THE INFLUENCE OF FORMAL CHARACTERISTICS ON THE TRANSMISSION OF JACOB VAN MAERLANT'S MEDIEVAL DUTCH *MARTIJN TRILOGY*: A COMPUTATIONAL STUDY

Sofie Moors (University of Antwerp – FWO Research Foundation – Flanders – 1182723N)
@SofieMoors
0000-0002-9444-9065

Mike Kestemont (University of Antwerp)
@Mike_Kestemont
0000-0003-3590-693X

1. Research Objective

In the Middle Ages, manuscripts were manually copied by scribes, resulting in variations among copies (Cerquiglini 1989; Driscoll 2010). The manipulation of texts by scribes in transmission is a topic of growing interest in research (Van Dalen-Oskam 2012; Kestemont 2018; Haverals & Kestemont *forthcoming*). However, little is known about the factors influencing this complex copying process. Especially the impact of text form on variations deserves further empirical research. Although evidence from psycholinguistics suggests that formal features (such as rhyme (scheme), stanza form, and text structure) could have a constraining influence (Rubin 1995), little is known about the impact of form on transmission. Slowly, computational research on this topic is shaping up (Thaisen 2014).

This paper presents a computational study of the transmission of the *Martijn trilogy*, a text by the well-known medieval Flemish author Jacob van Maerlant (ca. 1230-1235 – ca. 1288-1300). The trilogy is a dialogue in three parts (M1, M2, and M3) with a unique form, consisting of 13 verses per stanza with only two distinct rhyming sounds. Unlike other works by Maerlant, the *Martijn trilogy* deviates from the dominant paired rhyme scheme in Middle Dutch literature. Despite its special form presenting considerable challenges to copyists, little scholarly attention has been given to its transmission. In this paper, we aim to study the influence of formal characteristics on the written transmission of Maerlant's *Martijn trilogy*. To achieve this, all 17 manuscripts, fragments, and prints before 1500 are computationally aligned. Through the comparison of multiple versions of the same text, this study aims to empirically validate existing hypotheses, such as the constraining character of rhyme.

2. Corpus and Methodology: Data Collection and Preprocessing

For most of the text witnesses (Ant, C, E, GE, K, G, W, Br, A, Z, D2, Y), existing editions of the text could be digitized using Optical Character Recognition (via ABBYY FineReader). Of other texts we had an incomplete edition (M2 of the trilogy) which we completed with custom transcriptions based on photographs (B, F, O, D). Finally, Moors (2022) recently published a transcription of manuscript L. The transcription practices of these editions were not uniform. Therefore, we manually collated them all with photographic facsimiles and adjusted them to obtain a maximally diplomatic transcription. Text-structuring elements such as abbreviatory glyphs, whitespace, stanza structure, folio numbers, and lombardic capitals were faithfully reproduced. The transcriptions are available in XML (TEI-MVN framework developed by Boot & Brinkman).

3. Data analysis

The manuscripts were aligned twice at the character level, using the Needleman-Wunsch algorithm and the CollateX software (Dekker and Middell 2011), with options to keep or exclude the markup for capital letters, abbreviations, and punctuation. Some manuscripts are fragmentary: incomplete verses were annotated and left out of the comparison (Figures 1 and 2). The aligned corpus contains 15,311 verses in total. In a next step, we will also compare the verses on word, verse, or stanza level (as demonstrated in Figure 3).

Figure 1 *Pixel Plot showing the presence of lines in each text witness. O is for example complete (1824 lines), Ant is totally damaged and is therefore not included*

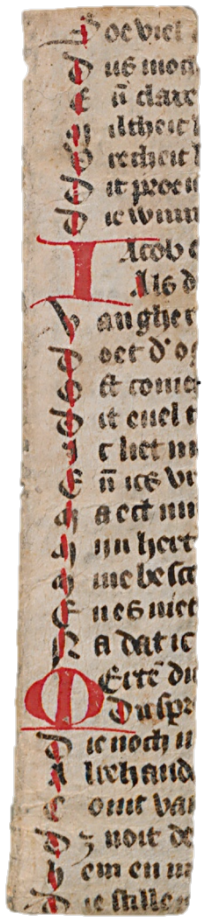

Figure 2 *Fragment Ant, Antwerp University Library, Special Collections, MAG-P 64.20*

A	hi	storte	dor	ons	sijn	bloet	root
B	hi	storte	doir	ons	zijn	bloet	root
BR	hi	storte	dor	ons	sijn	bloet	root
D	hi	stortede	doer	ons	sijn	bluet	root
D2	hi	stortte	voer	ons	sijn	bloet	root
E	hi	storte	doer	ons	sijn	bloet	root
F	hij	stortte	om	ons	siin	bloet	root
G	hi	storte	om	ons	sijn	bloet	roet
O	hij	storte	dor	ons	zijn	bloet	root
Z	hi	storte	dor	ons	-	bloet	roet

Figure 3 *Alignment at the word-level of verse line 316 of M2*

4. Initial Results

A first test of the main hypothesis, which explores the relationship between the rhyming words and transmission, was conducted. The variation in the rhyming words has been contrasted quantitatively with the variation in the unaffected textual parts. Results showed that the verse-final characters, which form part of the rhyming words, were subject to less spelling variation than the preceding regions in the line. We observe a clear decline in variation (Figure 4).

The graph reflects the average mismatch at the character level resulting from the alignment, both for Needleman-Wunsch (NW) and CollateX (CX). The x-axis is the relative position of a character in the verse: a letter with position 6 in a line of 10 letters gets the value 0.6 on the x-axis, for example. The same position in a verse of 11 letters gets a value of 6/11. On the y-axis, we then see the mean of mismatches between characters in these specific positions. As such, we can account for the differences in verse length. Future research will involve lemmatizing the corpus to determine if the rhyming words are also subject to less *lexical* variation.

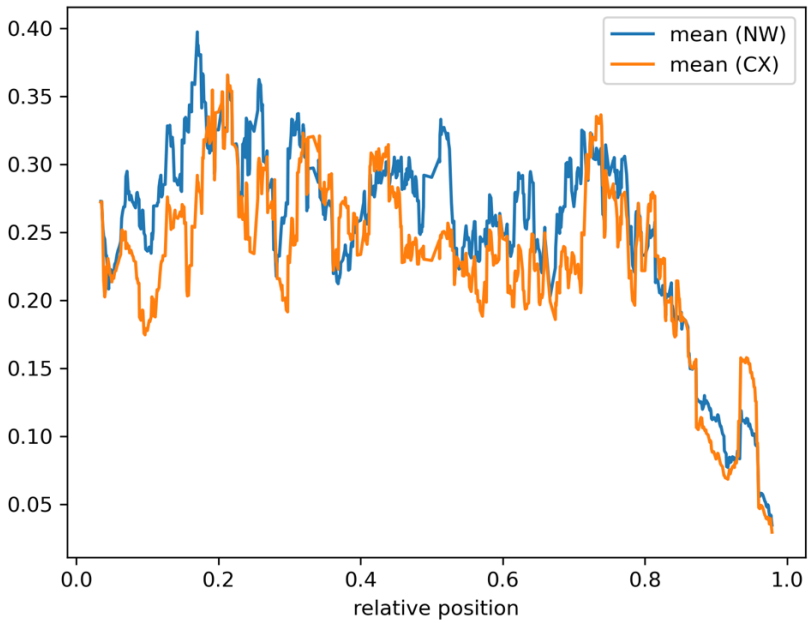

Figure 4 *Average character-level mismatches for the Needleman-Wunch (NW) and CollateX (CX) alignments*

In addition to the verse-level variation, we also looked at the macro-relationships between different manuscripts (visualized by a clustermap, Figure 5). We see for example that the variants of manuscript L, F, Br and G cluster. These fragments were mainly written in Brabant at the end of the fourteenth century (Moors 2022, p. 36). We also determined the abbreviation density of the manuscripts (Figure 6). In the presentation, we would like to elaborate on these clusters of manuscripts, and on whether there is a relationship between abbreviation density and the materiality/form, layout (for example D is a print, G is written in prose), location or dating of the manuscripts. In this sense, we not only want to measure the influence of text form on text transmission, but we also want to find out the influence of text transmission (dating, localization, materiality,... of the text carriers) on text form.

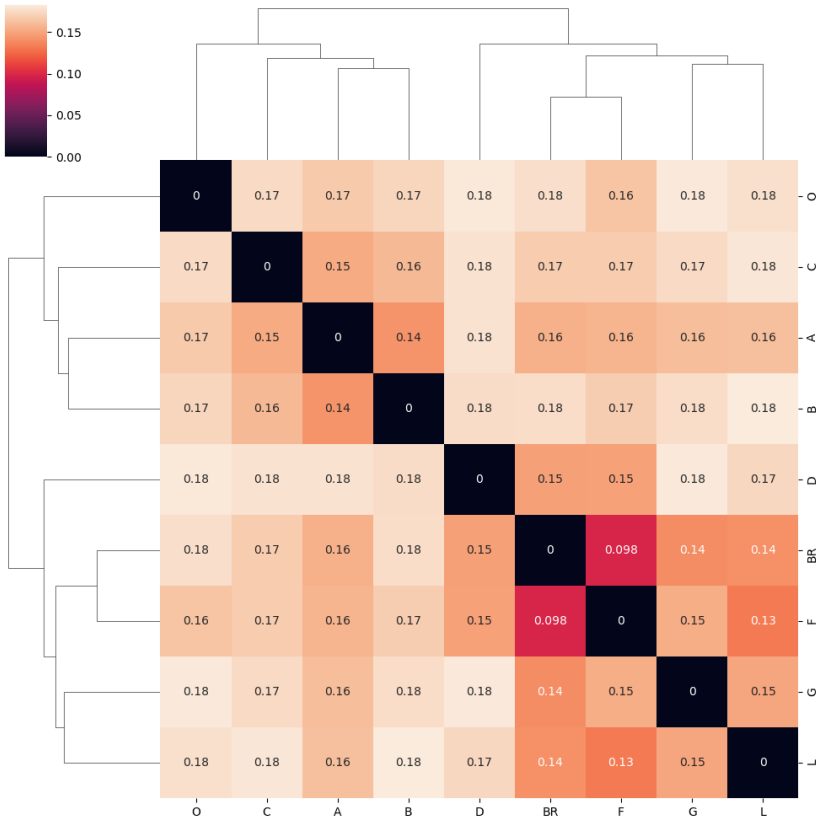

Figure 5 *Seaborn Clustermap*

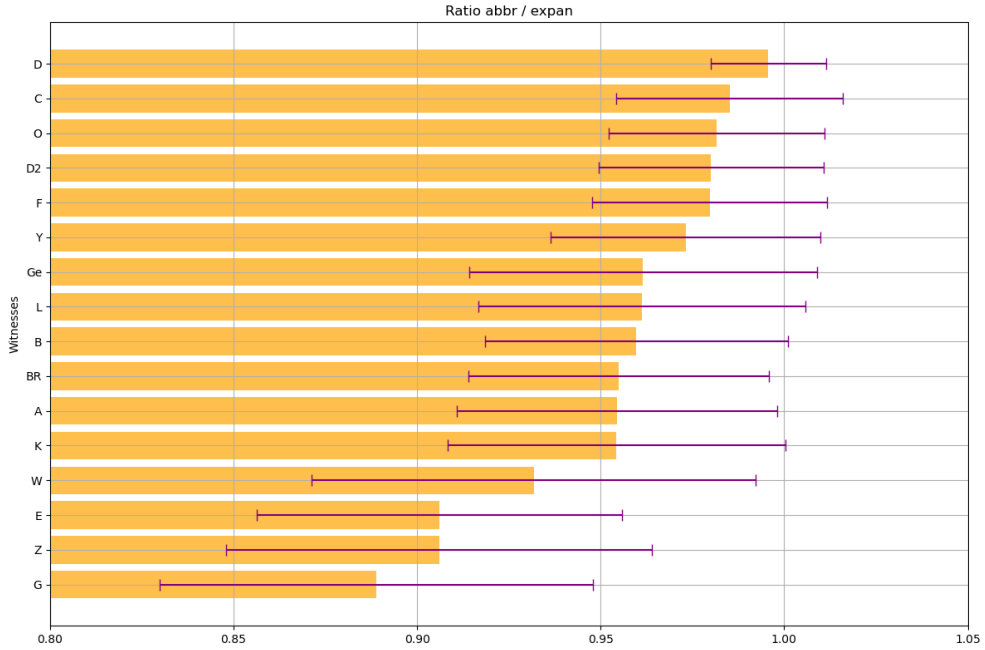

Figure 6 *Abbreviation density in each textual witness*

5. Bibliography

Boot, P. and Brinkman, H. (n.d.). *Richtlijnen voor digitale diplomatische edities in de reeks 'Middeleeuwse Verzamelhandschriften uit de Nederlanden'.*

Cerquiglini, B. (1989). *Éloge de la variante: Histoire critique de la philologie.* Paris: Seuil.

Dekker, R. H. and Middell, G. (2011). *Computer-Supported Collation with CollateX: Managing Textual Variance in an Environment with Varying Requirements.* University of Copenhagen, Denmark.

Driscoll, M. J. (2010). "The words on the page: Thoughts on philology, old and new." J. Quinn and E. Lethbridge (red.). *Creating the medieval saga: Versions, variability, and editorial interpretations of Old Norse saga literature*. Odense: Syddansk Universitetsforlag, pp. 85-102.

Haverals, W. and Kestemont, M. (*forthcoming*). From exemplar to copy: the scribal appropriation of a Hadewijch manuscript computationally explored. *Journal of Data Mining and Digital Humanities*.

Kestemont, M. (2018). Aan de taal kent men de hand. Talige kopiistherkenning en de scribenten van de Lancelotcompilatie. *Spiegel der Letteren,* **60**: 157-188.

Moors, S. (2022). Een bijzondere Brabantse tekstgetuige van de Martijntrilogie van Jacob van Maerlant. De fragmenten Lyon, Bibliothèque municipale, ms 6848. *Queeste,* **29**(1): 36-65.

Nichols, S. G. (1997). Why Material Philology? *Zeitschrift für deutsche Philologie,* **116**: 10-30.

Rubin, D. C. (1995). *Memory in oral traditions: The cognitive psychology of epic, ballads, and counting-out rhymes.* New York: Oxford University Press.

Thaisen, J. (2014). Initial position in the Middle English verse line. *English Studies*, **95**(5): 500-513.

Van Dalen-Oskam, K. (2012). The secret life of scribes. Exploring fifteen manuscripts of Jacob van Maerlant's Scolastica (1271). *Literary and Linguistic Computing,* **27**: 355-372.