Abbreviation Application: A Stylochronometric Study of the Abbreviations in the Oeuvre of Herne's Speculum Scribe

The Herne charterhouse in present-day Belgium had an impact on medieval manuscript production that is hard to overestimate (Kwakkel, 2002). Located approximately 30 kilometers southwest of Brussels, this Carthusian monastery realized a remarkable large output of manuscripts, in Latin as well as in the Dutch vernacular. A total of 46 production units were produced there, of which no less than 40 were written in Middle Dutch (Kwakkel, 2003). Those 46 units even contain works that are only known today because they were copied in Herne, such as Filip Utenbroeke's sizable contribution to the *Spiegel historiael*, the Middle Dutch adaptation of the *Speculum historiale*. However, Herne was not only renowned as a hub for manuscript production through copying, but also as a *translatorium*, a center of translation. This is underscored by the residency of the anonymous 'Bible Translator of 1360' within the monastery. What makes this vast output even more remarkable, is that most of these manuscripts were produced within the relatively short span of 1350 to 1400. In short, during the latter half of the 14th century, the Herne charterhouse quickly established itself as an influential center for the production and dissemination of vernacular literature in the Low Countries.

This study zooms in on Herne's most prolific resident: the Speculum Scribe. Employing methods from the field of computational scribal modelling, we aim to investigate how his use of abbreviations evolved over time and discern what insights this provides into the chronological order of his manuscript production. Scribal modelling, also known as scribal profiling, was first defined by Angus McIntosh and concerns itself with the identification or profiling of different scribes. Due to the absence of a standardized language, and by extension spelling variety, scribal variation is omnipresent in medieval manuscripts (Kestemont et al., 2010; McIntosh, 1975). Parallel to the hypothesis that every author possesses a unique fingerprint, McIntosh argues that scribes leave their personal mark on a copied text as well. Instead of viewing this scribal variation as an obstacle (e.g. for authorship attribution), scribal modelling embraces this variation as an opportunity to deepen scholarly knowledge about scribal practices.

The corpus used in our study encompasses Middle Dutch manuscripts associated with Herne in the period 1350-1400 (Haverals & Kestemont, 2023).[1] These manuscripts, nowadays scattered across various European libraries, were retrieved from their silo's and digitally aggregated by Haverals and Kestemont (2023), encouraging the data's reuse in case studies like this one. First, digital facsimiles of the manuscripts were created,[2] which were partially manually transcribed. Using those transcriptions and diplomatic editions that were already digitally available, a Handwritten Text Recognition (HTR) model was trained using Transkribus.[3] This model then transcribed all remaining material in the corpus. Given the research focus on scribal practices, particularly the use of abbreviations, it was of great importance that the transcriptions were as close to the original manuscript as possible.

[1] Freely available: https://zenodo.org/records/10005366
[2] Available on request: https://zenodo.org/records/10010382
[3] https://readcoop.eu/transkribus/

Accordingly, graphemic, hyper-diplomatic reproductions of the manuscripts (Robinson & Solopova 1993) were created, which involves closely replicating the original manuscript's spelling, including brevigraphs, letters, and other glyphs.

To model the text computationally, we transformed the graphemic transcription into a bag-of-words representation consisting of TF-IDF weighted character bigrams. Crucially, we restricted the vocabulary to bigrams that included at least one brevigraph or abbreviatory glyphs representing two or more characters (Honkapohja, 2021). For instance, the Middle Dutch word for 'and' was 'ende', which was often abbreviated as 'en̄'. 'En̄' consists of two characters including a brevigraph, and will thus be included in the bag-of-words as is. However, abbreviations also occur in longer words, for instance 'leidet' ('leads') could be abbreviated as 'leidʒ'. In that case, 'dʒ' would be included in the bag-of-words. Brevigraphs represent distinctive choices made by scribes. Accordingly, by focusing on bigrams that contain a brevigraph, we can more directly investigate the scribe's individual style, while also reducing the influence of the content of the texts.

Previous research (Vandyck et al., in press) has revealed an unexpected behavior of Vienna, ÖNB, Cod. 12.857 when compared to other works attributed to the Speculum Scribe. Through a *UMAP*[4] scatterplot, it was observed that this manuscript deviates from the main Speculum scribe's oeuvre, clustering with manuscripts written by different scribes, namely Brussels, KBR, 2979, and Saint Petersburg, BAN, O 256 (Fig. 1). Interestingly, all three of these manuscripts contain Middle Dutch translations of the evangeliaria. Kwakkel suggested that the Saint Petersburg manuscript could have served as the exemplar for the Viennese one (2002, pp. 47–52). However, since the analysis is based on character bigrams including a brevigraph, it is improbable that these manuscripts clustered due to similar content alone; which was confirmed by an additional experiment (Vandyck et al., in press). Yet, paleographic observations conclude that Vienna, ÖNB, Cod. 12.857 was certainly written by the Speculum Scribe, so why does it not cluster together with the rest of his works? The answer is most likely that the manuscript was one of his earliest works in which he closely adhered to his source.

To investigate this further, we trained a *random forest classifier* model which assessed each feature's importance when differentiating between two classes, which are in our case: Vienna, ÖNB, Cod. 12.857 and the other works written by the Speculum scribe. In the resulting boxplot (Fig. 2), an interesting observation can be made regarding the use of 'dʒ': although the Speculum Scribe uses it regularly in his other works, it is almost absent in the Vienna manuscript. Traditionally, 'dʒ' is used to abbreviate 'det', for instance at the end of a word (cfr. supra). However, the Speculum Scribe is known to also use it to abbreviate 'dit' and 'dat' ('this' and 'that'). To investigate this absence further, we extracted all words containing 'dʒ' from the Viennese manuscript. Remarkably, 'dʒ' primarily serves to abbreviate 'det' in the Viennese manuscript, not 'dit' and 'dat'. This observation is most likely compatible with the hypothesis that Vienna, ÖNB, Cod. 12.857 is a relatively early manuscript in the Speculum Scribe's oeuvre. If that is the case, an evolution becomes clear: the scribe used 'ʒ' in the traditional way in his early manuscripts: namely to abbreviate 'et'; only in his later works, he applies it to abbreviate 'it' and 'at' – we hypothesize that he gradually broadened the application of this brevigraph.

[4] https://umap-learn.readthedocs.io/en/latest

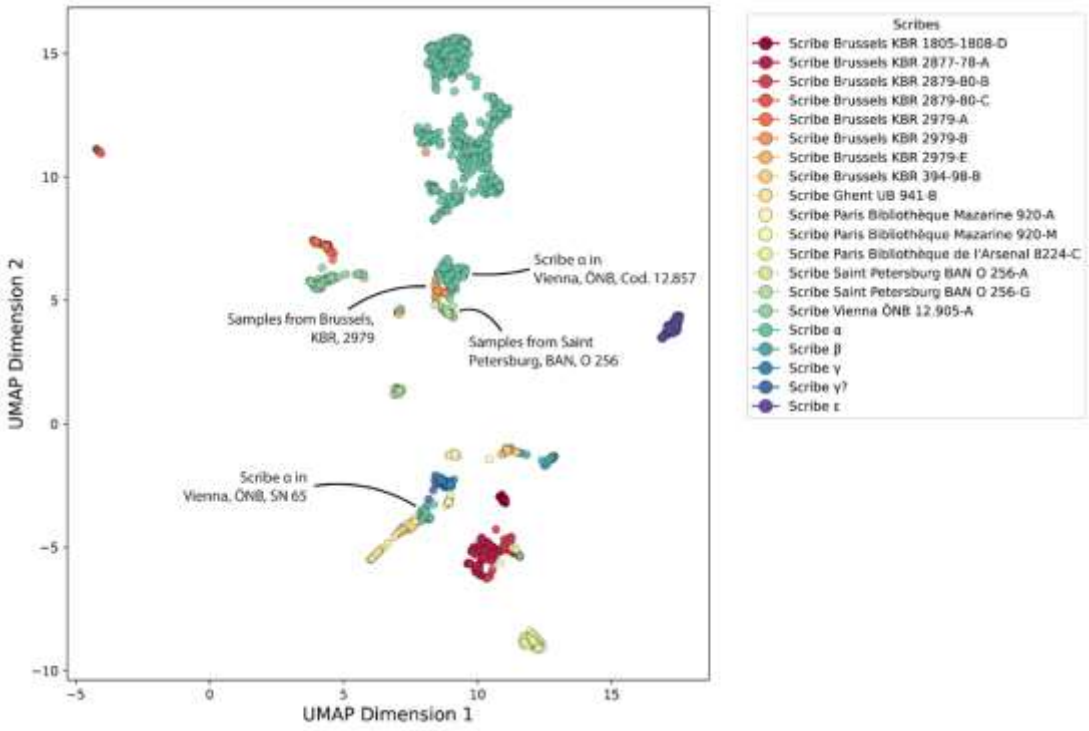


Fig. 1. Stylometric scatterplot of the Herne scribes, using character bigrams with brevigraphs. This scatterplot, created using PCA and UMAP dimensionality reduction techniques, illustrates the stylistic variation among scribes in the corpus. Each dot represents a 5,000-character segment, coloured uniquely to correspond to a specific scribe (only scribes are retained that can contribute at least 10 segments). The spatial proximity of the dots indicates stylistic similarity.

Upon closer inspection, 'dʒ' is not completely absent in the context of 'dat' in the Viennese manuscript.[5] However, it only appears rarely and consistently in the same circumstance: at the end of a line, presumably out of necessity when the scribe ran out of space when copying a verse. There are two places where 'dʒ' does not occur at the end of a line break, but those places support our hypothesis further. The first other place where 'dʒ' can be found, is in a pericope list at the start of the manuscript (Fig. 3a). According to Kwakkel, that list was added to the manuscript in a later stadium (2002, p. 275). The second place where 'dʒ' can be found, is under the text in a lighter ductus, next to a small frame (Fig. 3b). According to Kwakkel, additions as these were made later by scribe (2002, 277). Therefore, just as the pericope list, these instances of 'dʒ' were not present when the manuscript was originally written.

Concluding, one of the ways in which the scribes in the monastery of Herne were able to be this prolific, was by employing abbreviations. In this case study regarding the Speculum Scribe, an unexpected evolution in the use of abbreviations can be witnessed. In his early manuscripts, the Speculum scribe never used 'dʒ' to abbreviate 'dat' when he had the option. He only did it sparingly when he ran out of space. However, when later adding new parts to the manuscript, he did opt for 'dʒ' to replace 'dat'. What we would like to analyse further is whether

[5] But it is never used to abbreviate 'dit'.

this change of the use of ‘dʒ’ happened instantly or gradually; and whether this evolution can be used to chronologically order his other works in more detail.

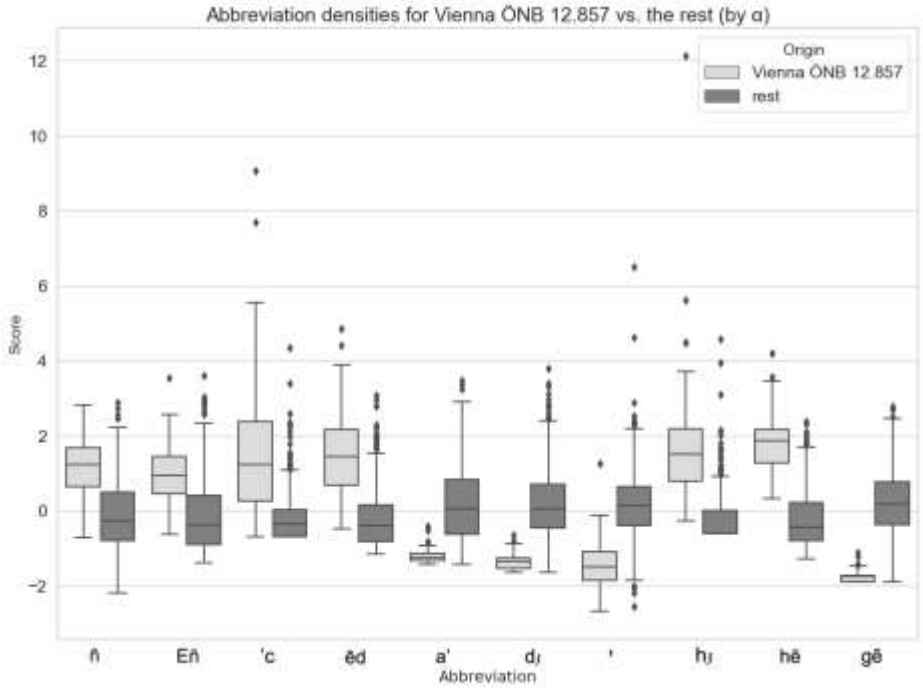
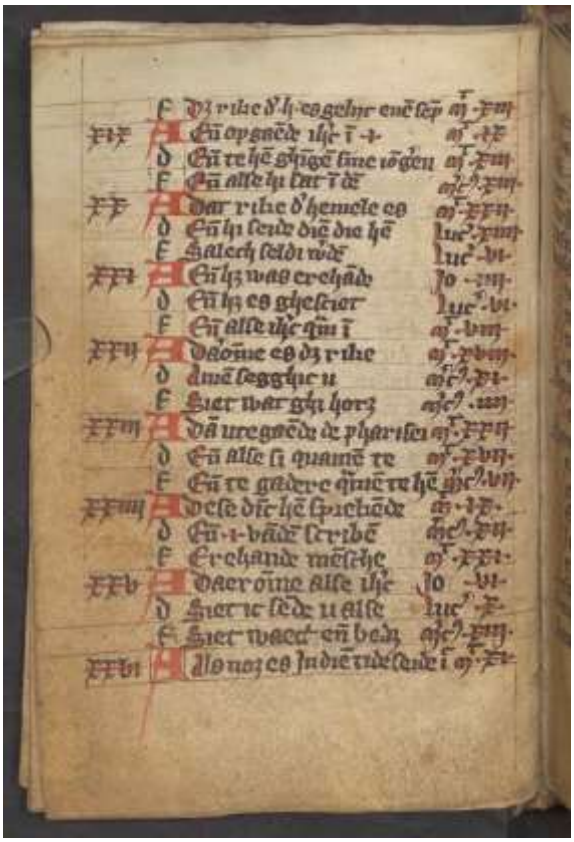


Fig. 2: Boxplot showing the distribution of the most important features when distinguishing between Vienna, ÖNB, Cod. 12.857 (light) and all other manuscripts by the Speculum Scribe (dark).

a



b

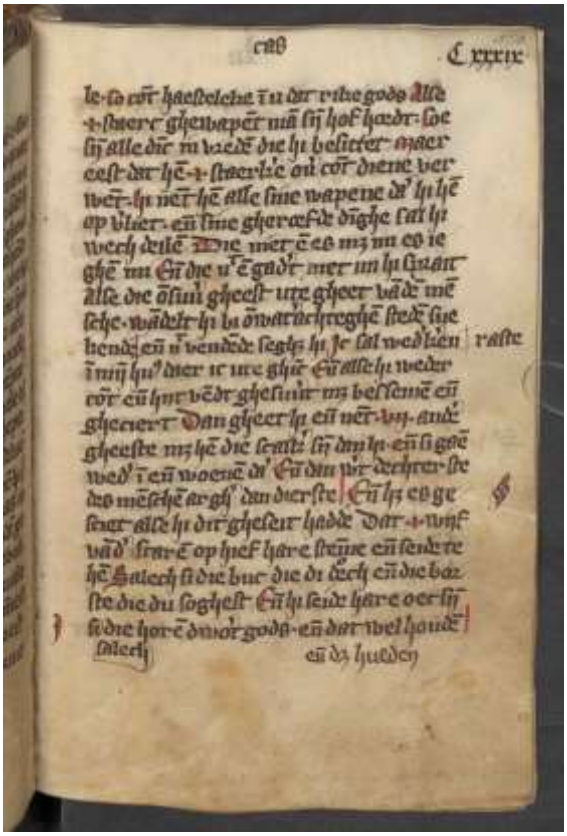


Fig. 3: Vienna, ÖNB, Cod. 12.857: (a) ‘Da’ōme es dʒ rike’ in the middle of page of the pericope list that was later added (fol. 7v). (b) ‘eñ dʒ hueden’ a later addition to the page, at the bottom (fol 157r).

Bibliography

Haverals, W., & Kestemont, M. (2023, December 6). *The Middle Dutch Manuscripts Surviving from the Carthusian Monastery of Herne (14th century): Constructing an Open Dataset of Digital Transcriptions*. CHR 2023: Computational Humanities Conference, Paris.

Honkapohja, A. (2021). Digital Approaches to Manuscript Abbreviations: Where Are We at the Beginning of the 2020s? *Digital Medievalist*, *14*(1), Article 1. https://doi.org/10.16995/dm.88

Kestemont, M., Daelemans, W., & De Pauw, G. (2010). Weigh your words—Memory-based lemmatization for Middle Dutch. *Literary and Linguistic Computing*, *25*(3), 287–301. https://doi.org/10.1093/llc/fqq011

Kwakkel, E. (2002). *Die dietsche boeke die ons toebehoeren de kartuizers van herne en de productie van middelnederlandse handschriften in de regio Brussel (1350-1400)*. Peeters.

Kwakkel, E. (2003). A meadow without flowers. What happened to the Middle Dutch manuscripts from the Charterhouse Herne? *Quærendo*, *33*, 191–211.

McIntosh, A. (1975). Scribal Profiles from Middle English Texts. *Neuphilologische Mitteilungen*, *76*(2), 218–235.

Robinson, P., & Solopova, E. (1993). Guidelines for Transcription of the Manuscripts of the Wife of Bath's Prologue. *The Canterbury Tales Project, Occasional Papers, I*, 19–52.

Vandyck, C., Haverals, W., & Kestemont, M. (2024). Making Characters Counts. A Computational Approach to Scribal Profiling in 14th-Century Middle Dutch Manuscripts from the Carthusian Monastery of Herne Monastery. In *Approaches to Digital Codicology: Interdisciplinarity and intersections*. Brepols Publishers.