Constrained. A Computational Study of the Influence of Formal Characteristics on the Transmission of the Middle Dutch Martijn trilogy by Jacob van Maerlant

Moors, Sofie
sofie.moors@uantwerpen.be
University of Antwerp, Belgium; FWO Research Foundation - Flanders

Research problem statement

In the Middle Ages, the transmission of texts depended on the transcription of manuscripts, with scribes making copies by hand (Driscoll 2010). This handwritten transmission affected the **stability** of texts, leading to each manuscript copy containing unique **variants** (Cerquiglini 1989). More recent philological research, or **material philology**, focuses on this variation *as such* (Nichols 1997). The interest in textual transmission has led to research into how copyists manipulate a text to their liking (Van Dalen-Oskam 2012; Kestemont 2018; Haverals / Kestemont *forthcoming*).

Little research has been done on the factors and underlying mechanisms involved in this complex copying process. An aspect of transmission that deserves further empirical research is the text form. Although evidence from psycholinguistics suggests that formal features (such as rhyme, stanza form, and text structure) could have a constraining character (Rubin 1995), little is known about the influence of form on variation. Slowly, more computational research on this topic is shaping up and being published (Thaisen 2014).

In this paper I will undertake a computational investigation of the transmission of the so-called *Martijn trilogy*, a text by the famous medieval Flemish author Jacob van Maerlant (ca. 1230-1235 - ca. 1288-1300). The *Martijn trilogy* is a text in the form of a dialogue that consists of three parts (labeled M1, M2, and M3). It is remarkable that in this argumentative poem, unlike the rest of his oeuvre, Maerlant did not opt for the dominant paired rhyme scheme, but rather used his own fixed form in which he uses only two rhyme sounds per stanza of 13 verses. Previous researchers paid less attention to the ways in which Maerlant's text was transmitted, even though it was precisely the special form of the poems that presented major challenges to copyists.

Corpus & methodology

Data collection & preprocessing

In this paper, I will investigate to what extent formal aspects might have influenced the written transmission of Maerlant's *Martijn trilogy*. To this end, **all 17 different manuscripts**, **fragments**, and **printsbefore 1500 will be computationallyaligned**. For the data collection, existing editions could be used, which were digitized using Optical Character Recognition (ABBYY FineReader software). I have assessed the possible risk of errors in the existing transcriptions by checking them manually after consulting the sources via photographs or autopsy. Subsequently, all diplomatic transcriptions were provided with **semantic markup**. In doing so, text-structuring elements such as abbreviations, whitespace, stanza structure, folio numbers, and lombardic capitals were marked in the transcription. These raw text files were converted to XML-files by using a custom programming script (TEI-MVN framework developed by Boot & Brinkman).

Data analysis

The alignment has been made with the Needleman-Wunsch algorithm [1] and with the specialized software *CollateX* (Dekker / Middell 2011). For the *CollateX* input, the XML-files were converted to JSON-files, with flags added to choose whether or not to retain (the markup for) capital letters, abbreviations, and punctuation. This aligned corpus consists of about 15,500 verses and allows an automated comparison down to the individual character, word, verse, or stanza (see for an example Figure 1).

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 1. *Alignment at the word-level of verse line 316 of M2.*

Preliminary findings

A first test of one of the main hypotheses, namely the correlation between the text form (and specifically the rhyming scheme) and the transmission, has been performed. The variation in the rhyming words has been contrasted quantitatively with the variation in the unaffected textual parts. The test showed that the last characters of a verse – i.e. characters that are part of the rhyming words – are less subject to (spelling) variation than other words in the text (there is a decline in the line of the graph in the last positions, see Figure 2). To test that this is not an alignment artifact, we inverted the strings. This resulted in an almost exact inversion of the initial graph, which confirms our hypothesis. For the next step, the corpus will be lemmatized to see if the rhyming words are also less subject to *lexical* variation.

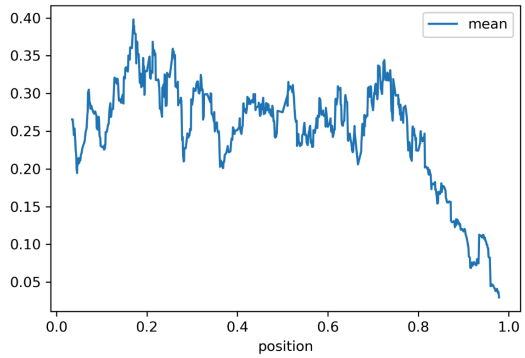

Figure 2. *Graph of variation. We see this decline in all conditions: with/without abbreviations/punctuation/capitals.*

By comparing different versions of the same text, this project empirically tests whether the data analysis is compatible with existing hypotheses (such as, for example, the idea that rhyme has a constraining character) and hopes to get a better sense of which views can be excluded.

Notes

1. https://gist.github.com/slowkow/06c6dba9180d013dfd82bec217d22eb5. Many thanks to Mike Kestemont, Wouter Haverals and Maud Goddefroy for their help in writing the code.

Bibliography

Boot, Peter / **Brinkman, Herman** (n.d.): *Richtlijnen voor digitale diplomatische edities in de reeks 'Middeleeuwse Verzamelhandschriften uit de Nederlanden'*.

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

Dekker, Ronald Haentjens / **Middell, Gregor** (2011): *Computer-Supported Collation with CollateX: Managing Textual Variance in an Environment with Varying Requirements.* University of Copenhagen, Denmark.

Driscoll, Matthew James (2010): "The words on the page: Thoughts on philology, old and new", in: Quinn, Judy / Lethbridge, Emily (eds.): *Creating the medieval saga: Versions, variability, and editorial interpretations of Old Norse saga literature*. Odense: Syddansk Universitetsforlag 85-102.

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

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

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

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

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

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