

The *Martijn Trilogy* Manuscripts: An Open Dataset for Analyzing Scribal Variation

Sofie Moors | ORCID: 0000-0002-9444-9065

Corresponding author

Department of Literature, University of Antwerp, Antwerp, Belgium

Research Foundation Flanders (FWO), Brussels, Belgium

sofie.moors@uantwerpen.be

Mike Kestemont | ORCID: 0000-0003-3590-693X

Department of Literature, University of Antwerp, Antwerp, Belgium

mike.kestemont@uantwerpen.be

Remco Sleiderink | ORCID: 0000-0003-0284-324X

Department of Literature, University of Antwerp, Antwerp, Belgium

remco.sleiderink@uantwerpen.be

Received 12 January 2024 | Revised 24 July 2024 |

Accepted 3 August 2024 | Published online 7 October 2024

Abstract

The authors present a dataset containing transcriptions of manuscripts of the Middle Dutch strophic poem *Martijn Trilogy* by the Flemish poet Jacob van Maerlant. Of his very large oeuvre, Maerlant's Strophic Poems had the longest tradition: originally written in the thirteenth century, copyists and printers continued to disseminate them until about 1500. These ten shorter poems on social, religious, and ethical issues stand out for their unusual and complex stanza form. The *Martijn Trilogy* was his most successful strophic poem: 17 text witnesses are extant, and the trilogy was imitated and even translated into French and Latin. This dataset contains hyperdiplomatic transcriptions of all witnesses, amounting to a total of 15,814 verses or 79,337 tokens. This open-access dataset abides by the FAIR principles, is licensed under a CC-BY-SA license, and is made available in multiple, complementary file formats. Since these transcriptions are strictly diplomatic, this corpus offers valuable possibilities for

research on scribal attributions (scribal profiling), abbreviations, stemmatology, textual stability and more.

Keywords

Middle Dutch literature – computational philology – digital manuscript studies – digital scholarly editing – stemmatology

- Related data set “martijnmanuscripts” with DOI [www.doi.org/10.5281/zenodo.10245412](https://doi.org/10.5281/zenodo.10245412) in repository “Zenodo”
- Also access the associated preprocessing scripts in Jupyter Notebooks for the Python language through the *Github* page: www.github.com/SofieMoors/martijnmanuscripts

1. Introduction

In pre-modern times, the transmission of texts depended on the transcription of handwritten documents or manuscripts (Driscoll, 2010). Scribes frequently produced inexact copies, which affected the stability of textual traditions. Manuscript copies thus commonly contain unique variants (Cerquiglini, 1989). While in older philological approaches, those scribal variations were often seen as obstacles to overcome in order to reconstruct the ‘original form’ of a text (*Urtext*), a shift is occurring in the so-called Material Philology (Gabriël, 2016; Nichols, 1997; Westra, 2014): material philologists increasingly call attention to punctuation, spelling variation and abbreviations (Gueville & Wrisley, 2022, 8; Honkapohja, 2021, §3, §6). The dataset presented here arises from the research project *Constrained*, where the focus lies precisely on these variations within the ‘real’ text as it has been ‘preserved, received, and annotated’, shifting our attention to scribes rather than authors (Andrews, 2013, 64; Palumbo, 2020, 98).¹

2. Problem

One text which has been predominantly studied in the context of reconstructing the authorial original, is the so-called *Martijn Trilogy*, a series of three

¹ Another research project of this kind is *The Canterbury Tales Project*. Retrieved from www.canterburytalesproject.org.

consecutive poems written by the renowned medieval Flemish author Jacob van Maerlant (ca. 1230–1235 – ca. 1288–1300).² Maerlant was active in the latter half of the thirteenth century in the regions of Holland, Zeeland, and Flanders. His texts were widely distributed in the Low Countries, both among the nobility, the bourgeoisie and the clergy, and exerted a great influence on other authors (van Oostrom, 1996, 7, 380). In the fourteenth century, Maerlant's oeuvre was ubiquitous: his works served as a source of inspiration, were imitated, and included in compilations. But above all, his texts were widely copied. The most telling testament to their popularity is perhaps their survival in around 200 distinct manuscripts, mostly fragmentary, although many complete copies are extant as well (Moors, 2023, 6).

Maerlant distinguishes himself from other authors of Middle Dutch texts by his versatility, both in terms of content and form. His immense oeuvre comprises a total of twelve major works in paired rhyme, which amount to over 300,000 verses. In the Middle Ages, no other writer is known to have equaled this number of vernacular verses (Moors, 2023, 6; van Oostrom, 1996, 11–12). Especially well-known and researched are his chivalric romances such as *Historie van Troyen*, his voluminous didactic works such as the *Rijmbijbel* and *Spiegel historiael* and his nature encyclopedia *Der naturen bloeme* (e.g., Biemans, 1997).

However, the longest tradition featured Maerlant's Strophic Poems, continuously disseminated by scribes and printers until around 1500 (Moors, 2022, 37; van Oostrom, 1996, 385).³ Together, these ten shorter poems on social, religious, and ethical issues amount to around 3,500 verses (van Driel, 2012, 59–60). The *Martijn Trilogy* was the most successful within the strophic poems. The text has been copied frequently and was imitated and translated by other authors (Besamusca, 2023, 225). The text is a dialogue between Jacob and Martijn consisting of 140 strophes of 13 verses each (1820 verses total), which have been divided over three parts (conventionally labeled M1, M2, and M3). M1 counts 75 stanzas and 975 verses, M2 consist of 26 stanzas and 338 verses and M3 of 39 stanzas, which are 507 verses. These three parts are usually – but not always – transmitted as a whole, which is why they are considered one poem (Warnar 2011, 2020).

It is remarkable that, in this argumentative poem, Maerlant did not opt for the dominant paired rhyme scheme. Instead, he used his own complicated stanza form, using only two rhyme sounds per strophe of 13 verses, the so-called

2 The most comprehensive study of Maerlant's life and oeuvre is that by van Oostrom (1996).

3 There is no certainty as to exactly when Maerlant wrote the *Martijn Trilogy*. Some researchers suggest that M1 was written around 1270, others believe that the Strophic Poems originated in the late 13th century. Besamusca (2023, 224) concludes that it is also still possible that Maerlant wrote these poems throughout his entire writing career.

‘clausule’. Dutch literary historian and Maerlant expert Frits van Oostrom calls it ‘a completely new text type in Dutch poetry’ (2006, 544). Bart Besamusca summarizes three reasons why the strophic poems occupy a ‘remarkable position’ in Maerlant’s oeuvre. Firstly, unlike his other works, they are quite a bit shorter. Secondly, they are not translated from French or Latin (i.e., they are indigenous), and thirdly, they were not written at the request of a patron, to the best of our knowledge (Besamusca 2023, 224).

With 17 text witnesses (with sigla A, Ant, B, Br, C, D, D2, E, F, G, Ge, K, L, O, W, Y and Z) still extant, the *Martijn Trilogy* is exceptionally well preserved for a Middle Dutch work. In their standard edition, Verdam and Leendertz (1918) took inventory of nine of the copies’ variants in the comprehensive *variant apparatus* (see Figure 5). The *Comburg manuscript* (A) served as *Leithantschrift* (Bosmans & Sleiderink, 2019, 82). However, this apparatus and its critical edition primarily aimed to determine which variants could be attributed to Jacob van Maerlant himself. Previous researchers paid less attention to how Maerlant’s text was transmitted, even though it was precisely the unique form of the poem *Martijn Trilogy* that presented considerable challenges to copyists. Typical in this regard is the stepmotherly treatment of fragment Z which, according to Verdam and Leendertz, would be ‘worthless’ for text criticism, resulting in its questionable exclusion from their apparatus (Sleiderink et al., 2020, 57; Verdam & Leendertz, 1918, xxxv). As, since the last edition over a hundred years ago, seven new witnesses have surfaced (Ant, K, L, Y, Br, D2 and Ge), a revision is long overdue.

3. Data

- **martijnmanuscripts deposited at Zenodo** – DOI:www.doi.org/10.5281/zenodo.10245412
- License – CC BY-SA
- **Temporal coverage:** 13th century-about 1500

To study the manuscript tradition of the *Martijn Trilogy*, we collected diplomatic, digital transcriptions of all 17 currently extant text witnesses.⁴ These

4 Not included in the corpus is the so-called *fragment Van Iterson* in the Hague manuscript, KB 73 G 26. At folio 75 recto the copyist copied stanza 10 from M2 (Mertens, 1978, 77–81; Bosmans & Sleiderink, 2019, 75). In addition, there are also 22 fourteenth-century sewing guards (stays) with remains of (at least) M1 and M2 in the manuscript Brussels, KB, hs. 77–78 (Deschamps & Mulder, 2003, 1–2). It cannot yet be ruled out that they belong to one of the other fragmentary manuscripts. These sewing guards were left out of the corpus because they are still inserted in the manuscript, making it difficult to transcribe them. Deschamps and

witnesses included (intact) manuscripts as well as fragmentary witnesses (with text loss) and two printed witnesses, amounting to 15,814 (parallel) verses or 79,337 words in total. Table 1 shows the total number of verses transmitted for each witness and how many of them are damaged (partial text loss) or missing (lacunose). To better visually understand precisely how many damaged or

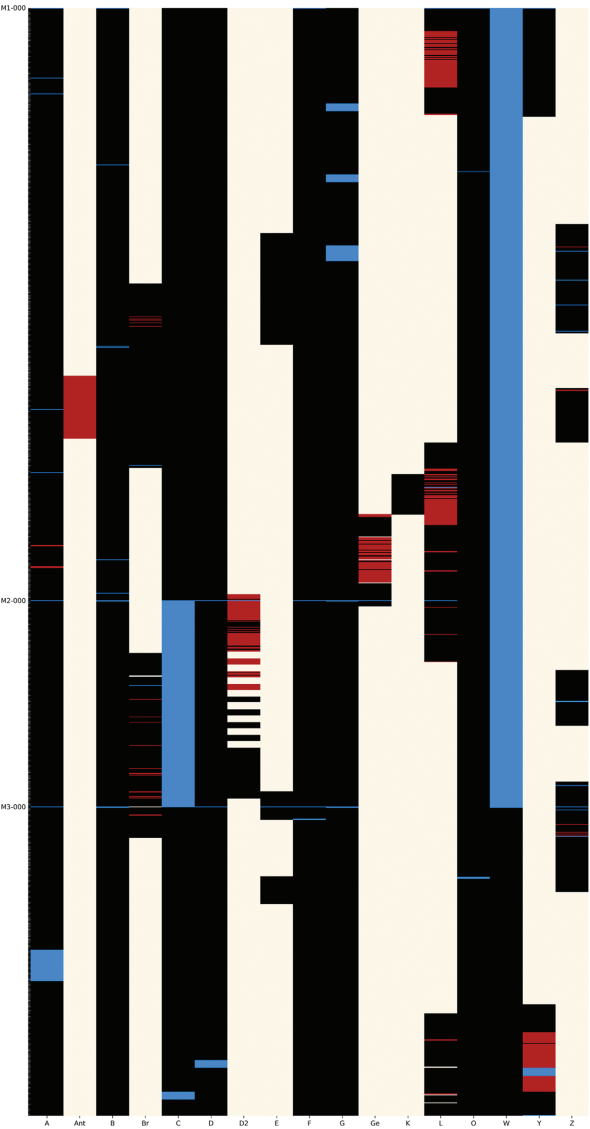
TABLE 1 Number of verses transmitted for each text witness of the *Martijn Trilogy*. The columns show the total number of verses transmitted for each witness, how many of them are damaged and how many undamaged verses remain.

Sigla	Total number of verses	Verses with damage	Verses without damage
A	1767	5	1762
Ant	104	104	0
B	1816	0	1816
Br	604	20	584
C	1472	0	1472
D	1811	0	1811
D2	247	100	147
E	276	0	276
F	1821	0	1821
G	1770	0	1770
Ge	148	70	78
K	67	0	67
L	700	162	538
O	1821	0	1821
W	508	0	508
Y	348	84	264
Z	534	7	527
TOTAL	15,814	552	15,262

Mulder (2003, 1–2) have already published everything legible, so we checked with the library to see if the stays could be unstitched for further transcription. These strips of parchment are stuck in the stitching of the manuscript to protect the spines of the quires. Removing them would mean taking apart the entire binding, which is otherwise in good condition, so it would be a very invasive job (personal communication Jan Pauwels: 18 July 2024).

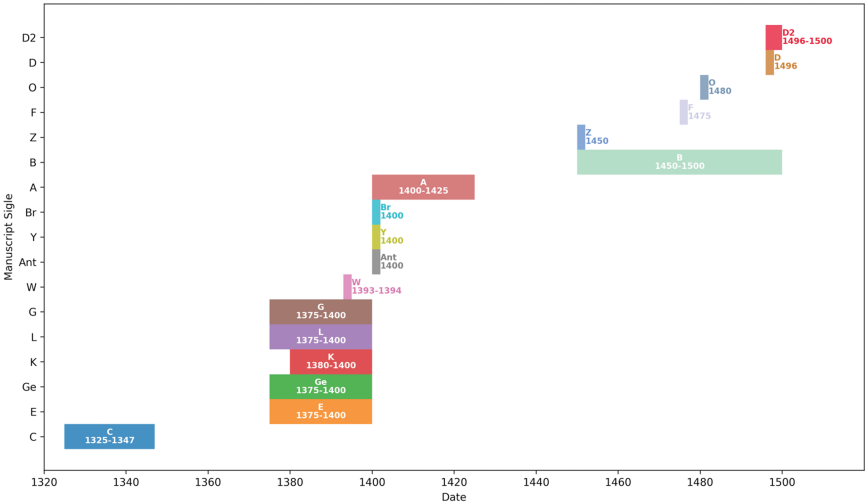
missing lines there are, we visualized the presence of lines in each text witness (see Figure 1). No fragments are known to originate from the same codices.

The witnesses are from 14 scribes: only one scribe occurs more than once in the corpus (G and W are copied by the so-called ‘Speculum scribe’ [Kwakkel, 2002, 45]). They date from the 14th and 15th centuries, but an exact dating is rarely available, which we have tried to visualize in a timeline (see Figure 2).



Note: The plot shows the presence of lines in each text witness. The witnesses have been aligned based on the verse identifiers taken from the standard edition. The beginning of each part of the trilogy is indicated on the vertical axis. Black means that a verse has been fully transmitted. Red means that a verse has been transmitted but is damaged. White means that a verse has not been transmitted because the text witness is fragmentary and has been cut. Blue represents verses that may never have been there and thus were omitted – intentionally or unintentionally – by the copyist. Ant is for example heavily damaged and contains no complete verse lines.

FIGURE 1
Presence of lines in each text witness of the *Martijn Trilogy*



Note: All dates are estimates. For example, manuscript C dates from about 1325–1347 (Biemans, 2020), manuscript Ant from about 1400, and manuscript F from about 1475. Only the printed witness D has a specific date. These dates are based on secondary literature (for an overview, see Bosmans & Sleiderink, 2019, 76–77). For manuscripts Y (Bosmans & Sleiderink, 2019), L (Moors, 2022) and Ant (Sleiderink et al., 2020), the expertise of Professor Erik Kwakkel was called upon. For this dataset paper, Kwakkel also looked at the manuscripts E, Ge and B (personal communication: 8 August 2023).

FIGURE 2 Dating of the textual witnesses of the *Martijn Trilogy*

We have witnesses from different regions, such as Hainaut (Herne) (G and W) and Flanders (A, E, and O). The ‘Brabant’-cluster of manuscripts (Ge, K, L, Y, Br, Z, F, D, and D2) (C: Brabant or Gelre) shows how Maerlant’s text also circulated beyond the county of Flanders, where the texts were presumably originally composed (Bosmans & Sleiderink, 2019, 79; Mertens, 1978; Moors, 2022, 59). D was printed in 1496 by Henrick die Lettersnider (Antwerp) and D2 is a reprint by the same printer from shortly after 1496 (Besamusca, 2023, 226).

4. **Methods**

4.1. *Data Gathering*

An overview of the corpus can be found in Table 2. We give in the schema for each text witness first the siglum and then the current repository and signature. For most of the witnesses (C, E, Ge, K, G, W, Ant, Y, Br, A, Z, D2), existing editions of the text could be digitized using Optical Character Recognition (via *ABBY*

TABLE 2 Overview of the sigla, repositories/signatures, the used editions, and publicly available images (if any)

Sigla (alternative title)	Repository & signature	Edition	Images
1. A (Comburg manuscript)	Stuttgart Württemberg State Library Cod. Poet. et phil. 2 ^o 22	Brinkman & Schenkel (1997)	Digital
2. Ant	Antwerp University Library Special Collections MAG-P 64.20	Sleiderink, Moors, & De Schepper (2020)	IIIF
3. B (Haagse Pelgrimage-handschrift)	The Hague National Library of the Netherlands 76 E 6	Verdam & Leendertz / (1918); Mertens (1978)	Digital
4. Br	Brussels Royal Library of Belgium MS 19570	Mertens (1978)	
5. C (Gronings-Zutphens Maerlant Manuscript)	Groningen University Library HS 405	Staring van den Wildenborch (1834); Meijer (1838)	Digital
6. D (Print, Hendrik den Lettersnider)	The Hague National Library of the Netherlands 150 F 40	Verdam & Leendertz (1918); Mertens (1978)	Digital
7. D2 (Print)	Leiden University Library LTK 1021	Breugelmans (1986) /	Digital
8. E	Heidelberg University Library HS 200 (olim: HS. 362a, 83)	Mone (1838); Mertens (1978)	
9. F (Clignett-Serrurehandschrift)	The Hague National Library of the Netherlands 73 F 19	Verdam & Leendertz / (1918); Mertens (1978)	IIIF
10. G (Heber-Serrure codex)	Ghent University Library HS 1374	Gabriël & Kestemont (2023)	

TABLE 2 Overview of the sigla, repositories/signatures, the used editions, and publicly available images (if any) (*cont.*)

Sigla (alternative title)	Repository & signature	Edition	Images
11. Ge	Cologne Fondation Martin Bodmer Cod. Bodmer 101	Gruijs & Mertens (1975)	IIIF
12. K	Koblenz Landeshauptarchiv Best. 701 Nr. 262	Overgaauw (1996)	Digital
13. L	Lyon Bibliothèque Municipale MS 6848	Moors (2022)	/
14. O	Oxford Bodleian Library Canonici Miscellanei 278	Verdam and Leendertz (1918); Mertens (1978)	/
15. W	Vienna Austrian National Library Cod. 13708	Gabriël & Kestemont (2023; Appendix 2)	Digital
16. Y	Montigny-le-Bretonneux – Archives départementales des Yvelines HS 1F 180	Bosmans & Sleiderink (2019)	/
17. Z	Ghent Royal Academy of Dutch Language and Literature (KANTL): 6	Gailliard (1905); Mertens (1978)	IIIF

*FineReader*⁵), which we refer to in the third column. These witnesses served as a useful basis for a re-collation with the original sources. For other witnesses, we could resort to a preexisting incomplete edition (M2 of the trilogy by Mertens, 1978), which we supplemented based on photographic facsimiles

5 Retrieved from <https://pdf.abbyy.com>.

(B, F, O, D).⁶ Finally, Moors (2022) published a transcription of manuscript L. The last column lists publicly available images (if any).

4.2. Quality Control

The transcription practices of the prior editions were not uniform. Therefore, they were manually collated with photographic facsimiles and normalized to obtain a uniform, maximally faithful diplomatic transcription of each witness. The photographs were collected via the libraries, the *Lijst van gedigitaliseerde Middelnederlandse handschriften en drukken in binnen- en buitenlandse bibliotheken* (*List of digitized Middle Dutch manuscripts and prints in domestic and foreign libraries* on Wikisource [Kuiper & Schoenaers, n. d.])⁷ or through *Google Books*. After collating with photographic copies, we further clarified some uncertain or damaged passages by *in situ* examination of the manuscripts (cf. Nury, 2018, 33).

4.3. Data Preprocessing

All the transcriptions were encoded using the TEI-MVN framework (via *Oxygen XML Editor*)⁸ developed by Boot and Brinkman (2020). MVN, which stands for *Middeleeuwse Verzamelhandschriften uit de Nederlanden*, is an edition series (1994–present) with guidelines for diplomatic editions in book form, specifically for Middle Dutch texts. These guidelines responded to the need for “editions that show the concrete historical form in which texts were read and used” (Mertens, 1994, 9 [our translation]). As of 2003, the framework was expanded to support the publication of digital editions in XML, for which MVN provides additional guidelines to the standard TEI.

As can be seen from Figure 6, which shows the first stanza of M2 according to these MVN guidelines, this framework is quite complex (Haverals & Kestemont, 2023b) (for the sake of comparison, see Figures 3, 4 and 5, all showing this specific stanza). Since manually creating these XML files would be very labor-intensive and time-consuming, a custom programming script was created to automatically generate the XML files. First, we saved the output of the OCR as 17 raw text files and enriched them with simpler descriptive markup. We give an example of such a TXT snippet in Figure 7. Table 3 is the legend with this markup.

⁶ This was done by copying and altering the transcription of a similar witness (Andrews, 2013, 67).

⁷ We were able to complete the list with eight digitizations of (fragments of) the *Martijn Trilogy* (personal communication Dirk Schoenaers).

⁸ Retrieved from www.oxygenxml.com.



FIGURE 3 Folia 7 verso and 8 recto from the Clignett-Serrure manuscript (F)
THE HAGUE NATIONAL LIBRARY OF THE NETHERLANDS, 73 F 19, PICTURE
TAKEN BY ED VAN DER VLIST



FIGURE 4 Cutout from the Clignett-Serrure manuscript (F),
specifically the first stanza of M2 (f. 7 verso)

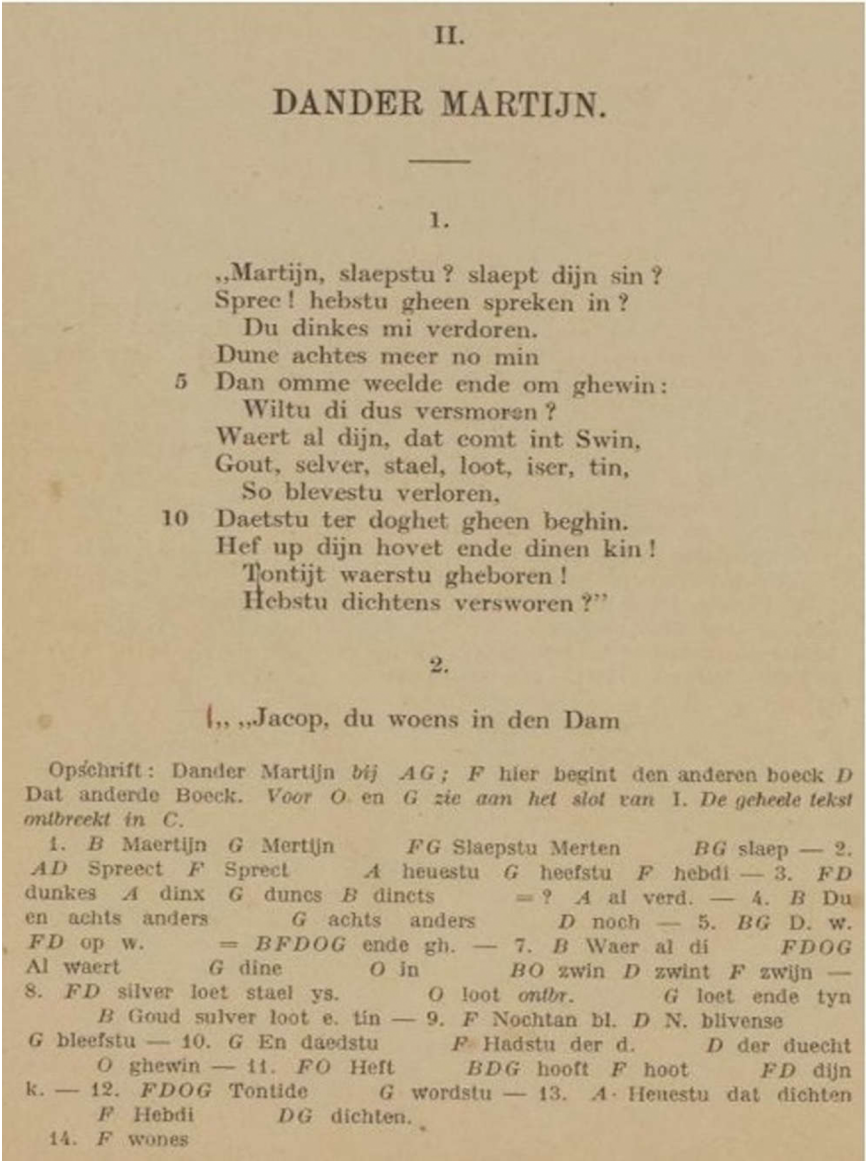


FIGURE 5 The first stanza of M2 as rendered in the standard edition by Verdam and Leendertz (1918), variants are listed at the bottom of the page
 PICTURE RETRIEVED FROM DELPHER: <https://resolver.kb.nl/resolve?urn=mmkbo2a:000032275:00007>

```

<text n="Tweede Martijn" xml:id="M2">
  <body>
    <p>
      <lg type="stanza" n="01">
        <lb n="5" xml:id="Ff7vb.5"/>
        <l n="F_M2_01_000"><hi rend="capitalsize1">H</hi>ier begint den anderen
          boeck</l>
        <lb n="6" xml:id="Ff7vb.6"/>
        <l n="F_M2_01_001"><hi rend="capitalsize2">S</hi>laepstu <choice>
          <abbr>m<g ref="#apomod"/>te<g ref="#bar"/></abbr>
          <expan>m<ex>er</ex>te<ex>n</ex></expan>
        </choice> slaept diin sin</l>
        <lb n="7" xml:id="Ff7vb.7"/>
        <l n="F_M2_01_002"><hi rend="capitalsize1">S</hi>precť hebdi <choice>
          <abbr>ghee<g ref="#bar"/></abbr>
          <expan>ghee<ex>n</ex></expan>
        </choice>
        <choice>
          <abbr>spreke<g ref="#bar"/></abbr>
          <expan>spreke<ex>n</ex></expan>
        </choice> in</l>
        <lb n="8" xml:id="Ff7vb.8"/>
        <l n="F_M2_01_003"><hi rend="capitalsize1">D</hi>u dunckes mi
          verdoren</l>
        <lb n="9" xml:id="Ff7vb.9"/>
        <l n="F_M2_01_004"><hi rend="capitalsize1">D</hi>une achtes meer no
          min</l>
        <lb n="10" xml:id="Ff7vb.10"/>
        <l n="F_M2_01_005"><hi rend="capitalsize1">D</hi>an op weelde ende
          ghewin</l>
        <lb n="11" xml:id="Ff7vb.11"/>
        <l n="F_M2_01_006"><hi rend="capitalsize1">W</hi>iltu dij dus
          versmoren</l>
        <lb n="12" xml:id="Ff7vb.12"/>
        <l n="F_M2_01_007"><hi rend="capitalsize1">A</hi>lwaert al diin dat comť
          int zwijn</l>
        <lb n="13" xml:id="Ff7vb.13"/>
        <l n="F_M2_01_008"><hi rend="capitalsize1">G</hi>out siluer loot yser
          tin</l>
        <lb n="14" xml:id="Ff7vb.14"/>
        <l n="F_M2_01_009"><hi rend="capitalsize1">N</hi>ochtan bleuestu
          verloren</l>
        <lb n="15" xml:id="Ff7vb.15"/>
        <l n="F_M2_01_010"><hi rend="capitalsize1">H</hi>adstu der doghet gheen
          beghin</l>
        <lb n="16" xml:id="Ff7vb.16"/>
        <l n="F_M2_01_011"><hi rend="capitalsize1">H</hi>eft op diin hoot <choice>
          <abbr>en<g ref="#bar"/></abbr>
          <expan>en<ex>de</ex></expan>
        </choice> diin kin</l>
        <lb n="17" xml:id="Ff7vb.17"/>
        <l n="F_M2_01_012"><hi rend="capitalsize1">T</hi>ontide waerstu
          gheboren</l>
        <lb n="18" xml:id="Ff7vb.18"/>
        <l n="F_M2_01_013"><hi rend="capitalsize1">H</hi>ebdi dichtens
          verzworen</l>
      </lg>
    </p>
  </body>
</text>

```

FIGURE 6 XML snippet of the first stanza of M2 (manuscript F) following the MVN guidelines

```
_M2_
§01§
0 <H>ier begint den anderen boeck
1 <S_2>laepstu m[er]te(n) slaept diin sin
<S>prect hebdi ghee(n) spreke(n) in
<D>u dunckes mi verdoren
<D>une achtens meer no min
5 <D>an op weelde ende ghewin
<W>iltu dij dus versmoren
<A>lwaert al diin dat comt int zwijn
<G>out siluer loot yser tin
<N>ochtan bleuestu verloren
10 <H>adstu der doghet gheen beghin
<H>eft op diin hoot en(de) diin kin
<T>ontide waerstu gheboren
<H>ebdi dichtens verzworen
```

FIGURE 7 TXT snippet from the first stanza of M2 (manuscript F). See Table 3, which defines the descriptive markup used.

TABLE 3 Legend for the descriptive text-structuring markup used added to the TXT transcriptions for elements such as abbreviations, stanza structure, folio numbers and Lombardic capitals

Transcription RICH_TXT	Example	Information
&FOLIO_COLUMN&	&2ra&, &112vb&	New page and column
TEXTPART	_M1_	New text part (M1, M2 or M3)
_____	_____	End of text part
§STROPHE§	§03§, §20§	New strophe
<CHARACTER_HEIGHT>	<O_3>m, <D>e	Capital letter and height of letter
{CHARACTER}	{per}semier, dor{per} hede, or{con}de	Special abbreviatory glyphs (3, p, 9, ...)
CHARACTER(CHARACTER)	en(de), co(ninc)	Macron
CHARACTER%CHARACTER%	n%i ^e t%, g% ^r oe% ^t	Abbreviation marked by superscript letter
[...]	es [...], m[...]	Damage gap
CHARACTER	*t*	Uncertain reading
CHARACTER[CHARACTER]	s[er]e	Abbreviation marked by ’

TABLE 3 Legend for the descriptive text-structuring markup (*cont.*)

Transcription RICH_TXT	Example	Information
Ø	Ø	Missing line (copyist error)
@CHARACTER@	@gaen@	Word crossed out
«CHARACTER»	«koen»	Word added

TABLE 4 Examples of abbreviations occurring in the *Martijn Trilogy* corpus

MVN char. ref. (xml)	Expansion(s)	Examples (abbrev.)	Examples (expand.)
#bar (macron)	de, n, m, ninc	eñ, dē, beclaghē, cō	ende, den, beclaghen, coninc
#apomod	<vowel>+r	s're, d', he'e	sere, daer, heere
superscript vowel	r+<vowel>(+e) u+<vowel>	t ^e ct, q ^a de, d ^a get, mesq ^a m, p ⁱ nche, t ⁱ niteit, g ^o t	trect, quade, draget, mesquam, prin- sche, triniteit, groet
superscript a	aer, ai	openb ^a ,w ^a	openbaer, waer
superscript t	iet, eit	n ^t , domh ^t	niet, domheit
#pbardes	per, paer	psonē, psemer, spt	personen, persemer, spaert
#pcondes	con, com	gde, gpaen	conde, compaen
#etfin	et, at, it, iet, eit, ijt	h3, ōtbo3, d3, n3, schoenh3, ontb3	het, ontboet, dat, niet, schoenheit, ontbijt
#pflour	pro	,ppheten	propheten
#usmod	us, as	d ⁹ , w ⁹	dus,was
#rrot	ur	nat ^{re} , auont ^{re}	nature, auonture

Note: See also Haverals & Kestemont (2023b), Table 3.

These raw text files were then automatically converted to XML files. As proposed by the TEI-MVN framework, text structuring elements such as abbreviatory glyphs (represented using the <choice> element, and then marked with the tag <abbr>), folio structure (marked with <pb>), and lombardic capitals (marked with <hi> and with an attribute “capitalsize” to show how many lines a capital letter takes up) were faithfully reproduced in the XML. For the visualization of the medieval abbreviatory glyphs (‘brevigraphs’), we used the encodings proposed by the *Medieval Unicode Font Initiative* (MUFI: abbreviation glyphs are encoded using a g-node with a ref attribute) [Haugen, 2009]). Table 4 shows some abbreviations in the corpus, the MVN guideline for referring to that abbreviation, and some examples (see also Table 3 in Haverals & Kestemont [2023b]). As for abbreviations, the resolved version can also be shown (marked with <expan> and <ex>). In the visualization, this expansion is italicized: ‘eñ’ becomes ‘ende’.

Some manuscripts are fragmentary due to missing pages or lacunose due to physical damage. Incomplete verses were annotated with <damage> and <gap>. Unclear passages were annotated in <unclear>. Spaces before punctuation (marked with <pc>) and before and after the typographical distinguishers in the form of dots around Roman numerals, were removed. New pages and columns were tagged with the page/folio number and ‘a’ or ‘b’ for the columns, for example [222ra]. The difference between i/j and u/v was encoded, and thus not adapted to the modern spelling, but the different allographs of s (normal or long), r and y were not represented, as customary in Middle Dutch studies (Mertens, 1994).

To ensure accurate comparison in a later stage of the research, we added a unique identifier to each verse, for example: <ln="A_M1_01_001">, containing the manuscript name (siglum), the specific part of the trilogy (M1, M2 or M3), the strophe number and verse number based on the standard edition by Verdam and Leendertz (1918) (the order of stanzas and verses in the witnesses can sometimes differ from this edition).

The layout of the page and line division was adopted from the standard edition rather than the original manuscript sources (Gueville & Wrisley, 2022, 5–6). This decision was made for pragmatic reasons, aiming to minimize discrepancies between versions (Andrews, 2020, 165). In the original sources, it is not uncommon to find multiple verses on a single line, posing challenges for comparison conducted at the verse level. Conversely, there are instances where a single verse spanned multiple lines.⁹

9 A later edition will strive to closely retain the layout of the manuscripts. However, for computational analysis, the recommendations of Guéville and Wrisley (2022, 5) regarding

5. Concluding Remarks

We offer these transcriptions in a series of complementary formats to facilitate further research and encourage the data's reuse. In addition to the richly encoded XML, we also provide formats such as HTML and XLSX, which are more accessible to scholars of a less digital orientation. Since the editions are diplomatic, this corpus is ideal for research on scribal attributions (e. g., Haverals & Kestemont, 2023a, and Kestemont, 2015, 2018), abbreviations (e. g., Haverals & Kestemont, 2023b, and Honkapohja, 2021), stemmatology (e. g., Roelli, 2020), textual stability (e. g., Thaisen, 2014), and more.

Author Contributions

Sofie Moors: conceptualization, formal analysis, investigation, methodology, resources, software, supervision, data curation, writing (original draft/review & editing); Mike Kestemont: conceptualization, formal analysis, methodology, resources, software, writing (review & editing); Remco Sleiderink: conceptualization, supervision, writing (review).

Acknowledgements

The dataset presented in this dataset paper results from a research project funded by the Research Foundation – Flanders (FWO), project number 1182723N.

Thanks to Wouter Haverals for his help in writing the scripts and to Nicky Voorneveld for her feedback on this dataset paper. Also, thanks to Erik Kwakkel for analysing the manuscripts' dating. We also thank Peter Boot for his help with the MVN framework.

transcription systems were followed: "one's choices for encoding must arise not only from the specific textual scenarios at hand but also from the ways that one wants to use such text downstream. If the use for such text is a screen-based, documentary digital edition for scholarly reading, perhaps the transcription can be as specific to the textual tradition as desired [...]. On the other hand, if the goal is to work with contemporary computational approaches to text, *the consistency, as well as the concision of transcription norms, becomes all the more important*" [emphasis added].

References

- Andrews, T. L. (2013). The third way: Philology and critical edition in the digital age. *Variants*, 10, 61–76. www.doi.org/10.1163/9789401209021_006.
- Andrews, T. L. (2020). Transcription and collation. In P. Roelli (Ed.), *Handbook of Stemmatology: History, Methodology, Digital Approaches* (pp. 160–175). De Gruyter. www.doi.org/10.1515/9783110684384-004.
- Besamusca, B. (2023). Jacob van Maerlant's Martijn Poems from a multilingual perspective. In C. van der Haven, J. Bloemendal, Y. Desplenter, & J. A. Parente (Eds.), *Literature without frontiers: Transnational perspectives on premodern literature in the Low Countries, 1200–1800* (pp. 223–237). Brill. <http://dx.doi.org/10.1163/9789004544871>.
- Biemans, J. A. A. M. (1997). *Onsen Speghele Ystoriale in Vlaemsche: Codicologisch onderzoek naar de overlevering van de Spiegel historiael van Jacob van Maerlant, Philip Utenbroeke en Lodewijk van Velthem, met een beschrijving van de handschriften en fragmenten*. Peeters.
- Biemans, J. A. A. M. (2020). Monastic scriptorium or municipal workshop. In E. C. Dijkhof (Ed.), *Medieval documents as artefacts: Interdisciplinary perspectives on codicology, palaeography and diplomatics* (pp. 55–92). Uitgeverij Verloren.
- Boot, P., & Brinkman, H. (2020). *Richtlijnen voor digitale diplomatische edities in de reeks 'Middeleeuwse Verzamelhandschriften uit de Nederlanden'*. www.github.com/HuygensING/mvn-xml/blob/main/docu/Richtlijnen%20MVN%20digitaal.pdf.
- Bosmans, G., & Sleiderink, R. (2019). Maerlant in Yvelines. Een nieuw fragment van de *Martijns* van Jacob van Maerlant (Montigny-le-Bretonneux, Archives départementales des Yvelines, 1F 180). *Queeste*, 26(1), 66–93.
- Breugelmans, R. (1986). Een fragment van een onbekende druk van de *Wapene Martijn*. Beschrijving, commentaar en transcriptie. In G. Van Eemeren, & F. Willaert (Eds.), *'t Ondersoek leert: Studies over middeleeuwse en 17^{de}-eeuwse literatuur ter nagedachtenis van prof. Dr. L. Rens* (pp. 125–136). Acco.
- Brinkman, H., & Schenkel, J. (Eds.) (1997). *Het Comburgse handschrift: Hs. Stuttgart, Württembergische Landesbibliothek, Cod. poet. et phil. 2° 22. Band I-II*. Uitgeverij Verloren.
- Cerquiglini, B. (1989). *Éloge de la variante: Histoire critique de la philologie*. Seuil.
- Deschamps, J., & Mulder, H. (2003). *Inventaris van de Middelnederlandse handschriften van de Koninklijke Bibliotheek van België*. Zesde aflevering. Koninklijke Bibliotheek van België.
- Driscoll, M. J. (2010). The Words on the page: Thoughts on philology, old and new. In J. Quinn, & E. Lethbridge (Eds.), *Creating the medieval saga: Versions, variability, and*

- editorial interpretations of Old Norse saga literature* (pp. 85–102). University Press of Southern Denmark.
- Gabriël, R. (2016). Introduction. Between stability and transformation: Textual traditions in the medieval Netherlands. *Queeste*, 23(2), 61–70.
- Gabriël, R., & Kestemont, M. (Eds.) (2023). *De Heber-Serrurecodex: Gent, Universiteitsbibliotheek, Hs. 1374*. Middeleeuwse Verzamelhandschriften uit de Nederlanden, 17. <https://hbsr.mvn.huygens.knaw.nl>. Huygens Instituut (KNAW).
- Gailliard, Edw. (1905). Een fragment van Maerlant's Wapene Martijn. In *Verslagen en mededeelingen der Koninklijke Vlaamsche Academie voor Taal- en Letterkunde* (pp. 561–607). A. Siffer.
- Gruijs, A., & Mertens, Th. (1975). A new fragment of Jacob van Maerlant's *Eerste* and *Tweede Martijn* (in Cod. Bodmer. 101, Bibliotheca Bodmeriana, Geneva). *Quaerendo*, 5(4), 346–352.
- Gueville, E., & Wrisley, D. J. (2022). *Transcribing medieval manuscripts for machine learning*. [preprint].
- Haugen, O. E. (2009). *Mufi character recommendation v. 3.0*. www.mufi.info/m.php?p=mufi.
- Haverals, W., & Kestemont, M. (2023a). From exemplar to copy: the scribal appropriation of a Hadewijch manuscript computationally explored. *Journal of Data Mining and Digital Humanities*, 1–21. www.doi.org/10.46298/JDMDH.10206.
- Haverals, W., & Kestemont, M. (2023b). The Middle Dutch manuscripts surviving from the Carthusian monastery of Herne (14th century): Constructing an open dataset of digital transcriptions. *CHR 2023: Computational Humanities Research Conference, December 6–8, 2023*. Paris.
- Honkapohja, A. (2021). Digital approaches to manuscript abbreviations: Where are we at the beginning of the 2020s? *Digital Medievalist*, 14(1), p.None. www.doi.org/10.16995/dm.88.
- Kestemont, M. (2015). A computational analysis of the scribal profiles in two of the oldest manuscripts of Hadewijch's Letters. *Scriptorium*, 69, 159–177.
- Kestemont, M. (2018). Aan de taal kent men de hand. Talige kopiïstherkenning en de scribenten van de Lancelotcompilatie. *Spiegel der Letteren*, 60, 157–188. www.doi.org/10.2143/SDL.60.3.3285821.
- Kuiper, W., & Schoenaers, D. (n. d.). *Lijst van gedigitaliseerde Middelnederlandse handschriften en drukken in binnen- en buitenlandse bibliotheken*. https://nl.wikisource.org/wiki/Wikisource:Lijst_van_gedigitaliseerde_Middelnederlandse_handschriften_en_drukken_in_binnen-_en_buitenlandse_bibliotheken.
- 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.

- Meijer, G. J. (1838). Van der Drievoudichede; van Jacob van Maerlant. In *Nalezingen op het leven van Jezus, enz: Verslag van den roman van Walewein, en andere bijdragen tot de oude Nederlandsche letterkunde* (pp. 119–152). J. Oomkens.
- Mertens, Th. (Ed.) (1978). *Den anderen merten. Synoptische archiefeditie van Jacob van Maerlant's tweede Martijn*. Alfa.
- Mertens, Th. (1994). *Richtlijnen voor de uitgave van Middeleeuwse Verzamelhandschriften uit de Nederlanden*. Uitgeverij Verloren.
- Mone, F. J. (1838). Aus Maerlants Wapen Martijn. In *Anzeiger für Kunde der deutschen Vorzeit*. Vol. 7, pp. 244–258. Germanisches Museum.
- 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. www.doi.org/10.5117/QUE2022.1.002.MOOR.
- Moors, S. (2023). Jacob van Maerlant: leven en werk van de auteur van de *Rijmbijbel*. In B. Caers, & J. Pauwels (Eds), *De Rijmbijbel van Jacob van Maerlant: Het oudste geïllustreerde handschrift in het Nederlands* (pp. 6–12). Amsterdam University Press.
- Nichols, S. G. (1997). Why material philology? *Zeitschrift für deutsche Philologie*, 116, 10–30.
- Nury, E. (2018). *Automated collation and digital editions: From theory to practice*. [Doctoral Thesis.] King's College.
- Overgaauw, E. A. (Ed.) (1996). Een nieuw fragment van Jacob van Maerlant's Eerste Martijn (Koblenz, Landeshauptarchiv, Best. 701: Nr. 262, fol. 173). *Queeste*, 3, 191–196.
- Palumbo, G. (2020). Criticism and controversy. In P. Roelli (Ed.), *Handbook of stemmatology: History, methodology, digital approaches* (pp. 88–109). De Gruyter. www.doi.org/10.1515/9783110684384-003.
- Roelli, P. (Ed.) (2020). *Handbook of stemmatology: History, methodology, digital approaches*. De Gruyter. www.doi.org/10.1515/9783110684384.
- Sleiderink, R., Moors, S., & De Schepper, M. (2020). Bandwerk. Een nieuw fragment van Wapene Martijn van Jacob van Maerlant (Antwerpen, Universiteitsbibliotheek, Bijzondere Collecties, MAG-P 64.20). *Queeste*, 27, 43–62. www.doi.org/10.5117/QUE2020.1.002.SLEI.
- Staring van den Wildenborch, A. C. W. (1834). De Wapene Martijn van Jacob van Maerlant, naar het Zutphense hs. medegedeeld, met eene inleiding en aantekeningen van M. Siegenbeek. In *Nieuwe Werken van de Maatschappij der Nederlandsche Letterkunde te Leiden*, deel III, 2^{de} stuk (pp. 81–225). Blussé & van Braam.
- Thaisen, J. (2014). Initial position in the Middle English verse line. *English Studies*, 95(5), 500–513. www.doi.org/10.1080/0013838X.2014.924275.
- van Driel, J. (2012). *Meesters van het woord: Middelnederlandse schrijvers en hun kunst*. Uitgeverij Verloren.

- van Oostrom, F. (1996). *Maerlants wereld*. Prometheus.
- van Oostrom, F. (2006). *Stemmen op schrift. Geschiedenis van de Nederlandse literatuur vanaf het begin tot 1300*. Bert Bakker.
- Verdam, J., & Leendertz, P. (Eds.) (1918). *Jacob van Maerlant's Strophische gedichten. Nieuwe bewerking der uitgave van Franck en Verdam*. A.W. Sijthoff's.
- Warnar, G. (2011). The discovery of the dialogue in Dutch medieval literature: A discourse for meditation and disputation. In K. Enenkel, & W. Melion (Eds.), *Meditatio – Refashioning the self: Theory and practice in late medieval and early modern intellectual culture* (pp. 69–88). Brill. www.doi.org/10.1163/ej.9789004192430.i-440.25.
- Warnar, G. (2020). De andere Mertens. Jacob van Maerlants Martijn-dialogen als drieluik. *Ons Geestelijk Erf*, 3, 231–247. www.doi.org/10.2143/OGE.90.3.3289106.
- Westra, J. H. (2014). What's in a name: Old, new, and material philology, textual scholarship, and ideology. In M. van der Poel (Ed.), *Neo-Latin philology: Old tradition, new approaches* (pp. 13–24). Leuven University Press. www.doi.org/10.2307/j.ctt14jxt24.4.