JS Journal of Historical Syntax
Volume 8, Article 7: 1–12, 2024 · DOI: 10.18148/hs/2024.v8i7.239

INTRODUCTION: TOWARDS A COMPARATIVE HISTORICAL DIALECTOLOGY

Ann-Marie Moser
University of Zürich

Lea Schäfer
University of Kassel

Sophie Ellsässer
University of Osnabrück

Abstract In this article, we highlight the challenges facing research in the field of comparative historical dialectology, with a focus on morphological and syntactic phenomena. We define the term "historical comparative dialectology" and use four different studies from different varieties to illustrate how the aims of historical comparative dialectology can be approached. The studies consider different varieties and language stages (different dialects of Old and Modern Catalan, the dialectal continuum of Yucatec Maya, historical dialects of High German, different Occitan dialects) and discuss various phenomena from a historical comparative perspective: differential object marking, numeral classifiers, number distinctions of nouns, and inflectional patterns.

1 The purpose of this Special Collection

The Special Collection is the result of a workshop at the conference of the *Deutsche Gesellschaft für Sprachwissenschaft* (German Society for Linguistics) which took place in Tübingen in February 2022. The aim of the workshop was to unite experts in different varieties, grammatical structures, and fields (dialectology, typology, and historical linguistics) to discuss three key questions towards a historical comparative dialect syntax and morphology:

i. Language comparison: Are there formal or functional similarities/ differences between cross-linguistic morpho-syntactic phenomena?

ii. Philological: How to identify historical dialects and which types of sources are suitable for a historical comparative dialectology?

iii. Methods: Which (geo-)statistical methods can help to model conclusions about language change processes? (state of the art)

The feasibility of a historical comparative dialectology is demonstrated by numerous phonological studies (cf. Cravens 2002). For syntax and morphology, however, we have just begun to identify and analyze historical oral varieties using fine-scaled geolinguistic, statistical, and philological methods. Dialect syntax was long considered the stepchild of dialectology (see Glaser 2000). In the meantime, this gap has been reduced, especially in relation to the modern Germanic and Romance dialects. It is therefore time to leave the comfort zone of modern dialects and to establish a historical dialect syntax and morphology. An overarching goal would be to create a comparative compilation of these results from the individual varieties to gain general knowledge about language change.

In the following, we will first introduce the term "comparative historical dialectology" (section 2) and then point out the challenges of a historical dialectology (section 3). In section 4 we present the results of the Special Collection, discuss them in relation to our three key questions mentioned in section 1, and finally point out future research perspectives.

2 Comparative historical dialectology: terms and concepts

We speak explicitly of a historical dialectology and not of historical linguistics as dialects are, from a diachronic perspective, more natural than standard/written languages. Weiß (2004: 182–185) argues that languages such as German or English were once languages without native speakers and were secondarily learned written languages. The linguistic situation in European languages, up to the 19th century, "was characterized by the existence of only a written standard and spoken dialects" (Weiß 2004: 183). Dialects are therefore an important source for linguistics as "data from standard languages may not always qualify as evidence in linguistics" (Weiß 2004: 202). Diachronically speaking the "true descendants of the older stages of the language are the modern dialects" and it is the standard language (e.g., Standard German, Standard English) "that is the deviant variety that needs to be especially explained" (Simon 2004: 210).

We define a dialect as being consistent "of groups of actually communicating individuals" (Weiß 2009: 257); a language, on the other hand, "consists of groups of mutually comprehensible dialects, which are communicatively isolated from other such groups" (Weiß 2009: 257). The relevant criterion, which is "standardly assumed in sociolinguistics" (Weiß 2009: 258), is thus mutual intelligibility. In practice, speaker's perception also plays a role: For example,

Serbo-Croatian should be, based on linguistic findings, one language with two dialects (Weiß 2009: 258–259). By its speakers it is perceived as being two languages, and due to political developments (Weiß 2009: 258–259), we find today a Croatian and a Serbian national language.

The concept of *dialect* is closely linked to that of language variation as "[t]he input to any form of dialectology is language variation" (Laing & Lass 2006: 418). Areal patterns represent different linguistic systems (Weinreich 1954: 390) or "minimal gradations between grammars" (Seiler 2004: 384). Synchronic variation reveals diachronic aspects, as it shows what is possible in history (see e.g. the High German Consonant Shift). While dialectology studies language variation with a spatial emphasis, historical dialectology studies "historical linguistics with a spatial emphasis" (Laing & Lass 2006: 418). Historical dialectology brings together spatial and temporal aspects, and combines typology, dialectology, and linguistic reconstruction. It shows the range (and limits) of linguistic change and serves as a test field for mechanisms of language change such as innovation, diffusion, grammaticalization processes, morphological and syntactic change. The diachronic perspective shows particularly well the interplay of internal and external factors (Gerritsen & Stein 1992: 7).

By *internal factors* we understand "those inherent in, and arising out of, any given synchronic state of the language system" (Gerritsen & Stein 1992: 7). *External factors*, on the other hand, are those forces which arise "out of the location and use of language in society" (Gerritsen & Stein 1992). Diachronic variation is even more complex than synchronic variation as language change must also be modeled.

Comparative historical dialectology is all the more challenging because its focus requires not only description, analysis, and modeling, but also comparison with other dialects and languages. Laing & Lass (2006: 418) name three goals or tasks of an historical dialectology: to describe and analyze (a) how linguistic forms and structures change through time; (b) how they vary across space – that is, the country or region where they were spoken and written; and (c) how the situation and intentions of the speakers and writers of the language engender this variation (cf. Laing & Lass 2006: 418). A comparative historical dialectology now requires, in addition, a (rather challenging) fourth goal: the comparative compilation of these results from the individual varieties to gain general knowledge about language change.

3 Challenges of a historical dialectology

The literature mentions numerous challenges that historical dialectology entails (cf. e.g. Laing & Lass 2006; Labov 2010: 10–11; Becker & Schallert 2021):

- Text samples "depend on the contingent survival of text witnesses" (Laing & Lass 2006: 418), they "survive by chance [and] not by design" (Labov 2010: 11). Even if we therefore have numerous grammars for older language stages in some languages such as German, these often only offer a very limited picture of regional variation (Becker & Schallert 2021: 199–200). The further back one goes, the fewer sources generally are available, so that for most European languages we have a more detailed picture of the regional variants in modern times than in the Middle Ages or in ancient times.

- The further back we go, the less information we have about our textual witnesses or authors: Laing & Lass (2006: 418) speak of the "increasing opacity of social milieu with the passing of time". The variation does not necessarily have to be regional but can also be the result of text genre or simply a copied/adopted form from another region (Becker & Schallert 2021: 199–200).

- When working with older editions (instead of diplomatic editions or electronic facsimiles), it is possible that the editor has intervened to correct and "normalize" texts, at least this is known to be the case in German philology: "Many older editions, including the canonical texts of the courtly period (c. 1170–1250 CE), present a somewhat artificial version of Middle High German and its complexity" (Becker & Schallert 2021: 200).

- The text samples represent written rather than spoken data: "The 'native speakers' of past stages of a language are writers and copying scribes. Their output is our only source material; there is no recourse to language data of any other kind" (Laing & Lass 2006: 418). In addition, there are standardization processes, i.e. signs of an emerging supra-regional writing habit. The authors/writers are often oriented towards a prestigious variety, which is why many written texts cannot be considered representative of the "native language" of the author/writer (Fleischer & Schallert 2011: 30–31).

- We cannot interview speakers, that is, we can only describe the structures which are attested. We cannot, however, say which structures are ungrammatical in a certain variety (Fleischer & Schallert 2011: 31).

- In historical dialectology, "it is rarely possible to achieve the fineness of resolution typical of studies of contemporary language states. Variables will differ according to which historical vernacular is under scrutiny and at what period" (Laing & Lass 2006: 418).

- Syntax is a data-hungry science, and we need vast amounts of texts to make (reliable) statements about forms and functions (Fleischer & Schallert 2011: 72–73).

4 Results and further perspectives

In this section, we will first present the individual contributions. We will then collect the results of the contributions and summarize them in relation to our three key questions. We will finally identify perspectives for future research.

Anna Pineda's article is about the dialectal-historical development of Differential Object Marking (DOM) in different dialects of Old and Modern Catalan. The dialects can be, broadly speaking, divided into western (Valencian Catalan and North-Western Catalan) and eastern dialects (Central Catalan, Rossellonese Catalan, Balearic Catalan, and Alguerese Catalan). Her study comprises a sample of 69 text extracts, starting from the earliest texts written in the 11th to 13th centuries and continuing to the 18th century. This corresponds to the periods usually classified as Old Catalan (11th–16th centuries) and Modern Catalan (17th–18th centuries). The selection includes various text types from different regions, representing a representative choice of the diachronic, diatopic, diaphasic and diastratic variation of Old/Modern Catalan. In her study, Pineda focuses on DOM with proper nouns, with definite NPs and with indefinite NPs. She suggests that texts belonging to Western Catalan, and especially to the Valencian area, show a much more advanced stage of grammaticalization of DOM than dialects belonging to the eastern part. In these texts, DOM appears earlier and also spreads faster to the different types of objects, from the more prominent ones in terms of animacy and definiteness to the less prominent ones, that is, from personal pronouns to proper names, then to definite NPs and, eventually, even to indefinite NPs. Her study highlights how syntactic change can be tracked more precisely by taking historical dialectal varieties into account. Since DOM is also documented in many other languages, it would be desirable for future studies (in other languages) to also include historical dialects and thus show spatial variation and change.

The variation that we observe in space today can also be the result of processes that took place much earlier: The study by Barbara Blaha Pfeiler and Stavros Skopeteas is based on data from the Atlas of Yucatec Maya and draws conclusions about diachronic language change on the basis of the observed microvariation. Yucatec Maya is spoken on the peninsula of Yucatán (Mexico) and dialectal variation is manifested as a continuum in geographical space. The study centers on the use of numeral classifiers: Some speakers employ the general classifier instead of a specific sortal classifier, and other

speakers insert the general classifier in measure expressions although these constructions could be formed with an expression of measure alone (without a difference in meaning). The question of their study is therefore whether these phenomena reflect distinct diachronic processes or just result from a single process. Based on data collected in 85 locations in the peninsula of Yucatán, the authors identify a partially overlapping pattern in the distribution of these constructions. They show that this overlapping pattern can be accounted for by two diachronic processes (and not by one single source of variation). Their data further reveal that the two diachronic processes are partially related, but still vary independently in various respects. These findings underline the relevance of examining the dimensions of microvariation in order to test hypotheses about syntactic change. Their research highlights that empirical data, as obtained in the context of microvariational studies, challenge previous generalizations: the need for granularity increases along with the size of the data set.

Nathalie Fromm uses historical reference corpora in order to initially investigate dialectal dimensions of a prominent phenomenon in German morphology: the strengthening of number distinctions. Fromm focuses on nouns of neuter a-stems which did not show formal distinctions of singular and plural in Old High German (sg. *wort* 'word', pl. *wort* 'words') but do so (at least in Standard) New High German (sg. *Wort* 'word', pl. *Worte* or *Wörter* 'words'). So far, there have been studies on the diachronic development of the phenomenon towards Standard German on the one hand and synchronic descriptions of dialects that lack number strengthening in certain parts on the other. Fromm deduces that the strategies of establishing number distinctiveness differ between (different) dialects and Standard German. She focuses on the period from the early 9th century to the mid-17th century, a time in which no German Standard language was established. Based on a contrastive analysis of reference corpora (Referenzkorpus Althochdeutsch, Referenzkorpus Mittelhochdeutsch, Referenzkorpus Frühneuhochdeutsch), she comes to the conclusion that unmarked plural forms (showing no number strengthening) predominate in all dialects for the nouns with neuter *a*-stems in her investigation period. If plural is marked at these lexemes, the individual dialect areas differ in the strategies used. While Upper German as well as Eastern Central German dialects prefer *-er* as a plural marker, Western Central German dialects prefer *-e*. Fromm links these strategies to different phonological characteristics in the dialects (in particular to the phonological loss of the suffix *-e*). In some cases these characteristics seem to have an influence on morphology, in other cases they do not. Despite these differences, there are common developments in the dialects: Fromm finds evidence that the animacy fac-

tor is central to the development of number strengthening. The results show that it is worth examining diachronic processes in spatial dimensions. On the one hand, this allows differences in recent varieties to be explained, and on the other hand, different factors influencing morphological change and the resulting variations can be analyzed.

The article by Louise Esher, "Areal continuities and discontinuities emerge from parallel studies of inflection in diachrony," explores the diachronic development of inflectional patterns in Occitan dialects. Through a longitudinal study, the research examines analogical changes in preterite forms across different Occitan-speaking regions, offering insights into historical dialectology and inflectional analogy. The study identifies four distinct groups of speech varieties, each showing unique developmental trajectories and internal consistency over several centuries. For example, Gascon varieties exhibit a suite of unique changes, while certain Lengadocian varieties around Toulouse follow an entirely separate developmental path. In a large group of Lengadocian, Provençal, and Alpine varieties, changes occur simultaneously across these regions, indicating near-simultaneous development. In contrast, northern Occitan varieties such as Lemosin and Auvergnat display the same changes at a later period, suggesting diffusion from southern varieties or independent parallel development. This research highlights how morphological characteristics can reveal historical dialectal divisions and continuity, complementing studies of lexical and phonological features. The findings suggest that longstanding dialect realities can be better understood through the study of historical linguistic changes.

5 Conclusion

The following conclusions can be drawn with regard to our questions:

- Language comparison: Are there formal or functional similarities/ differences between cross-linguistic morpho-syntactic phenomena?
  - Based on the work by Pineda we can see that different diachronic development paths can be identified in the development of DOM: Western dialects are more advanced in the grammaticalization and diffusion of DOM than Eastern dialects. Pineda herself suggests that DOM is a very interesting candidate for future studies on historical dialectology: It is present in a great number of languages of the world and is subject to variation even within a given language family (Romance in this case). In addition, we think that the comparison with other languages/varieties is suit-

able for obtaining information about (i) language change in morphosyntax and (ii) about the interplay of language change mechanisms with typological generalizations such as the Animacy Hierarchy.

- The comparative approach of Esher's study of different dialects within Occitan can shed light on how similar morphological changes can occur independently in different regions or spread across regions, revealing patterns of morphological changes, especially analogy, its direction, nature and motives.
- Fromm has also shown that although phonological influences sometimes trigger variation in morphological marking, this is not necessarily the case. Strategies of plural marking can be influenced by phonology, but there are also varieties that are affected by the same phonology but have different morphological strategies. However, the article has also shown that there are semantic affected that affect morphological structures in all dialects and are therefore probably stronger than phonological characteristics.
- Dialectal variation in Maya is manifested as a continuum in geographical space, with the major source of diversity being the East-West axis. The study by Blaha Pfeiler and Skopeteas confirms this finding by showing that differences in the use of sortal classifiers are due to differences between areas (and not due to differences between ages or between urban vs. rural centers). Based on data from contemporary Yucatec Maya they infer two historical processes that underlie the attested synchronic variation between dialects: (a) the generalization of the classifier *p'éel* and (b) the loss of the dual nature of mensural classifiers. To summarize, their study is thus an example of how syntactic variation in space reflects diachronic processes.

- Philological: How to identify historical dialects and which types of sources are suitable for a historical comparative dialectology?

  - The data offered in the paper by Pineda comes from two online available corpora, they were searched manually. Her corpus includes 69 texts, with a total of 1,500,000 words. For each work the first 30,000 words (for Old Catalan) and the first 15,000 words (for Modern Catalan) have been analyzed (human direct objects appear quite frequent). The texts can be assigned to different

genres (fiction, administrative prose, legal and juridical documents), and it must remain open to what extent one can speak of oral varieties (Pineda does not discuss this aspect).

  - Since no tagged electronic corpora are currently available for the Occitan historical dialects from the 14th to the 19th century, Esher's paper relies on manual consultation of published texts from the period in question (cf. Esher 2021a, Esher 2021b, Esher 2021c, Esher 2021d). The available historical documents, accessible through published and digitized editions, vary greatly in type and quantity across different regions and periods. In the 14th and 15th centuries, the focus was on administrative documents, historical chronicles, and religious mystery plays. The 16th and 17th centuries saw Occitan used for religious and literary purposes, including hagiographies, dramas, and narrative poetry. By the 18th century, narrative and lyric poetry, religious doctrine, and political propaganda in Occitan were prevalent, aimed at reaching the general populace.
  - Fromm can rely on annotated and digitally accessible reference corpora that date from a period before the establishment of a German standard variety and show a degree of geographical variation. Fromm also does not go into detail about the extent to which these mirror oral varieties, but it can be assumed that we are dealing here with regional writing rather than conceptual orality.
  - The contribution by Skopeteas and Blaha Pfeiler is not based on historical dialects, but on spoken language data from the 21st century. Yucatec Maya is an indigenous endangered language (still) spoken on the peninsula of Yucatán (Mexico); earlier written records of the culture ("historical dialects") are still difficult to access. With their study on microvariation, they show that conclusions about historical processes can also be drawn from contemporary variation and that, in turn, the dialectal perspective helps understanding diachronic change.

- Methods: Which (geo-)statistical methods can help to model conclusions about language change processes? (state of the art)

  - Pineda conducts a qualitative and quantitative study and traces grammaticalization processes with the help of (changing) percentages over time. This method is also used by Fromm in the analysis of number strengthening. In so doing, she can use the

assignments to larger dialect areas within her corpus data as a guide.

- Esher's study examines the geographical distribution of morphological changes, highlighting areas of continuity and discontinuity within the Occitan dialect continuum. This geographical perspective helps understanding how dialects have evolved over time and space.
- Blaha Pfeiler and Skopeteas use statistics to analyze syntactic change: they test the interaction effect between space and measure type in the analysis of mensural classifiers. In order to estimate the effect of space, they use Generalized Additive Models (GAM), which assess the factorial effects by fitting non-parametric smoothers. The significance of the effects at issue is assessed with model comparison, based on a backwards-elimination procedure. The calculations were made in R, with the *itsadug* package for model comparison and the package *mgcv* for generalized additive models.

In summary, our Special Collection shows that a comparative historical dialectology is feasible and that it refines or even corrects findings on language change in individual languages and varieties. At the same time, however, we see clear challenges with regard to suitable data ("oral historical varieties"). As far as geostatistical methods are concerned, we believe we are just at the beginning and can (and should aim) to cooperate with relevant experts in this area. All in all, the future of comparative historical dialectology lies not only in intensive philological work, but also in interdisciplinary projects and goals.

References

Becker, Carsten & Oliver Schallert. 2021. Areal variation in Middle High German: A perspective from charters. *North-Western European Language Evolution* 74(2). 199–241.

Cravens, Thomas D. 2002. *Comparative historical dialectology: Italo-Romance clues to Ibero-Romance sound change*. Amsterdam/New York: John Benjamins.

Esher, Louise. 2021a. Hypercorrection and velar-to-labial change in Occitan preterites. *Zeitschrift für romanische Philologie* 137(1). 61–98.

Esher, Louise. 2021b. La diachronie du prétérit occitan à la lumière des sources textuelles. *Linguistica Pragensia* 31(2). 115–136.

Esher, Louise. 2021c. Les prétérits nord-occitans à consonne thématique /t/. *Revue des Langues Romanes* 125(1). 145–162.

Esher, Louise. 2021d. Prospettive comparative e storiche sulla flessione del preterito nelle parlate occitane della provenza. *Vox Romanica* 80. 203–238.

Fleischer, Jürg & Oliver Schallert. 2011. *Historische Syntax des Deutschen: Eine Einführung*. Tübingen: Narr.

Gerritsen, Marinel & Dieter Stein. 1992. Introduction: On 'internal' and 'external' in syntactic change. In Marinel Gerritsen & Dieter Stein (eds.), *Internal and external factors in syntactic change*, 1–16. Berlin/New York: De Gruyter Mouton.

Glaser, Elvira. 2000. Erhebungsmethoden dialektaler Syntax. In Dieter Stellmacher (ed.), *Dialektologie zwischen Tradition und Neuanstößen*, 258–276. Stuttgart: Steiner.

Labov, William. 2010. *Principles of linguistic change. volume 1: internal factors*. Oxford/Cambridge (Mass.): Blackwell.

Laing, Margaret & Roger Lass. 2006. Early Middle English dialectology: Problems and prospects. In Ans van Kemenade & Bettelou Los (eds.), *The handbook of the history of English*, Oxford/Cambridge (Mass.): Blackwell.

Seiler, Guido. 2004. On three types of dialect variation, and their implications for linguistic theory: Evidence from verb clusters in Swiss German dialects. In Bernd Kortmann (ed.), *Dialectology meets typology: Dialect grammar from a cross-linguistic perspective*, 367–399. Berlin/New York: De Gruyter Mouton.

Simon, Horst. 2004. The relevance of Variation. Remarks on Weiß's standard-dialect-problem. In Martina Penke & Anette Rosenbach (eds.), *What counts as evidence in linguistics? The case of innateness*, 209–213. Amsterdam/New York: John Benjamins.

Weinreich, Uriel. 1954. Is a structural dialectology possible? *Word* 10. 388–400.

Weiß, Helmut. 2004. A question of Relevance. Some Remarks on Standard Languages. In Martina Penke & Anette Rosenbach (eds.), *What counts as evidence in linguistics? The case of innateness*, 648–674. Amsterdam/New York: John Benjamins.

Weiß, Helmut. 2009. How to define dialect and language: A proposal for further discussion. *Linguistische Berichte* 219. 251–270.

Dr. Ann-Marie Moser
Deutsches Seminar
Schönberggasse 9
CH-8001 Zürich
ann-marie.moser@ds.uzh.ch

Prof. Dr. Lea Schäfer
Insitut für Germanistik
Kurt-Wolters-Straße 5
D-34125 Kassel
lea.schaefer@uni-kassel.de

Juniorprof. Dr. Sophie Ellsäßer
Germanistik
Neuer Graben 40
D-49074 Osnabrück
sophie.ellsaesser@uni-osnabrueck.de