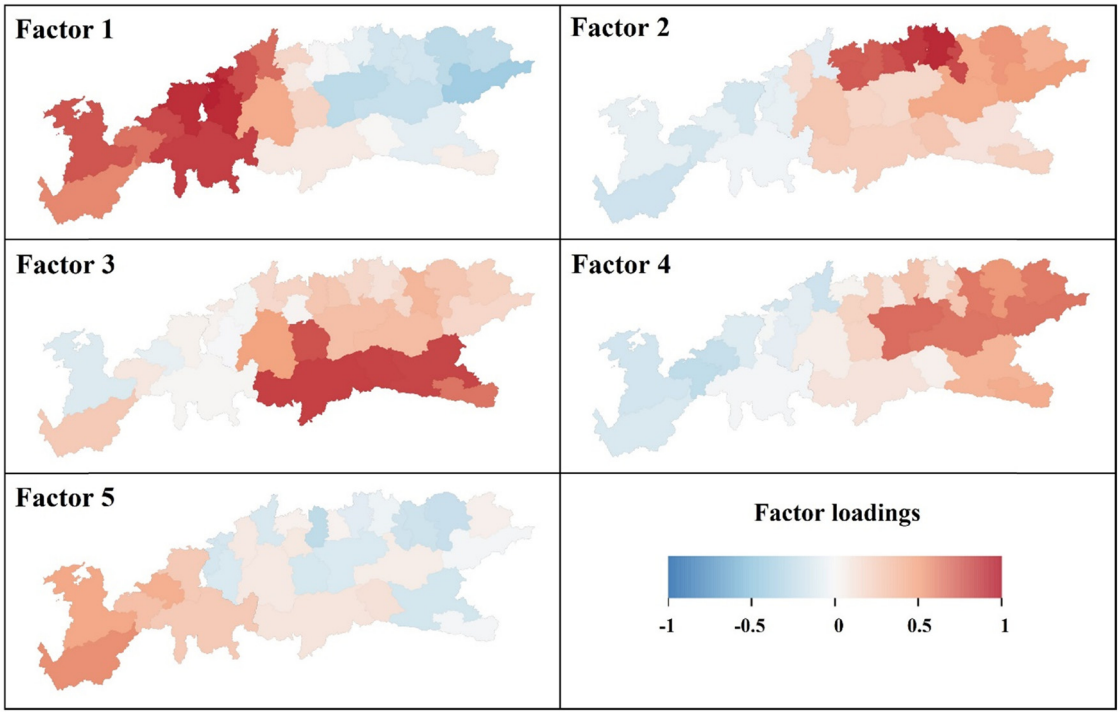

Figure 8: Individual factor loadings.

studied by analysing the factor scores, which indicate the associations between each factor and the linguistic variants. Table 4 displays the top five positive factor scores for each factor.

The results presented in Table 4 provide further support for the connection between the geolinguistic patterns revealed by the factor loadings and the traditional dialect regions:

- The factor scores for F1 are highest for linguistic features that are typically associated with the Alemannic dialects in the Alpine region but are absent in Bavarian (see e.g. Jutz 1931, Hotzenköcherle 1986, Christen 2019, Streck 2019, Lenz 2019). This includes the pronunciation of word-final *-en* as schwa, even after alveolar/dental sounds (V12, cf. spellings like *mischte* for *misten* [‘muck out’]), the realization of the diminutive MHG *-lîn* as *-le* or *-li* (V13, cf. spellings like *hüsli* for *Häuslein/Häuschen* [‘(little) house’]), the retention of a front vowel for MHG *æ* (V2, cf. spellings like *chääs* for *Käse* [‘cheese’]), the preservation of intervocalic lenis plosives such as /b̥/ (V16, cf. spellings like *chübel* for *Kübel* [‘bucket’]), as well as the retention of long monophthongs for MHG *û* (V3, cf. spellings like *huus* for *Haus* [‘house’]).
- Regarding F2, we find both general Bavarian dialect features and some (West-)Central Bavarian dialect features (see Kranzmayer 1956, Koch 2019). The Bavarian dialect features include the realization of /ɔɐ̯/ as a reflex of MHG *ei* (V6, cf. spellings like *goaß* for *Geiß* [‘goat’]), the realization of /ɔ/ or /ɒ/ as a reflex of MHG *a* (V1, cf. spellings like *gowe* for *Gabel* [‘fork’]), and the realization of /a:/ as a reflex of MHG *æ* (V2, cf. spellings like *kaas* for *Käse* [‘cheese’]). The (West-)Central Bavarian dialect features comprise the pronunciation of word-final *-en* as schwa after velar and palatal sounds (V11, cf. spellings like *kocha* for *kochen* [‘(to) cook’]) and the realization of a diphthong for MHG *e/ë* before *l* due to the vocalisation of /l/ (V8, cf. spellings like *meicha* for *melken* [‘(to) milk’]).
- F3 is closely linked with features that are specific to the South Bavarian dialects and not present in Central Bavarian (see, e.g. Kranzmayer 1956, Wiesinger 1990, Lenz 2019). Furthermore, it encompasses some features that are typical of Bavarian dialects in general, in contrast to the Alemannic dialects. Typical for South Bavarian is the preservation of intervocalic fortis plosives such as /t/ (V17, cf. spellings like *fuitto* for *Futter* [‘fodder’]), the consonantal realization of postvocalic /l/ (V7 and V8, cf. spellings like *pelz* and *solz* for *Pelz* [‘fur’] and *Salz* [‘salt’]), and the retention of /n/ in word-final *-en* after velar and palatal sounds (V11, cf.

Table 4: Variants with the highest factor scores per factor (value of the factor scores indicated in brackets)

	F1 'Alemannic'	F2 '(South) Central Bavarian in Germany'	F3 'South Bavarian'	F4 '(South) Central Bavarian in Austria'	F5 'Alemannic in Switzerland'
1	<a, e, ä> for final *-en* after alveolars/dentals (V12) (1.8)	<oa> for MHG *ei* (V6) (2.0)	<t(t)> for intervocalic *t* (V17) (1.8)	<uo, ua> for MHG *uo* (V5) (2.5)	<ch> for initial *k* (V18) (2.3)
2	<li/le> for the diminutive MHG *-lîn* (V13) (1.7)	<a, e, ä> for final *-en* after velars/palatals (V11) (1.8)	<el, äl> for MHG *e/ë* before *l* (V8) (1.8)	<st> for postvocalic *st* (V15) (1.6)	<ue, uä> for MHG *uo* (V5) (1.9)
3	<ä(ä), e(e)> for MHG *æ* (V2) (1.6)	<ai, äi, ei> for MHG *e/ë* before *l* (V8) (1.6)	<i> for MHG *ü* (V4) (1.7)	<n> for MHG *-e* in nominative feminine singular (V14) (1.4)	<a, e, ä> for MHG *-e* in nominative feminine singular (V14) (1.7)
4	<b(b)> for intervocalic *b* (V16) (1.4)	<o> for MHG *a* (V1) (1.6)	for MHG *a* before *l* (V7) (1.6)	<ö> for MHG *e/ë* before *l* (V8) (1.4)	<(e)r, (a)r> for final *er* (V10) (1.2)
5	<u(u)> for MHG *û* (V3) (1.3)	<a(a)> for MHG *æ* (V2) (1.6)	<(e)n> for final *-en* after velars/palatals (V11) (1.5)	<(e)n> for final *-en* after velars/palatals (V11) (1.2)	<t(t)> for intervocalic *t* (V17) (1.0)

spellings like *melchn* for *melken* [‘(to) milk’]), while the derounding of MHG *ü* to /ɪ/ is typical for the whole Bavarian area (V4, cf. spellings like *hitt* for *Hütte* [‘hut’]).

- F4 is associated with the co-occurrence of South-Central Bavarian, East-Central Bavarian, and some general Bavarian features (see, e.g. Kranzmayer 1956, Wiesinger 1990, Lenz 2019). Examples of (more or less) general Bavarian features include the occurrence of [ʊ̯] as a reflex of MHG *uo* (V5, cf. spellings like *jauchngruam* for *Jauchegrube* [‘cesspool’]), the realization of *-n* instead of *-e* due to analogical levelling in the nominative singular of many feminine nouns (V14, cf. spellings like *hittn* for *Hütte* [‘hut’]), and the realization of postvocalic *st* as /st/ (V15, cf. spellings like *ausmistn* for *ausmisten* [‘muck out’]). The retention of the nasal in word-final *-en* (V11, cf. spellings like *mochn* for *machen* [‘make’]) is typical for South-Central (and South) Bavarian, while the vocalization of MHG *e/ë* before *l* to /œ:/ (V8, cf. spellings like *möchn* for *melken* [‘(to) milk’]) is a feature of East-Central Bavarian.
- Finally, F5 is characterised by the co-occurrence of some High and Highest Alemannic variants and general Alemannic features (e.g. Jutz 1931, Hotzenköcherle 1986, Christen 2019). This includes the realization of prevocalic *k* as a fricative (V18, cf. spellings like *chäs* for *Käse* [‘cheese’]), the more centralized offset in the reflex of the diphthong MHG *uo* (V5, cf. spellings like *fuetr* for *Futter* [‘fodder’]), the preservation of the ending *-e* in the nominative singular of many feminine nouns (V14, cf. spellings like *hüttä* for *Hütte* [‘hut’]), the retention of the rhotic in word final *-er* (V10, cf. spellings like *uuter* for *Euter* [‘udder’]), and the preservation of intervocalic fortis plosives such as /t/ (V17, cf. spellings like *buttere* for *buttern* [‘(to) butter’]).

6 Discussion and conclusion

Using the VerbaAlpina crowdsourcing data, the aim of this article is to investigate the geographical structures in dialect phonology in the German-speaking Alpine region. By applying factor analyses, we have shown geographical patterns which are generally consistent with Wiesinger’s (1983) widely accepted dialect classification (see Section 2). As to be expected, the most important difference confirmed by the VerbaAlpina data – captured by Factor 1 (accounting for about 40% of the variability in the data) – is the distinction between Bavarian and Alemannic dialects. Consequently, Factor 1 is associated with (morpho-)phonological features (displayed in the written responses) typically found in Alemannic dialects within the Alpine region, while being absent in Bavarian. We also found evidence for a Bavarian-Alemannic transition zone in Tyrol and Bavaria. Moreover, within Bavarian, we could identify a South Bavarian area, which corresponds to Factor 3 and exhibits a strong association with linguistic features traditionally attributed to the South Bavarian dialects.

Interestingly, our findings did not reveal a distinct Central Bavarian region but rather two (South-)Central Bavarian dialect areas, separated by the Austrian-German border (Factor 2 and Factor 4). While the traditional dialect classifications do not take into account border effects between Germany and Austria, previous research indicates that differences between Austrian and German dialects have arisen due to more recent changes (cf. e.g. Scheuringer 1990, Scheutz 2007). Our study shows the relevance of these divergence processes for the geolinguistic structuring of the Alpine region. For instance, a characteristic feature associated with Factor 2 (representing (South-)Central Bavarian in Germany) is the realization of a diphthong (<ai, äi, ei>, Table 4) for MHG *e/ë* before *l* (V8), resulting from the vocalization of /l/. Conversely, the vocalization to /œ:/ (<ö>, Table 4) serves as a distinctive feature for Factor 4 (representing (South-)Central Bavarian in Austria). Interestingly, this variable has undergone significant change over past decades: Originally, the traditional (South-)Central Bavarian dialects of western Austria, much like many dialects in neighbouring Bavaria, utilized diphthongs (cf. e.g. Kranzmayer 1956, 29). However, these diphthongs have been replaced by long monophthongs, which diffused from eastern Austria, where Vienna, the capital city of Austria, is located (for the vocalization of /l/ in the West Central Bavarian dialects of Salzburg see Kaiser and Bülow 2022, 455–6; for the influence of Vienna on the Austrian dialects see Vergeiner et al. 2023). Conversely, in Bavaria, there is a reverse trend, with local dialect features being replaced by more prestigious western variants from the metropolitan area around

Munich and other cities (cf. e.g. Scheuringer 1990, Scheutz 2007, 41–4).[6] Similar patterns of divergence can be observed for other variables, such as the reflexes of MHG *ei* (V6), where many Austrian dialects adopted the variant /a:/ spreading from Vienna and its vicinity (Scheutz 1985, 242–3, Scheuringer 1990, 416). Likewise, the deletion of the nasal in word-final *-en* after velar or palatal sounds (V11) has potentially diminished in the more contemporary Central Bavarian dialects of Austria due to the influence of the vernacular language spoken in Vienna and other urban centres (Wiesinger 1989, 24, but cf. Vergeiner and Wallner 2022), whereas the feature has been retained in Bavaria, where it is generally more accepted (Scheuringer 1990, 379).

Similar effects at national borders have been noted in other dialects, such as Alemannic along the French-German border (cf. Auer et al. 2015, 2017). It is probable that these divergence processes are not primarily a consequence of 'communication barriers' along the border but are ultimately influenced by attitudinal factors (as discussed, e.g. in the study by Auer et al. 2017, 41–2). In this context, it is interesting to note that Kleene (2020) found that German and Austrian speakers nowadays perceive the political border between the two countries as a linguistic border, and they no longer share a common linguistic identity.

It is important to note, however, that Factor 2, despite being non-dominant, exhibits relatively high loadings also in Bavarian Austria; thus, there is (yet) no sharp boundary between Bavarian in Austria and Germany but a continuum. Moreover, both Factor 3 and Factor 4 display positive factor loadings throughout Bavarian, which indicates a certain coherence among the Bavarian language area. Both the capacity to uncover the linguistic basis of the identified geographical structures and the capacity to reveal dominant and latent geographical structures are significant advantages of the non-aggregative dialectometric measures employed in this study (cf. also Pickl 2016).

Interestingly, for the Alemannic dialects, there might be a similar border effect, as indicated by the non-dominant Factor 5, which separates the Swiss German dialects from the rest of the Alemannic region. However, as most Alpine regions in Switzerland belong to the Highest Alemannic dialect group, this intra-Alemannic division aligns more closely with the traditional dialect classification (cf. e.g. Bohnenberger 1953, Wiesinger 1983). It is worth noting that we did not find similar border effects between Austria and Italy, as South Tyrol fits well within the South Bavarian dialect group (see the results for Factor 3). In addition, the Austrian-German border does not appear to be significant concerning the Alemannic dialects.

Eventually, the differences that emerge for the different borders are related to the different sociolinguistic situations with regard to the respective dialect-standard constellation (for the different constellations, see Auer 2005, Kehrein 2019). In general, the Alemannic and Bavarian dialects in Germany are more affected by levelling and degradation than the Bavarian dialects in Austria and South Tyrol and, in particular, more than the Alemannic dialects in Austria and German-speaking Switzerland (cf. Auer 2018, Lenz 2019, Christen 2019). As a result of levelling and degradation, the traditional dialects are transforming into more widespread regiolects, especially among younger speakers, which is why it is to be expected that such forms will also be reflected in crowdsourcing data. For the border effects, it is also important that there is evidence that speakers orient themselves more strongly to regional centres at the level of the regiolect, such as Munich for Bavarian in Germany (cf. e.g. Kleene 2020, 83) and Vienna for Bavarian in Austria (cf. e.g. Scheuringer 2001, 100).

With respect to our research questions, the following findings are apparent: Firstly, geographical patterns for dialect phonology in the Alpine region can be identified based on indirectly collected crowdsourcing data (Research Question 1). We have shown that the written VerbaAlpina crowdsourcing data are, in principle, suitable for studying the phonological variation in our research area. The fact that the data are useful for analysing the phonological structure of dialects is not a matter of course. It reflects that the participants, i.e. the VerbaAlpina-crowd, have a good sense for writing down their dialects. It certainly plays a role that in active dialect communities – as we find them in the German-speaking Alpine region – dialectal writing has become popular on a mass scale, especially via messaging services like WhatsApp and social media platforms like

6 Very similar results are likely to be found for Middle High German *i* before *l* (see Scheuringer 1990, 184–7) – however, for the present study, only the lexeme *Milch* 'milk' could be analysed for this variable, with *Milch* showing different patterns due to the frequent occurrence of vowel epenthesis in Central Bavarian in Austria and Germany (cf. spellings like *milli* or *mille*; note that epenthesis blocks *l*-vocalization).

Instagram and Facebook (cf. e.g. Huber and Schwarz 2017, Glaznieks and Frey 2018, Glück and Glaznieks 2019). These dialectal writing practices not only contribute to expressing proximity and regional identity within dialectal communities but also offer great potentials for scientific research and data collection.

Secondly, our findings suggest that the traditionally established dialect areas in the German-speaking Alpine region generally remain valid, despite being established on data from a century ago and within a structuralist (qualitative) framework (Research Question 2). Thirdly, even though we observe that dialect areas do still not consistently coincide with state borders, state border effects leading to divergences are evident for certain variables and regions, especially for the Bavarian dialects along the German-Austrian border (Research Question 3). Thus, taking both a cross-dialectal and a cross-national perspective within the Alpine region, we were able to highlight the significance of border effects for the geolinguistic structures in our research area.

Ultimately, our study suggests that it is very fruitful to look at dialects across dialect and state borders. Such an approach not only enables a more comprehensive perspective but is also the only way to adequately reveal effects of state borders, which are undoubtedly present. Although the use of crowdsourcing data to answer our research questions has proven to be fruitful, the results need to be confirmed by findings from studies using other methods (e.g. sociolinguistic interviews). In the context of this study, we were only able to focus on phonological variables that can be represented in the written form. Of course, there are some variables whose variants cannot be clearly represented by linguistic 'laypersons' in the graphemic system. For example, our study cannot account for phenomena such as the variation between [ɔ] (in Austrian Bavarian) and [ɒ] (in German Bavarian), which is regarded as a very salient feature for the dialect differences at the Austrian-German border (cf. e.g. Scheuringer 1990, 210–7). Despite these limitations, our study contributes to the growing body of knowledge on how written crowdsourcing data can be used in order to answer dialectological research questions.

Abbreviations

F1	Factor 1
FA	Factor analysis
MHG	Middle High German
NUTS	Nomenclature of Territorial Units for Statistics
V1	Variable 1

Acknowledgements: We would like to thank the VerbaAlpina team for providing the data. First and foremost, we would like to thank Thomas Krefeld, Stephan Lücke, Christina Mutter and Florian Zacherl for their support and valuable advice. We would also like to thank our student assistants Janina Böhlen and Anne-Marie Lutgen for data preparation.

Funding information: VerbaAlpina has been funded from 2014 through 2023 by the German Research Foundation (DFG; see http://gepris.dfg.de/gepris/projekt/253900505).

Conflict of interest: The authors state no conflict of interest.

Data availability statement: The datasets analysed during the current study are available via the VerbaAlpina platform, http://verba-alpina.gwi.uni-muenchen.de/?page_id=162&db=222.

References

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

Auer, Peter. 2018. “Dialect change in Europe – Leveling and convergence.” In *The handbook of dialectology*, edited by Charles Boberg, John Nerbonne, and Dominic Watt, p. 159–76. Oxford: Wiley.

Auer, Peter and Christian Schwarz. 2015. “Dialect-to-standard advergence: The relevance of compound borrowing.” In *Linguistic variation. Confronting fact and theory*, edited by Rena Torres Cacoullos, Nathalie Dion, and André Lapierre, p. 263–82. New York & London: Routledge.

Auer, Peter, Julia Breuninger, Dominique Huck, and Martin Pfeiffer. 2015. “Auswirkung der Staatsgrenze auf die Sprachsituation im Oberrheingebiet (Frontière linguistique au Rhin Supérieur, FLARS).” In *Regionale variation des Deutschen. Projekte und Perspektiven*, edited by Roland Kehrein, Alfred Lameli, and Stefan Rabanus, p. 323–47. Berlin/Boston: De Gruyter Mouton.

Auer, Peter, Julia Breuninger, and Martin Pfeiffer. 2017. “Neuere Entwicklungen des Alemannischen an der französischdeutschen Sprachgrenze im Oberrheingebiet.” In *Räume, Grenzen, Übergänge. Akten des 5. Kongresses der IGDD*, edited by Helen Christen, Peter Gilles, and Christoph Purschke, p. 27–44. Stuttgart: Steiner.

Bauer, Werner. 1967. “Lautlehre der Mundart des Lungaus mit besonderer Berücksichtigung von Zederhaus.” *Dissertation*. Vienna: University of Vienna, Austria.

Bohnenberger, Karl. 1953. *Die Alemannische Mundart: Umgrenzung, Innengliederung und Kennzeichnung*. Tübingen: Mohr.

Bülow, Lars, Philip C. Vergeiner, Hannes Scheutz, Dominik Wallner, and 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.

Christen, Helen. 2019. “Alemannisch in der Schweiz.” In *Language and Space. An International Handbook of Linguistic Variation, Volume 4: Deutsch*, edited by Joachim Herrgen and Jürgen Erich Schmidt, p. 246–78. Berlin & Boston: de Gruyter.

Colcuc, Beatrice and Christina Mutter. 2021. “Cartography and interoperability of linguistic data. Digital geolinguistics illustrated by the project VerbaAlpina.” *Études Digitales* 11, 23–43.

Eller-Wildfeuer, Nicole, Paul Rössler, and Alfred Wildfeuer, eds. 2018. *Alpindeutsch. Einfluss und Verwendung des Deutschen im alpinen Raum*. Regensburg: Edition Vulpes.

Elspaß, Stephan and Robert Möller. 2015. “Atlas zur deutschen Alltagssprache (AdA).” In *Regionale Variation des Deutschen. Projekte und Perspektiven*, edited by Roland Kehrein, Alfred Lameli, and Stefan Rabanus, p. 519–39. Berlin & Boston: de Gruyter.

Gaeta, Livio and Guido Seiler. 2021. “A specter is haunting Europe: the Alps as a linguistic area?” *STUF - Language Typology and Universals* 74(1), 1–16.

Glaznieks, Aivars and Jennifer-Carmen Frey. 2018. “Dialekt als Norm? Zum Sprachgebrauch Südtiroler Jugendlicher auf Facebook.” In *Jugendsprachen. Aktuelle Perspektiven internationaler Forschung*, edited by Arne Ziegler, p. 859–89. Berlin & Boston: de Gruyter.

Glück, Alexander and Aivars Glaznieks. 2019. “Geschriebener Dialekt in Südtiroler Facebook-Texten.” *Linguistik online* 99(6), 79–94.

Grieve, Jack. 2014. “A comparison of statistical methods for the aggregation of regional linguistic variation.” In *Aggregating dialectology, typology, and register analysis. Linguistic variation in text and speech*, edited by Benedikt Szmrecsanyi and Bernhard Wälchli, p. 53–88. New York: de Gruyter.

Hotzenköcherle, Rudolf. 1986. *Dialektstrukturen im Wandel. Gesammelte Aufsätze zur Dialektologie der deutschen Schweiz und der Walsergebiete Oberitaliens*. Aarau: Sauerländer.

Huber, Judith and Christian Schwarz. 2017. “SMS-Kommunikation im mehrsprachigen Raum. Schriftsprachliche Variation deutschsprachiger SMS-Nutzer/innen in Südtirol.” *Networx* 76. http://www.mediensprache.net/networx/networx-76.pdf (15.08.2018).

Jutz, Leo. 1931. *Die alemannischen Mundarten (Abriß der Lautverhältnisse)*. Halle (Saale): Niemeyer.

Kaiser, Irmtraud and Lars Bülow. 2022. “Der Dialekt der jüngsten Generation. Wandel, Remanenz und Abbau dialektaler Merkmale bei Kindern im westmittelbairischen Raum Salzburgs.” In *Struktur von Variation zwischen Individuum und Gesellschaft. Akten der 14. Bayerisch-Österreichischen Dialektologietagung 2019 (Zeitschrift für Dialektologie und Linguistik – Beihefte 189)*, edited by Philip C. Vergeiner, Stephan Elspaß, and Dominik Wallner, p. 437–67. Stuttgart: Franz Steiner.

Kehrein, Roland P. 2019. “Areale Variation im Deutschen ‚vertikal‘.” In *Language and Space. An international handbook of linguistic variation, Volume 4: Deutsch*, edited by Joachim Herrgen and Jürgen Erich Schmidt, p. 121–58. Berlin & Boston: de Gruyter.

Kleene, Andrea. 2020. *Attitudinal-perzeptive Variationslinguistik im bairischen Sprachraum. Horizontale und vertikale Grenzen aus der Hörerperspektive*. Mannheim: IDS-Verlag.

Koch, Günter. 2019. “Bairisch in Deutschland.” In *Deutsch (Language and Space. An International Handbook of Linguistic Variation 4)*, edited by Joachim Herrgen and Jürgen Erich Schmidt, p. 279–318. Berlin & Boston: de Gruyter.

Kranzmayer, Eberhard. 1956. *Historische Lautgeographie des gesamtbairischen Dialektraumes*. Graz & Wien: Böhlau.

Krefeld, Thomas and Stephan Lücke eds. 2014. *VerbaAlpina. Der alpine Kulturraum im Spiegel seiner Mehrsprachigkeit, München*, online. doi: 10.5282/verba-alpina.

Krefeld, Thomas and Stephan Lücke. 2014. “VerbaAlpina – Der alpine Kulturraum im Spiegel seiner Mehrsprachigkeit.” *Ladinia* 38, 189–211.

Krefeld, Thomas and Stephan Lücke. 2017. “22 Monate – Arbeitsbericht des Projekts VerbaAlpina (31.7.2016).” *Ladinia* 40, 259–72.

Krefeld, Thomas and Stephan Lücke. 2019. “54 Monate VerbaAlpina – auf dem Weg zur FAIRness.” *Ladinia* 43, 139–55.

Kruijt, Anne, Stefan Rabanus, and Marta Tagliani. 2023. “The VinKo-Corpus: Oral data from Romance and Germanic local varieties of Northern Italy.” In *Neue Entwicklungen in der Korpuslandschaft der Germanistik: Beiträge zur IDS-Methodenmesse 2022 (Korpuslinguistik und interdisziplinäre Perspektiven auf Sprache 11)*, edited by Marc Kupietz and Thomas Schmidt, p. 203–12. Tübingen: Narr.

Kunzmann, Markus and Christina Mutter. 2020. “Dialektdatenerhebung neu gedacht: Vom Nutzen des Netztes für die Sprachwissenschaft.” *IDS Sprachreport* 36(2), 32–7.

Kunzmann, Markus and Christina Mutter. 2021. "Zur Verwertbarkeit indirekt erhobener Sprachdaten: Erfahrungen aus dem Forschungsprojekt ‚VerbaAlpina'." In *Linguistik grenzenlos: Berge, Meer, Käse und Salamander 2.0 – Linguistica senza confini: montagna, mare, formaggio e salamandra 2.0 (Korpus im Text 14. Version 1)*, edited by Stephan Lücke, Noemi Piredda, Sebastian Postlep, and Elissa Pustka.

Leemann, Adrian and Nanna Haug Hilton, eds. 2021. *Using Smartphones to Collect Data for Linguistic Research* (Linguistics Vanguard 7 (s1)).

Leino, Antti and Saara Hyvönen. 2008. "Comparison of component models in analysing the distribution of dialectal features." *International Journal of Humanities and Arts Computing* 2(1–2), 173–87.

Lenz, Alexandra N. 2019. "Bairisch und Alemannisch in Österreich." In *Language and Space. An International Handbook of Linguistic Variation, Volume 4: Deutsch*, edited by Joachim Herrgen and Jürgen Erich Schmidt, p. 318–63. Berlin & Boston: de Gruyter.

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

Pickl, Simon. 2016. "Fuzzy dialect areas and prototype theory. Discovering latent patterns in geolinguistic variation." In *The future of dialects. Selected papers from Methods in Dialectology XV*, edited by Marie-Hélène Côté, Remco Knooihuizen, and John Nerbonne, p. 75–97. Berlin: Language Science Press.

Pickl, Simon and Simon Pröll. 2019. "Ergebnisse geostatistischer Analysen areal-sprachlicher Variation im Deutschen." In *Language and Space. An international handbook of linguistic variation, Volume 4: Deutsch*, edited by Joachim Herrgen and Jürgen Erich Schmidt, p. 861–79. Berlin & Boston: de Gruyter.

Pickl, Simon, Simon Pröll, Stephan Elspaß, and Robert Möller. 2019. "Räumliche Strukturen alltagssprachlicher Variation in Österreich anhand von Daten des 'Atlas zur deutschen Alltagssprache (AdA)'." In *Dimensions of Linguistic Space: Variation – Multilingualism – Conceptualisations*, edited by Lars Bülow, Ann Kathrin Fischer, and Kristina Herbert, p. 39–60. Berlin: Lang.

Rabanus, Stefan, Ermenegildo Bidese, and Silvia Dal Negro. 2019. "Deutsch als Minderheitensprache in Italien." In *Language and Space. An international handbook of linguistic variation, Volume 4: Deutsch*, edited by Joachim Herrgen and Jürgen Erich Schmidt, p. 1096–1114. Berlin & Boston: de Gruyter.

Russ, Charles V. J. 1990. "High Alemannic." In *The Dialects of Modern German. A Linguistic survey*, edited by Charles V. J. Russ, p. 364–93. London: Routledge.

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.

Scheuringer, Hermann. 2001. "Die deutsche Sprache in Österreich." In *Regionale Standards. Sprachvariationen in den deutschsprachigen Ländern*, edited by Elisabeth Knipf-Komlósi and Nina Berend, p. 95–119. Budapest: Dialóg Campus Kiadó.

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

Scheutz, Hannes. 2007. *Drent und herent. Dialekte im salzburgisch-bayerischen Grenzgebiet*. Salzburg et al.: EuRegio.

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

Schirmunski, Viktor M. 2010 [1962]. *Deutsche Mundartkunde. Vergleichende Laut- und Formenlehre der deutschen Mundarten*, edited and commented by Larissa Naiditsch. Frankfurt am Main: Lang.

Seiler, Guido. 2010. "Investigating language in space: Questionnaire and interview." In *Theories and Methods (Language and Space. An international handbook of linguistic Variation 1)*, edited by Peter Auer and Jürgen Erich Schmidt, p. 512–27. Berlin & Boston: de Gruyter.

Streck, Tobias. 2019. "Alemannisch in Deutschland." In *Language and Space. An international handbook of linguistic Variation, Volume 4: Deutsch*, edited by Joachim Herrgen and Jürgen Erich Schmidt, p. 206–45. Berlin & Boston: de Gruyter.

Vergeiner, Philip C., Lars Bülow, and Dominik Wallner. 2021. "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. and 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., Jan Luttenberger, Lars Bülow, Dominik Wallner, and David Britain. 2023. "Revisiting areal and lexical diffusion. The case of the Viennese Monophthongization in Austria's base dialects." *Linguistics* 61(4), 915–957.

Wieling, Martijn and John Nerbonne. 2015. "Advances in Dialectometry." *Annual Review of Linguistics* 1(1), 243–64.

Wiesinger, Peter. 1983. "Die Einteilung der deutschen Dialekte." In *Dialektologie*, edited by Werner Besch, Ulrich Knoop, Wolfgang Putschke, and Herbert E. Wiegand, p. 807–900. Berlin & New York: de Gruyter.

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

Wiesinger, Peter. 1990. "The Central and Southern Bavarian Dialects in Bavaria and Austria." In *The Dialects of Modern German. A linguistic survey*, edited by Charles V. J. Russ, p. 438–519. London: Routledge.

Language Atlas projects

Bayerischer Sprachatlas = Hinderling, Robert, Werner König, Ludwig Maximilian Eichinger, Hans-Werner Eroms, Horst Haider Munske, and Norbert Richard Wolf, eds. 1998-2014. *Bayerischer Sprachatlas*. 6 Regionalteile. Heidelberg: Winter.

Deutscher Sprachatlas = Wrede, Ferdinand, Walther Mitzka, and Bernhard Martin, eds. 1927-1956. *Deutscher Sprachatlas auf Grund des Sprachatlas des Deutschen Reiches von Georg Wenker*. Marburg: Elwert.

Sprachatlas der deutschen Schweiz = Hotzenköcherle, Rudolf, Robert Schläpfer, Rudolf Trüb, and Paul Zinsli, eds. 1962-1997. *Sprachatlas der deutschen Schweiz* (SDS). 8 volumes. Bern & Basel: Francke.

Sprachatlas des Deutschen Reichs = Wenker Georg. 1888–1923. *Sprachatlas des Deutschen Reichs*. Marburg. Hand-drawn maps.

Sprachatlas von Oberösterreich = Adalbert-Stifter-Institut des Landes Oberösterreich, eds. 1998-. *Sprachatlas von Oberösterreich*. Linz: Adalbert-Stifter-Institut des Landes Oberösterreich.

Tirolischer Sprachatlas = Klein, Karl Kurt, Ludwig Erich Schmitt, and Egon Kühebacher, eds. 1965-1969. *Tirolischer Sprachatlas*. 2 volumes. Marburg: Elwert.

Vorarlberger Sprachatlas mit Einschluss des Fürstentums Liechtenstein, Westtirols und des Allgäus = Gabriel, Eugen, ed. 1985-2006. *Vorarlberger Sprachatlas mit Einschluss des Fürstentums Liechtenstein, Westtirols und des Allgäus*. 5 volumes. Bregenz: Vorarlberger Landesregierung.