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 Unlike most other historical documents, charters can usually be located exactly on the basis of textual evidence.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; Girnth 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