The One and Only? Authorship Verification on Jan van Boendale and the Middle Dutch Antwerp School

Caroline Vandyck[1,2]

[1] Department of Literature (ACDC), University of Antwerp, Antwerp, Belgium
[2] Research Foundation Flanders (FWO), Brussels, Belgium

Abstract

This study investigates the authorship of the so-called Antwerp School, a cluster of eleven interrelated historiographic and didactic Middle Dutch texts produced in early fourteenth-century Antwerp. Although traditionally attributed at least in part to Jan van Boendale, the extent of his contribution remains contested. To assess whether these texts (or a subselection of them) could have been written by a single author, computational authorship verification methods are applied. To this end, lemmatised rhyme-word bigrams are represented as TF-IDF vectors, and their stylistic similarity is measured using cosine distance. The analysis combines four complementary approaches: threshold-based authorship verification, dimensionality reduction, authorship ranking in the form of a bootstrapped nearest neighbour evaluation, and intertextual similarity detection. Across all approaches, consistent stylistic patterns emerged within the Antwerp School — especially between *Brabantsche yeesten* and *Van den derden Eduwaert*, as well as between *Melibeus* and *Dietsche doctrinale* — supporting the maximalist hypothesis of single authorship.

Keywords: authorship verification, stylometry, Middle Dutch

1 Introduction

In the early fourteenth century, no surrounding area or city in the Duchy of Brabant came close to Antwerp in terms of literary production in Middle Dutch. Although the city was gradually growing, it was still considerably smaller — demographically, as well as politically and economically — than Brussels and Louvain, the leading cities in Brabant [7]. Yet, in less than half a century, eleven remarkably similar and interrelated works were produced there. The works in question are: *Brabantsche yeesten*, *Sidrac*, *Korte Kroniek van Brabant*, *Lekenspiegel*, *Jans teesteye*, *Van den derden Eduwaert*, *Melibeus*, *Boec Exemplaer*, *Dietsche doctrinale*, *Boec van der wraken* and *Hoemen ene stat regeren sal*. Together, these works make up the so-called Antwerp School. The name for this corpus, first introduced by Van Mierlo in 1940 [26] and later used by Heymans in 1989 [7], therefore does not refer to a literal "school" consisting of unified authors with similar ideals, but to a corpus of texts that are strikingly similar.

The texts are similar in various ways, the most important factor being their genesis: Antwerp in the early fourteenth century. Almost all texts of the Antwerp School mention both the year and the place of their composition. These references are often accompanied by a note that the author himself lived and worked in Antwerp. Accordingly, all works in the School testify of a remarkable urban consciousness [7]. *Brabantsche yeesten* is even generally regarded as the first Middle Dutch historical work to originate in a distinctly urban environment [23]. Furthermore, the works also show similarities in terms of content, as they are all historiographical, moralistic and didactic in nature. History often functions as a means of educating a lay audience in morality and ethics in

Caroline Vandyck. "The One and Only? Authorship Verification on Jan van Boendale and the Middle Dutch Antwerp School." In: *Computational Humanities Research 2025*, ed. by Taylor Arnold, Margherita Fantoli, and Ruben Ros. Vol. 3. Anthology of Computers and the Humanities. 2025, 149–171. https://doi.org/10.63744/hhieMxHypx67.

this corpus. Additionally, the texts also show strong similarities in norms and values. Furthermore, almost all of the works appeal to the oeuvre of Jacob van Maerlant and show admiration for his work — in *Lekenspiegel* he is even named *vader der dietsche dichteren algader* [father of all Middle Dutch authors] [5, vs. 119–120, p. 163]. Last, the corpus shows stylistic similarities, with certain verses and words recurring, indicating a tight intertextual network (see e.g. [5; 7; 15; 20; 28]).

The existence of the Antwerp School naturally raises questions about the authorship of the texts in question, especially since there is little chance that all eleven works were written by different authors in the relatively small urban area of Antwerp [7]. Most of the works were transmitted anonymously, but there is a clear main candidate author in literary history: Jan van Boendale. The most significant reason for this is that in *Jans teesteye*, the author introduced himself extensively as Jan van Boendale [20, vs. 1–8, p. 137] [own translation]:

Alle die ghene die dit werc	To everyone who
Sien lesen ende horen	sees, reads, or hears this work
Die gruetic Jan gheheten Clerc	I greet you, Jan, called the clerk,
Vander Vueren gheboren	born in Tervuren
Boendale heetmen mi daer.	where they call me Boendale.
Ende wone te Andwerpen nu	I now live in Antwerp
Daer ic ghescreven hebbe menech jaer	where I have written for several years
Der scepenen brieve dat segghic u	the aldermen's letters, that is what I tell you.

In this prologue, Boendale states that he was born in the town of Tervuren near Brussels, where he is called Boendale after his hometown – but that he now lives in Antwerp, where they call him Jan de Clerc [Jan the Clerk], after his profession. This profession is confirmed by two fourteenth-century city accounts of Antwerp, where he appears alongside his occupations [6]. Accordingly, Boendale can be followed through the years as he makes his rise as a high-ranking secretary in the city of Antwerp. He is one of the first Middle Dutch authors with substantial biographical data, making him an incredibly interesting figure (see Figure 1).

Jans teesteye is not the only work that refers to a clerc named Jan. The author of *Lekenspiegel* refers to himself as *Jan, u arme clerc* [Jan, your poor clerk] [20, vs. 24, p. 137], and a rubric in manuscript BRUSSELS, KBR, IV 865 (p. 28) of *Brabantsche yeesten* reads *Dit dichte meester Jan van Andwerpen* [This was composed by master Jan of Antwerp] [11].[1] Accordingly, Jan van Boendale is the most common candidate author for the Antwerp School, though his precise contribution remains debated. In this discussion, two schools of thought have emerged: the "minimalist" and the "maximalist" school [1]. Minimalists, e.g. De Vries and Appelmans, attribute the Antwerp School only partially to Boendale, most often *Lekenspiegel*, *Jans teesteye*, *Brabantsche yeesten* and *Van den derden Eduwaert* [1; 5]. However, the number of works and which works exactly are attributed to him, can differ across studies. Maximalists, e.g. Van Anrooij and Reynaert, ascribe nearly the entire Antwerp School to Boendale, adding *Korte Kroniek van Brabant*, *Sidrac*, *Melibeus*, *Dietsche doctrinale*, *Boec van der wraken*, *Boec Exemplaer*, and *Hoemen ene stat regeren sal* [15; 23].[2] In short, between the minimalist and maximalist poles lies a spectrum of intermediate positions, e.g. Van Oostrom [27]. Table 1 offers a brief, non-exhaustive selection of existing hypotheses. Altogether, the Antwerp School encompasses roughly 60,000 verses, which makes the question of authorship complex.

Furthermore, the authorial question is complicated by an additional feature that links the works of the Antwerp School: many of the texts were revised, updated and continued at different times,

[1] It is not sure whether this rubric was written by the author or added by a scribe [11].
[2] Reynaert and Van Anrooij do not include *Boec Exemplaer*, Van Oostrom does not rule it out [15; 23; 27].

Figure 1: The first folium (f. 1r) of *Dietsche doctrinale* in codex The Hague, KB, 76 E 5. The miniature portrays the author of the text, though it remains uncertain whether the figure represents the original author Albertan of Brescia, or Jan van Boendale, who possibly adapted the work to Middle Dutch. Image made by the Royal Library of the Netherlands, available in the public domain via Wikimedia Commons.

Title (estimated date)	**De Vries** [5]	**Van Anrooij** [24]	**Appelmans** [1]	**Reynaert** [15]	**Van Oostrom** [27]
Brabantsche yeesten (1316–1351)	YES	YES	YES	YES	YES
Sidrac (1318)	/	NO	NO	PLAUS.	DOUBT
Korte kroniek van Brabant (1322–1332/3)	/	YES	NO	YES	DOUBT
Lekenspiegel (ca. 1325–1330)	YES	YES	YES	YES	YES
Jans teesteye (in between 1330 and 1334)	YES	YES	YES	YES	YES
Van den derden Eduwaert (shortly after 1340)	YES	YES	YES	YES	YES
Melibeus (1342)	/	NO	NO	YES	YES
Boec Exemplaer (before *Dietsche doctrinale*)	/	NO	/	/	DOUBT
Dietsche doctrinale (1345)	NO	NO	NO	YES	DOUBT
Boec van der wraken (1346–1351)	/	YES	NO	YES	YES
Hoemen ene stat regeren sal (before ca. 1350)	/	YES	/	YES	/

Table 1: An overview of the Antwerp School, (estimated) dates and attributions across a sample of studies. The columns indicate different scholars' hypotheses and the presence of a work in their analyses. Cells marked "YES" indicate that the scholar attributes a work to Boendale, "NO" means they do not. "/" indicates that the work is not mentioned or discussed by the scholar. Titles in bold are generally considered part of the minimalist school.

and consequently consist of different versions.[3] The role of the author in this dynamic process is the subject of much debate; it is unclear whether the interventions were made by the author himself, or by scribes or later editors [17; 21; 22]. The situation is further complicated by the fact that the texts often reuse parts of other works. For example, the first three books of *Brabantsche yeesten* are largely borrowed from Jacob van Maerlant's *Spiegel historiael* [11; 21], and *Korte Kroniek van Brabant* is possibly a kind of trailer for the *Brabantsche yeesten* [11; 21]. This is further evidence of the large degree of intertextuality present within the Antwerp School.

The authorship of the Antwerp School poses a difficult question, which is why I aim to approach it using computational methods. Using methods from the field of computational authorship verification, this paper investigates whether the texts of the Antwerp School (or a subselection of them) could have been written by the same author. After preprocessing the data, four different analyses will be conducted. First, the cosine distances between the works of the Antwerp School and a control corpus will be calculated; second, dimensionality reduction will produce a scatterplot representing stylistic similarity; third, nearest neighbour verses will be identified and compared

[3] It was not unheard of for medieval authors to revise or expand their works over time; Jean Froissart, for instance, is a well-documented example of this practice [4].

between text pairs; and fourth, parallel verses will be extracted.[4]

2 Materials

2.1 Corpus

The corpus used in this study is the freely available digital *Corpus Middle Dutch* [9]. The corpus encompasses a collection of 350+ Middle Dutch rhymed texts from the period 1300–1550 in XML format. All of the works of the Antwerp School are included, as well as works written by Boendale's contemporaries and works in similar genres. This makes the corpus a well-suited basis for authorship verification purposes, as it provides all texts necessary for a target as well as control corpus. However, most of the texts in the corpus are based on older, critical editions of the works. Such editions introduce editorial interference, which should be kept in mind when interpreting the results. Nevertheless, previous research has shown that computational authorship methods — especially those relying on robust stylometric features — are generally able to overcome noise and still yield reliable results [12]. Another challenge using this corpus is that the included editions do not indicate the different revisions (*cfr. supra*) made in *Brabantsche yeesten* or other works.

The target corpus encompasses all works of the Antwerp School, with a few adjustments compared to the *Corpus Middle Dutch*. First, it excludes the very short works: *Boec exemplaer* (298 verses), *Hoemen ene stat regeren sal* (18 verses), and *Korte Kroniek van Brabant* (374 verses).[5,6] Second, the work *Brabantsche yeesten* was divided into three separate parts. The first three books (*Brabantsche yeesten* (B1-3)) constitute a Brabant-oriented anthology of the *Spiegel historiael* (P1-4), authored by Jacob van Maerlant.[7] Previous research has shown that these first three books exhibit greater stylistic affinity to Maerlant than to Boendale, as Boendale largely copied his source text.[8] Accordingly, this section is treated separately. In addition, Boendale copied the fourth and the beginning of the fifth book of the *Brabantsche yeesten* (*Brabantsche yeesten* (B4-5)) in a similar fashion, but based on a different source text, namely *Chronica de origine ducum Brabantiae*. This passage is therefore also not representative of Boendale's writing style, and was separated from the rest of the fifth book. Only after verse 900 in the fifth book (*Brabantsche yeesten* (B5)), Boendale starts to write new, original text [11; 21]. Third, all intertextual passages — passages that are borrowed from other works — present in any of the works are left out of the analysis, so that two texts do not get mistaken to be from the same author solely because of an intertextual passage. The passages are marked based on the overview given by Vandyck and Kestemont [29].[9]

There will be two separate control corpora: one containing the oeuvres of Boendale's contemporaries Jacob van Maerlant and Lodewijk van Velthem — authors with established oeuvres — and one containing all other didactic as well as historiographic works outside of the Antwerp School, included in the *Corpus Middle Dutch*.

[4] Code available on GitHub via https://github.com/Caroline-Vandyck/authorship-boendale.

[5] *Sidrac* is left out of the analyses in the main text, but is included in Appendix B. This work is the only text of the Antwerp School mainly written in prose and is therefore difficult to compare to the works written in rhyme. The pro- and epilogue of the work are written in rhyme, but only consist of 222 verses.

[6] *Korte Kroniek van Brabant* also has a version of almost 2,000 verses, but its authorship is contested. Van Anrooij and Van Oostrom suggest ca. 1,800 verses were added by the Herald of Beyeren (ca. 1345–1414, Brabant), thereby making it unrepresentative of Boendale's writing style [14; 23].

[7] The "B" after *Brabantsche yeesten* stand for "book", the "P" after *Spiegel historiael* stand for "part".

[8] Figure 3 shows that *Brabantsche yeesten* (B1-3) clusters with *Spiegel historiael* (P1-4) and that *Brabantsche yeesten* (B4-5) appears in proximity as well.

[9] An overview is included in the code.

2.2 Preprocessing

The texts in the corpus underwent several preprocessing steps. First, each token was assigned a lemma and a part of speech, using the freely available model *hug-tdn-1400-1600* on the platform GaLAHaD [10]. The model was trained on a ground truth, human-annotated corpus of over 140,000 words from sources dating to the 15th and 16th century. Depending on the benchmark corpus, it obtains a macro F1 score ranging from 0.61 to 0.74 and a micro accuracy score ranging from 0.85 to 0.91 for lemmatisation; and a macro F1 score ranging from 0.52 to 0.89 and a micro accuracy score ranging from 0.93 to 0.96 for part-of-speech tagging. For lemmatisation, the model adheres to the GiGaNT lexicon, and for part-of-speech tagging, to the TDN (Tagset voor Diachroon corpusmateriaal van het Nederlands).[10] To overcome scribal variation, only lemmatised tokens are used in the analysis, and to reduce the influence of content, all proper nouns were converted to their part-of-speech label (*NOU-P*).

From the preprocessed texts, bigrams of rhyme words were extracted. A rhyme word is defined as the final word of a verse. The bigrams represent consecutive combinations of line-final words, but without assuming a formal rhyme relationship. While this method does not explicitly identify rhyme pairs according to a rhyme scheme (such as AABB), true rhyme pairs will occur more frequently and thus dominate the top-ranked bigrams. As a result, most of the extracted bigrams are likely to reflect actual rhyme relationships rather than arbitrary or transitional combinations. The order of the rhyme word combination is disregarded in the analysis, *hertog-hoog* [duke-high] is the same as *hoog-hertog* [high-duke]. Previous research has confirmed the importance of rhyme words and reliability of using them as features in Middle Dutch rhymed texts [12; 13; 29]. All texts were divided into segments of 1,800 verses.[11] The segmented texts were turned into a TF-IDF weighted bag-of-words, only including the 500 most frequent rhyme bigrams. By restricting the analysis to the most frequent rhyme bigrams, a stable basis for comparison across all segments is ensured, as these bigrams occur relatively consistently throughout the corpus. This approach also minimizes noise and reduces the influence of rare or coincidental word pairs that could otherwise distort the analysis [29].

3 Analysis and Results

3.1 Authorship Verification

The first step in the analysis was to apply authorship verification by comparing the stylistic similarity between the works of the Antwerp School to the stylistic similarity between established works of Boendale's contemporaries. To this end, all segments of the works of Boendale's contemporaries Jacob van Maerlant and Lodewijk van Vethem were compared to each other exhaustively (the first control corpus). This resulted in two types of text segment combinations: combinations of segments from different texts written by the same author (SA) (Maerlant–Maerlant or Velthem–Velthem) and combinations of segments written by different authors (DA) (Maerlant–Velthem or Velthem–Maerlant). *Jans teesteye*, the only work that can be attributed to Boendale with certainty and is therefore certainly not written by Maerlant or Velthem, was also included in this comparison. For each combination, the cosine distance between the two text segment pairs was calculated. The idea behind this is that SA pairs will consistently show more stylistic similarity than DA pairs. Based on the similarity distribution of the DA and SA pairs, an optimal threshold can then be calculated to differentiate between the two sorts of pairs. Values above the threshold would indicate different authorship, whereas values below the threshold indicate shared authorship. After this

[10] Detailed information can be found on the platform itself: https://portal.clarin.ivdnt.org/galahad/home.

[11] This way at least one segment of *Van den derden Eduwaert* is included, as well as two segments of *Melibeus*; the two shortest works in the target corpus.

calibration, the same procedure is applied to the works of the Antwerp School to assess whether the similarity distribution within the Antwerp School is more similar to the SA or DA pairs.

At first glance, the distribution of distance of the Antwerp School seems to be even lower than that of the SA pairs, which hints in the direction of shared authorship (see Figure 2). Furthermore, all text combinations are situated below the optimal threshold. However, some works seem to be more similar to each other than others. For instance, the combinations *Melibeus–Dietsche doctrinale*, as well as *Brabantsche yeesten–Van den derden Eduwaert* are located far below the lowest DA whisker of the boxplot. Accordingly, their stylistic similarity is remarkably high and could point at shared authorship. Slightly higher in the boxplot, *Jans teesteye*, *Boec van der wraken* and *Lekenspiegel* also show great similarities to the combination *Melibeus–Dietsche doctrinale*. *Van den derden Eduwaert* and *Brabantsche yeesten* only get linked to the other works higher up in the plot, but still below the optimal threshold and in a range that is more similar to SA pairs than DA pairs. This finding seems to favour the maximalist approach, but more analyses are desired.

3.2 Dimensionality Reduction

To obtain more insight into the works and how they relate to each other, a combination of PCA and UMAP dimensionality reduction was applied to the same vectors as before, and generated a scatterplot (see Figure 3). The dots represent the segments of 1,800 verses per text. The further they are apart, the more different they are; the closer they are together, the more similar they are. The corpus used here is slightly different than before, as it encompasses the Antwerp School, as well as all other historiographic or didactic works in the *Corpus Middle Dutch* (the second control corpus). The scatterplot demonstrates a clear separation between the Antwerp School and other texts, with all Antwerp works forming a coherent and distinct cluster. This means that the texts of the corpus are similar enough to each other for the model to recognize, and different enough from the other works to cluster away from them. Once again, *Melibeus* and *Dietsche doctrinale* are extremely close together, accompanied by *Jans teesteye*, *Boec vander wraken* and *Lekenspiegel*. *Brabantsche yeesten* and *Van den derden Eduwaert* are a bit more separated from the rest, but are still part of the cluster. Two works that are typically not associated with the Antwerp School also show up in the vicinity to the main cluster: *Grimbergse oorlog* and *Der ystorien bloeme*. Based on this plot, it is hard to tell whether these works share authorship, although they seem to be a bit more separated.

Grimbergse oorlog is an anonymous historiographical work, dated around 1266-1268 based on the described events [3].[12] Its content ties in well with the Antwerp School, as it recounts the war between the Dukes of Brabant and the Lords of Grimbergen. Serrure and Blommaert, who edited and published the work, even suspect that Boendale, the main candidate author for the Antwerp School, read *Grimbergse oorlog* [18]. Therefore, it is not a complete surprise that this work clusters closely together with the Antwerp School: it is similar in genre and content, but also in origin.

Der ystorien bloeme is the first Middle Dutch translation and adaptation of the *Legenda aurea* (1275–1300) by Jacobus de Voragine, the most popular medieval collection of hagiographies. Although less biographical than historiographical texts, *Der ystorien bloeme* remains a historical and narrative work. Claassens suspects that this translation dates from the late thirteenth or early fourteenth century. Interestingly, the translator of the work seemed to be relatively self-aware for his time. He provides a title to his work and presents himself as a clerc [2]. It is uncertain where the text was written, but the language is a mix of Eastern Flemish and West Brabantian [8]. In their computational research, Hogenbirk and Kestemont also noted similarities between works of the

[12] The epilogue notes that the final 1,400 verses were added by a continuator, so the work was split in two for the analysis; "*Grimbergse oorlog*" as it is used here, refers to the first part. The prologue (vs. 1–500) was excluded due to uncertain authorship [3]. Although Boendale, in theory, could have written the continuation based on its dating (after 1326), that section clusters further away from the Antwerp School if the segment length is reduced to 1,400.

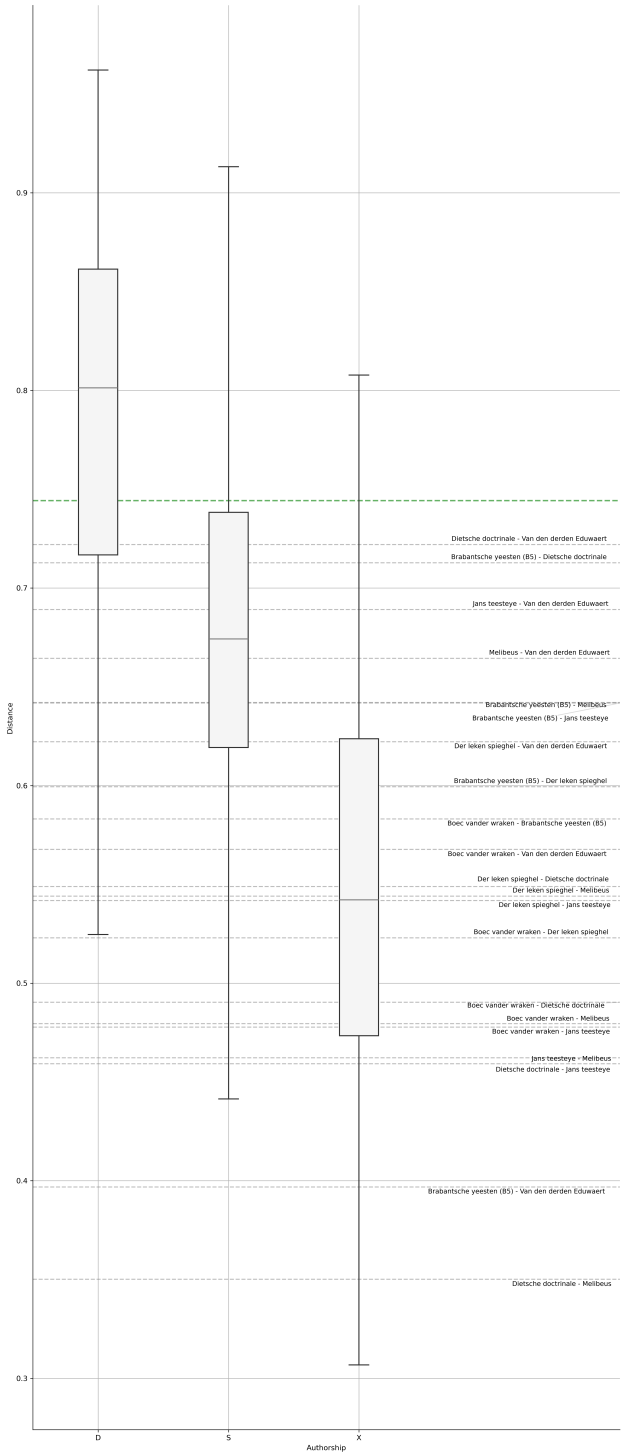

Figure 2: Boxplot showing the distributions of the stylistic difference (cosine distance) based on rhyme bigrams. “D” represents different-author pairs, “S” represents same-author pairs, and “X” refers to combinations within the Antwerp School. The label of each text combination represents the mean cosine distance between its segments. The green dashed line marks the optimal threshold to differentiate between different and same author pairs. The Antwerp School combinations fall below the threshold, with a distribution more closely resembling the “S” boxplot than the “D” boxplot, suggesting strong internal stylistic consistency.

Antwerp School and *Der ystorien bloeme*, but concluded that the work was written by Boendale's contemporary Lodewijk van Velthem [8].[13]

3.3 Authorship Ranking

In the two previous analyses, the results are dependent on the control corpora, to a certain extent. To counter this, I applied a bootstrapped analysis that searches for the nearest neighbour of a text segment in the entire *Corpus Middle Dutch*. Once again, the same vectors are used; however, instead of using them as is, a bootstrapped procedure is applied. For each segment of the texts ("focus segment"), 250 searches for the nearest neighbour are performed, but with two variations per iteration: (1) half of the 500 rhyme bigrams will be randomly selected; (2) the focus segment will be compared to only one segment per text in the control corpus. This reduces the influence of longer texts, but also reduces the influence of coincidence [29]. For each iteration, the similarity between the focus segment and the randomly selected segments from other texts is measured, identifying which text contains the most similar segment. After all iterations, the proportion of times each text was identified as containing the most similar segment to the focus segment is calculated, providing a percentage-based representation of these relationships.

The results are compelling: all of the texts of the Antwerp School have another text from the Antwerp School as nearest neighbour (see Table 2). Especially *Brabantsche yeesten*, *Van den derden Eduwaert*, *Melibeus*, and *Dietsche doctrinale* have a very clear Antwerp neighbour in the first rank. *Lekenspiegel*, *Boec van der wraken*, and *Jans teesteye* have a less clear favourite, yet the neighbours are spread evenly throughout works of the Antwerp School. *Lekenspiegel* and *Boec van der wraken* seem to be the most stereotypical texts of the corpus, as its nearest neighbours are relatively evenly spread throughout the Antwerp School.

For completeness, *Der ystorien bloeme* and *Grimbergse oorlog* were also included in the table. *Der ystorien bloeme* is most similar to *Boec van der wraken*, and *Van den derden Eduwaert* and *Brabantsche yeesten* (B5) can also be found in the top five. However, the percentages of the Antwerp School account for approximately 40% of the nearest neighbours in the top five, whereas this is 90% or more for the works traditionally associated with the Antwerp School. The percentages are also more scattered and lower in general, indicating that there is no clear favourite text or group of texts for *Der ystorien bloeme*. *Grimbergse oorlog* also has two works of the Antwerp School in its top five, with *Brabantsche yeesten* (B5) in the lead, but the ranking seems to be more genre-based. After all, the work seems to gravitate more toward the other historiographic works in the scatterplot. Thus, the two latter works display a different nearest neighbour distribution compared to the other texts traditionally associated with the Antwerp School.

3.4 Intertextuality

The final step of the analysis focuses on examining stylistic similarities in greater detail. Here, distant reading serves as a guide for close reading. Based on a method laid out by Kestemont, it is possible to identify and retrieve parallel verses between two texts [13]. Identifying corresponding verses has often been a way of linking various texts to each other, or even ascribing them to the same author [13]. Many traditional philologists have relied on their reading memory to identify parallel verses, also for the Antwerp School, e.g. Reynaert [15]. Kestemont laid out a method to perform this task computationally, for which the observations of traditional philologists formed the calibration corpus [13].

[13] I repeated the authorship verification analysis with *Grimbergse oorlog* and *Der ystorien bloeme* included in the target corpus and present it in Appendix A. I also repeated the analysis with only Lodewijk van Velthem and *Der ystorien bloeme* in the target corpus, for comparison.

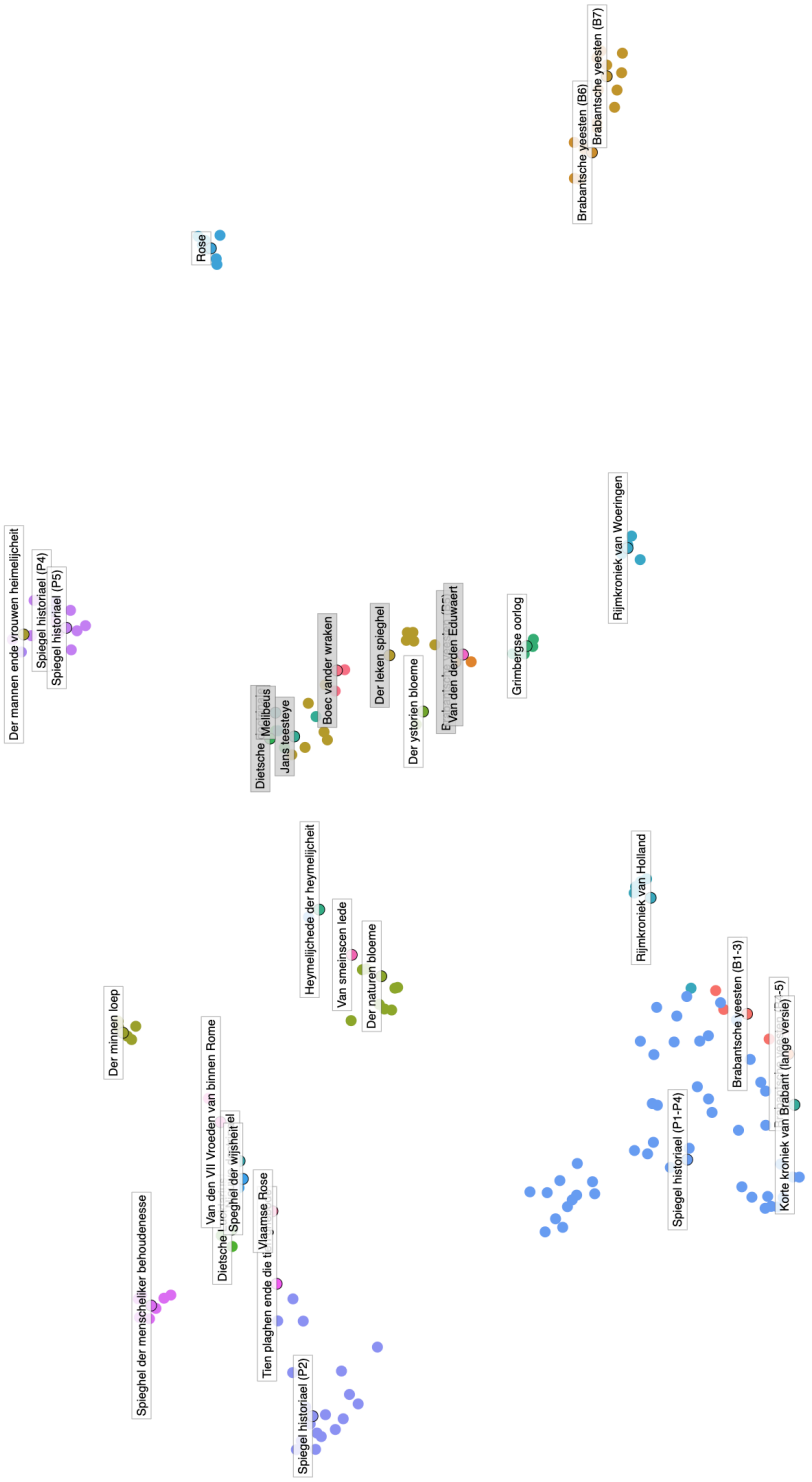

Figure 3: Scatterplot based on a combination of PCA and UMAP, illustrating the stylistic differences between the Antwerp School and all other historiographic and didactic works in the *Corpus Middle Dutch*. Each dot represents a segment of 1,800 verses from a text and the "middle" dot of each text is labeled. The larger the distance between dots, the more stylistic difference there is. The shorter the distance, the more similar the texts are. The Antwerp cluster (with grey labels) is located in the middle of the plot, suggesting a distinct but cohesive stylistic profile.

Target	Rank 1	Rank 2	Rank 3	Rank 4	Rank 5
Brabantsche yeesten (B5)	***Van den derden Eduwaert* (86.40%)**	***Brabantsche yeesten* (B1-3) (3.80%)**	*Roman der Lorreinen II* (2.20%)	*Spiegel historiael (P4)* (1.20%)	***Boec vander wraken* (1.00%)**
Van den derden Eduwaert	***Brabantsche yeesten* (B5) (88.40%)**	***Brabantsche yeesten* (B1-3) (5.20%)**	***Boec vander wraken* (3.60%)**	*Spiegel historiael (P4)* (0.80%)	*Rijmkroniek van Woeringen* (0.40%)
Der leken spiegel	***Boec vander wraken* (22.67%)**	***Dietsche doctrinale* (20.63%)**	***Jans teesteye* (15.73%)**	***Melibeus* (7.87%)**	***Brabantsche yeesten* (B4-5) (7.33%)**
Jans teesteye	***Dietsche doctrinale* (49.40%)**	***Melibeus* (20.80%)**	***Boec vander wraken* (18.40%)**	***Der leken spiegel* (10.60%)**	*Brabantsche yeesten* (B6) (0.20%)
Boec vander wraken	***Jans teesteye* (33.74%)**	***Melibeus* (21.33%)**	***Dietsche doctrinale* (16.93%)**	***Der leken spiegel* (11.47%)**	***Van den derden Eduwaert* (8.27%)**
Melibeus	***Dietsche doctrinale* (88.40%)**	***Boec vander wraken* (5.80%)**	***Jans teesteye* (3.20%)**	***Der leken spiegel* (1.80%)**	*Der mannen ende vrouwen heimelijcheit* (0.40%)
Dietsche doctrinale	***Melibeus* (84.00%)**	***Jans teesteye* (8.27%)**	***Der leken spiegel* (4.67%)**	***Boec vander wraken* (1.87%)**	*Enaamse Codex* (0.53%)
Der ystorien bloeme	***Boec vander wraken* (21.20%)**	***Van den derden Eduwaert* (12.00%)**	*Spiegel historiael* (P4) (7.20%)	***Brabantsche yeesten* (B5) (5.80%)**	*Grimbergse oorlog* (5.60%)
Grimbergse oorlog	***Brabantsche yeesten* (B5) (18.08%)**	*Renout van Montalbaen* (11.36%)	*Roman der Lorreinen II* (11.04%)	***Van den derden Eduwaert* (8.72%)**	*Rijmkroniek van Woeringen* (8.24%)

Table 2: The Antwerp School, *Der ystorien bloeme* and *Grimbergse oorlog* along with their top five highest-ranking texts, based on the bootstrapped analysis. Titles in bold belong to the Antwerp School.

In order to find parallel verses computationally, the nearest neighbour of every verse pair gets identified. The verse pair then gets retrieved if the similarity is higher than a previously defined similarity threshold. This threshold is the optimal boundary of separating identical from different verses, but it is impossible to exclude all mismatches. Kestemont calibrated this threshold based on aforementioned calibration corpus, containing the observations of traditional philologists [13]. The verses are compared based on their lemma's, but with a bit more weight given to the rhyme words. Table 3 shows the top eight parallel verses retrieved from the combination *Melibeus–Dietsche doctrinale*. The verses are highly similar and there are many more except for the ones included. This is a striking example; yet other text combinations are also highly similar. For instance, *Brabantsche yeesten* and *Van den derden Eduwaert*, but also *Lekenspiegel* and *Jans teesteye* show many parallel verses.[14] Overall, a remarkably large number of verses is shared between works of the Antwerp School.

The same method can also serve as a basis for identifying *where* similar verses occur in two texts, and accordingly for detecting longer intertextual passages. To compare two texts, the cosine distance between each verse and its nearest neighbour is calculated again. A rolling window of 100 verse pairs then moves through the texts, comparing these distances (see Figure 4, left). The

[14] More tables are included in the code.

Tokens ***Melibeus***	**Tokens** ***Dietsche doctrinale***	**Lemmas** ***Melibeus***	**Lemmas** ***Dietsche doctrinale***	**Cosine Distance**
Seneca seghet noch dit woert / Dat niemand bat toe en hoort	Noch seit seneca dit woert / Dat niemanne bat toe en hoort	NOU-P zeggen dit woord / dat niemand bet toezenden ne hoeren	nog zeggen NOU-P dit woord / dat niemand bet toezenden ne hoeren	0.000000
Gheeft mi op die wrake / Ic sal lonen die sake	Hi sprect gheeft mi op die wrake / Ende ic sal lonen die sake	geven ik op die wrake / ik zullen lonen die sake	hij spreken geven ik op die wrake / en ik zullen lonen die sake	0.039398
Die van vele lieden ontsien es / Moet vele lieden ontsien weder	Vele lieden hi moet van dien / Vele lieden weder ontsien	dan van veel lieden ontzien zijn / veel lieden ontzien weder	veel lieden hij moeten van dat / veel lieden weder ontzien	0.050010
Ende so wien men ontsiet / Ende mach gheminct wesen niet	Soe wien datmen ontsiet / En mach ghemint wesen niet	en zo wie men ontzien / en mogen minnen wezen niet	zo wie dat men ontzien / ne mogen minnen wezen niet	0.060991
Haddic enen voete inden grave / Nochtan so woudic leren	Haddic enen voet int graf / Nochtan soudic willen leren	hebben ik een voet in de graf / nochtan zo willen ik leren	hebben ik een voet in het graf / nochtan zullen ik willen leren	0.063768
Die wise paus Innocentius / Scrijft in sine boeke aldus	Die paus Innocentius / Scrijft in sinen boeke aldus	die wijs paus NOU-P / schrijven in zijn beuk aldus	die paus NOU-P / schrijven in zijn beuk aldus	0.074965
Salomon seghet die hoede sijn mont / Hoedt sijn ziele talre stont	Soe wie hoeden can sinen mont / Hoet sine ziele talre stont	NOU-P zeggen die hoeden zijn mond / hoeden zijn ziel te al stonde	zo wie hoeden kunnen zijn mond / hoeden zijn ziel te al stonde	0.090034
Wat si u segghen ende tonen / Hoedt u altoes jeghen honen	Watsi segghen ofte tonen / Hoedt u altoes jeghen honen	wat zij u zeggen en tonen / hoeden u altoos jegen hoon	watsie zeggen ofte tonen / hoeden u altoos jegen hoon	0.105172

Table 3: Parallel verse pairs between *Melibeus* and *Dietsche doctrinale*, retrieved by calculating the cosine distances based on lemmatised verses, giving more weight to verse-final (rhyme) words.

lower the value, the smaller the distance between a verse and its nearest neighbour — and the more similar they are. A steep drop thus indicates multiple highly similar verses after one another, and could indicate an intertextual passage, as observed between the *Brabantsche yeesten* and *Van den derden Eduwaert*.[15]

To gain more insight into this, the parallel verses were visualised in an additional way. Every dot in Figure 4 (right) represents a nearest neighbour verse pair with a cosine similarity above the defined similarity threshold.[16] The larger the dot, the more similar the verses are, and the greater the chance they represent an actual match instead of an accidental mismatch. The scattered dots indicate isolated similar verse pairs — occasional instances of comparable lines. The "trains" of dots show longer sequences of similar verses, and where they occur in their source texts. For instance, approximately verses 4000–4500 of *Brabantsche yeesten* (B5) correspond with verses

[15] More text combinations are included in the code. All texts of the Antwerp School were exhaustively compared and the important ones were enclosed.

[16] The code repository contains html files which are interactive and show the parallel verses when hovering over the dots.

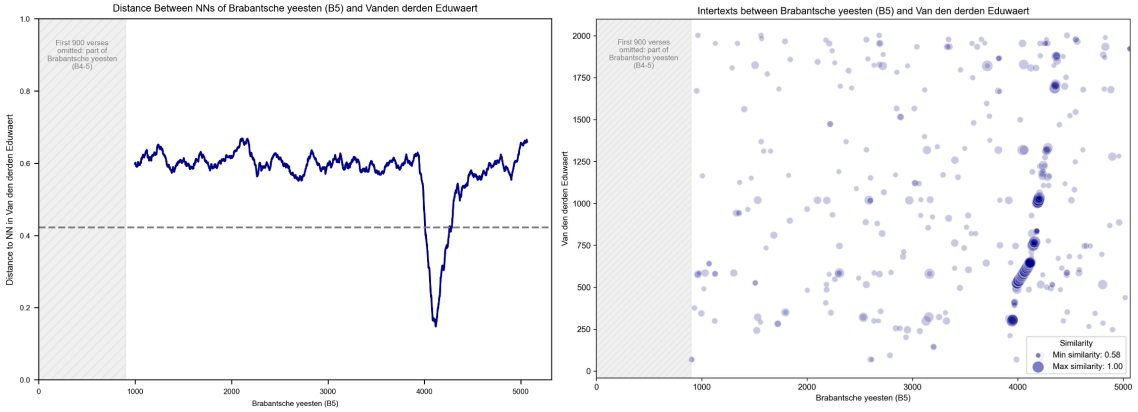

Figure 4: Rolling cosine distance between the nearest neighbour verses of *Brabantsche yeesten* (B5) and *Van den derden Eduwaert* (left); and possible parallel verses and passages between aforementioned works (right).

250–1250 of *Van den derden Eduwaert*. After its completion in 1316, *Brabantsche yeesten* was expanded five times over the following decades to include more recent historical events. Stein demonstrated that verses 3910 up until the end were added in the fourth version ca. 1348 [21]. He also remarked that a substantial part of this addition to *Brabantsche yeesten* was based verbatim on *Van den derden Eduwaert*. It is reassuring that the algorithm is able to identify this intertextual passage.

This passage was initially not encoded in the XML file, but could potentially influence the high similarity between *Brabantsche yeesten* and *Van den derden Eduwaert*. For this reason, it was excluded from *Brabantsche yeesten* (B5) in all of the described analyses. As shown, however, this did not influence their clustering. The analysis also confirmed an intertextual passage between the endings of *Jans teesteye* and *Boec van der wraken* [16]. Similarly, the passage from *Boec van der wraken* was excluded.

4 Discussion

In all analyses, *Van den derden Eduwaert* and *Brabantsche yeesten* (B5) consistently search for each others company, even with their shared passage removed. Their persistent proximity strongly suggests shared authorship, particularly in light of the numerous intertextual verses connecting the two works. In addition to the longer shared passages, numerous scattered verses in the works also coincide. This is not the way in which Middle Dutch compilers or authors would typically reuse source material. Instead, they would straightforwardly borrow large chunks of text more or less verbatim, rather than selectively or purposefully reuse isolated verses or verse pairs [15]. In this case, however, the author seems to consciously or unconsciously revert to the same formulations across different works. Accordingly, the intertexts between the two works seem to suggest not borrowing, but shared authorship. As Table 1 shows, this hypothesis has frequently and relatively consistently been put forward in previous research. In *Brabantsche yeesten* (B5), the author refers the reader to another of his own works on the life of King Edward III (*Eduwaert den derden*): *Diet al wille weten, vore ende na / Ic rade hem dat hi ten boeke ga / Daer ic dhistorie al te male / In hebbe gheset, redenlic wale* [Whoever wants to know all, before and after / I advise him that he go to the book / Where I have the story entirely / Set down orderly] [30, p. 299, own translation]. It is therefore generally assumed that the author of *Brabantsche yeesten* (B5) refers here to his earlier work, most likely *Van den derden Eduwaert*, both of which are commonly attributed to Jan van Boendale. The computational analysis appears to corroborate this conclusion.

There is more controversy regarding the authorship of *Melibeus* and *Dietsche doctrinale* (Table 1). However, in a relatively recent study, Reynaert argues that both works should be attributed to Boendale [15]. The findings seem to confirm this, as the two texts form the most closely related pair across all of the analyses. Both are Middle Dutch translations of Albertan of Brescia's *Liber Consolationis et Consilii* and *De amore et dilectione Dei*, respectively (Figure 1). Partially, their high degree of similarity can be explained by the works having the same source author. Yet, the intertexts – both on micro and macro level – again seem to also point in the direction of shared authorship in Middle Dutch. Additionally, Reynaert noted that both of these works often succeed each other in manuscripts, which leads him, amongst other reasons, to believe that the same person translated both texts [15].

Melibeus and *Dietsche doctrinale* additionally show close resemblance to *Jans teesteye*, *Boec van der wraken* and *Lekenspiegel* in the analysis. While *Van den derden Eduwaert* and *Brabantsche yeesten* form a somewhat distinct pair within the larger cluster, they nevertheless still remain remarkably close to the other Antwerp School texts — especially when compared to the control corpora. This pattern suggests that, at the very least, all Antwerp School texts were produced within a comparable context, likely by authors of similar intellectual and ideological backgrounds, and that they functioned within the same climate.[17] Yet, in light of both previous scholarship and the present findings, the evidence seems to strengthen the case for a single authorial hand behind the corpus, most plausibly that of Jan van Boendale. The recurrence of parallel verses across the corpus is particularly striking and offers compelling support for this interpretation.

Two texts showed up in the proximity of the Antwerp School that have not been linked to the corpus before: *Grimbergse oorlog* and *Der ystorien bloeme*. However, due to the dating of the texts and previous research, it is unlikely that Boendale wrote any of them. *Grimbergse oorlog* is dated ca. 1266-1268, which is too early to be authored by Boendale. The affinity to the Antwerp School can be explained by its similarity in genre, content, time and space to *Brabantsche yeesten* and *Van den derden Eduwaert*. It has been thought that Boendale read *Grimbergse oorlog*, but a further connection has not been explored [18]. It seems likely that the author of *Grimbergse oorlog* had a profile similar to that of Boendale. However, given the ranking of the other texts in Table 2, it seems that genre and content play an important role in the similarity of the text to the Antwerp School, so it is too speculative to assume shared authorship.

Although research by Hogenbirk and Kestemont concluded that *Der ystorien bloeme* was not authored by Boendale, previous studies had noticed the stylistic similarities with the Antwerp School. Like many Antwerp School works, *Der ystorien bloeme* appears to be aimed at a secular, listening audience. Possibly related to the aural context, the author of *Der ystorien bloeme* frequently applies formulaic or filler verses — verses that contribute little to the story content wise, but include emphasis formulas, source references, truth claims, or references to the act of speaking, often applied to maintain rhyme [8]. These occur so frequenctly — about one in every seven or eight verses — that Claassens labels the text a "literary failure" and its author a "bad poet" [2]. Claassens' critique parallels Van Driel's assessment of Boendale's work, who also relied on filler verses, likely to make his writing more accessible to a lay audience [2; 25]. Claassens further observes that parts of *Der ystorien bloeme* were reused in *Lekenspiegel*, though he believes the latter demonstrates greater poetic skill. Van Driel, by contrast, highlights *Lekenspiegel* and *Brabantsche yeesten* as examples of extensive formulaic writing, citing an excerpt where nearly half the verses are formulaic [25]. Hogenbirk and Kestemont however investigated these stylistic parallels computationally and concluded that the author was likely Lodewijk van Velthem, Boendale's contemporary [8]. *Der ystorien bloeme* tends to cluster with texts that also contain many formulaic verses, without this necessarily implying shared authorship. Further study of these formulas could

[17] All Antwerp School texts included in the analysis, so except from *Boec Exemplaer*, *Hoemen ene stat regeren sal*, *Korte kroniek van Brabant* and *Sidrac*.

clarify the relationships between the texts.

5 Conclusion

The results of this study support the so-called maximalist position: the stylistic similarities across the Antwerp School suggest a shared authorship. In particular, *Van den derden Eduwaert* and *Brabantsche yeesten* consistently cluster together across all analyses, even in the absence of their shared passage. The presence of numerous parallel verses throughout both texts strongly suggests shared authorship, rather than mere intertextual borrowing. A similar pattern is observable in *Melibeus* and *Dietsche doctrinale*, which, despite their similar Latin source texts, exhibit a degree of internal textual overlap that strongly suggests shared authorship. More broadly however, all works – included in the analysis – traditionally associated with the Antwerp School also share a very similar style and share similar rhyme words, implying that they were likely produced by the same author, presumably Jan van Boendale. The extent and distribution of parallel verses across the corpus further underscore this hypothesis. In further research, I intend to quantify these parallel verses and compare whether there are more of them within the Antwerp School than in the rest of the texts in the *Corpus Middle Dutch*. This way, I am also able to include the shorter works (*Boec Exemplaer*, *Korte Kroniek van Brabant*, *Hoemen ene stat regeren sal* and *Sidrac*) in the research. Finally, although it is unlikely that Boendale authored *Grimbergse oorlog* or *Der ystorien bloeme*, their notable proximity to the Antwerp School warrants consideration in further research. A close reading of their parallel verses could yield more insights as to why they cluster together closely with the Antwerp School.

Acknowledgements

I would like to express my gratitude towards the anonymous reviewers of the CHR Program Committee, investing their time and sharing their expertise. Furthermore, I would like to thank Dr. Dirk Schoenaers for sharing his insights on the draft of the paper, as well as my supervisors Prof. Dr. Mike Kestemont and Prof. Dr. Remco Sleiderink for their constructive comments and helpful suggestions. This research is part of the research project "Boendale's Many Faces: Modeling Historical Positionality in Jan van Boendale's (c. 1280-1351?) Oeuvre", funded by the Research Foundation Flanders (FWO), project number (1133925N).

References

[1] Appelmans, Janick. "Van Thomas' bijen tot Jans teesteye. De Brabantse historiografen en hun didactische bekommernis (dertiende en veertiende eeuw)". In: *In de voetsporen van Jacob van Maerlant: liber amicorum Raf De Keyser: verzameling opstellen over middeleeuwse geschiedenis en geschiedenisdidactiek*, ed. by Raoul Bauer. Leuven: Leuven University Press, 2002, pp. 256–283.

[2] Claassens, Geert. H.M. "Der ystorien bloeme. Een mislukte Legenda aurea-bewerking in het Middelnederlands?" In: *Ons Geestelijk Erf 70*, no. 2 (1996), pp. 99–120.

[3] Croenen, Godfried. "Het dubbele auteurschap van de Grimbergse oorlog". In: *Maar er is meer. Avontuurlijk lezen in de epiek van de Lage Landen. Studies voor Jozef D. Janssens*, ed. by Remco Sleiderink, Veerle Uyttersprot, and Bart Besamusca. Amsterdam: Davidsfonds/Literair Amsterdam University Press, 2005, pp. 131–151.

[4] Croenen, Godfried. "La guerre en Normandie au XIVe siècle et le problème de l'évolution textuelle des Chroniques de Jean Froissart". In: *La guerre en Normandie (XIe-XVe siècle)*, ed. by Anne Curry and Véronique Gazeau. Symposia. Caen: Presses universitaires de Caen, 2018, pp. 111–147. DOI: 10.4000/books.puc.11862.

[5] De Vries, Matthias. *Der leken spieghel. Leerdicht van den jare 1330*. Leiden: Du Mortier en Zoon, 1844.

[6] Gordeau, Era. “Boendale en de koning van Engeland”. In: *Tijdschrift voor Nederlandse Taal- en Letterkunde 136*, no. 4 (2020), pp. 210–223. DOI: https://doi.org/10.5117/TNTL2020.4.002.GORD.

[7] Heymans, J. “Geschiedenis in Der Leken Spiegel”. In: *Wat duikers vent is dit!*, ed. by G. R. W. Dibbets and P. W. M. Wackers. Uitgeverij Quarto, 1989, pp. 25–40.

[8] Hogenbirk, Marjolein and Kestemont, Mike. “Het werk van een klerk. Een stylometrische verkenning van het auteurschap achter Der ystorien bloeme”. In: *Tijdschrift voor Nederlandse Taal- en Letterkunde 129*, no. 44 (2014), pp. 246–273.

[9] INT. “Corpus Middelnederlands”. 2021. URL: http://hdl.handle.net/10032/tm-a2-j6.

[10] INT. “hug-tdn-1400-1600: A Computational Language Model for Middle Dutch”. Available on the GaLAHaD platform. Developed by the Instituut voor de Nederlandse Taal (INT). 2025.

[11] Kestemont, Mike. “Een stylometrisch onderzoek naar Jan van Boendales auteurschap voor de Brabantse yeesten”. In: *Revue belge de Philologie et d’Histoire 89*, no. 3 (2011), pp. 1019–1048. DOI: 10.3406/rbph.2011.8349.

[12] Kestemont, Mike. *Het gewicht van de auteur. Stylometrische auteursherkenning in Middelnederlandse literatuur.* Gent: Koninklijke Academie voor Nederlands Taal- en Letterkunde, 2013.

[13] Kestemont, Mike. “Zwerfverzen. De computationele detectie van intertekstualiteit in de Middelnederlandse epiek”. In: *Tijdschrift voor Nederlandse Taal- en Letterkunde 141*, no. 2 (2025), pp. 87–125.

[14] Oostrom, F.P. van. *Het woord van eer. Literatuur aan het Hollandse hof omstreeks 1400.* 5th ed. Amsterdam: Ooievaar, 1996.

[15] Reynaert, J. “Boendale of “Antwerpse School”? Over het auteurschap van Melibeus en Dietsche doctrinale”. In: *Al t’Antwerpen in die stad. Jan van Boendale en de literaire cultuur van zijn tijd*, ed. by Wim van Anrooij. Amsterdam: Prometheus, 2002, pp. 127–157.

[16] Schoenaers, Dirk. “Een verloren hoofdstukje van de Brabantsche yeesten? Vanden parlementen die te Mechlen lagen ende ander dingen”. In: *Queeste 30*, no. 1 (2023), pp. 6–42. DOI: https://doi.org/10.5117/QUE2023.1.002.sCho.

[17] Schoenaers, Dirk. “My is leet dat ics vergat. De Vijfde Partie van Lodewijk van Velthem en het vierde boek van Jan van Boendales Brabantsche yeesten”. In: *De boeken van Velthem. Auteur, oeuvre en overlevering*, ed. by Bart Besamusca, Remco Sleiderink, and Geert Warnar. Hilversum: Verloren, 2009, pp. 231–250.

[18] Serrure, C.P. and Blommaert, Ph. *De Grimbergsche oorlog. Ridderdicht uit de XIVe eeuw*. Vol. 2. Gent: C. Annoot-Braeckman, 1852.

[19] Sleiderink, Remco. *De stem van de meester. De hertogen van Brabant en hun rol in het literaire leven (1106-1430).* Amsterdam: Prometheus, 2003.

[20] Snellaert, F.A. *Nederlandsche gedichten uit de veertiende eeuw van Jan van Boendale, Hein van Aken e.a.* Brussel: M. Hayez, Drukker der Koninklijke Akademie, 1869.

[21] Stein, Robert. “Wanneer schreef Jan van Boendale zijn ‘Brabantsche yeesten’?” In: *Tijdschrift voor Nederlandse Taal- en Letterkunde 106* (1990), pp. 262–277.

[22] Van Anrooij, Wim. “Boendales ‘Boec van der wraken’: datering en ontstaansgeschiedenis”. In: *Queeste 2* (1995), pp. 40–53.

[23] Van Anrooij, Wim. “Literatuur in Antwerpen in de periode van ca. 1315-1350, een inleiding”. In: *Al t’Antwerpen in die stad. Jan van Boendale en de literaire cultuur van zijn tijd*, ed. by Wim Van Anrooij. Amsterdam: Prometheus, 2002, pp. 9–16.

[24] Van Anrooij, Wim. “Stand van Zaken; Jan van Boendale en de Antwerpse School”. In: *Nederlandse Letterkunde 5* (2000), pp. 86–98.

[25] Van Driel, Joost. *Meesters van het woord. Middelnederlandse schrijvers en hun kunst*. 1st ed. Middeleeuwse Studies en Bronnen. Hilversum: Uitgeverij Verloren, 2012.

[26] Van Mierlo, Jozef. *Geschiedenis van de letterkunde der Nederlanden*. Vol. 2. Malmberg, ’s Hertogenbosch / Brussel: Teulings’ Uitgevers-maatschappij L.C.G. Malmberg / Standaard Boekhandel, 1940.

[27] Van Oostrom, Frits Pieter. *Wereld in woorden. Geschiedenis van de Nederlandse literatuur 1300-1400*. 4th ed. Amsterdam: Bert Bakker, 2013.

[28] Van Uytven, R. “Het Antwerpen van Jan van Boendale”. In: *Al t’Antwerpen in die stad. Jan van Boendale en de literaire cultuur van zijn tijd*, ed. by Wim Van Anrooij. Amsterdam: Prometheus, 2002, pp. 17–29.

[29] Vandyck, Caroline and Kestemont, Mike. “Een auteur van formaat. Een attributie-onderzoek naar het oeuvre van de Limborch-dichter”. In: *Spiegel der Letteren 66*, no. 2 (2024), pp. 111–167. DOI: 10.2143/SDL.66.2.3293546.

[30] Willems, J.F. “Van den Derden Eduward, Koning van Engeland.” In: *Belgisch museum voor de Nederduitsche tael- en letterkunde en de geschiedenis des vaderlands. 4* (1840), pp. 298–366.

A Appendix: Authorship Verification including *Grimbergse oorlog* and *Der ystorien bloeme*

The exact same experiment as described in Section 3.1 was repeated, but with *Grimbergse oorlog* and *Der ystorien bloeme* included in the target corpus (see Figure 5). As can be seen, the similarity distribution of *Grimbergse oorlog* is very similar to the DA text pairs and the cosine distances all fall around the threshold. This implies, just as the dating of the text, that the work was not written by Boendale.

Der ystorien bloeme seems to behave more as the SA text pairs and all text combinations fall below the threshold. However, when the experiment was repeated with the works of Lodewijk van Velthem in the target corpus, the combinations behave in the same way (see Figure 6). This confirms the suspicion that *Der ystorien bloeme* clusters with works with many filler verses.

B Appendix: *Sidrac*

Sidrac is a text often included in the Antwerp School, but hard to compare to the rest of the works since it is the only one written in prose. Only the pro- and epilogue are rhymed. The results were not included in the main text since they were indecisive; accordingly, I will also keep this discussion brief. It is hard to determine whether differences observed between *Sidrac* and the other works should be attributed to an authorial difference, or the difference in form (prose vs. rhyme). The code contains a notebook with the option to analyse the text in several ways compared to the rest of the Antwerp School; it includes: (1) the option to remove the prose part of *Sidrac* and only load in the (rhymed) pro- and epilogue; (2) the option to remove the rhyme words in all of the

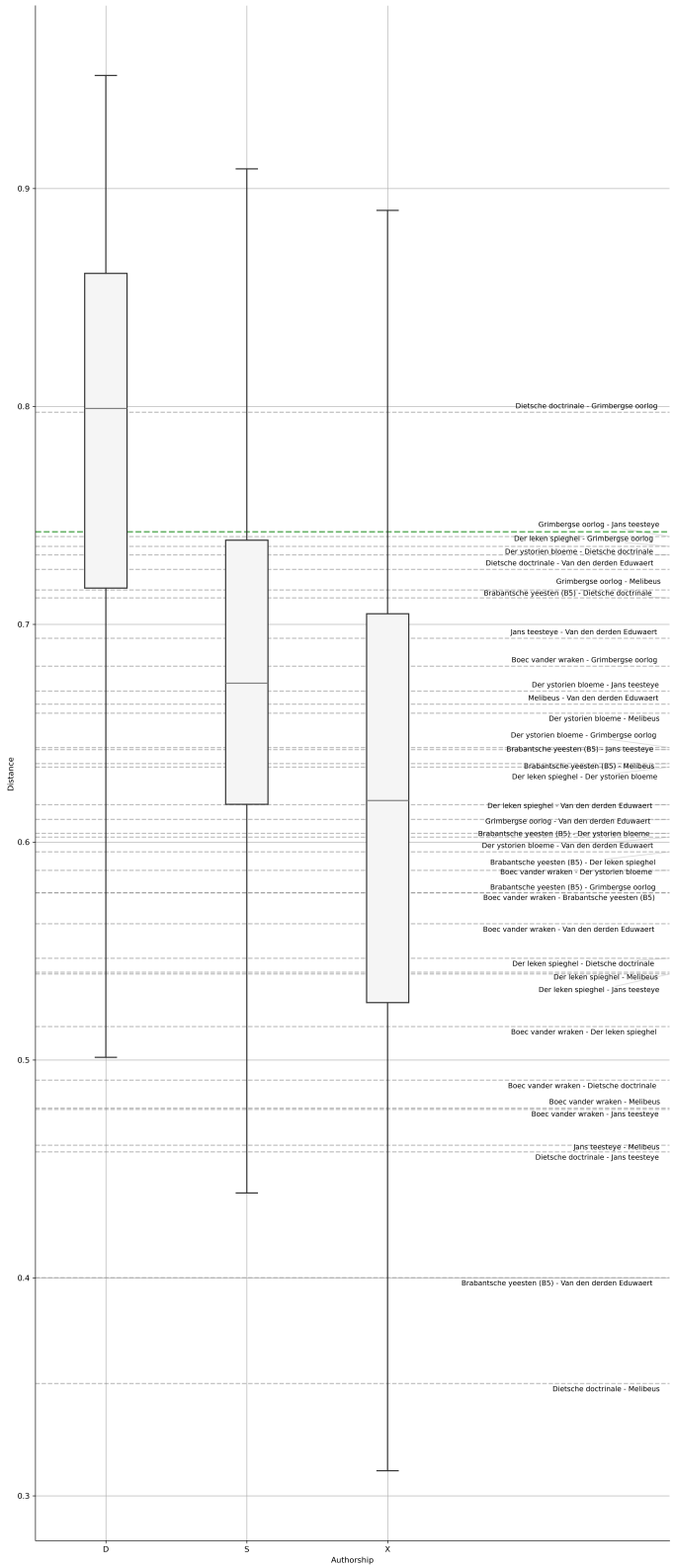

Figure 5: Boxplot showing the distributions of the stylistic difference, this time including *Grimbergse oorlog* and *Der ystorien bloeme* in the "X" boxplot. The former seems to fit in more with the DA pairs, whereas the latter seems to fit in more with the SA pairs.

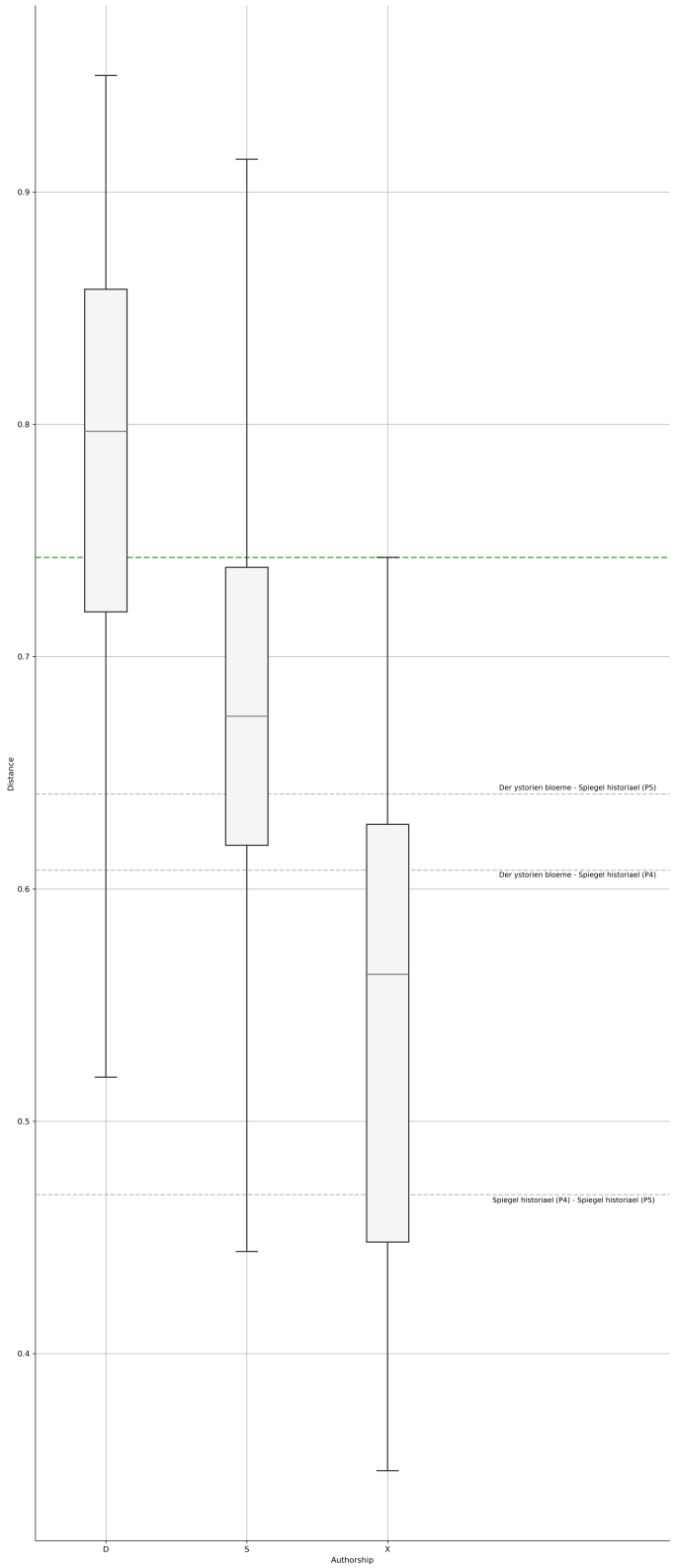

Figure 6: Boxplot showing the distributions of the stylistic difference, this time with the works of Lodewijk van Velthem and *Der ystorien bloeme* in the "X" boxplot. *Der ystorien bloeme* seems to be relatively similar to the other works by Velthem.

other works; (3) the option to limit the amount of segments per title in the scatterplot (otherwise e.g. *Spiegel Historiael* would dominate the plot).

Since the pro- and epilogue contain only 222 verses in total, it is hard to draw conclusions when applying option (1). There is too little data for the analyses, which is why it is not reported on. Option (2) did not have much influence and was thus not included, but can be tried out when running the code. The idea behind removing all rhyme words, was that this might be more fair towards *Sidrac*, since the factor of the rhyme constraint is then partially taken away. Option (3) was consistently applied.

Authorship verification First, I applied authorship verification to the Antwerp School including *Sidrac*. Yet, since *Sidrac* has no rhyming words, I had to revert to different features. First, I segmented the texts into segments of 5000 tokens each. Then, vectors were created based on the 100 most frequent words to decrease the influence of content.[18] Again, I applied TF-IDF weighting to this. The resulting similarity distribution of *Sidrac* seems to behave as a DA text pair (see Figure 7). Only three combinations including *Sidrac* are located below the threshold. However, it is hard to estimate whether *Sidrac* differs from the Antwerp School because of a different authorship, or because the writing style in prose differs from rhyme.

Dimensionality reduction I also recreated the scatterplot based on aforementioned vectors (see Figure 8). As can be seen, it is more difficult to produce clear clusters based on the most frequent words than based on solely the rhyme bigrams. This coincides with the hypothesis that rhyme words are important features when looking for the author of a Middle Dutch text. The Antwerp cluster is still somewhat visible on the left side of the plot. *Sidrac* is located above the Antwerp School. It somewhat moves away from the rest of the works and forms a cluster of its own, but perhaps less so than expected given its different nature. It is hard to conclude anything regarding the authorship based on this plot.

Authorship ranking I also carried out the bootstrapped authorship ranking with *Sidrac* included (see Table 4). Overall, the rankings based on the new feature vectors appear relatively stable compared to the main analysis (Table 2), which is positive.[19] Two Antwerp School texts appear in the top five of *Sidrac*, namely *Boec van der wraken* and *Lekenspiegel*.

Target	Rank 1	Rank 2	Rank 3	Rank 4	Rank 5
Sidrac	*Heymelijchede der heymelijcheit* (16.73%)	***Boec vander wraken* (8.98%)**	*Dietsche Lucidarius* (7.71%)	***Der leken spieghel* (7.53%)**	*Brabantsche yeesten (B6)* 4.80%)

Table 4: *Sidrac* along with its top five highest-ranking texts, based on the bootstrapped analysis. Titles in bold belong to the Antwerp School.

Intertexts Last, I looked into the intertexts. Intertextual passages were not retrieved, which is to be expected given the prose nature of *Sidrac*. Yet, when comparing *Sidrac* to *Lekenspiegel*, as well as *Dietsche doctrinale* and *Jans teesteye*, the prologue (up to verse 191) of *Sidrac* can be observed in the graph, as it shows a smaller distance. However, this similarity is also there when for instance comparing it to *Die rose*, another rhymed work, but not written by Boendale [29]. This shows that the form demonstrably matters when comparing works.

I also compared the rhymed pro- and epilogue to the other texts in the Antwerp School in order to retrieve parallel verses. This yielded some surprising parallels, for instance between *Sidrac* and *Boec van der wraken* (see Table 5). For such a short rhyming part in *Sidrac*, there seem to be many parallel verses. Especially the last included parallel verse catches the eye, as it seemed

[18] This is also the reason why option (2) did not have a lot of influence. The chance of rhyme words being in the top 100 most frequent words, is small.

[19] The full table can be viewed in the code. It is not included to limit the length of this appendix.

less conventional than the other ones. Accordingly, I searched for other verse pairs containing the lemmas “goud”, “menigvout” and “gesteente” (see Table 6). Four out of five texts these verses appear in belong to the Antwerp School. Note however, that although *Brabantsche yeesten* (B6) is a continuation of *Brabantsche yeesten*, it was continued by Wein van Cotthem, not Boendale [19]. It is interesting that this is the only verse of the five that does not contain “zilver”, forming a weaker match.

Tokens ***Boec van der wraken***	**Tokens** ***Sidrac***	**Lemmas** ***Boec van der wraken***	**Lemmas** ***Sidrac***	**Cosine Distance**
Nu seght allegader Amen / Amen	Amen segt gemeenlike / Amen	nu zeggen algader amen / amen	amen zeggen gemeenlijk / amen	0.155850
Daer om so biddic met trouwen / Hen die dit boec selen scouwen	Ende ic bidde hem allen met trauwen / Die desen bouc selen scauwen	daar om zo bidden ik met trouw / zij die dit boek zullen schouwen	en ik bidden hij al met trouw / die deze boek zullen schouwen	0.189874
Amen seghet allegader Amen / Ay hoe menichwerven heb ic gheseit	Amen segt gemeenlike / Amen	amen zeggen algader amen / ai hoe menigwerf hebben ik zeggen	amen zeggen gemeenlijk / amen	0.282956
MI quam een out boec in de hant / Daer ic in ghescreven vant	Soe quam my een boec ter hant / Daer ic in bescreven vinden	ik komen een oud boek in de hand / daar ik in schrijven vinden	zo komen ik een boek ter hand / daar ik in beschrijven vinden	0.312447
Des onne ons die hemelsche Vader / Amen seghet allegader Amen	Amen segt gemeenlike / Amen	des on ons die hemels vader / amen zeggen algader amen	amen zeggen gemeenlijk / amen	0.338080
Ghesteinte silver ende gout / Ende diere juwele menichfout	Die beter es menichfout / Danne ghesteente zelver of gout	gesteente zilver en goud / en die juweel menigvout	die goed zijn menigvout / dan gesteente zelf of goud	0.361816

Table 5: Parallel verse pairs between *Boec van der wraken* and *Sidrac*, retrieved by calculating the cosine distances based on lemmatised verses, giving more weight to verse-final (rhyme) words.

Title	**Tokens**	**Lemmas**	**Rhyme**
Boec vander wraken	Ghesteinte silver ende gout / Ende diere juwele menichfout	gesteente zilver en goud / en die juweel menigvout	goud menigvout
Brabantsche yeesten (B6)	Perlen ghesteinte ende gout / Ende chierheit menichfout	parel gesteente en goud / en sierheid menigvout	goud menigvout
Melibeus	Dat wijf es beter menechfout / Dan ghesteynte silver of gout	dat wijf zijn goed menigvout / dan gesteente zilver of goud	menigvout goud
Sidrac	Die beter es menichfout / Danne ghesteente zelver of gout	die goed zijn menigvout / dan gesteente zelf of goud	menigvout goud
Dietsche doctrinale	Vriends troest es beter menichfout / Dan ghesteinte siluer oft gout	vriend troost zijn goed menigvout / dan gesteind zilver of goud	menigvout goud

Table 6: Verse variants featuring the rhyme pair *goud* – *menigvout* and containing the lemma *gesteente* across multiple Middle Dutch texts.

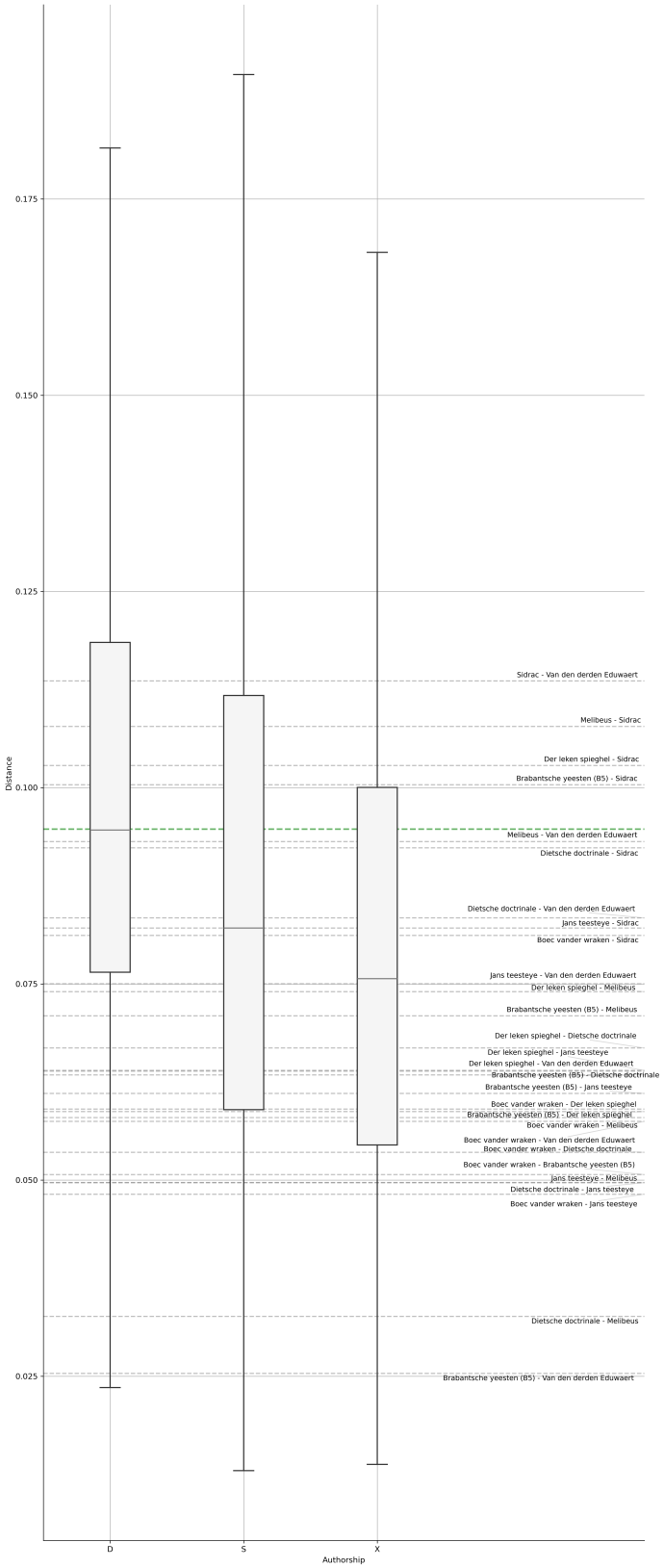

Figure 7: Boxplot showing the distributions of the stylistic difference, this time including *Sidrac*. It seems to behave more as though it is from a different author than the other works in the Antwerp School.

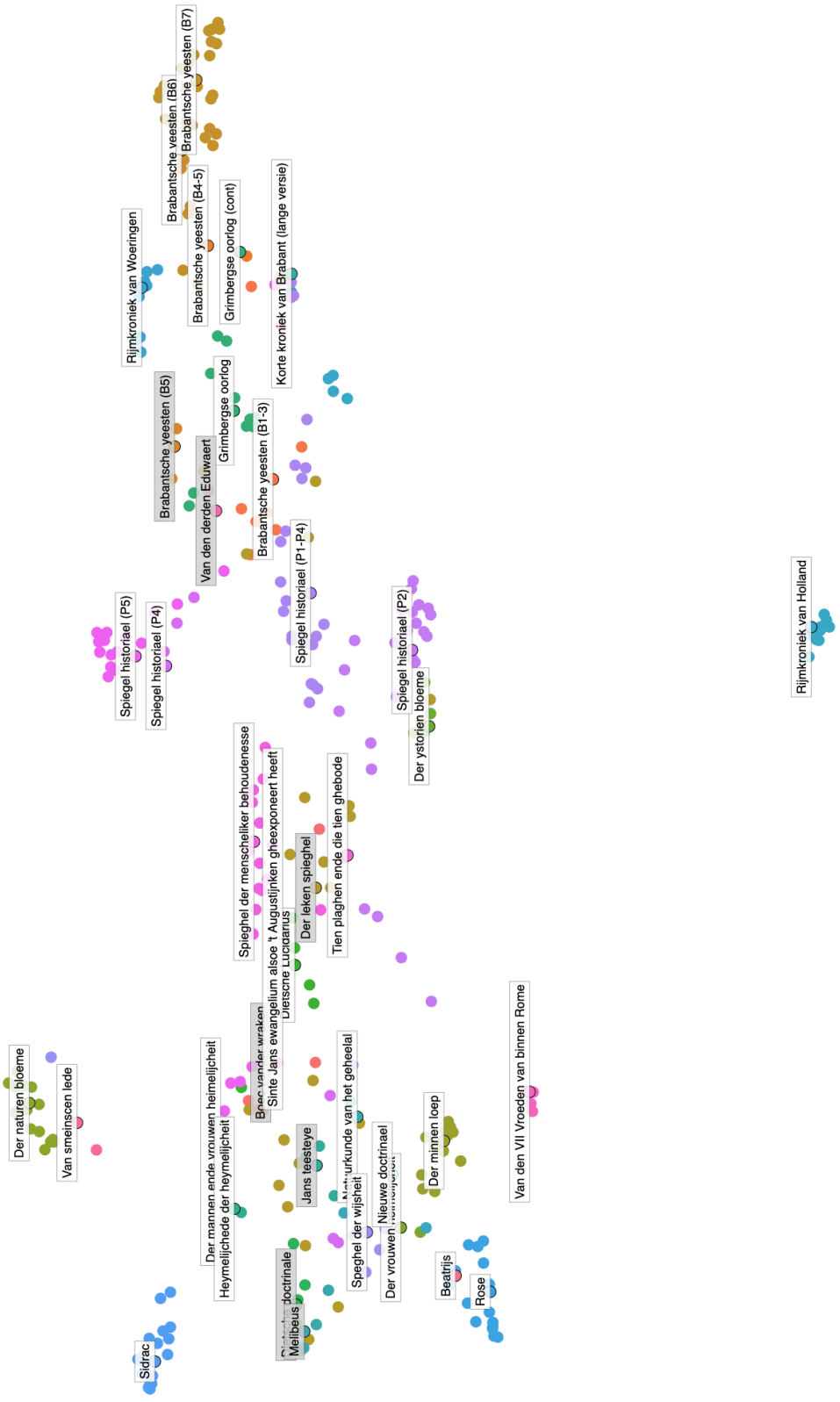

Figure 8: Scatterplot based on a combination of PCA and UMAP, now based on the 100 most frequent words and including *Sidrac*. Each dot represents a segment of 5000 tokens from a text.