languages

MDPI

*Editorial*

# Dialectal Dynamics—An Introduction

**Alfred Lameli [1,*], Simonetta Montemagni [2] and John Nerbonne [3,4]**

[1] Research Center Deutscher Sprachatlas, Marburg University, 35032 Marburg, Germany
[2] CNR, Istituto di Linguistica Computazionale "Antonio Zampolli", 56124 Pisa, Italy; simonetta.montemagni@cnr.it
[3] Center for Language and Cognition, University of Groningen, Broerstraat 5, 9712 CP Groningen, The Netherlands; j.nerbonne@rug.nl
[4] Deutsches Seminar, University of Freiburg, 79098 Freiburg im Breisgau, Germany
* Correspondence: lameli@uni-marburg.de

**Abstract**

The study of dialects leads very naturally to the study of their geographic distribution and the nature of the distribution, e.g., by examining whether the distribution is based simply on geographic distance or on relatively distinct dialect regions. Dialectal dynamics poses the further question of why the distribution takes the form it does. Does variation arise through migration, i.e., due to the relative lack of communication among people who live far from one another? Sociolinguists have shown convincingly that variation is often employed to indicate identification with others, leading to the adoption of speech habits and changes in the distribution of variation. Purely linguistic processes may push some varieties toward change while others are more resistant, and contact with other languages and dialects, including particularly standard languages, almost inevitably results in changes. This volume examines studies in the area of dialectal dynamics, including studies focused on methods that promise to illuminate this complex field.

**Keywords:** dialectology; dialectometry; language change; language contact; historical linguistics



Received: 26 September 2025
Accepted: 26 September 2025
Published: 15 October 2025

**Citation:** Lameli, A., Montemagni, S., & Nerbonne, J. (2025). Dialectal Dynamics—An Introduction. *Languages*, *10*(10), 265. https://doi.org/10.3390/languages10100265



## 1. Introduction

We would like to introduce this volume of papers on dialectal dynamics, first noting its relation to other topics in linguistics, especially in language variation, historical linguistics and language contact. Second, the topic is now timely, which motivated us to issue a call for papers focusing on this in 2023. Third, we will introduce the papers and how they contribute to the volume. Fourth and finally, we wish to suggest prospects for future work in this direction.

## 2. Dialectal Dynamics

All languages show variation—the variable use of different linguistic forms. These may be different pronunciations, different lexical items, different morphological forms, and even different syntaxes, phrasal or sentential. There are also indications that pragmatic variation might be added to this list. When the variation depends on geography, we speak of dialects, and other forms of variation depend on social class, personal style and the situation of use, where the term 'dialect' is also sometimes used. We concentrate here on geographic variation, and we refer to the varieties as 'dialects' (Chambers & Trudgill, 1998). We hasten to add that, while the volume is focused on how the geographic varieties change, this introduction looks broadly at the influences on this change, including the influence of other varieties, in particular standard varieties.

The study of dialects is dialectology, and the primary research question in dialectology is how variation is distributed geographically, e.g., in roughly discrete areas or rather in a continuum of differences. It is a small step from characterizing the distribution of variation to asking what influences the variation, and to noting that the distribution is subject to change. This is why dialect dynamics is central to dialectology. Dialect dynamics focuses on these two questions: what influences geographic variation in language and how does it change over time? We launched a call for papers for a Special Issue on dialectal dynamics suggesting that progress in this aspect of dialectology would likely arise in reflections about social and geographical factors and their interaction, a path advocated by Chambers and Trudgill.

The subject is not new, but the past two decades (since Chambers and Trudgill) have seen progress in the application of more exact methods in dialectology, in particular the development of measures of differences among varieties, i.e., dialectometry. This development began with Séguy (1971), whose work shifted the focus from differing individual features to aggregate differences, enabling more robust characterizations, and dialectometry has gained considerable ground in the last two decades (see Wieling & Nerbonne, 2015, and references there). A second big advantage of dialectometric studies over more traditional ones is the fact that they produce numerical characterizations of dialect differences, enabling the application of sensitive regression techniques when it comes to examining influences (Nerbonne, 2009). So the time seems ripe to examine a well-established research question anew.

We add that the call for papers emphasized that work on less-studied languages would be particularly welcome. We are gratified that several research groups responded to this aspect of the call (see below), and we are especially pleased that some papers not only report on variation in less-studied languages but also examine questions of dynamics in them. Dialectologists are in great agreement that ideas about the distribution of linguistic variation should be general, i.e., applicable in principle to all languages.

## 3. The Scholarly Landscape

The study of dialectal dynamics is closely intertwined with several other linguistic disciplines and subfields, all of which require careful consideration. In its current form, dialectology, historical linguistics, sociolinguistics, and contact linguistics can no longer be regarded as entirely independent fields. Rather, they jointly contribute to a more comprehensive understanding of language variation and change, grounded in empirical data and informed by both linguistic and extralinguistic factors.

In recent years, computational linguistics has also begun to play an increasingly important role in the study of dialectal dynamics, particularly with the increasing availability of large and complex datasets and the advent of machine learning methods. These approaches have enabled the development of tools that can model and account for dialectal differences more effectively, offering new insights into patterns of linguistic variation across regions and social groups.

The following subsections will offer several observations and reflections on each of these disciplines, with particular emphasis on their respective contributions to the study of dialectal dynamics.

### *3.1. Historical Linguistics*

Historical linguistics studies how languages change. Historical linguistics does not distinguish sharply between “languages” and “dialects”, following the lead of general linguistics, which also does not, so studies of how dialects change might even be classified as a subdiscipline of historical linguistics. Historical linguistics teaches us that changes

may arise through (social) differentiation, a subject we will discuss below (sociolinguistics), but also through contact to other varieties, which is also discussed separately. In addition, historical linguistics has demonstrated that languages can change as children learn them anew in each generation (Crain et al., 2006), and that both system pressures, e.g., promoting the simplification of morphological paradigms, and frequency in use, in which, e.g., reduced forms may dominate, tend to lead to language change. Neither of the last two factors plays a role in the papers here, however.

Proceeding from the analogy to historical linguistics we should additionally note that changes in dialects are often more difficult to detect than changes in unrelated languages. Changes due to contact with unrelated languages often result in pairs of words from source and target languages in which the phonology of the word in the target language is quite different from the rest of the language. For example, plurals ending in /-i/ in English are often borrowings, such as BLINI ('pancakes') or APPARATCHIKI ('bureaucrats'), both from Russian, or PAPARAZZI ('importunate photographers'), from an Italian name. Naturally, loan words are easier to spot than other effects of language contact. All the effects become increasingly difficult to detect the further in time one looks after the period of contact (because contact effects tend to become absorbed in the target language). If one has samples of cognate words in many dialects then a borrowed word usually differs to a greater degree from the words in the rest of a sample (when one compares aggregate levels of difference). So French borrowings still stand out when one compares a sample of Germanic varieties. See English MAGAZINE vs. German ZEITSCHRIFT, Dutch TIJDSCHRIFT, Swedish TIDNING/TIDSKRIFT, etc.

Although historical linguistics studies language change in general, it is nonetheless often focused on historical reconstruction, inferring earlier stages of language history from later realizations. Bowern (2018) reviews a substantial body of the literature in this area.

Historical linguists themselves often emphasize parallels with biology, citing no less an authority than Darwin, who commented on the similarities in *The Origin of Species* (1859) as well as elsewhere. Atkinson and Gray (2005) comment extensively on the line of thought demonstrating that both languages and species pass on their properties even while modifying (some of) them. This is often referred to as "descent with modification".

Dialectology has also noted further parallels to biology, more specifically population genetics, in the geographic dispersal of variation. Mayr (1954) pointed out that once a group is isolated from the larger population it issues from, then further generations are limited in their variation to that of the selection in the subpopulation, a phenomenon he dubbed the "founder effect". Mufwene (2008) has demonstrated a similar restriction in the linguistic variation among colonial varieties of English and French. The effect of migration on linguistic variation has also been noticed earlier, resulting in ideas on the possible effects of the Migration Period (*Völkerwanderung*) on differentiation of dialects and languages. This perspective is particularly evident in some early 19th-century dialect classifications, which were heavily oriented toward reconstructing the settlement patterns of Germanic tribes. In this context, dialect geography was often closely linked to ethnolinguistic reconstructions of tribal distribution.

A further dimension of this connection lies in the differential documentation of the Germanic language branches. While North and East Germanic are attested through textual sources, West Germanic is largely unattested and must be reconstructed. Nevertheless, modern dialects preserve phonological features that are thought to reflect the sound system of Proto-West Germanic. One of the most striking differences, for instance, between Low German and Upper German dialects within present-day Germany is that Low German varieties have retained the consonantal system of West Germanic to a much greater extent—apart from a few exceptions. It was thus a natural step for Wenker (1876) to use dialect

data as a means of reconstructing features of West Germanic and, by extension, to infer linguistic characteristics associated with historical Germanic tribes. Today, such approaches are viewed more critically, particularly due to their underlying assumption of a quasi-linear continuity between prehistoric linguistic stages and modern dialects.

Further, Sewall Wright first noted that genetic variation tends to increase with distance, but not linearly, rather instead, roughly logarithmically (Wright, 1940). Nerbonne (2010) speculated that the same relation could be seen in the geographic distribution of linguistic variation, confirming this in samples from Norwegian, German, Dutch, Bulgarian, the Bantu spoken in Gabon, and the American English of the Eastern seaboard. It has since been confirmed often in dialectometric studies, with occasional reports that linear fits were slightly better than sublinear ones.

The debate on the proper perspective on the biology–linguistic analogy is still proceeding in a lively fashion. Croft (2021) contributes to an extended exchange with Mufwene (2008), discussing not only the role of transmission from one generation to the next but also the role of "horizontal" exchange, which we touched on above in noting Bowern's (2018) work.

### *3.2. Sociolinguistics*

Labovian sociolinguistics began with a now famous study on the social motivation of sound change (Labov, 1963), contributing to varietal dynamics directly, and it has flourished in innumerable demonstrations that speakers use linguistic variants as a means of identifying themselves (including distinguishing themselves from other groups). Labov (2001) documents and systematizes a good deal of this work. Within this line of research there is an interesting controversy regarding the relative influence of categorical factors such as social class, age, gender (sex), or ethnicity as opposed to the sheer mass (or density) of linguistic experience (Heath, 2003). This ought to be examined using the dialectometric methods mentioned in the last section.

There is good reason to follow Chambers and Trudgill's (1998) suggestion to view dialectology and sociolinguistics as two perspectives on variationist linguistics, a term they championed. This integration of spatial and social dimensions was also claimed to be essential for a comprehensive understanding of linguistic variation by Britain (2002) and Nerbonne (2009). From this perspective it is only natural to assume that dialectal dynamics should involve the same social forces as those studied primarily in sociolinguistics.

Dialectology and sociolinguistics are thus two closely related subfields of linguistics that both investigate language variation and change, but which developed with distinct focuses and methodologies. Sociolinguistics, from its earliest stages, has adopted a quantitative approach to study how social variables influence linguistic behavior. Over time, statistical tools such as logistic regression (Cedergren & Sankoff, 1974) and, more recently, mixed-effects regression models (e.g., Johnson, 2009; Tagliamonte & Baayen, 2012) have become standard in the field, enabling researchers to model complex interactions between linguistic and social variables. In contrast, dialectology has moved only in recent decades toward quantitative analysis through dialectometry (see above) although with significant differences: for example, dialectometry typically emphasizes aggregate variation patterns, whereas sociolinguistics rather focuses on individual linguistic variables.

Over time, the two disciplines have increasingly converged, leading to integrative approaches in the study of linguistic variation and change. While sociolinguistics excels in capturing social dynamics of language use, dialectology offers insights into geographic and historical patterns of variation.

Recent research, aiming to combine the strengths of both approaches, moved toward a more unified model of linguistic variation that incorporates both geographic and social

influences, as well as the contribution of individual linguistic features (Lameli, 2019). Wieling et al. (2011) proposed the use of a generalized additive model in combination with a mixed-effects regression approach to simultaneously account for the effects of geographical, social, and linguistic variables in modeling dialect variation. This approach has been applied to various languages, including Dutch, English, and Italian (Montemagni et al., 2013; Wieling et al., 2014). In contrast to approaches that define dialect areas solely based on aggregate dialect distances, Wieling and Nerbonne (2011) introduced a novel method that simultaneously identifies geographical dialect areas and uncovers the underlying linguistic features.

Despite progress, full integration of the dialectological and sociolinguistic perspectives remains challenging. This challenge is tackled in some of the papers in this volume.

### *3.3. Languages in Contact*

We noted the frequent focus of historical linguistics on historical reconstruction above, inferring earlier stages of language history from later realizations. From the point of view of reconstruction, language contact confounds the enterprise (Thomason & Kaufman, 1988) and therefore must be taken very seriously. The more recent work using phylogenetic inference attempts reconstruction even while being aware that loan words (perhaps the most frequent effect of contact) may infect the enterprise (Bowern, 2018). We do not know of work applying these methods to dialects, however.

As noted above, dialect contact is often difficult to detect, but we mention it here because it is probably ubiquitous. Trudgill (1986) is an early recognition of the importance of contact in studying dialectal dynamics, while Jeszenszky et al. (2023) is an attempt to sort conceptual issues in the study of the role of dialects in language contact. Several papers in this volume continue the research direction of examining the role of contact.

Special attention must be paid to the relation of dialects to the standard languages in the countries they are spoken in. Auer and Hinskens (1996) emphasized the importance of this in a seminal article. In many European countries, standard languages are taught throughout the educational system, are used in radio and television, and are encountered frequently due to modern mobility. The result is not only that it is rare to find speakers without at least a passive command of the standard but also that fewer and fewer speakers have confident command of local dialects (Lameli, 2025 for German dialects). A further, striking result of dialect-standard contact is the emergence of so-called regiolects, ways of speaking adopted by speakers that use some dialect features, perhaps in an effort to show local allegiance (Schmidt, 2010). One recent example of this is Buurke's (2025) examination of the vitality of two varieties spoken in the north of the Netherlands, Frisian and Low Saxon, which takes care to take the influence of standard Dutch into account.

### *3.4. Computational Linguistics*

Computational linguistics (CL) is sometimes regarded as a subdiscipline of linguistics, but it emphasizes the development of applications—useful products and processes—more than most linguistics subdisciplines. Popular topics include automatically adapting computational tools to linguistic subvarieties, automatically analyzing the input to digital assistants (Krückl et al., 2025), normalizing text written in dialect for the purpose of search, etc.

The relationship between computational linguistics and the study of dialects has gained increasing prominence in recent years, driven by advances in data availability and computational modeling techniques. In principle, these approaches—employing methods such as machine learning (including deep learning), statistical modeling, and natural language processing (NLP)—enable large-scale analyses of diatopic, diastratic, and

diachronic dialectal variation. Among these methods, NLP has come to play an increasingly central role. When applied to the study of dialectal variation, NLP tools facilitate the large-scale analysis of both written and spoken language, including non-standard varieties such as dialects. In practice, however, the development of dialect-aware NLP tools—tailored to process non-standard dialectal input—remains a significant challenge, largely due to the limited availability of suitable resources and language technologies.

The growing interest in processing non-standard dialectal varieties is evidenced by several initiatives dedicated to this area. Notably, the Association for Computational Linguistics (ACL) has hosted the VarDial workshop series twelve times, co-located with major international NLP conferences. These workshops focus on computational methods and language resources for closely related languages, language varieties, and—most relevant here—dialects. VarDial has also featured shared tasks aimed at developing or adapting models for dialect processing (Gaman et al., 2020), which have contributed valuable dialectal datasets to the research community. Among existing resources, the recently released DIALECTBENCH represents the first large-scale benchmark specifically designed for evaluating NLP systems on dialectal varieties (Faisal et al., 2024).

One of the most relevant applications of NLP in this area is automatic dialect identification, where machine learning models are trained to classify texts or speech according to geographic or social dialects. This task has attracted increasing attention, particularly with the rise of large dialectal corpora collected from social media platforms and other online sources (see, e.g., Abdul-Mageed et al., 2024 for Arabic dialects). Another important application involves the integration of dialectal variation into machine translation systems, enabling translation between different dialects of the same language or between a dialect and its corresponding standard variety—for instance, across the many dialects of Arabic.

NLP also contributes to variationist and sociolinguistic research by enabling the automatic extraction and quantification of structured linguistic features from large-scale dialect corpora, often derived from user-generated content such as social media. It is also the source of the technology for automatic sequence alignment that has played a useful supporting role in dialectometry (Heeringa, 2004), and it hosts work on morphological and syntactic analysis of the sort that might be used more in measuring differences between geographical (and other) varieties. The paper of Tang, van Heuven, Heeringa and Gooskens in this Special Issue adopts precisely this approach, demonstrating the potential of NLP to support empirical dialectal inquiry.

## 4. The Papers Rethinking Dialectal Dynamics

Linguistic variation has long been studied as a function of geography, with dialect boundaries traditionally understood as the result of physical distance and natural barriers between speech communities. However, recent advances in quantitative dialectology, computational modeling, and interdisciplinary sociolinguistics have revealed a far more intricate picture (Wieling et al., 2011). Dialects are not simply distributed according to geography but are shaped by a complex interplay of historical migration, trade networks, social identity, digital communication, and cognitive constraints (Lameli, 2019). While some dialects undergo rapid convergence due to modern connectivity, others persist despite intense external pressures, displaying surprising patterns of resilience.

This Special Issue builds on the progress made in dialectometry and computational sociolinguistics to explore dialectal dynamics, and our call for papers specifically asked for contributions on varieties of less-studied languages. The research presented here moves beyond the traditional focus on well-documented European linguistic landscapes, incorporating case studies from Inner Asia Minor Greek, Mennonite Low German, Basque, Sinitic languages, and regional varieties of Dutch, Japanese, German and English. In

addition to expanding the empirical base of dialectology, the goal is to develop a deeper understanding of why linguistic variation is distributed as it is, how historical and social factors interact with linguistic structures, and what role modern influences—such as digital communication and economic dependencies—play in shaping dialectal change.

Addressing dialectal dynamics from a comparative-structural standpoint, Chaoju Tang, Vincent van Heuven, Wilbert Heeringa and Charlotte Gooskens explore how linguistic distances manifest across Sinitic and European varieties, contributing to the debate on the language–dialect distinction. Using commensurable lexical and syntactic data from 32 varieties—covering Germanic, Romance, Slavic, and Chinese—they measure both lexico-phonetic and syntactic distances, revealing that some so-called Chinese dialects (e.g., Cantonese and Mandarin) are as lexically distinct as language pairs like Dutch–German or Portuguese–Romanian. However, in syntax, the Sinitic varieties form an unusually cohesive cluster, with much smaller internal variation than their European counterparts. This duality in the results supports the idea that dialectal dynamics cannot be captured by lexical distance alone, and shows the value of integrating multidimensional metrics when evaluating linguistic relatedness.

Their approach shares a methodological affinity with that of Jonathan Dunn and Sidney Wong, who also quantify structural similarities, albeit focusing exclusively on syntactic constructions within global English varieties. Dunn and Wong address dialectal dynamics from a socio-interactional perspective, investigating how patterns of grammatical similarity across English dialects are shaped by language contact and population contact at the city level. Using computational construction grammar and geo-referenced social media corpora, the authors construct a high-dimensional similarity network based on the usage of grammatical constructions. They find that while geographic distance and linguistic landscape (as a proxy for language contact) are strong predictors of grammatical similarity, population contact—as measured by air travel—is less consistently explanatory. Importantly, the study confirms that dialectal similarity is not uniformly distributed across all levels of grammar: more concrete constructions show higher regional variation, while more abstract schematic structures tend to be shared across dialects. This work thus expands methodology for modeling grammar in use and adds nuance to our understanding of how interactional environments shape grammatical convergence and divergence.

The traditional view that dialect variation is primarily dictated by geography is complicated by findings from several of the studies included in this issue. The analysis of Wilbert Heeringa and Fumio Inoue on Japanese dialects reveals that contemporary dialect boundaries align not only with geographical regions but also with historical political and cultural divisions. The persistence of the East–West dialectal divide in Japan, despite efforts at national linguistic standardization, demonstrates the long-term influence of historical settlement patterns and political history on linguistic variation. These findings support a growing body of research suggesting that linguistic change follows a non-linear trajectory, where older settlement and migration histories can override more recent shifts in transportation and communication networks.

One particularly unexpected insight from this collection comes from the study of Floris Nijhuis, John L. A. Huisman and Roeland van Hout on the dialects surrounding the former Zuiderzee in the Netherlands. Traditionally, bodies of water have often been considered linguistic barriers, reinforcing dialect divergence by limiting communication between speech communities. However, the research in this issue demonstrates that the Zuiderzee, rather than isolating communities, actually functioned as a linguistic conduit, facilitating dialect mixing due to extensive maritime trade. Dialect similarity in the region correlates not only with overland proximity, but also with historical trade connections, suggesting that economic interdependence can sometimes exert a greater influence on

linguistic convergence than physical geography. This finding complicates long-standing assumptions about the role of geographic isolation in dialect formation, highlighting the need to incorporate economic and mobility patterns into models of dialect contact.

Beyond geographical and historical constraints, social interaction continues to play a crucial role in dialectal stability and change. Traditional models of dialect diffusion emphasize face-to-face interaction as the primary mechanism of linguistic transmission. However, in their research on English dialect usage in digital communication, Dunn and Wong suggest that social media now serves as an additional arena for dialectal influence. Contrary to expectations that digital communication would lead to widespread dialect leveling, the findings indicate that social media may in fact reinforce certain dialectal features, as regional grammatical structures gain visibility and become markers of social identity in online discourse (Eisenstein et al., 2014). While some features undergo erosion due to exposure to dominant linguistic norms, others persist due to algorithm-driven linguistic reinforcement, creating a paradoxical effect in which digital spaces simultaneously promote both convergence and preservation.

A similarly differentiated perspective on dialectal variation is offered by Göz Kaufmann's study of Mennonite Low German across five countries in the Americas (USA, Mexico, Bolivia, Paraguay, Brazil). Rather than focusing on social identity or linguistic stability, the study highlights how internal grammatical factors—such as auxiliary choice, clause type, and the syntactic relationship between verbs—shape the formation of two-verb clusters in subordinate clauses. Among other things, the analysis reveals that speakers employ structurally distinct variants to signal different degrees of clausal integration. This highlights that even in contact settings, internal grammatical mechanisms drive syntactic differentiation. A different picture emerges in the study of Inner Asia Minor Greek by Stavros Bompolas and Dimitra Melissaropoulou, which examines the influence of Turkish on Greek dialects in a contact setting. While lexical borrowing follows predictable patterns of geographical distance, grammatical borrowing appears to be shaped more by economic and social relationships, suggesting that different linguistic subsystems exhibit different levels of permeability to external influence.

Cognitive constraints also play a role in shaping the dynamics of dialectal change, a factor we might have added to contact and social affinity. Research by Gotzon Aurrekoetxea on Basque dialects introduces the perspective that variation exists not only across speakers but also within individual linguistic behavior. Intra-speaker variation, in which individuals unconsciously shift between different grammatical structures, suggests that dialect stability may be influenced by the cognitive flexibility of speakers themselves. This challenges conventional models that assume that dialect change occurs uniformly across a speech community, instead highlighting the role of individual linguistic behavior in either accelerating or slowing down broader linguistic shifts. While this study emphasizes intra-speaker flexibility, it finds an intriguing parallel in Dunn and Wong's work, where construction-based similarity is influenced not just by usage frequency but by structural abstraction—highlighting how both individual behavior and systemic structure interact to shape dialectal patterns.

Returning to more concrete patterns of change, the question of why some dialectal features change more readily than others is also addressed in this collection. While lexical items are frequently borrowed and adapted across dialects, grammatical structures tend to be more resistant to change. Kaufmann's study of Mennonite Low German verb clusters highlights the role of internal grammatical dynamics in shaping syntactic variation, showing that structural differentiation emerges from within the system rather than as a direct response to external linguistic influence. In contrast, findings from Bompolas and Melissaropoulou on Inner Asia Minor Greek suggest that structural borrowing tends

to occur only under conditions of deep and sustained social contact between linguistic communities. These studies reinforce the idea that linguistic change operates at different rates across different components of language, with vocabulary shifting more fluidly than core syntactic and morphological structures.

While the study of Asia Minor Greek underscores how deep socio-cultural ties can preserve grammatical structure in contact settings, another line of research broadens the lens by asking how dialects and everyday culture co-pattern at a regional scale. Addressing dialectal dynamics from a spatial–cultural perspective, Philip Vergeiner explores how linguistic and non-linguistic patterns relate across Austria, offering a methodological contribution to understanding the interaction between language and culture. Using a multivariate dialectometric approach, Vergeiner compares patterns of phonological variation with regional distributions of traditional cultural practices drawn from ethnographic data. His findings show that while linguistic variation forms relatively coherent and large-scale regional clusters—aligning with traditional dialect classifications—the patterns of everyday culture are more fragmented and small-scale. This asymmetry challenges earlier assumptions about the tight coupling of dialect and folk culture, suggesting that different historical and environmental forces shape language and culture in distinct ways. Vergeiner's work demonstrates the value of combining computational dialectometry with ethnographic sources, even as it reveals the complexity and limits of their alignment. At the same time, the findings complement those from the Zuiderzee study (Nijhuis et al. this volume), which also decouples linguistic change from mere geographic proximity, emphasizing instead the role of socio-economic and historical networks in shaping dialect boundaries. Here, trade centers serve as focal points for dialect mixing, while rural inland communities retain more conservative linguistic features. The case of Inner Asia Minor Greek further supports this idea, showing that economic reliance on Turkish-speaking communities played a role in grammatical borrowing patterns. Taken together, these studies suggest that economic exchange can override geographic distance in determining linguistic convergence, adding a critical dimension to our understanding of dialect contact.

Taken together, the research presented in this Special Issue highlights the need for a more nuanced and interdisciplinary approach to dialectology—one that incorporates insights from quantitative modeling, computational analysis, historical linguistics and cognitive science. By including work on lesser-studied languages and dialects, this collection moves beyond traditional European-centric perspectives, providing fresh evidence for the ways in which geography, economy, media, and social networks interact to shape linguistic diversity. These studies offer not only a deeper understanding of specific dialect landscapes but also broader theoretical insights into the mechanisms of linguistic change. As the field of dialectology continues to evolve, integrating these diverse approaches will be crucial in developing more comprehensive models of language variation and stability in an increasingly interconnected world.

Our wish to include more research on lesser-studied varieties is fulfilled here, e.g., by the papers on Asia Minor Greek, Mennonite Low German in Brazil, and to a lesser extent by the paper incorporating tone into the study of Chinese dialect pronunciations. In this vein we are also pleased that we can include a paper by Aidas Gudaitis on Lithuanian, a language with a rich history of research on dialects that has unfortunately often been less than optimally accessible (Mikulėnienė et al., 2019). The paper describes perhaps with more cartographic rigor than any found elsewhere how maps are being created for the work on Lithuanian. Gudaitis reconstructs shifting Lithuanian–Slavic boundaries in Southeastern Lithuania, showing how dialectal dynamics respond to long-term patterns of migration, urbanization, and socio-economic change. Applying GIS methods the study reveals both the erosion of Lithuanian in rural areas and its expansion in the Vilnius–

Trakai region. This approach parallels Vergeiner's spatial analysis of language and culture. Like the Zuiderzee study, Gudaitis highlights how mobility and economic centrality—not geography alone—shape patterns of linguistic convergence and retreat.

In a comparable effort to broaden methodology, and paralleling the structural focus seen in Tang et al. and Dunn and Wong, Matthew Sung and Jelena Prokić address a longstanding gap in dialectometric analysis by focusing on tonal variation in Chinese. While Chinese dialects have often been studied, Sung and Prokić further an important foray into an issue that has yet to yield to general dialectometric analysis, namely the assessment of tonal differences. This issue doubtlessly contributes to the distribution of language variation today (in China and in other areas where tone languages are spoken), and it may well contribute to our understanding of dialectal dynamics in the future. There is general consensus that dialects differ in their tones and that speakers of Chinese are sensitive to this, and there is also scholarly consensus in linguistics that tones belong to a suprasegmental level of phonology. The latter view of course suggests that it would be appropriate to combine the measure of tonal and segmental differences to a single measure of pronunciation difference, but a first step is the development of a measure of tone difference that can be applied to languages where single words and syllables are normally pronounced with a sequence of tones. Sung and Prokić apply edit distance to sequences of tones in Yue and Southern Pinghua, which is indeed a less-studied variety, using onset–contour–offset designations of the tones. They offer an analysis of 104 varieties based on 130 monosyllabic words as a proof of concept, but leave efforts towards combining tonometry (their term) and measures of segmental differences as well as validation for future work.

## 5. Prospectus

Where might progress lie in the future? We begin by noting a technical aspect of Buurke's (2025) work that offers a perspective on dealing with situations where two (and perhaps more) potential sources of change are to be compared. Buurke compared pronunciations, represented as sequences of phonetic segments, and then applied a three-dimensional version of Levenshtein distance developed by Heeringa and Hinskens (2015). Buurke could compare two varieties, e.g., a basal Frisian variety and standard Frisian, from the perspective of a third, e.g., standard Dutch (Buurke, 2025, pp. 111–115), opting for more operations with lower costs, usually due to greater similarity (which offers a perspective on attributing changes to their correct sources that clearly deserves further development and evaluation).

We turn then to more general issues. We unfortunately could not include papers on lectal (or varietal) coherence, an interesting new concept gaining favor in sociolinguistics (Guy & Hinskens, 2016), which we bring up here because it has been claimed to be a sign of linguistic health (Beaman, 2020), in which case it would clearly indicate directions in dialectal dynamics. One of Beaman's hypotheses was "that 'less coherent' lects are more vulnerable to change while 'more coherent' lects are more resistant to change" (Beaman, 2020, p. 261). The research line in lectal coherence shares with quantitative dialectology an emphasis on multivariate analysis as well as an emphasis on validation. In dialectometry this has been pursued in studies examining the degree to which analyses correlate with dialect speakers' intuitions of similarity (Gooskens & Heeringa, 2004) and by the emphasis in both undertakings on the need to examine many variables.

Lects or varieties are regarded as coherent to the degree that representative linguistic variables tend to correlate in their values. The research line then asks whether there are important differences in the coherence of varieties, and this is definitely a new question in variationist linguistics. Beaman and Guy (2022) are clear in their introduction to a collection

of papers on this topic that there is still a lively ongoing discussion of exactly how to measure the coherence. In the spirit of the Chambers–Trudgill fusion we would like to encourage a broader bandwidth of communication between the two communities (which of course overlap a bit already). Dialectologists have developed techniques for identifying the most characteristic elements in varieties (Prokić et al., 2012). Given a variety to be contrasted with others, one identifies characteristic elements such as lexicalizations (e.g., TONIC vs. SODA), sound realizations (e.g., [ø] vs. [ɪ] in coda position such as CAR), or construction differences (whether the modal need must combine with an infinitive vs. a passive participle, i.e., NEEDS TO BE WASHED vs. NEEDS WASHED). Elements are characteristic if they are representative, i.e., tend to occur throughout the variety, and also distinctive, i.e., tend not to occur otherwise. The lectal coherence discussion aims to extend this sort of work in measuring the degree of coherence of variables and even more, in comparing varieties with respect to this coherence, but collaboration might accelerate progress.

Simulation is a second, very general tool that might see much more use in studying dialectal dynamics. Kretzschmar and Juuso (2014) describe a cellular automaton they have used to study and illustrate the value of simulation in studying dialectal dynamics, and we should like to underscore that point here. Most variationists work quite empirically, gathering data carefully and then analyzing it statistically to confirm or refute more abstract hypotheses about variation. Computer simulations offer an alternative path to study variation. One of the better-known papers on the influence of simple distance included an (overly) simple simulation of the influence of likely contact on dialect similarity (Nerbonne, 2010). Kretzschmar (2010) is a book-length study on applying complex dynamic systems theory to dialectology.

In general, simulations offer a great deal of freedom in exploring, e.g., the relative importance of the different influences on variation—geographic, social, linguistic, and historical—all of which in turn range over a myriad of potential forms. For example, geographical influence may be primarily based on regional divisions, or on simple distance (perhaps enhanced with topography or overlaid with information on accessibility, such as travel time), but there may be demographic forces at work as well. We will not attempt to illustrate further the variety of forms other influences may take, since it should suffice to say that dialectal dynamics is not at a loss for potential factors. But empirical studies are expensive, time-consuming and ever exposed to the dangers of confounds, imperfect operationalizations, and, of course, errors in data capture and measurement. Computer simulations not only offer an alternative method for examining dynamics and suggesting new hypotheses; they are also much more easily revised!

So, we also encourage more research exploring simulation. This will involve attention to yet another heretofore unmentioned field of collaboration, that of physicists studying language dynamics. One highly cited paper is Abrams and Strogatz (2003), in which they examine language death, surely an important type of event in language and dialect dynamics. They used a single differential equation to achieve this, but a lot of the field seems to have moved on to simulations. Zankoc et al. (2024) use simulations to study the diffusion of linguistic traits, and their paper illustrates how well this research line is able to derive hypotheses from its simulations, e.g., focus on interactions at different distances, show the effect of self-organization, and examine the degree of diffusion over time. They introduce the interesting question of how large the clusters of similarity tend to be, one we have not seen mentioned in the linguistic literature. We think their paper would have improved with the collaboration of linguists because they run their simulations from an initially random distribution of language similarity and then model how attraction (convergence) between individuals with similar languages might lead to regions of similarity like ones seen in dialectology. We know historically that divergence has been the primary dynamic, and that

similar varieties often later diverged (this is the history of the large language families such as Romance, Germanic or Bantu), but a lot of the processes in later dialectal development are processes of convergence, meaning that their simulations might be directly used to study these.

As a third point in this prospectus, we suggest that energy might be devoted to more and better analytical tools for studying dynamics; i.e., we would like to signal aspects of change instead of merely detecting differences in real or apparent time. We would like to see probes of synchronic language samples which detect changes and potential sources of those changes or which predict a language's susceptibility to change. This motivated us to introduce the work on lectal coherence (above). If lectal coherence indeed signals a decreased susceptibility to change in a given variety, then we should wish to measure it reliably, i.e., defining which (sorts of) variables to inspect and how to gauge coherence statistically (part of the volume Beaman & Guy, 2022 introduces). Further tests of its predictive power with respect to dynamics are naturally the goal of such preliminaries.

Finally, and as a fourth point, Artificial Intelligence (AI) methods will have an increasing impact on dialectological research and its applications in the future. Recent advances in AI technology such as Large Language Models (LLMs) provide an unprecedented array of capabilities for understanding, generating, translating, and analyzing linguistic productions across a wide variety of contexts. However, despite their broad applicability, the effectiveness of LLMs in dialect-specific settings—especially in the case of low-resource dialects—remains an underexplored area. This research gap presents a critical challenge, which is currently being addressed from multiple angles, in a variety of tasks such as dialect classification, machine translation, and speech synthesis to mention only a few.

Bartelds et al. (2022) propose a method for measuring pronunciation differences (of the same words or concepts in different dialects) using the activation patterns in an LLM. The authors used a 24-level LLM that had been trained on nearly 1000 h of speech (Librispeech) using the customary task of predicting the next sound. The training consumed 150 days of a state-of-the-art processor. Their results not only hold the promise of enabling measurements without recourse to phonetic transcription, but, in an evaluation based on data used in the best transcription-based processing study to date, the LLM-based technique reduced error by roughly 40%!

Generative AI opens up new possibilities for further tasks such as dialect generation, simulation and adaptation in human–computer interaction, where dialect-sensitive responses could improve user engagement and inclusivity. Recent advances in speech synthesis and "dialect cloning"—the generation of speech output in region-specific accents or sociolects—show how AI can replicate subtle phonetic and prosodic features of dialects (Fischer et al., 2025). LLMs are also increasingly being applied to dialectal machine translation. Notable examples include systems translating between specific dialects and the standard language, as in the case of Arabic dialects and Modern Standard Arabic (Al-Khalifa et al., 2024), or between different—typically low-resource—dialects (Ondrejová & Šuppa, 2024).

These developments point to a future in which AI systems will not only be able to recognize and analyze dialectal differences, but also interact in dialect-specific ways, giving them greater cultural and linguistic salience. However, this advance also raises critical questions regarding the authenticity, ethical implications and sociolinguistic impact of AI-mediated dialect use, all of which need to be carefully considered by the research community.

Of course, there may be other excellent probes for detecting ongoing or recent changes that we do not yet see on the horizon. Blaha Pfeiler and Skopeteas (2022, p. 19) creatively appeal to demographic information to develop a "distribution index" in a situation where

Mayan language varieties are changing due to the influence of migration. We are undoubtedly not aware of all the relevant ongoing work, but we look forward to further progress!

**Author Contributions:** J.N. suggested the Special Issue and approached A.L. and S.M. about collaboration. The authors collaborated on inviting contributions, and on conceptualizing and writing the introduction. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Acknowledgments:** As editors we are indebted to the authors of the papers in this volume and also to the referees who advised us and the authors on optimizing the papers. We were supported professionally by the MDPI office.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

Abdul-Mageed, M., Keleg, A., Elmadany, A., Zhang, C., Hamed, I., Magdy, W., Bouamor, H., & Habash, N. (2024). NADI 2024: The fifth nuanced arabic dialect identification shared task. In *The second Arabic natural language processing conference* (pp. 709–728). Association for Computational Linguistics.

Abrams, D. M., & Strogatz, S. H. (2003). Modelling the dynamics of language death. *Nature*, *424*, 900. [CrossRef]

Al-Khalifa, H., Darwish, K., Mubarak, H., Ali, M., & Elsayed, T. (Eds.). (2024). *The 6th workshop on open-source Arabic corpora and processing tools (OSACT) with shared tasks on Arabic LLMs hallucination and dialect to MSA machine translation*. Available online: https://aclanthology.org/events/ws-2024/#2024osact-1 (accessed on 20 June 2025).

Atkinson, Q. D., & Gray, R. D. (2005). Curious parallels and curious connections—Phylogenetic thinking in biology and historical linguistics. *Systematic Biology*, *54*(4), 513–526. [CrossRef]

Auer, P., & Hinskens, F. (1996). The convergence and divergence of dialects in Europe. New and not so new developments in an old area. *Sociolinguistica*, *10*(1), 1–30. [CrossRef]

Bartelds, M., de Vries, W., Sanal, F., Richter, C., Liberman, M., & Wieling, M. (2022). Neural representations for modeling variation in speech. *Journal of Phonetics*, *92*, 101137. [CrossRef]

Beaman, K. V. (2020). *Coherence in real- and apparent-time: A sociolinguistic study of Swabian* [Ph.D. thesis, Queen Mary University of London].

Beaman, K. V., & Guy, G. R. (2022). The coherence of linguistic communities: Orderly heterogeneity and social meaning. In K. Beamon, & G. Guy (Eds.), *The coherence of linguistic communities* (pp. 1–13). Routledge.

Blaha Pfeiler, B., & Skopeteas, S. (2022). Sources of convergence in indigenous languages: Lexical variation in Yucatec Maya. *PLoS ONE*, *17*(5), e0268448. [CrossRef] [PubMed]

Bowern, C. (2018). Computational phylogenetics. *Annual Review of Linguistics*, *4*(1), 281–296. [CrossRef]

Britain, D. (2002). Space and spatial diffusion. In J. K. Chambers, P. Trudgill, & N. Schilling-Estes (Eds.), *The handbook of variation and change* (pp. 603–637). Blackwell.

Buurke, R. (2025). *Frisian and low Saxon in flux* [Ph.D. dissertation, University of Groningen].

Cedergren, H. J., & Sankoff, D. (1974). Variable rules: Performance as a statistical reflection of competence. *Language*, *50*(2), 333–355. [CrossRef]

Chambers, J. K., & Trudgill, P. (1998). *Dialectology*. Cambridge University Press.

Crain, S., Goro, T., & Thornton, R. (2006). Language acquisition is language change. *Journal of Psycholinguistic Research*, *35*(1), 31–49. [CrossRef]

Croft, W. A. (2021). A sociolinguistic typology for languages in contact. In E. Aboh, & C. Vigouroux (Eds.), *Variation rolls the dice* (pp. 23–56). John Benjamins.

Eisenstein, J., O'Connor, B., Smith, N. A., & Xing, E. P. (2014). Diffusion of lexical change in social media. *PLoS ONE*, *9*(11), e113114. [CrossRef]

Faisal, F., Ahia, O., Srivastava, A., Ahuja, K., Chiang, D., Tsvetkov, Y., & Anastasopoulos, A. (2024). DIALECTBENCH: An NLP benchmark for dialects, varieties, and closely-related languages. In *Proceedings of the 62nd annual meeting of the association for computational linguistics (volume 1: Long papers)* (pp. 14412–14454). Association for Computational Linguistics.

Fischer, H., Lameli, A., Schubert, M., & Siegert, I. (2025, July 20–23). *Cloning dialects: Recreating and localizing dialectal voices*. IEEE International Professional Communication Conference (ProComm) (pp. 358–367), Sønderborg, Denmark.

Gaman, M., Hovy, D., Ionescu, R. T., Jauhiainen, H., Jauhiainen, T., Lindén, K., Ljubešić, N., Partanen, N., Purschke, C., & Zampieri, M. (2020). A report on the VarDial evaluation campaign 2020. In *Proceedings of the 7th workshop on NLP for similar languages, varieties and dialects* (pp. 1–14). International Committee on Computational Linguistics.

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

Guy, G. R., & Hinskens, F. (2016). Linguistic coherence: Systems, repertoires and speech communities. *Lingua*, *172–173*, 1–9. [CrossRef]

Heath, J. (2003). Review of W. Labov, principles of linguistic change. Vol. 2: Social factors. *Anthropological Linguistics*, *45*(4), 466–472.

Heeringa, W. (2004). *Measuring dialect pronunciation differences using Levenshtein distance* [Ph.D. dissertation, University of Groningen].

Heeringa, W., & Hinskens, F. (2015). Dialect change and its consequences for the Dutch dialect landscape. How much is due to the standard variety and how much is not? *Journal of Linguistic Geography*, *3*(1), 20–33. [CrossRef]

Jeszenszky, P., Hasse, A., & Stöckle, P. (2023). Dialect areas and contact dialectology. In R. van Gijn, H. Ruch, M. Wahlström, & A. Hasse (Eds.), *Language contact: Bridging the gap between individual interactions and areal patterns* (pp. 135–177). Language Science Press.

Johnson, D. E. (2009). Getting off the GoldVarb standard: Introducing rbrul for mixed-effects variable rule analysis. *Language and Linguistics Compass*, *3*(1), 359–383. [CrossRef]

Kretzschmar, W. A., Jr. (2010). Language variation and complex systems. *American Speech*, *85*(3), 263–286. [CrossRef]

Kretzschmar, W. A., Jr., & Juuso, I. (2014). Cellular automata for modeling language change. In *International conference on cellular automata* (pp. 339–348). Springer International Publishing.

Krückl, X. M., Blaschke, V., & Plank, B. (2025). Improving dialectal slot and intent detection with auxiliary tasks: A multi-dialectal Bavarian case study. In *Proceedings of the 12th workshop on NLP for similar languages, varieties and dialects* (pp. 128–146). Association for Computational Linguistics.

Labov, W. (1963). The social motivation of a sound change. *Word*, *19*(3), 273–309. [CrossRef]

Labov, W. (2001). *Principles of linguistic change (vol. 2): Social factors*. Blackwell.

Lameli, A. (2019). Spatial distribution of regional dialects: Interplay of language, community and action. In S. D. Brunn, & R. Kehrein (Eds.), *Handbook of the changing world language map* (pp. 1–19). Springer.

Lameli, A. (2025). Gesprochenes deutsch in den regionen. Eine standortbestimmung für die bundesrepublik deutschland. In N. Proske, T. Weber, M. Dannerer, & A. Deppermann (Eds.), *Gesprochenes deutsch.* Struktur, variation, interaktion (pp. 51–79). De Gruyter.

Mayr, E. (1954). Change of genetic environment and evolution. In J. Huxley, A. C. Hardy, & E. B. Ford (Eds.), *Evolution as a process* (pp. 157–180). Allen & Unwin.

Mikulėnienė, D., Čepaitienė, A., Bakšienė, R., Geržotaitė, L., Kardelytė-Grinevičienė, D., Leskauskaitė, A., Meiliūnaitė, V., & Vyniautaitė, S. (2019). *Dialektometrinis tradicinių lietuvių tarmių klasifikacijos pjūvis: Žvalgomasis tyrimas [Dialectometric classification of traditional Lithuanian dialects: An exploratory study]*. Lietuvių Kalbos Institutas.

Montemagni, S., Wieling, M., De Jonge, B., & Nerbonne, J. (2013). Synchronic patterns of Tuscan phonetic variation and diachronic change: Evidence from a dialectometric study. *Literary and Linguistic Computing*, *28*(1), 157–172. [CrossRef]

Mufwene, S. S. (2008). Colonization, population contacts, and the emergence of new language varieties: A response to Peter Trudgill. *Language in Society*, *37*(2), 254–259. [CrossRef]

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

Nerbonne, J. (2010). Measuring the diffusion of linguistic change. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1559), 3821–3828. [CrossRef]

Ondrejová, V., & Šuppa, M. (2024). Can LLMs handle low-resource dialects? A case study on translation and common sense reasoning in Šariš. In *Proceedings of the eleventh workshop on NLP for similar languages, varieties, and dialects* (pp. 130–139). Association for Computational Linguistics.

Prokić, J., Çöltekin, C., & Nerbonne, J. (2012). Detecting shibboleths. In *EACL 2012 joint workshop of LINGVIS & UNCLH. Visualization of linguistic patterns and uncovering language history from multilingual resources* (pp. 72–80). Association for Computational Linguistics.

Schmidt, J. E. (2010). Language and space: The linguistic dynamics approach. In P. Auer, & J. E. Schmidt (Eds.), *Language and space: Theories and methods* (pp. 201–225). de Gruyter Mouton.

Séguy, J. (1971). La relation entre distance spatiale et distance linguistique. *Revue de Linguistique Romane*, *35*, 335–357.

Tagliamonte, S., & Baayen, H. (2012). Models, forests and trees of York English: Was/were variation as a case study for statistical practice. *Language Variation and Change*, *24*(2), 135–178. [CrossRef]

Thomason, S. G., & Kaufman, T. (1988). *Language contact, creolization, and genetic linguistics*. University of California Press.

Trudgill, P. (1986). *Dialects in contact*. Basil Blackwell.

Wenker, G. (1876). *Über die verschiebung des stammsilben-auslauts im germanischen*. Bonn.

Wieling, M., Montemagni, S., Nerbonne, J., & Baayen, R. H. (2014). Lexical differences between Tuscan dialects and standard Italian: Accounting for geographic and socio-demographic variation using generalized additive mixed modeling. *Language*, *90*(3), 669–692. [CrossRef]

Wieling, M., & Nerbonne, J. (2011). Bipartite spectral graph partitioning for clustering dialect varieties and detecting their linguistic features. *Computer Speech and Language*, *25*(3), 700–715. [CrossRef]

Wieling, M., & Nerbonne, J. (2015). Advances in dialectometry. *Annual Review of Linguistics*, *1*(1), 243–264. [CrossRef]

Wieling, M., Nerbonne, J., & Baayen, R. H. (2011). Quantitative social dialectology: Explaining linguistic variation geographically and socially. *PLoS ONE*, *6*(9), e23613. [CrossRef]

Wright, S. (1940). Breeding structure of populations in relation to speciation. *The American Naturalist*, *74*(752), 232–248. [CrossRef]

Zankoc, C., Heinsalu, E., & Patriarca, M. (2024). Language dynamics model with finite-range interactions influencing the diffusion of linguistic traits and human dispersal. *The European Physical Journal B*, *97*(6), 66. [CrossRef]