

Areal variation in Middle High German: Methodological and quantitative aspects

Oliver Schallert¹, Carsten Becker², and Helmut Schmid³

¹LMU Munich, Departement of German
⟨oliver.schallert@lmu.de⟩

²Marburg University, Institut für Deutsche Philologie des Mittelalters
⟨carsten.becker@staff.uni-marburg.de⟩

³LMU Munich, Centrum für Informations- und Sprachverarbeitung
⟨schmid@cis.lmu.de⟩

July 4, 2022

Using a large, electronically available corpus of Middle High German charters (originally compiled by Friedrich Wilhelm; cf. Wilhelm 1932–2004), we demonstrate that this text genre offers fine-scaled information on areal variation in the language period in question. Unlike most other historical documents, charters can usually be located exactly on the basis of textual evidence. We discuss and evaluate two methods for retrieving such information, i. e. (1) regular expressions (on the basis of spelling variants extracted from the *Wörterbuch der mittelhochdeutschen Urkundensprache* (WMU) ‘dictionary of Middle High German language of deeds’ (Kirschstein et al. 1994–2004) and (2) queries based on semi-automated annotations (with the aid of Schmid’s 2020 *RNNTagger*). Patterns of areal variation are indispensable for gaining a deeper understanding of fine-scaled grammatical variation (Bresnan, Deo, and Sharma 2007) or the diffusion of linguistic innovations (Nerbonne 2010), in particular the layering nature of grammaticalization processes that manifest in competing constructional variants (Hopper 1991; Girth 2000; Becker and Schallert 2021, 2022). What is more, they can be useful for localizing manuscripts whose provenience is unknown and/or controversial (cf. Fleischer 2019 who demonstrates that already the 100 most frequent word forms per manuscript allow for a general assessment).

We focus on so-called “contraction forms”, i. e. cases of shortening that occur with several verbs in this language period, e.g. *lâzen/lân* ‘let’, *haben/hân* ‘have’, etc. and are assumed to show a clear areal profile (Paul et al. 2007: 280–284; Klein, Solms, and Wegera 2018: 948–949), which is also confirmed by our data (cf. map 1 on *haben* [1PL.PRES]). Such forms cannot be solely explained by phonological deletion processes but are also sensitive to morphological factors like e. g. paradigmatic position and even syntactic effects (Nübling 1995). As cases of (weak) suppletion, they are also of a more general interest for morphological theory and typology (Nübling 2000; Corbett 2007). Contractions are also attested in modern dialects of German (which are usually well-documented in regional atlases, grammatical descriptions, or dictionaries) so that it is possible to compare and relate our findings to these data.

In order to obtain more general information on the written varieties of MHG, we performed an analysis of trigram tokens for all charters with unique issue location.¹ We compared the respective frequency profiles via cosine distance and used two techniques for visualizing/analyzing the ensuing distance matrix (Birkenes 2019):

- Multidimensional scaling (Heeringa 2004) yields a tripartite division that reflects the traditional difference between West and East Upper German quite clearly, whereas the central and northern regions show more overlaps (map 2).

¹ We thank Magnus Birkenes for his help with these analyses.

- This west-east divide in the southern regions is corroborated by a hierarchical cluster analysis (Ward's method with two clusters), as illustrated by map 3.

References

- Becker, Carsten and Oliver Schallert (2021). "Areal variation in Middle High German: A perspective from charters". *North-Western European Language Evolution* [NOWELE] 74.2, 199–241.
- (2022). "Urkunden als Quelle (schreib-)sprachlandschaftlicher Variation: Bestandsaufnahmen und Perspektiven". *Zeitschrift für deutsches Altertum und deutsche Literatur* [ZfdA] 151, 143–178.
- Birkenes, Magnus B. (2019). "North Frisian dialects: A quantitative investigation using a parallel corpus of translations". *Us Wurk* 68.3–4, 119–168.
- Bresnan, Joan, Ashwini Deo, and Devyani Sharma (2007). "Typology in variation: a probabilistic approach to *be* and *n't* in the Survey of English Dialects". *English Language and Linguistics* 11.2, 301–346.
- Corbett, Greville (2007). "Canonical Typology, Suppletion, and Possible Words". *Language* 83.1, 8–42.
- Fleischer, Jürg (2019). "Zur sprachlichen Einordnung mittelhochdeutscher Zeugnisse. Grundsätzliche und quantitative Explorationen anhand der 'Kaiserchronik'-Überlieferung". *Zeitschrift für deutsches Altertum und deutsche Literatur* 148.2, 258–284.
- Girnth, Heiko (2000). *Untersuchungen zur Theorie der Grammatikalisierung am Beispiel des Westmitteldeutschen*. (Reihe Germanistische Linguistik; 223). Tübingen: Niemeyer.
- Heeringa, Wilbert (2004). *Measuring Dialect Pronunciation Differences*. (Groningen dissertations in linguistics; 46.) Groningen: Rijksuniversiteit Groningen.
- Hopper, Paul J. (1991). "On some principles of grammaticalization". *Approaches to grammaticalization*. Ed. by Elisabeth C. Traugott and Bernd Heine. Vol. 1: Focus on theoretical and methodological issues. (Typological Studies in Language; 19.1). Amsterdam, Philadelphia: Benjamins, 17–35.
- Kirschstein, Bettina, Ursula Schulze, Sibylle Ohly, and Peter Schmitt, eds. (1994–2004). *Wörterbuch der mittelhochdeutschen Urkundensprache; auf Grundlage des Corpus der altdeutschen Originalurkunden bis zum Jahr 1300*. [3 Vols.] Berlin: Erich Schmidt.
- Klein, Thomas, Hans-Joachim Solms, and Klaus-Peter Wegera (2018). *Mittelhochdeutsche Grammatik. Teil 2: Flexionsmorphologie*. Berlin, Boston: De Gruyter.
- Nerbonne, John (2010). "Measuring the diffusion of linguistic change". *Philosophical Transactions of the Royal Society B: Biological Sciences* 365, 3821–3828.
- Nübling, Damaris (1995). "Kurzverben in germanischen Sprachen: Unterschiedliche Wege – gleiche Ziele". *Zeitschrift für Dialektologie und Linguistik* 62.2, 127–154.
- (2000). *Prinzipien der Irregularisierung: eine kontrastive Analyse von zehn Verben in zehn germanischen Sprachen*. Linguistische Arbeiten; 415. Tübingen: Niemeyer.
- Paul, Hermann, Thomas Klein, Hans-Joachim Solms, Klaus-Peter Wegera, Ingeborg Schröbler, and Heinz-Peter Prell (2007). *Mittelhochdeutsche Grammatik*. 25th ed. (Sammlung kurzer Grammatiken germanischer Dialekte, A. Hauptreihe; 2.) Tübingen: Niemeyer.
- Schmid, Helmut (2020). "Deep Learning-Based Morphological Taggers and Lemmatizers for Annotating Historical Texts". *Proceedings of Digital Access to Cultural Heritage*. New York, NY: Association for Computing Machinery, 133–137. doi: <https://doi.org/10.1145/3322905.3322915>.
- Wilhelm, Friedrich et al., ed. (1932–2004). *Corpus der altdeutschen Originalurkunden bis zum Jahr 1300*. Lahr: Schauenburg [Vol. 1–4]; Berlin: Erich Schmidt [Vol. 5].