Delta for Middle Dutch—Author and Copyist Distinction in *Walewein*

Karina van Dalen-Oskam and Joris van Zundert
Huygens Instituut, The Hague, The Netherlands

Abstract

The Middle Dutch Arthurian romance *Roman van Walewein* ('Romance of Gawain') is attributed in the text itself to two authors, Penninc and Vostaert. Very little quantitative research into this dual authorship has been done. This article describes our progress in applying different non-traditional authorship attribution methods to the text of *Walewein*. After providing an introduction to the romance and an overview of earlier research, we evaluate previous statements on authorship and stylistics by applying both Yule's measure of lexical richness and Burrows's Delta. To find out whether these new methods would confirm or even enhance our present knowledge about the differences between the two authors, we applied an adapted version of John Burrows's Delta procedure. The adapted version seems to be able to distinguish the double authorship of the romance. It also helps us to confirm some and to reject other earlier statements about the position in the text where the second author started his work.

Correspondence:
Karina van Dalen-Oskam, Huygens Instituut, Postbus 90754, NL-2509 LT Den Haag, The Netherlands.
E-mail:
karina.van.dalen@ huygensinstituut.knaw.nl

1 Introduction[1]

Until now, non-traditional authorship attribution techniques have hardly been applied to medieval texts. The fact that we do not and probably never will know the author of many texts, may play a role in this. There are cases, however, in which the researcher does not necessarily want to identify a specific author but wishes to distinguish one (perhaps anonymous) author from another. In these cases, it could indeed be very interesting to turn to modern authorship attribution techniques and find out of how much help they actually can be.

In our research on medieval Dutch literature, we have long been intrigued by a Middle Dutch Arthurian romance that seems to be the perfect case for an experiment like this. In this article, we first present the case and the questions we started out with (Section 2). We then summarize our first results and present the new questions to which these lead (Section 3). In Section 4, we investigate

two specific questions. We then focus on the possibilities of Burrows's Delta for distinguishing the authors as well as the scribes of *Walewein* (Section 5). After presenting our results, we sum up our conclusions and describe how we plan to test our results and to refine the method we are developing.

2 The Case: *Roman van Walewein*

One of the few Middle Dutch romances that were not translated from Old French is the Arthurian romance known as *Walewein* or *Roman van Walewein* ('Romance of Walewein'). The main character of this verse text in rhyming couplets is Arthur's nephew Walewein—the 'Gawain' of English texts and the 'Gauvain' of the French. *Walewein* was probably written around the year 1260, but the only surviving complete manuscript dates from almost a century later. The manuscript was written by two scribes. The first wrote lines

1	Court scene: a costly flying chess set appears then disappears
2	Walewein sets out to find it, in order to present it to King Arthur
3	Walewein finds the chess set at the court of King Wonder
4	Walewein will receive it in return for the 'Sword with the two rings'
5	Walewein finds the sword at the court of King Amoraen
6	Walewein receives it and has to bring Princess Ysabele to King Amoraen
7	Walewein finds the princess at the court of King Assentijn
8	Return to King Amoraen, Walewein receives (princess and) sword
9	Return to King Wonder, Walewein receives the chess set
10	Return to King Arthur, Walewein presents the chess set to King Arthur

Fig. 1 Storyline of *Walewein*

1–5,783 and the second—who stated that the job was finished in 1350—lines 5,784–11,202 (van Es, 1957, 2: p. 403). Apart from the complete manuscript, there are two extant fragments of another manuscript from roughly the same time, which together contain almost 400 lines of text. The digital file we used for our research was based on van Es's edition (van Es, 1957, 1).[2]

The narrator of the text states that the romance was written by two authors. Penninc started the text[3] but did not finish it, much to the regret of Pieter Vostaert, who decided to pick up on the story. At the end of the work, the narrator tells his audience that Vostaert has added 'about' 3,300 lines to Penninc's work.[4] Exactly *when* Vostaert added the end to the story is uncertain; this could be relatively shortly (a few years at the most) after Penninc's writing around 1260, or after a much longer time period. Theoretically, it would be possible that Vostaert wrote his text shortly before (or in) 1350, the year in which the manuscript was finished. Researchers up till now favoured the first option and have not seriously taken the second possibility into account.

The structure of the story is presented in Fig. 1. Several *Walewein* researchers have presented their opinion about the exact place in the text where Vostaert took over from Penninc. The text has 11,202 lines. The prologue comprises the first 32 lines, the epilogue lines 11,173–11,198. The last four lines are an addendum by the second scribe, who dates his work to the year 1350. Vostaert may have added the prologue or reworked it.[5] We do not know whether he included these 32 lines in his 'about 3,300' lines. In fact, it is even uncertain whether the manuscript from 1350 contains the same number of lines as the supposed manuscript that Vostaert himself completed. For the sake of convenience, however, we started from the hypothesis that changes in this respect were relatively minor.

It is possible that Vostaert added lines at several places in the text, although none of our predecessors seems to have thought of that possibility. However, we think it is highly implausible that Vostaert would then have been able to estimate that he had added about 3,300 lines. Because he mentions such a number and discusses his work as 'continuing'

Prologue ends with line	32
Main text ends with line	11,172
Epilogue ends with line	11,198
(scribe's addendum ends with line	11,202)

1. Vostaert also counted the complete prologue and epilogue as his and as being part

of the 'about 3,300' lines: 11,198 minus (3,300 minus 32) = *about* line 7930

2. Vostaert did not include the complete prologue and epilogue as his and as being part

of the 'about 3,300' lines: 11,173 minus 3,300 = *about* line 7872

Fig. 2 The broadest demarcation of the area in which Vostaert says that he took over from Penninc

Walewein lines 1 - 2: Vanden coninc Arture / Es bleven menighe avonture [...]

('Concerning King Arthur / there is many an adventure [...]', Johnson and Claassens

2000: 28, 29)

Vanden	lemma1 **van**, lemma2 **die**
coninc	lemma **koning**
Arture	lemma **arthur**
Es	lemma **zijn**
bleven	lemma **blijven**
menighe	lemma **menig**
avonture	lemma **avontuur**

Fig. 3 An example of the lemmatization of *Walewein*—the first two lines

a work that 'lacked a conclusion', we begin with the hypothesis that he added most of his work *after* the text of Penninc. We think that Vostaert estimated how many lines he had contributed by counting how many folia he had added and multiplying this figure by the mean number of lines he wrote on a page. To delineate the area in which Vostaert supposedly took over (Fig. 2), we calculated the number of lines for two possibilities, namely that Vostaert included the prologue and the epilogue in the 3,300 lines, and that he did not include them.

This yields the lines 7,872 and 7,930, respectively, as the area around which Vostaert's work starts.[6]

In the story outline, this area is part of episode number 7, which tells of Walewein's arrival at the castle where the princess lives, and describes the fights he has to engage in to get into the castle and then out of it again.

The first editor of *Walewein*—W. J. A. Jonckbloet—pointed to line 7,835 as the place where Vostaert 'roughly' started his work (Jonckbloet, 1846–1848, 2: p. 303). He based this

assumption on the usage of loan words, which clearly differs from one part of the text to the other (Jonckbloet, 1846–1848, 2: p. 182). In 1936, Maartje Draak in her dissertation on the structure of *Walewein* as being derived from a well-known folktale, argues for the takeover happening at line 7,840 or 7,843. She does so based on the first occurrence of 'a typical Vostaert word' in line 7,844 as *terminus ad quem* and of 'a typical Penninc rhyme' in lines 7,837–7,838 as *terminus a quo* (Draak, 1936, pp. 205–6). The second editor of the complete text—G. A. van Es—quotes the opinions of his predecessors and states that it is not possible to pinpoint the exact location, because it is entirely plausible that Vostaert edited the last lines of Penninc's text. He thus suggests that the break occurred somewhere between the lines mentioned by Jonckbloet and Draak (i.e. lines 7,835–7,842). He proposes that these lines may have been partly rewritten by Vostaert on the basis of Penninc's lines, and continues by presenting an overview of the differences between both authors (more on that in Section 4). In a study of the Middle Dutch Arthurian romance *Ferguut*, our colleague Willem Kuiper—one of the few researchers of Middle Dutch to use quantitative methods—states that based on spelling variants, Vostaert took over from Penninc at line 7,835 (Kuiper, 1989, p. 245 n. 84).

In 2001, a totally new interpretation was put forward by A. M. Duinhoven (2001, pp. 57–60), who states that it would not be logical to look for the author change in the area around line 7,840. As mentioned, these lines are in the middle of a long and coherent text episode, and this fact and the differences between the two authors as presented by earlier researchers do not convince Duinhoven. He disqualifies their choice for the area around line 7,842 as being the result of a silly miscalculation: they obviously subtracted 3,330 lines from 11,172 in stead of 3,300. Duinhoven's new solution is to assume a transcription error in the lines in which Vostaert mentions the number of lines he has added. A scribe, Duinhoven states, incorrectly transcribed the original 'three hundred and thirty lines' as 'thirty-three hundred lines'. Thus, Duinhoven considers line 10,872 (11,202 minus 330) as the place where the author changeover occurred. He argues that the text

has a logical break at this place, as there it is written that 'Of this I shall say no more, rather I shall continue the story for you' (10,872–10,873, Johnson and Claassens, 2000, pp. 481).

3 First Experiments

Our first question was a bold one: could we, by making use of quantitative, non-traditional authorship attribution techniques, find the exact place where Vostaert took over from Penninc and prove which of our predecessors was right? In choosing between the most well-known authorship attribution techniques, we were guided by our wish to follow-up on measurement results with research on word and sentence level, resulting in stylistic descriptions of the differences between the two authors. We believe that this is a good way not only to gain an insight into the craft of authors but also to discover more about how the techniques really work and what they actually measure. In addition, it is a way to evaluate the soundness of our approach and can therefore help us to refine the methods.[7]

To be able to present stylistic descriptions based on measurement results, the unit of information yielded by the measurements should fit the logic of philology. For our purposes, the most useful unit of information is (at least, at the moment) the word. We could compare authors by looking at the lexical preferences their work shows. We could group words, abstracting from word preferences to emphasized concepts, and through these perhaps in due time also themes or motifs, and we could show other scholars exactly how we do that: our work could be checked and repeated by others. Thus, in view of our present goals, very useful techniques that measure strings but neglect word boundaries do not qualify.

Our wish to be able to look for words, concepts, etc. also resulted in our decision to lemmatize the complete text and consequently base our measurements on the lemma's (headwords). To give an example: the Middle Dutch word for 'sword' occurs in different spellings in *Walewein*: *swert*, *swart* (both uninflected) as well as in inflected forms like *swerde*, *swerden* and *swarde*, *swarden*. All these different

word forms denote the same word. Because we have chosen to research vocabulary and not spelling we have to tag each occurrence with the same head word—modern Dutch ZWAARD in this case. But the word forms *swert* and *swart* are also used as the adjective 'black', so these occurrences need to be set apart from all the 'sword's with another headword, ZWART. And to complete the picture: the word form *swert* is used once as an enclitic form of the first person singular in the present tense of the verb *sweren* 'to swear' and the pronoun *dat*. So this word form gets two headwords: ZWEREN and DAT.

We tagged every word form (token) with the appropriate headword (type). For inflected nouns, adjectives, and pronouns, this is the nominative form, while for conjugated verb forms, this is the infinitive form of the verb. Furthermore, all spelling variants were normalized in the headword using as a guideline the modern Dutch spelling (or the spelling the word would have had, had it survived in our current vocabulary). See Fig. 3 for an example.

Lemmatizing the text is a first interpretation of the material. However, as we mentioned, the tagging of the unit of the word (here, the type) is necessary to be able to talk or write in a comprehensible way about the material or the results of the measurements. To return to the example given earlier: lemmatization makes it possible for us to make statements about how many times the word ZWAARD 'sword' is used in whatever spelling, and how often the word ZWART 'black', etc. It creates the opportunity to say something about authors' preferences for certain words instead of for certain spellings that may be used for altogether different words.[8]

We developed a web application that enabled us to semi-automatically tag the text. This application (called Autonom) generates a frequency list for a given text. Subsequently, any word form in the frequency list can be annotated by multiple tags and properties. The work of annotating is conveniently facilitated through the web interface, which obviates the need to use laborious and potentially error-prone handcrafted XML. Annotations and annotation properties for each user are stored as separate stand off XML hierarchies to support multiple views and interpretations of the text by different researchers.[9]

For our first experiments, we decided to use Udny Yule's Characteristic K to measure lexical richness and Burrows's Delta to examine significant differences in the use of high frequency words throughout *Walewein*. In order to be able to detect any changes in the course of the text, we devised a walking window of 2,000 lines moving in one-line steps through the text. We measured Yule's K or Burrows's Delta as shown in Fig. 4.

We used the formula for Yule's K, shown in Fig. 5.[10]

Yule's K measures the amount of word repetition in a text: the higher the K is, the more the repetition; the lower the K is, the richer the vocabulary. Applying this formula in walking window fashion yielded the results presented in Fig. 6, which for every line shows the difference between Yule's K in the 1,000 lines before and the 1,000 lines after that line.

Figure 6 represents only the area in *Walewein* in which we would expect to find the author change, although we did of course do measurements for the complete text (more about Yule in Section 4). Suffice it to say, the measurements did not yield any sign of significant changes in the final part of the text, which means that this approach neither confirmed nor repudiated Duinhoven's opinion that Vostaert took over from Penninc at line 10,872. In the area shown in Fig. 6, however, there is a clear, inverted arch-like trend that is broken by

Lines 1 - 2,000	(result plotted on line 1,000)
Lines 2 - 2,001	(result plotted on line 1,001)
Lines 3 - 2,002	(result plotted on line 1,002)
[....]	
Lines 9,201 - 11,200	(result plotted on line 10,200)
Lines 9,202 - 11,201	(result plotted on line 10,201)
Lines 9,203 - 11,202	(result plotted on line 10,202)

Fig. 4 Walking window through Walewein

$$10^4 \cdot \left[-\frac{1}{N} + \sum_{i=1}^{N} V_{(i,N)} \left(\frac{i}{N} \right)^2 \right]$$

Fig. 5 Udny Yule's Characteristic K

a localized peak around line 7,880. This may indicate the place where the author change took place, so we kept this in mind during the subsequent measurements. Comparable localized peaks were not found in any other part of the text.

Yule's K for the parts of the text before and after line 7,880 differ a lot, as is shown in Fig. 7. Penninc repeats words less often, which means that his lexicon is richer and his wording more elaborate. The difference is statistically significant.

In his Busa Award presentation and in two later articles, John Burrows introduced and detailed what he named the 'Delta procedure' (Burrows 2002, 2003). This method is an attempt to identify, with some degree of confidence, authorship in 'open games' (i.e. cases in which there is little or no text-external evidence to guide researchers to likely candidates). In such cases, one can only try to identify likely authors from a large set or database of texts or textual features.

The Delta procedure is based on a comparison of the differences between the frequencies of the most frequent words in a set of texts, and the frequencies of those words in the text we are trying to attribute to authors represented by texts in the set. The first step in the process is to compute the overall mean token frequencies for a large set of texts. By computing z-scores, these mean frequencies can then be compared with the frequencies of the same tokens in a test text and also with the frequencies of these tokens in the subset of texts of a particular author represented in the main text database. Thus, for each frequency the difference from the overall mean frequency is expressed by a number of standard deviations for that frequency in the main set. This gives us two z-scores for each frequency: one for the frequency of the token in the test text, and one for the frequency of the same token in the subset of a given author in the main set. Subsequently, the absolute difference between the two z-scores is
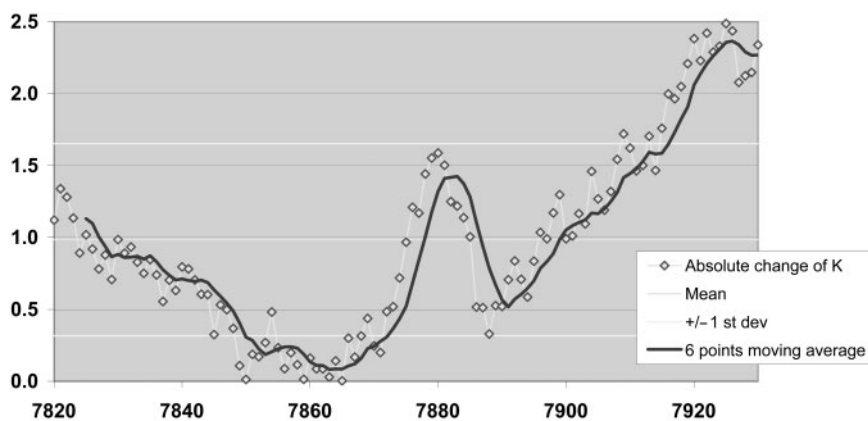
Fig. 6 Yule's K for the area in *Walewein* in which we would expect the author change

Complete text (lines 1 – 11,202) $K = 173.98$

Penninc, lines 1,000 – 7,500 $K = 171.40$ −2.58 (−1.48%) compared with complete text

Penninc, lines 4,581 – 7,880 $K = 167.42$ −6.56 (−3.77%) compared with complete text

Vostaert (lines 7,880 – 11,181) $K = 183.07$ +9.09 (+5.22%) compared with complete text

Fig. 7 Yule's K for different parts of *Walewein*

computed for all frequencies. Finally, the mean of these absolute differences gives us 'Delta' for the test text, that is, a unitary measure that provides us with an indication of how far apart the token frequencies lie that are characteristic of the test text and the frequencies that are characteristic of all the combined text of a certain author in the main set.

The lower the Delta value for a test text, the higher the possibility that both the test text and the subset of texts within the main set to which it is compared were written by the same author. Thus, if the Delta 'between' a certain test text and the subset of a given author in the main set has a low value, we attribute the text to that author. However, if none of the test texts scores a value for Delta that is significantly lower than the delta computed for all other test texts, we consider the possibility that the actual author of the test text was not represented within the main set of texts.

The Delta procedure is grounded in the availability of a large set of texts of considerable length representing a pool of authors from which a likely candidate may be selected. In our case, we do not have such a set at our disposal because of the scarceness of known Middle Dutch authors and the lack of other texts that can be ascribed to Penninc or Vostaert. We, therefore, had to devise a testing environment enough like the Delta procedure to be able to apply Delta to the *Walewein* text. It may be that the specific nature and design of this environment overstretches the feasible application of Delta. In addition, because of the scarce nature of our material (about 68,000 tokens) we have only a feeble statistical basis on which to build. Nevertheless, we thought that it might be interesting and possibly insightful to do some 'Delta work' on *Walewein*.

To devise some kind of text database from our 68,000 or so tokens, we developed an application that is able to base a main set on a given division of the main text. The number of divisions can be chosen arbitrarily, but to resemble the testing environment and text sizes Burrows used we chose to divide the main text into parts each comprising 12,000 tokens, which gave us just five 'texts' of 12,000 tokens and a smaller 'text' of 7,946 tokens as our main set. We instructed the application to select

a specific portion of the main text as the 'subset' of texts from either Penninc or Vostaert for Delta comparison. The application can walk a sliding window through the text line by line and, given a certain subset of the text as a comparison base, can compute a Delta for each window of text. We can arbitrarily size the window to contain as many or as few lines as we want. For the set of tests we ran for this contribution, we chose to use window sizes of between 1,000 and 2,000 lines. Cf. Fig. 8 for a schematic explanation.

We took measurements in this way throughout the text. We used two different parts of the text as the 'subset' to which we compared each window of 2,000 lines. The first set of measurements compared the frequencies of the 150 most frequent words in the 3,000 lines of subset text from Penninc (i.e. lines 1,000–4,000) (Fig. 9). The second set took lines 8,000–11,000 as the subset, as according to Vostaert and most previous researchers they were almost certainly written by the second author. Line 8,000 is well after the area in which we would expect the author change, so this part of the text is supposed to function as a relatively stable representation of Vostaert's work (Fig. 10). Both figures present the results through the complete text and mark the location of line 7,880, which we found to be significant in our measurement of lexical richness with Yule's *K*.

Figure 9, in which part of Penninc's text is used as the basis for the measurements, is difficult to analyse. With part of Vostaert's text as a subset, however, the measurements yield a clearer picture. In Fig. 10, there are clearly two levels of Delta in the text, which meets our expectations: we assume the author change to have occurred somewhere halfway down the sloping line. This agrees more or less with what the measurements of Yule's *K* yielded, namely that the change takes place somewhere around line 7,880. Figure 11 presents both graphs superimposed; the lines for both measures cross at about line 7,147. At this point in the text, there is no clear break in the story.[11]

However, these measurements do not help us to pinpoint the exact location in the text. We do not have enough information to give credit to one of our predecessors as being right about the
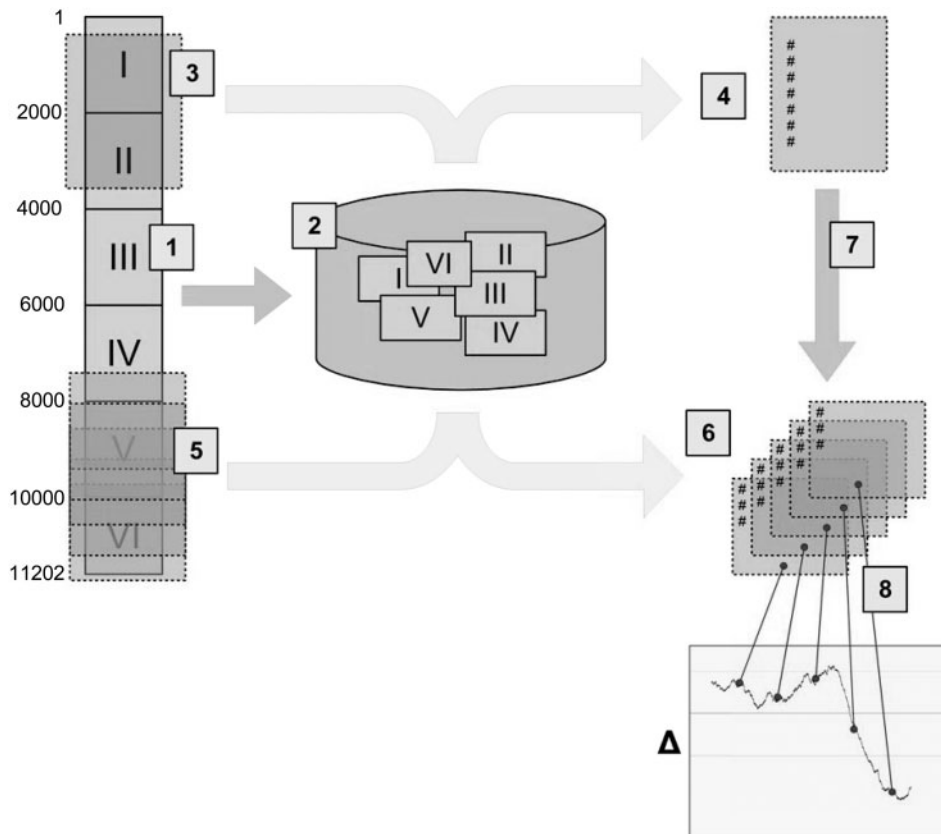
Fig. 8 (1) The main text is divided into six consecutive fragments. (2) The six fragments constitute the main set for the Delta procedure. (3) An 'undisputed' part of the text is chosen as a subset of the texts as a comparison base for the procedure. (4) From main set and subset a frequency z-score list for the subset is computed. (5) A window is constructed and moved through the text line by line. (6) For each window a frequency z-score list is computed against the frequencies from the main set. (7) For each window a delta is computed given the z-score list of the subset. (8) Each delta is plotted in a graph

line(s) where Vostaert took over from Penninc. Although our first question was too bold, the measurements did provide us with enough information to examine the possible differences between the two authors.

For the time being, we focused on line 7,880 as the break, assuming that Penninc's influence is uppermost before that line and Vostaert's after it. We then surveyed the vocabulary of the text before and after line 7,880 and generated a list of words that were used significantly more by Penninc and significantly more by Vostaert, respectively. We arranged them in order of significance, with the

word with the highest significance for each author at the top (Table 1).

The list in Table 1 gives rise to some interesting observations. As to the methodological aspect, the word at the top of Penninc's list shows that we would also need tagging for part of speech, as the lemma 'zijn' is homophonous: it denotes either a form of the verb 'to be' or a form of the possessive pronoun 'his'. Of course we would want to know for which of these Penninc has a predilection (if any, after disambiguating them). Tagging for part of speech would be necessary in order to be able to distinguish the different functions of 'zijn' in a list
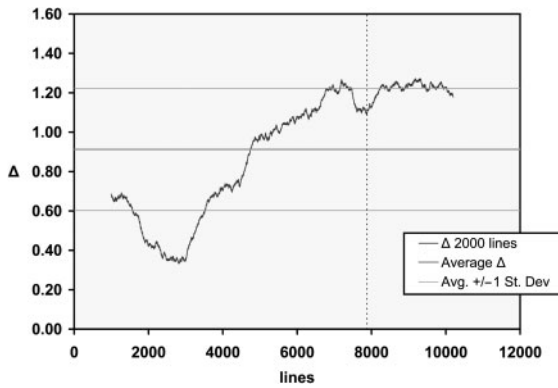
Fig. 9 Burrows's Delta. The text corresponding to lines 1,000–4,000 in Penninc's part of the romance was used as a subset for the procedure. From this subset the top most 150 frequently used words were used for comparison
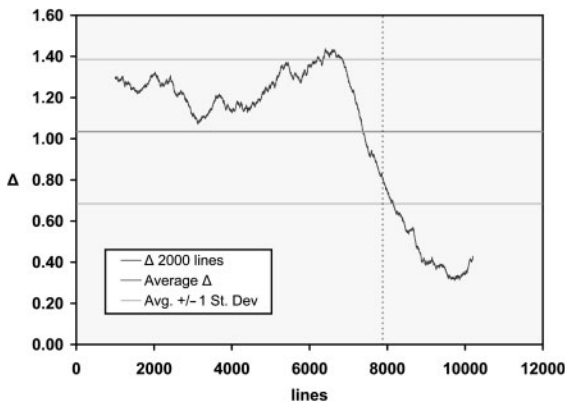

Fig. 10 Burrows's Delta. The text corresponding to lines 8,000–11,000 in Vostaert's part of the romance was used as a subset for the procedure. From this subset the top most 150 frequently used words were used for comparison

like this (but a PoS tagger for Middle Dutch is not yet available).

Apart from that, the first thing we noticed is that Penninc makes significantly more use of first- and second-person personal pronouns, in contrast to Vostaert's significantly greater use of the third person. Penninc also uses many modal verbs. But why? Are there several reasons for these differences, or can all be explained by only one or two 'special
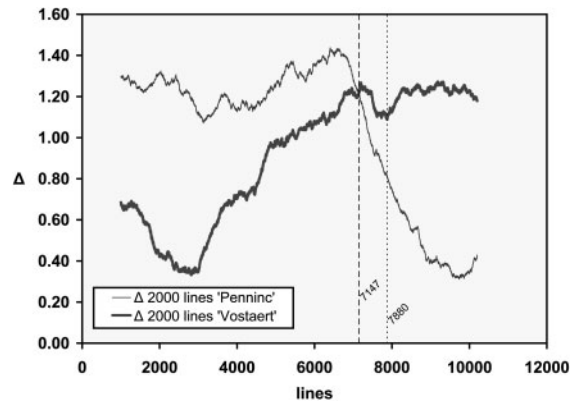
Fig. 11 Figures 9 and 10 in one graph

effects' of the individual authors? In 1957, van Es mentioned that Penninc uses more dialogue than Vostaert. Could that be the explanation for the statistical differences between the two authors for the pronoun differences *and* those in the modal verbs (expected to be used more in dialogue than in otherwise narrated text)? This guided us back to the work of van Es and the differences he found based on a selection of quantitatively analysed lexical elements.

4 Dialogue and Court versus Sword

For the introduction to his edition of *Walewein* in 1957, van Es made use of quantitative research on a select group of lexical and grammatical items conducted by G. S. Overdiep at the beginning of the twentieth century and reported on in his masters thesis (van Es, 1957, 2: p. 344). This research compared the usage of the selected items in the text parts of Penninc and Vostaert. As mentioned, van Es assumed that Vostaert reworked the last lines of Penninc's text and assumed that Vostaert's work starts to be visible in the range of lines 7,835–7,842. The items on which van Es reports concern rhyme words, some syntactical aspects, usage of past perfect and historical present tense and the differences in their frequencies at the end of the line and/or in the line, usage of adverbs,

Table 1 Significant words per author, taking line 7,880 as the break

	Penninc	Stdev Mean	0.05242999 0.0166		Vostaert	Stdev Mean	0.03920838 0.0167
			z-score				z-score
be, his	zijn	0.8413	15.7293	the, this	die	0.6234	15.4755
I	ik	0.8042	15.0217	he	hij	0.4112	10.0614
me	mij	0.6790	12.6328	to	te	0.3670	8.9353
you	gij	0.5059	9.3325	knight	ridder	0.3659	8.9071
my, mine	mijn	0.4223	7.7364	large	groot	0.3406	8.2613
may	mogen	0.3158	5.7060	duke	hertog	0.3051	7.3573
it	het	0.2957	5.3222	very, pain	zeer	0.2951	7.1002
stand	staan	0.2665	4.7663	they, she	zij	0.2886	6.9355
we	wij	0.2514	4.4775	Walewein	walewein	0.2823	6.7757
lord	heer	0.2328	4.1224	there	daar	0.2748	6.5846
that	dat	0.2195	3.8692	so, thus	zo	0.2260	5.3397
yonder	gene	0.2137	3.7587	of	van	0.2242	5.2924
your	uw	0.2131	3.7465	Isabele	isabele	0.1844	4.2767
you	u	0.2095	3.6793	maiden	jonkvrouw	0.1813	4.1977
say	zeggen	0.2022	3.5387	hit, slay	slaan	0.1607	3.6728
god	god	0.1903	3.3124	in	in	0.1382	3.0998
live	leven	0.1774	3.0663	horse	hors	0.1349	3.0160
come	komen	0.1702	2.9290	how	hoe	0.1348	3.0117
need	moeten	0.1653	2.8359	self	zelf	0.1334	2.9774
gate	poort	0.1650	2.8300	other	ander	0.1330	2.9662
see	zien	0.1599	2.7316	fox	vos	0.1228	2.7068
squire	knaap	0.1524	2.5898	no	geen	0.1196	2.6245
then	doe	0.1485	2.5157	to	toe	0.1171	2.5612
give	geven	0.1485	2.5150	man	man	0.1131	2.4601
well, rather	wel	0.1479	2.5043	many	menig	0.1074	2.3153
over	over	0.1474	2.4931	black	zwart	0.1023	2.1845
king	koning	0.1454	2.4555	also	ook	0.0985	2.0859
thus	dus	0.1396	2.3445	begin	beginnen	0.0980	2.0739
stay	blijven	0.1392	2.3375	because	want	0.0969	2.0465
inside	binnen	0.1267	2.0992	brave	stout	0.0961	2.0252
not	ne	0.1229	2.0275	speak	spreken	0.0957	2.0155
at	aan	0.1147	1.8707	to	tot	0.0942	1.9779
shall	zullen	0.1038	1.6623	helmet	helm	0.0925	1.9352
you	jij	0.1034	1.6550	(some)one	men	0.0918	1.9169
loyal	trouw	0.1011	1.6111	sweet	lief	0.0912	1.9009
go	gaan	0.1009	1.6075	on	op	0.0910	1.8953
serpent	serpent	0.0958	1.5093	blood	bloed	0.0884	1.8290
allow	laten	0.0954	1.5030	and	en	0.0873	1.8027
desire	begeren	0.0915	1.4280	walk	lopen	0.0852	1.7485
day	dag	0.0878	1.3569	merciful	goedertieren	0.0820	1.6672
where	waar	0.0821	1.2481	hour	stonde	0.0812	1.6466
all	al	0.0807	1.2211	do	doen	0.0804	1.6262

and aspects of the vocabulary used by Penninc and by Vostaert, especially the frequencies of loan words of French origin. In addition, van Es investigated the preferences of both authors for certain scenes in the story (van Es, 1957, 2: pp. 341–72). In a concluding section, van Es presents a short description of the character of both authors and the relationship he assumes between them. He describes Penninc as calm, deliberate, contemplative, elaborate in his wording, focused on courtly details and as using a lot of dialogue, while he describes Vostaert as tempestuous, passionate, and focused on action,

especially on fights. Van Es assumed that these differences could be accounted for by the age difference between the two authors and by their possibly different backgrounds (e.g. working in different regions or at different courts). Van Es thinks that Penninc was older than Vostaert, but believes that they must have known one another (van Es, 1957, 2: pp. 372–3). This implies that van Es dates Vostaert's work not very long after Penninc's.

Based on the list of lexical preferences in the two parts of the text as presented at the end of Section 3, we decided to investigate two aspects. The first was the dialogue: could we confirm van Es's remark that Penninc made more use of dialogue than Vostaert did, and would this help us to largely explain the results of our measurements of lexical richness and of Delta? The second question we wanted to answer was whether van Es was indeed right in his intuition that Penninc was more elaborate in his courtly scenes, and Vostaert more elaborate when there was any fighting to be done.

We tagged the text for dialogue, marking direct and indirect speech as well as narrated monologue (*erlebte Rede*). Because the last two occur only very occasionally, we grouped all three types of speech together to get an idea of the amount of dialogue both authors use. We again used line 7,880 as the point of division. We measured two things: the mean number of words used in dialogue and the mean number of clauses. By a 'clause' we mean a set of words spoken or thought by a character in the text surrounded by words of the narrator or author. Thus, if a character's speech is interrupted once by such a phrase as *seit hi* ('he said'), we counted two clauses. For these measures we also used a walking window, this time of only 1,000 lines.

Figure 12 presents a graph showing the percentage of words in dialogue of the text for a 1,000-line window.

This graph has a significant peak around the line where the change of copyists takes place. However, there is another explanation for this: lines 5,328–5,805 contain the long monologue by the fox Roges telling Walewein his life story. That this must be the reason for the peak

becomes clear when we look at the graph presented in Fig. 13, which shows the number of clauses for the 1,000-line walking window.

Neither of these two figures reveals a significant difference either in the mean number of words or in the number of clauses in dialogue between the two parts of the text. Bar charts totalling the numbers for both parts provide some more insight. Figure 14 presents the mean percentage of words used for dialogue in every 1,000 lines for both parts,
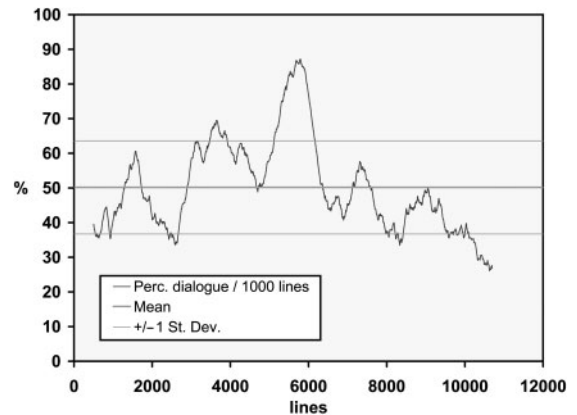

Fig. 12 Percentage of words used in dialogue. A 'walking window' of 1,000 lines of text was used, calculating the percentage for every window while moving the window through the text line by line
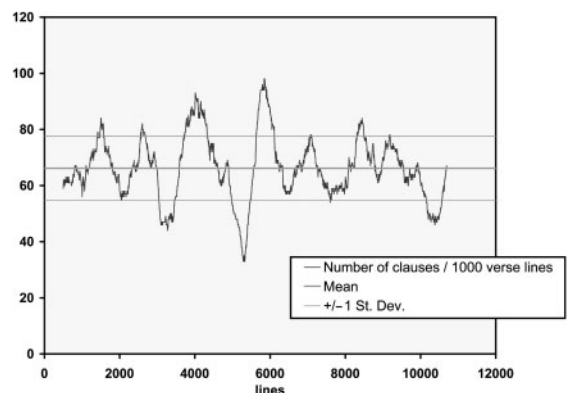

Fig. 13 The number of clauses in dialogue. A 'walking window' of 1,000 lines was used, counting the number of lines of dialogue for every window while moving the window through the text line by line

and Fig. 15 the mean number of clauses used per 1,000 lines for both parts.

We conclude from these graphs that Penninc uses more words in dialogue in about the same number of clauses. Thus, Penninc seems to be more elaborate in his wording than Vostaert is, or/and he uses dialogue for other purposes than Vostaert does, for example, for long monologues, such as the one by the fox Roges. This is not the place, however, to go into this more deeply. It will be clear, though, that van Es's remark that Penninc's part contains substantially more dialogue cannot unambiguously be confirmed by our measurements.

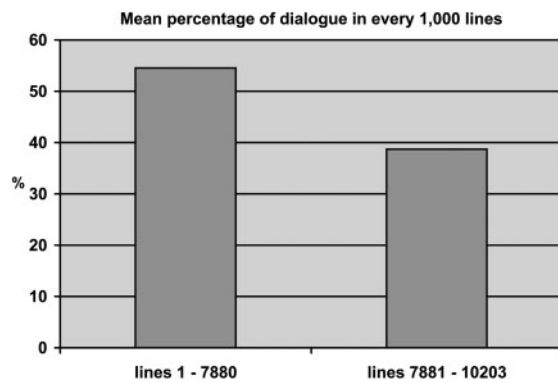Our second question concerns the contrast that van Es saw between 'court' (Penninc) and

'sword' (Vostaert). Could it be that Penninc is more elaborate than Vostaert when he writes about scenes in courtly surroundings, and could Vostaert use more words for fight scenes than Penninc does? To measure this, we tagged the text for scenes that take place at a castle and for fight scenes. As these two types of scenes can occur at the same place in the text, we defined a court scene as starting at the moment Walewein enters a castle (or a pavilion in which a lord and his retinue are temporarily staying) and ending when he departs from that castle or pavilion, and a fight scene as starting with the first physically hostile act committed by Walewein or his adversary (regardless of who is 'right') and ending when one of the fighting parties surrenders, is taken prisoner, or dies. Both the first part of the text (up to line 7,880) and the second part (from line 7,880 onwards) contain the starts of seven court scenes. The first part contains nine fight scenes, while the second part has six.

We measured the number of words used for court scenes in the first and in the second part of the text and for the fight scenes in both parts. The measurements for both types are given in Fig. 16. Per type, the darker bar on the left gives the results for the first part of the text (Penninc) and the lighter bar on the right gives those for the second part (Vostaert).

Fig. 14 Mean percentage of words in dialogue per 1,000 lines of text for both parts of the main text
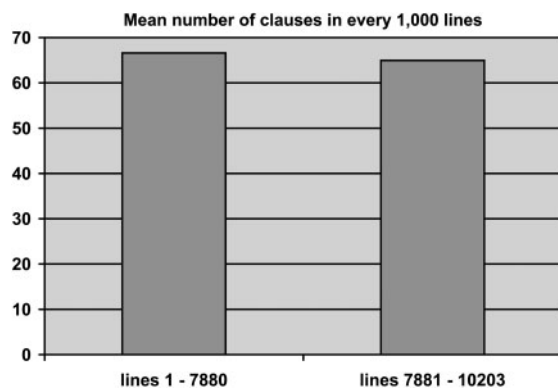
Fig. 15 The number of clauses used on average in every 1,000 lines for both text parts
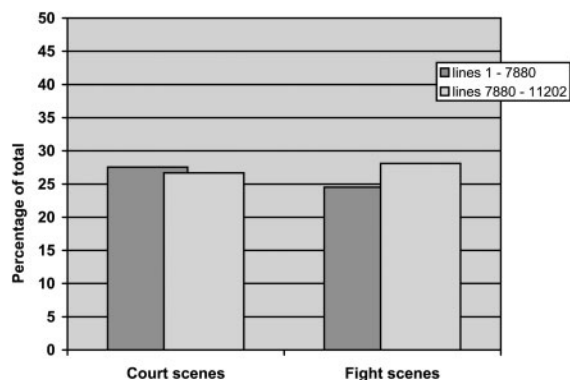

Fig. 16 The number of words used in court scenes and in fight scenes for the parts of the text before and after line 7,880 as a percentage of the total number of words in those parts. Darker bars: first part (Penninc); lighter bars: second part (Vostaert)

Per type of scene, the graph shows only a small difference between the two parts of the text. The authors are comparable in their attention to both 'court' and 'sword' when we look at the number of words. The maximum amount of distinguishable difference is that Penninc uses about 3.6% fewer words for fight scenes than Vostaert does. Put in other terms: of every hundred words, Vostaert devotes less than four words more to fighting than Penninc does, which is not a very notable difference. If we look at the authors separately, Vostaert pays just a little less attention—1.4 words less in every hundred words—to 'court' than to 'sword', while Penninc uses 3.0 words less in every hundred words for 'sword' than for 'court'. The overall differences are minimal, though, and do not confirm the significant differences that van Es thought he detected.

Because we wanted to know a little bit more about the way Penninc and Vostaert described court scenes and fight scenes (can we again say something about the possible elaborateness of Penninc compared with Vostaert?), we used Yule's *K* to measure the lexical richness of court and fight scenes in both parts of the text. This led to the bar charts shown in Fig. 17.

It is certainly interesting to see that Penninc's court scenes contain less repetition and thus have a higher lexical richness than Vostaert's court scenes
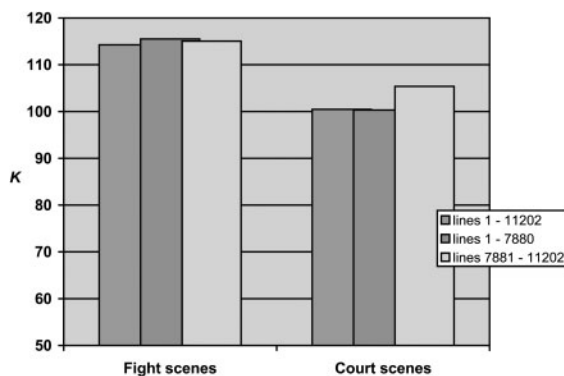
Fig. 17 Lexical richness (Yule's *K*) for fight scenes and for court scenes for the parts of the text before and after line 7,880 and for the complete text. Darkest bars: first part (Penninc), lightest bars: second part (Vostaert)

do. This may explain why van Es thought that he detected a larger attention for court scenes on Penninc's part. But we see a different picture when there is fighting involved: the lexical richness is more or less the same for both authors. This is contrary to what van Es thought he had seen. When we look at the bars representing Vostaert's text part, we even see that his work shows more word variation in court scenes than in fight scenes. This is highly unexpected, seen from van Es's point of view.

Both the subject of dialogue and of a possible difference in preference for court scenes or fight scenes need to be investigated further. For the goal we have set ourselves for this article, however, the results of our measurements seem to show that van Es's intuitions proved to be difficult to support with statistically significant measurement results. Above all, the elements we looked at gave us no help in finding out in what exact ways Penninc and Vostaert differ: they do not seem to explain the majority of the differences in vocabulary concerning pronouns and modal verbs. In addition, the results of the measurements do not provide clues regarding in which direction to look for them.[12] We, therefore, returned to Burrows's Delta, hoping to discover as to which part of the vocabulary of both the authors reveals most effectively how the two differ.

5 A Delta Procedure Analysis of *Walewein*

Delta is primarily an author-distinguishing measure based on differences in the usage of high frequency words by authors. Thus, it is to be expected that two texts written by two different authors will show differing values for Delta. Because *Walewein* has always been assumed to have been written by two authors, we predicted that we would find two roughly stable but distinctly different values for the parts of the text written by each of the two authors. If we were to plot a graph depicting the value of Delta for each text window of, for example, a thousand lines at any given point in the text of *Walewein* (comparing it with part of Vostaert's text), we would expect it to look something like the graph in Fig. 18. Thus, we would expect to find two

consistent Delta values that would slide over into each other around the line where we suspect the change of authors took place (i.e. at line 7,880).

Indeed this is what we found as a result of our preliminary experiments described in Section 3 (cf. Fig. 10). In his critical examination of Burrows's Delta, David Hoover shows that Delta is not only quite apt at distinguishing authorship, but that taking into account an increased number of high-frequency words, it does the job even better (Hoover, 2004). Indeed, repeated tests with an ever larger number of high-frequency words (up to the 150th ranking word) showed that the graph in Fig. 10 became increasingly clear-cut, with two increasingly stable values for Delta and the expected drop of the Delta value stable around line 7,880.

As pointed out in the previous section, however, we are interested in the Delta procedure as a means not only of distinguishing between authors, but also of telling us more about the qualitative differences between the vocabulary of Penninc and Vostaert. For that reason, we wanted to know which *part* of the high-frequency vocabulary discriminated most clearly between both the authors, or in other words: could Delta perhaps guide us to the set of high-frequency words that most clearly set Penninc apart from Vostaert? Supposedly the study of that part of the vocabulary of both the authors would reveal content-related, thematic, and/or stylistic differences between the work and methods of Penninc and Vostaert.
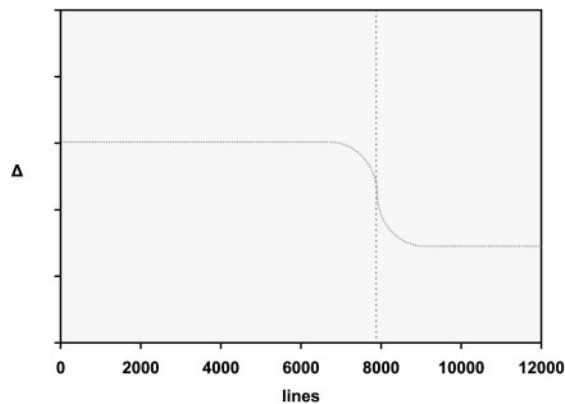
Hence, we decided to run a series of tests in which we used consecutive subsets of the high-frequency vocabulary to calculate Delta. We hypothesized that the Delta graph for each subset would closely resemble the overall view found in Fig. 10, and we expected that the graph for the most discriminating subset would have the largest drop between otherwise rather stable Delta values for Penninc's part on the one hand and Vostaert's part on the other.

First, we computed a Delta graph based on only the fifty most frequent words (Fig. 19). Rather contrary to our expectations, the most prominent feature of this graph is the peak around line number 6,000. We quickly discerned that this peak coincides closely with the change of scribes that takes place around that line (cf. Section 2 of this article). Extensive application of the Delta procedure to other situations where a change of scribe is known could further test this finding.

Figure 20 shows the Delta graph we computed for the top 100 through 150 most frequently used words. We found that the measurements for the range of the 100 through 150 highest frequency words fitted closest with our predictions. We, therefore, think that a closer look at the lexicon that is defined by this frequency range will reveal the most telltale differences between Penninc and Vostaert. The quantitative and qualitative analysis of

Fig. 19 'Walking' Delta. The text corresponding to lines 8,000–11,000 in Vostaert's part of the romance was used as a subset for the procedure. From this subset the topmost fifty frequently used words were used for comparison with the text in a 'walking window' of 1,000 lines

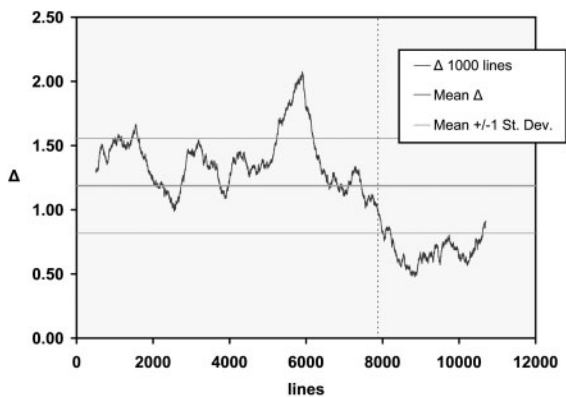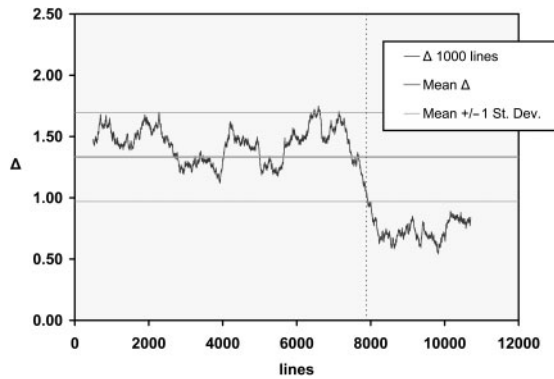**Fig. 18** Hypothesis for the 'walking' Delta graph for a window of 1,000 lines

Fig. 20 'Walking' Delta. The text corresponding to lines 8,000–11,000 in Vostaert's part of the romance was used as a subset for the procedure. From this subset the 100–150 ranking frequently used words were used for comparison with the text in a 'walking window' of 1,000 lines
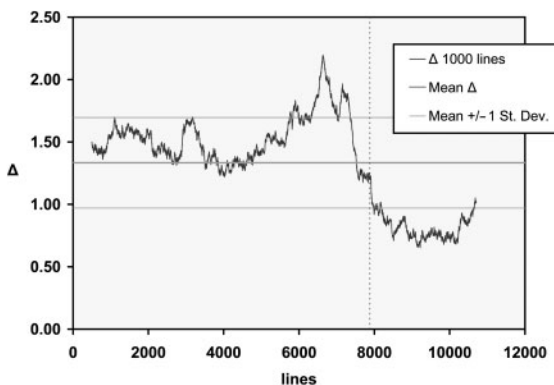

Fig. 21 'Walking' Delta. The text corresponding to lines 8,000–11,000 in Vostaert's part of the romance was used as a subset for the procedure. From this subset the 50–100 ranking frequently used words were used for comparison with the text in a 'walking window' of 1,000 lines

this subset of the high-frequency vocabulary will be part of our further research.

By far the most intriguing graph was produced when we took the Delta measurement for the range of 50 through 100 highest frequencies (Fig. 21). Again we found high Delta values for Penninc and lower Delta values for Vostaert. The peak characterized as an indication for a change of scribe in the

first graph (Fig. 19) is again visible. Interestingly, a close look reveals a plateau just before line 7,880 (where Vostaert supposedly took over from Penninc). Here we find a rather constant value for Delta for about 400 lines. The value is almost exactly halfway between Penninc's and Vostaert's 'plateaus'. It is very tempting to think of this short halt in the freefall of Delta from Penninc's rather stable value to Vostaert's rather stable value as a part of the text that was originally written by Penninc and later adapted by Vostaert. It might be the case that we see Vostaert at work creating the transition from the already existing text into his own work.[13] As we are particularly interested in the way the styles of writing of both authors differ, it will be interesting to follow-up the Delta procedure analysis of the complete text of the *Walewein* with a closer analysis of the vocabulary of the parts that are undisputedly attributed to either Penninc or Vostaert. But given this intriguing result and for exactly the same reason it will be most interesting to see in what ways the vocabulary of this specific part differs from the parts adjoined to it.

6 Conclusions and Next Steps

The most important result of our experiments is that at least in *Walewein* we seem to be able to distinguish scribes as well as authors by using Burrows's Delta for different sets of word frequencies. Measurements of the group of fifty words with the highest frequencies clearly point to the paleographically visible change of scribe in *Walewein*. The frequencies from the range of 100–150 give the sharpest visualization of the change of authors. And perhaps we even can point out the area where the work of the first author was edited by the second one: the frequencies 50–100 result in a third level in the text, distinct from the copyist and the author change and at the place in the text between the two distinguishable authors. This seems a confirmation of van Es's suggestion that Vostaert edited the last lines of Penninc before starting out 'on his own', although the number of edited lines our measurements might indicate is much higher, namely about 400 rather than only eight lines.

As of yet, it is uncertain whether the measures are seemlessly applicable to other texts. We are currently planning new research projects to test the measures on other texts that were written by more than one copyist and/or more than one author, and on different Middle Dutch texts that were written by the same scribe. To give a few examples: we will select episodes from Jacob van Maerlant's Middle Dutch verse translation of the medieval Latin *Historia scholastica* and compare the vocabulary of these episodes in the fifteen different manuscripts that have come down to us. This may lead to more knowledge about what kind of lexical freedom different scribes allowed themselves in this genre—the next step would of course be to compare this with other kinds of texts. We are also looking for a test case that includes recognized exemplar changes so we may gain more insight in lexical consequences in subsequent copies of a text. With the results of these and of other experiments, we will return to *Walewein* and then also have a look at the two short fragments which survive from another manuscript (cf. Section 2). Because of the smallness of these fragments we did not include them in our measurements up till now.

On a methodological and technical level, we want to advance the application we have developed to compute Delta graphs for a single text. It would be very valuable if we were able to compute which subset of the vocabulary creates the most significant 'Delta contrast' between two parts of a text (in this case, between the parts supposedly written by different authors), as we think that this may shed more light on how the work of one author differs from that of another. However, although the current application is able to compute a Delta graph for a given text, it is not able to evaluate the general characteristics of such a graph, nor to statistically validate those general characteristics. In addition, the current application cannot easily and automatically generate and compare a large collection of Delta graphs based on an iterative subselection of a word frequency list. Both these capabilities are needed if we want to calculate more conclusively the most significant Delta contrast in a text. Furthering our method and technology in this direction poses an interesting challenge, as quite sophisticated statistical approaches will be needed, as well as some amount of grid computing given the large number of graphs and validations that would have to be computed.

It seems plausible that the application of Burrows's Delta to these sorts of questions can be used not only for texts in Middle Dutch, but also for younger Dutch texts and for texts in other languages. We hope that colleagues working in these areas will be interested in collaborating in this; that is, that they will be interested in testing the methods on other time periods and languages and, together with the present authors, in refining it for more general use.

A very important aspect of our work is the tools we have developed. We should like to provide the international research community with these tools, as they will be useful not only for the research of colleagues but also for the repeatability of our own work. This, after all, is a very important aspect of good research.

References

Burrows, J. F. (2002). 'Delta': a Measure of Stylistic Difference and a Guide to Likely Authorship. *Literary and Linguistic Computing*, **17**(3): 267–287.

Burrows, J. F. (2003). Questions of Authorship: Attribution and Beyond. *Computers and the Humanities*, **37**(1): 5–32.

Draak, A. M. E. (1936). *Onderzoekingen over de Roman van Walewein*. Haarlem (Reprint, 'Met aanvullend hoofdstuk over "Het Walewein onderzoek sinds 1936"', Groningen: Bouma's Boekhuis/Amsterdam: Bert Hagen, 1975).

Duinhoven, A. M. (2001). De bron van Walewein. *Nederlandse letterkunde*, **6**: 33–70.

Hoover, D. L. (2004). Testing Burrows's Delta. *Literary and Linguistic Computing*, **19**(4): 453–475.

Johnson, D. F. and Claassens, G. H. M. (2000). *Dutch Romances I. Roman van Walewein*. Cambridge: D. S. Brewer (Arthurian Archives, VI).

Jonckbloet, W. J. A. (1846–1848). *Roman van Walewein door Penninc en Pieter Vostaert*, **2**: Leiden: D. du Mortier en zoon (Werken uitgegeven door de vereeniging ter bevordering der oude Nederlandsche letterkunde, 3 and 5.).

Kuiper, W. T. J. M. (1989). *Die riddere metten witten scilde. Oorsprong, overlevering en auteurschap van de Middelnederlandse Ferguut, gevolgd door een diplomatische editie en een diplomatisch glossarium*. Amsterdam: Schiphouwer and Brinkman.

Sonnemans, G. H. P. (1995). *Functionele Aspecten van Middelnederlandse Versprologen*, 2: S.l.: s.n..

Tweedie, F. J. and Baayen, H. (1998). How variable may a constant be? Measures of lexical richness in perspective. *Computers and the Humanities*, **32**: 323–352.

Van Dalen-Oskam, K. and van Zundert, J. (2004). Modelling features of characters. Some digital ways to look at names in literary texts. *Literary and Linguistic Computing*, **19**(3): 289–301.

Van Dalen-Oskam, K. and van Zundert, J. (2005). De list van het lexicon. Auteursonderscheiding met behulp van computer-ondersteunde woordenschatanalyse. *Nederlandse Letterkunde*, **10**: 212–233.

Van Es, G. A. (1957). *De jeeste van Walewein en het schaakbord van Penninc en Pieter Vostaert*, 2: Zwolle.

Yule, G. U. (1944). *The Statistical Study of Literary Vocabulary*. Cambridge.

Notes

1 This article is based on, and builds upon, the papers the authors read at the ALLC/ACH joint conferences in 2005 (van Dalen-Oskam and van Zundert, *Mining the Differences between Penninc and Vostaert*) and 2004 (van Dalen-Oskam, *Penninc versus Vostaert: Contrasting co-authors by Means of Authorship Attribution Techniques*), at the XXI International Arthurian Conference in 2005 (van Dalen-Oskam and van Zundert, *The Voice of the Sword or the Voice of the Court?*) and at the eleventh International Medieval Congress in Leeds in 2004 (van Dalen-Oskam, *Contrasting Themes in Penninc and Vostaert's* Roman van Walewein).

2 Van Es adapted the use of u/v and i/j to modern Dutch usage and modernized interpunction and capital initials. The minor editorial changes in the text he opted for are listed and accounted for in the edition. The type of edition did not interfere with the results of our measurement because of our choice to lemmatize the text, cf. Section 3.

3 Cf. Prologue, lines 23–27 (in the translation of Johnson and Claassens, who present their English translation on the opposite pages of their edition of the Middle Dutch text): 'Thus prayed Penninc, who made this book. And many a night he sat sleepless over it before he considered in his mind that he had completed the book upon which he had started' (Johnson and Claassens, 2000, p. 29).

4 Cf. Epilogue, lines 11,173–11,188: 'Penninc, who composed this book – I do not care who hears it – was unwise not to have brought it to an end. Pieter Vostaert continued it as best he could according to the text that he found which Penninc left behind: it seemed to him a pity for it to be left unfinished, but seeing as how it lacked a conclusion, I believe little credit would have been accorded the poet; what is more without an ending all of his pains and effort would not have been worth a cherry. Approximately thirty-three hundred verses were composed by Pieter as skilfully as he could, in simple rhyme.' (Johnson and Claassens, 2000, p. 495)

5 Van Es, 1957, 2: pp. 342–3. Only implicitly mentioned in Sonnemans, 1995, 1: pp. 110–2.

6 The two other options would be that Vostaert included the prologue but not the epilogue in his 3,300 and that he included the epilogue but not the prologue, but these options do not extend the area as aforementioned.

7 We reported on our first experiments in van Dalen-Oskam and van Zundert, 2005.

8 Another benefit of using lemmatized texts is that the results of any measurements can easily be checked against the raw text. If only the raw texts are tested, however, any further checking requires additional normalization and analysis.

9 Cf. van Dalen-Oskam and van Zundert, 2004.

10 Cf. Yule, 1944. This specific form of the formula of K is taken from Tweedie and Baayen, 1998, p. 330 (16).

11 We are aware of the fact that we are comparing parts of the same texts. Because we are fairly certain that this text is written by two authors, we think it is valid to make these comparisons. In theory, it is possible that the graph reaches the horizontal axis at a point where this text part is compared with itself. To check whether this could be the case, we plan to experiment with e.g. removing sections from the complete text in a next phase of our research.

12 We measured Delta for the dialogue and non-dialogue text in *Walewein* with a sliding window, but this did not result in graphs that were in any way insightful or significant with regard to differences in the use of high-frequency words by either Penninc or Vostaert in dialogue or non-dialogue.

13 How exactly this process came about will have to be the subject of further analysis, study, and debate. It could be that Vostaert initially tried to model his style of writing on what his predecessor had written. But as he progressed, he may have given up on the idea of imitating, and fell back on his own style of writing. Alternatively, Vostaert might have deliberately or rather intuitively tried to create a smooth transition from the already existing text into his own work, by adapting the latter part of Penninc's text. For now any such qualitative interpretations of this intriguing part remain speculative.