CHAPTER 2

In Praise of the Variant Analysis Tool: A Computational Approach to Medieval Literature

Karina van Dalen-Oskam

Introduction

When I first read Bernard Cerquiglini’s Éloge de la variante (1989), one of the first and most influential publications related to the New Philological “movement,” I was working as a lexicographer on the Dictionary of Early Middle Dutch (1200–1300). The year was 1993. Our dictionary project was one of the first to be fully computer-assisted. We had created a digital corpus of literary and non-literary texts enriched with lemmas and part-of-speech tags. By means of an intricate system that enabled us to query the corpus and to sort, classify and mark the exact texts to use as quotations in our dictionary entry, we could input our lexicographical descriptions into a digital form from which the actual dictionary could be generated in unlimited kinds of visualizations. Our project had been running for four years, and for several years we had put our year’s work of entries before a committee of wise scholars for comment. Our lexicographical and linguistics peers in The Netherlands and Flanders gave us much support and useful criticism. We had a difficult time, however, in explaining two of our most fundamental methodological choices to our literary colleagues: our corpus consisted only of those Middle Dutch manuscripts of which it was certain that they were written in the Thirteenth Century, and we only used the texts of these manuscripts in a diplomatic edition. Literary scholars did not understand why the most famous of our Middle Dutch literary texts, of which it was certain that they were originally written in the Thirteenth Century, were excluded from our corpus only because their manuscripts were from the Fourteenth Century – much of the lexicon must still have been Thirteenth Century. We explained that if you want to find out how much of the lexicon is indeed from the Thirteenth Century, you have to be able to compare it to a reference corpus that was as rigidly selected as ours was. If such a Fourteenth Century copy was included in the Thirteenth Century corpus, the corpus would be polluted and would exclude the possibility of a comparison. And as to our fanatical adherence to diplomatic texts: we explained that if you want to know which spellings and variants Thirteenth Century scribes used in which
geographical area or in which decade (for example), you again needed editions reproducing the exact spellings and variants as represented in the manuscripts. Therefore, diplomatic editions are the most suitable, as they mark resolved abbreviations with italics and do not adapt spelling, punctuation and case use. We also preferred editions in which emendations were not applied to the main text, but were suggested by the editors in (e.g.) footnotes, in order to maintain a purely Thirteenth Century text which would give the scribes due credit for having presented a linguistically correct text. We found it difficult to explain to our literary peers that these choices were necessary to guarantee a good, verifiable dictionary, which in its turn could be used by scholars editing a text in the correct manner, editions which in their turn would be trustworthy sources for even better dictionaries in the future.¹

As a matter of fact Cerquiglini, in his Éloge de la variante (only translated into English in 1999, under the title In praise of the variant), eloquently described the philological side of what we were doing in our lexicographical project. He proposed to abandon the old approach to medieval text and text editions, and instead to apply a New Philological perspective to medieval texts. Instead of looking at variation as the result of alterations which polluted and degraded “the original,” each copy of a text (with all its partly unique, partly shared variants) should be seen as a new and unique text, functioning in its own, unique way, in its own, unique context. Each copy should be researched based upon this approach.

I was elated. One of my lexicographical colleagues, Katrien Depuydt, and I wrote a paper describing our methodological approach, with reference to Cerquiglini’s book. It was presented at the first International Medieval Congress at the University of Leeds in 1994, published in a volume of proceedings and ignored by our literary peers.²

The Role of the Computer: Cerquiglini’s Predictions

In his 1989 book, Cerquiglini admitted that the methodological changes he proposed would not be easy to apply in practice. He expected that the computer would play a very important role in leading us back to medieval textual variance. It is now 2012, and what I propose to do in this paper is look back at

---

¹ A point well understood, however, by Masters (1991) 284. Masters’ article – a publication I found thanks to an explicit reference in Busby (1993) – takes a comparable stand to Cerquiglini in his Éloge de la variante.

² Van Dalen-Oskam and Depuydt (1997).
Cerquiglini’s specific predictions and to show where his words concur with current practice, and where his vision and our current expectations differ. In the latter case, I will show how current, as well as anticipated, technical developments may influence editorial practices in the near future. In this, I will pay special attention to the discipline of Stemmatology: the building of family trees of manuscripts. This discipline has an intriguing status when we consider it from Cerquiglini’s point of view. In the “old” days, a stemma was one of the first things an editor would establish. Building a stemma helped to find out which of the existing manuscripts of a text seemed to be the closest representative of the original text by the original author. This would usually become the manuscript to be edited, while the variants in other manuscripts could be either ignored by the editor as unimportant, or mentioned in the apparatus, or chosen as a better reading than the one in the manuscript being edited and promoted/elevated into the edition of the text. From the New Philological point of view, however, the building of the stemma is not necessary at the beginning of the edition project – we merely want to have a transcription of all manuscripts. So why has Stemmatology gone through such a big technical development on the one hand, while on the other hand most text editors have ignored this new potential? I will end my paper with a short reflection on this issue.

To show exactly how Cerquiglini thought about the role of the computer in 1989, I will quote rather extensively from the last pages of his book. Here, he states that the computer:

allows the reader to see and consult not only the totality of the manuscripts of a particular medieval work but also the editions (empirical, Lachmannian, Bédierist, etc.) which took these manuscripts as their objects. Moreover, it can provide a great many minor bits of information, which should remain virtual so they will not get in the way of reading but which one needs to be able to locate: makers of concordances, frequency lists, tables of rhymes, every sort of calculation, codicological and paleographic data, and so on – everything that a printed edition usually abandons or from which it makes a painful choice, everything that the hyper-scholarly edition hypostatizes to the point of unreadability.3

Cerquiglini further states that:

---

3 Cerquiglini (1999) 79–80.
the computer is able to help us detect the dynamics of the text by making visible the connections prepared and suggested by the editor. It is less a question, therefore, of providing data than of making the reader grasp this interaction of redundancy and recurrence, repetition and change, which medieval writing consists of – and to do so according to the two axes that we have brought to light. Vertically, along the thread that leads through the work, it can bring back all the things that each noteworthy utterance constantly echoes but which modern memory no longer hears; the screen unrolls the infinity of memorable context. Horizontally, it can compare the utterances within a persistent and chosen range of variants that are paraphrases of one another from one manuscript to the next, even indicating by some symbol or note what the characteristics of this relationship are.4

Cerquiglini knew very well what all this would mean:

That truly would be publishing on a grand scale, editing on a scale never before realized, yet indispensable, and only the information technology of today can provide us with the means, probably even the idea, of doing so. Because the computer, through its dialogic and multidimensional screen, simulates the endless and joyful mobility of medieval writing as it restores to its reader the astounding faculty of memory – the memory that defines its aesthetic reception and is basic to the pleasure taken by a reader.5

The Role of the Computer: Realized Potential

Only now, more than 20 years after Éloge de la variante, is the type of digital text editions Cerquiglini had in mind slowly starting to emerge. Peter Robinson has initiated several important digital editions, such as in the Canterbury Tales project I will go into more deeply in the last section of this contribution and the very impressive Codex Sinaiticus website.6 So too Huygens ING and the Van Gogh Museum have published the widely acclaimed website Vincent van Gogh: The Letters,7 and Barbara Bordalejo has published the Online Variorum

4 Cerquiglini (1999) 80.
5 Cerquiglini (1999) 80–81.
6 Cf. http://codexsinaiticus.org/en/
7 Cf. http://www.vangoghletters.org/vg/
of Darwin’s *Origin of Species*. These are only a few of a new generation of (digital) editions. Much work has been done on a more theoretical level as well, by scholars such as Peter Shillingsburg and Peter Robinson who both contributed to an interesting volume of articles entitled *Text and genre in reconstruction* on the subject of, respectively: “Digital editions for everyone” (Robinson), sketching a new way of involving many more people in editing than scholars alone, and: “How literary works exist: implied, represented, and interpreted” (Shillingsburg), which, amongst other things, deals with several key theoretical and technical aspects of digital editing. Elena Pierazzo gives a solid description of the state of the art of digital editing in her article “A rationale of digital documentary editions.” For an overview of the history of digital editions before 2010, Edward Vanhoutte’s article “Defining electronic editions: a historical and functional perspective” is extremely useful. Joris van Zundert and Peter Boot look ahead at the possible task distribution between scholarly editors and digital libraries in their paper “The Digital Edition 2.0 and the Digital Library: Services, not Resources.” And these are only a few of the important publications of the last couple of years. This is not the place to give an elaborate overview of the current state of the art. My aim here is, before I go into the reasons for the mentioned time lapse between Cerquiglini’s ideas and the emergence of the digital edition more deeply, to present some digital editions with which I am familiar in more detail, to sketch the ways in which current digital editions fulfil some of Cerquiglini’s predictions, and diverge from others. The first two examples are online digital editions published by Huygens ING. In dealing with these two editions I would advise the reader to open a browser on a computer, go to the sites described, and to follow my instructions. This will bring about a much clearer grasp of new potentialities than pages and pages of words could achieve. The third example is from a digital edition on CD-ROM made by Andrea de Leeuw van Weenen.

Example 1: Cerquiglini imagined the parallel presentation of different witnesses on screen. An example of this is the online text edition of an essay by Lodewijk van Deyssel, *Menschen en bergen*. Go to the navigation bar on the left and choose “Bekijk de editie” [view the edition]. Under the button “Versies en varianten” the user can choose from an edition of all different versions of the essay. Under “Versies vergelijken” [compare versions], the user can select a

---

8 Cf. http://darwin-online.org.uk/Variorum/index.html
9 Robinson (2010), Shillingsburg (2010).
10 Pierazzo (2011).
11 Vanhoutte (2010).
12 Cf. http://menschenenbergen.huygens.knaw.nl/path
part of the text, and decide which of the versions of that part he or she wishes to compare.

Here (Figure 2.1), we compare the text in the manuscript (left text column) with the first printed edition (right text column). Additions are marked with a green mark and are underlined; replacements with a purple mark.

Example 2: How many types of annotations can be shown without having them interfere with the reading of the text? Cerquiglini saw here the possibility of an infinite amount of information. For this, we go to the online edition of a Middle Dutch Arthurian romance, Walewein ende Keye [Gawain and Kay].

We open “11. De editie” in the navigation column on the left, then choose “Overzicht per kolom, met afbeeldingen” and go to 178r and 178ra. When we click on this, several columns are opened: a scan of the manuscript column, the transcription, and the annotations. There are three types of annotations in this edition, each marked with a different colour in the transcription. But the user can choose which ones to see and which ones to ignore by using the Options button at the top of the screen to the right (Figure 2.2). Unchecking all the boxes will result in a very simple and clear visualization of the text. It is also possible to show or not to show the italics in the text that denote abbreviations, the line numbers or a reference to the line numbers in an earlier edition.

13 http://www.waleweinendekeye.huygens.knaw.nl/path

Figure 2.1 Text version comparison in the Menschen en bergen online edition
In Praise Of The Variant Analysis Tool

of this text. Note that the user can also click away any of the columns he or she does not need, e.g. to give the scan of the manuscript and the transcription more space on the screen.

Example 3: Cerquiglini also mentioned the desirability of having concordances, indexes and linguistic information at hand in the digital edition. The two online editions presented above have simple search options that yield concordances of the hits. At the moment, this is about the extent of what we can do, but we want to do more to realize Cerquiglini’s idea. I will give an example from a recent edition of the Old-Icelandic *Alexanders saga* on CD-ROM by Andrea de Leeuw van Weenen, to show one of the things that I would like to achieve in the future.\footnote{De Leeuw van Weenen (2009).} In this edition, when the user moves the mouse over a word, a small box shows the most linguistically relevant information: the form of the word in the facsimile edition, the normalized word form, the lemma (dictionary entry), the linguistic form and function (e.g. present tense, third person singular for a verb form), and the page and line of the manuscript where the word occurs. This is exactly the information, by the way, with which we had enriched the corpus of Thirteenth Century Dutch on which the *Dictionary of Early Middle Dutch* was based.
Why Did It Take So Long to Publish this Kind of Edition?

The grand scale of Cerquiglini's ideas was recognized by other medievalists to require an enormous amount of work. Keith Busby expressed this most clearly in a reaction to Cerquiglini's book:

"Consider what would be involved in Cerquiglini's ideal edition: not just transcription of the texts, but the coding of abbreviations, different types of letters (for example, long and round s), majuscules and minuscules, varieties of capitals and other forms of decoration, word-division, line-division, enclisis, proclisis, scribal punctuation, etc. And because of considerable orthographical variation both between and within manuscripts, transcriptions would have to be uniformized in order to permit many kinds of linguistic analysis. For anything other than a short text extant in a small number of copies, the difficulties involved and the amount of time that would be required are enormous.15"

15 Busby (1993) 42.
Busby was quite right: each particular type of information and every category that can be made visible or clicked away in the examples I have presented above is encoded with special tags and this took the editors a great deal of time. In 1994, the TEI (Text Encoding Initiative) encoding scheme for textual scholarship started to be used in various projects. Editors needed to learn this markup language to make the kind of digital editions Cerquiglini had in mind. Encoding a text in TEI takes an enormous amount of work, and in practice is not done very often, or with only a very limited Tag Set. The example I have given of the essay Menschen en bergen is based on material that was encoded in this way. Most scholars deemed the learning curve for TEI to be too steep, or they did not see any immediate benefits, and decided not to spend time on learning and using it. They continued to make their editions in the traditional way, using the computer merely as a word processor, and using commercial software such as Word or WordPerfect. This resulted in traditional printed editions. The Word or WordPerfect files used for these editions are extremely hard to convert into files that can be used for a digital edition, because one needs not only to change the format of the files into an accurately tagged format, but also to think of the edition in a different way and use different types of tags to those required for a printed edition. Even in digital form, these editions will still look and feel like printed ones. As Cerquiglini so clearly perceived: print text editions are held back by the technique of book printing, and traditional choices in text editing are inspired by the technical possibilities and limitations of printing and reading printed text. The computer has a different set of possibilities and limitations, and for most of these limitations the computer also has a solution (e.g. the clicking away of information that is not useful to the user at this moment, but is available if needed later).

In recent years, several tools have been developed which have enabled the creation of a digital edition in an online environment. A good example is the tool for the crowd-sourcing project of the collected works of Jeremy Bentham (Causer et al. 2012). This tool requires the user to apply a simple encoding scheme. Huygens ING has taken another approach in the online work environment it has developed for transcribing and editing texts online, alone or with a “crowd” of volunteers. eLaborate gives textual scholars the possibility to create their editions online without the need of learning an encoding schema or a markup language. Scholars can upload scans, transcribe the text, and annotate their transcriptions with exactly the annotation categories they deem relevant for the text they edit. They can work on an edition together with colleagues all over the World: one only needs a computer with access to the Internet and a good browser (i.e. Mozilla Firefox). The text edition, as input in the online work environment, is very easily converted by our programmers to an online
digital edition, which can be updated any time the editor chooses to do so. The Walewein ende Keye edition I demonstrated earlier is an example of this way of working. This approach has proved to be much more user-friendly than TEI, and it makes the entire organisation of a large and complex edition project a lot more manageable, as became apparent when a number of users and moderators were interviewed. The eLaborate environment also offers the opportunity to add the transcription of (for example) other manuscripts or other editions at a later time, and to subsequently enrich the existing edition with a new set of annotations, e.g. for a very specific research question.

This way of working inspires textual scholars to rethink the traditional print-bound choices for a text edition by working in a digital environment and experiencing what “works” and what does not “work” in terms of creating a successful, usable, edition. Editors working in Word or WordPerfect have a much harder time exploring technical possibilities and demands (and are furthermore “imprisoned” in proprietary software which limits their freedom considerably). They are dependent on existing technology (book printing; perhaps pdf). Even using new technology that is in its infancy generates new ideas and new technological demands – these demands are, of course, always far ahead of what can be done at the present time. This can be frustrating, but it is the way forward: new potential will not be developed if there is no call for it or if nobody dares to look to the future. So using technology itself generates ideas about which technologies we might want to develop. This is the principal reason why it took so long to publish the kind of edition Cerquiglini had in mind. Only now have easily usable tools become available to textual scholars, inspiring them to rethink the traditional print edition without the need to learn a totally new tool (everyone can use a browser nowadays!). On the other hand, technical developments happen very fast. Ideas about how to prepare digital editions are, equally in a state of constant flux. We are slowly working towards the development of international technical standards that will hopefully lead to a surge in the publication of digital editions in the next decade. One of the challenges to be overcome is the need to make tools available for different platforms, for which the “microservices” as developed for the Interedition project seem promising. Also the architecture of the data and the way of storing annotations are non-trivial technical puzzles that still need an adequate solution.

---

16 Beaulieu, Van Dalen-Oskam and Van Zundert (2013).
17 Van Zundert (2012).
18 Van Dalen-Oskam and Van Zundert (2012).
The Next Step: A Computational Way of Looking at Variance

What the user of the digital editions I referred to cannot do as yet is to compare the edited versions of a text when they are not enriched with (underlying, usually not visible) special tagging. So Busby’s fear of having to manually add myriads of links etc. is still very valid. But there is light at the end of this tunnel; because tools have been and are being developed that can help the scholar who performs this kind of research and tagging. They will also have to be included for use in the digital editions themselves, which will not be easy, but certainly can be done. To give a better idea of the kind of text analysis that I anticipate to be achievable within editions in the future (not possible as yet) I will present a couple of examples from my own research into the fifteen manuscripts of Jacob van Maerlant’s *Scolastica* or *Rijmbijbel* (“Rhyming Bible,” a Middle Dutch translation/adaptation of Peter Comestor’s *Historia Scholastica*) written in 1271. A more elaborate description of this research has been published in *LLC: The Journal of Digital Scholarship in the Humanities* (Van Dalen-Oskam, 2012).

The fifteen manuscripts, dating from ca. 1285 until the end of the Fifteenth Century, have all functioned in their own unique context: they differ as to their geographical location, patrons, the cultural environment of the patron as well as of the scribe, and in the time period in which they were copied. What I want to know is the extent to which the copies differ, and I especially want to know how widely they differ on a content level and whether we can find out why the content is different. For the moment, I am not very interested in a family tree, a stemma, because I do not expect this to tell me what I want to know.

The first thing I want to do is to find out which of the copies seem to be broadly similar and which of them stand apart, but I do not have the time to personally evaluate all of the words and lines. The hypothesis, therefore, is that: if it is possible to have a computer calculate the vocabulary of, say, one episode in all of the fifteen copies, it should be very easy to observe any deviations. Imagine for example fifteen small bags, one for the same episode in each manuscript, with all the tokens (individual occurrences of words) of this episode in it. A bag representing a copy that has a longer or shorter text will be bigger or smaller, so visibly different. And since we are looking at copies, any visible difference in vocabulary or word frequencies is interesting. I want to make graphs showing the degree of similarity between the copies (so-called

---

19 Cf. for more information about usability of tools Van Zundert (2012).
20 The oldest manuscript was edited diplomatically by M. Gysseling (1983). A list of the more or less complete manuscripts is given in Postma (1991).
cluster analysis) and, one step deeper into the statistics, show the copies that differ from the others on statistical components (principal components analysis). The results of these graphs will then point me towards those manuscripts that promise to be most interesting for further analysis. 21

My example here is the apocryphal story about Judith, who killed the general of an alien army, Holofernes, when it beleaguered the city in which she lived. Since Middle Dutch did not yet have a fixed spelling, the copies contain many different spellings, even within the same copy, which can hardly ever be linked to lexical differences between dialects. If I were to do my measurements on the diplomatic edition of the text, the resulting graphs would be expected to show me the amount of similarities and differences on the level of spelling (I will not present the graphs here, but rest assured that they look very different from the ones shown below). In the tradition of corpus linguistics and computational lexicography, I tagged all fifteen episodes concerning Judith with lemmas, representing the word form that would be the dictionary entry if you were to look it up in a dictionary, and a code describing the part of speech. This is standard corpus procedure, but not often practised with respect to text editions. Using a statistical package called Minitab to calculate the similarities between the 250 highest frequency lemmas and their frequencies in the fifteen episodes results in the graph in Figure 2.4.

This cluster observation shows that fourteen of the fifteen manuscripts are calculated to be relatively similar, with only one, “I,” deviating significantly. When we read the Judith episode in this manuscript, it immediately becomes clear that the manuscript presents a much longer text about Judith, with much rephrasing, and even many of the rhyme words do not agree with the other fourteen manuscripts. This episode clearly has to be seen as a new text, with a new author (the scribe (or the scribe of his/her exemplar (or the etc.))).

Another multivariate measurement is principal components analysis. This shows the manuscript in a scatter plot, in which the horizontal axis shows the component on which most of the difference between the placed manuscripts is plotted, and the vertical axis, showing the second most important component (more components can be calculated and shown, but I will keep it to two). The Judith episodes now show up as in Figure 2.5.

In the graph in Figure 2.5 we see manuscript I differing extremely on the first component, but there is also another manuscript outside of the group on the second component, but in a different way: E. Reading the Judith episode in

21 The first results of this analysis were published in Kestemont and Van Dalen-Oskam (2009). The research into scribes was triggered by the results of earlier research published in Van Dalen-Oskam and Van Zundert (2007).
Figure 2.4  Cluster analysis made with Minitab 15 of the Judith episode in all fifteen manuscripts of the Rijmbijbel, for the 250 highest frequency lemmas.

Figure 2.5  Principal components analysis made with Minitab 15 of the Judith-episode in all fifteen manuscripts of the Rijmbijbel, for the 250 highest frequency lemmas.
E, we find that this scribe (or the scribe of his/her exemplar (or the etc.)) regularly rewrites phrases in the text which could be seen as difficult. The scribe, however, “corrects” the text in such a way that it loses the original meaning – which would be the meaning as we can find it in the Medieval Latin source text by Peter Comestor and the Medieval Latin Bible. These micro-changes are expected to be partly responsible for E’s place outside the group, but since the multivariate, top-down approach shows us an overview picture, further bottom-up research is needed to find out more about the details. This graph, therefore, leads us to also take a closer look at E, which we might not have done when approaching the texts through close reading only. Applying these tools yields more information about the possible next steps to be taken for further research than close reading would do, which is one of the reasons why using them is so fruitful (more about this “secret life of scribes” in Van Dalen-Oskam, 2012). Furthermore, manuscript I differs so much from the fourteen others, that the statistics may be severely distorted. If we remove manuscript I from the selection and conduct the same measurements on the fourteen others, we can gain a better grasp of the relative differences between the other manuscripts for this episode. There are still other manuscripts which draw our attention other than E: namely manuscripts H and C (Figure 2.6).

I have looked at five different episodes from the *Rijmbijbel* and all of them show different clusters and different patterns. This indicates that the study of scribal variation has to be dealt with on an episodic level, taking the content of the story into account as well as the context of the scribe, patron, location and date of copying. My next research into the *Rijmbijbel* manuscripts will focus on manuscript I and will seek to address the following questions: from where it originates, whether or not we can find out who copied it and for whom, and why Judith was a special figure for the scribe or the patron.

Once it is possible to offer statistical tools such as the one used to produce these graphs to users of a digital text edition of parallel texts, this will influence editorial choices in the making of the edition in a number of ways. Instead of investing time in listing variants or linking to variants, time would be better spent on e.g. providing a lemma and part-of-speech tag for all words. Tools to lemmatize historical Dutch will hopefully be available shortly. So we should, I think, provide a combination of complete transcriptions of all copies and smart computer tools that can be applied repeatedly. The scholarly advantage is that the base texts can be kept as “clean” as possible from interpretations by the editors, and can be reused for each new tool that becomes available. The “old” way of selecting variants and promoting them into the edition in fact closes the edition down for certain types of research. For example: a critical edition (the Dutch term for a transcription with adapted spelling, added
punctuation, capitalization, and the occasional emendation) is unsuitable for spelling research. But if we always provide the base (diplomatic) edition next to it, the edition as a whole is reusable for each new research question. So in this new digital context, the edition can have a much more durable and much wider value for all types of users, making a good investment of the (public) money spent on it. Printed editions are in fact not reusable: if a critical edition should need to be adapted for e.g. spelling research and changed into a diplomatic one, it would have to be completely redone. Digital diplomatic editions can easily be changed into, or supplemented with, a critical edition. Or it would be very easy to generate a version aimed at a wider audience which then could be presented on a separate website if required. But the basic digital diplomatic edition can also be reused by adding other information layers to the edition for a relatively small amount of funding, since one can build on what is already available instead of having to start all over again. Since the amount of work to be done is still extremely large, any way to reuse available information without the investment of extra time is worthwhile from a practical (and financial) point of view. In this way, the knowledge concerning a text can really grow, instead of having to be rebuilt from scratch in each new edition (or never getting off the ground in the first place).

Figure 2.6 Principal components analysis made with Minitab 15 of the Judith-episode in all fifteen manuscripts excluding manuscript I of the Rijmbijbel, for the 250 highest frequency lemmas.
Another major scientific advantage of this approach is also apparent: the combination of a digital working environment including transcription and annotation tools as well as text analysis tools will enable the user to repeat and verify measurements and visualizations. An edition of one text and a list of variants in other versions would not have this option. And if more smart tools are developed, we need not change our text edition: the sound and trustworthy basis will also be there for new approaches, which in turn will be easily verifiable by other scholars. My expectation is that this approach will open up a whole range of new research questions and that research into medieval text transmission will get an enormous boost, bringing results into our reach that we could never have achieved before. Also, a new task division will arise: textual scholars of all kinds will focus on the texts themselves; their transcription and annotation; their interpretation and which other questions to ask; and ICT developers will create new tools to further this interpretation on a scientific basis. Together they will create a new form of Computational Humanities – in which researchers from the “hard” sciences are eager to participate.

Stemmatology: Back to the Future?

At this point, I want to return to Stemmatology, by way of Busby’s 1993 reaction to Cerquiglini’s *Éloge de la variante*. Earlier I quoted Busby’s worries about the impracticalities of Cerquiglini’s ideas. In a footnote, he adds:

I have become less sceptical since having written this article of the possibilities of the Hypertext edition, especially in view of the remarkable tool that is Peter Robinson’s *Collate*. Two problems now present themselves, one purely material (the equipment required is expensive), and the other of attitude (how do we convince our colleagues to use the computer for something other than basic word-processing?).

The computer program *Collate* which Busby mentions, was developed by Peter Robinson for the collation, analysis and publication of texts preserved in multiple witnesses. It had to be installed on the scholars’ personal computers, not being available online. It was used, among others, for Robinson’s own Canterbury Tales Project, which aims to transcribe, collate and analyse all the Fifteenth Century witnesses of Chaucer’s *Canterbury Tales*. Apart from preparing the digital editions of all witnesses, the project aims to explore the

---

22 Busby (1993) 42–43, in footnote 41; in square brackets.
relationships between the witnesses. For this, it uses software originally developed for evolutionary biology (PAUP – Phylogenetic Analysis Using Parsimony and SplitTrees).23 The results of applying this method are stunning: beautiful “trees,” appearing where the different manuscripts could be located in the family tree of manuscripts, even indicating the distance a manuscript seems to have from its “parent” in the tree. It is clear that Robinson’s approach includes a great deal from New Philology: all manuscripts are to be edited, and a tool (Collate) is applied to tag variance, which will result in family trees that help to direct scholars to interesting manuscripts or parts of manuscripts and interesting methodological problems such as exemplar change and contamination.24

So in the chain of tasks of the textual scholar, the stemmas are now based on the transcription of all the text witnesses, and thus come after the editing is done, at the end, leading to new inroads and insights, instead of at the beginning, leading to the choice of one manuscript to be edited with a limited list of variants, and to a process of closing down texts for research instead of opening them up.

As I stated in my introduction, the impressive technical and methodological developments in Stemmatology have not been embraced with a great deal of enthusiasm by textual scholars. This is closely linked, in my perception, to the fact that textual scholars have not yet committed themselves wholeheartedly to the editing of all witnesses of a text: largely – I think, in accordance with Busby’s remark – because the technology (and the related knowledge about the technology) was not easily available to them. I expect that in the next few years, editors will finally begin to take the new philological direction, assisted by the tools that are available online and are easy to use, and that editors will soon be able to move on to the New Stemmatology of Robinson et al. The stemmatologists were simply far ahead of us in their way of thinking about texts and the need to have all versions digitally available. What we have to do now is to make their tools also available in online work environments, so that textual scholars can, in a sense, go back to the future of computational humanities.

References

Beaulieu, A., K. van Dalen-Oskam and J. van Zundert, “Between tradition and Web 2.0: eLaborate as social experiment in humanities scholarship,” in Social Software and the Evolution of User Expertise: Future Trends in Knowledge Creation and Dissemination,

23 Cf. http://www.canterburytalesproject.org/index.html
24 Cf. Spencer c.s. (2006); Windram c.s. (2008).
Busby, K., “Variance and the politics of textual criticism,” in *Towards a synthesis? Essays on the New Philology*, ed. K. Busby (Amsterdam, 1993), pp. 29–45.

Causer, T., J. Tonra and V. Wallace, “Transcription maximized; expense minimized? Crowdsourcing and editing The Collected Works of Jeremy Bentham,” in *LLC: The Journal of Digital Scholarship in the Humanities* 27(2) (2012), pp. 119–37. doi: 10.1093/llc/fqs004. First published online: March 28, 2012.

Cerquiglini, B., *Éloge de la variante. Histoire critique de la philology* (Paris, 1989).
—–, *In praise of the variant. A critical history of philology* (Baltimore, 1999).

De Leeuw van Weenen, A., *Alexanders saga. AM 519a 4° in The Arnamagnæan Collection, Copenhagen* (Copenhagen, 2009). Museum Tusculum Press. Manuscripta Nordica. Early Nordic Manuscripts in Digital Facsimile Volume 2.

Dictionary of Early Middle Dutch: *Pijnenburg, W.J.J., van Dalen-Oskam, K.H., Depuydt, K.A.C. and Schoonheim, T.H. c.s. 2001. Vroegmiddelnederlands Woordenboek. Woordenboek van het Nederlands van de dertiende eeuw in hoofdzaak op basis van het Corpus-Gysseling. Leiden/Groningen. 4 Vols. Digitally available at http://gtb.inl.nl/?owner=VMNW*

Gysseling, M., “Rijmbijbel/tekst,” in *Corpus van Middelnederlandse teksten (tot en met het jaar 1300)*, ed. M. Gysseling, series II, vol. 3 (Leiden, 1983).

Kestemont, M. and K. van Dalen-Oskam, “Predicting the past: Memory-based copyist and author discrimination in medieval epics,” in *Proceedings of the twenty-first Benelux conference on artificial Intelligence (BNAIC 2009)* (Eindhoven, 2009), pp. 121–28.

Masters, B., “The distribution, destruction and dislocation of authority in medieval literature and its modern derivatives,” in *Romanic Review* 82 (1991), pp. 270–85.

Pierazzo, E., “A rationale of digital documentary editions,” in *LLC: The Journal of Digital Scholarship in the Humanities* 26(4) (2011), pp. 463–77. doi: 10.1093/llc/fqr033. First published online: July 28, 2011.

Postma, A., “Overzicht van Scolastica-handschriften,” in *Scolastica willec ontbinden. Over de Rijmbijbel van Jacob van Maerlant*, eds. J. van Moolenbroek and M. Mulder (Hilversum, 1991), pp. 145. (Middeleeuwse Studies en Bronnen xxv)

Robinson, P. “Electronic editions for everyone,” in *Text and genre in reconstruction. Effects of digitization on ideas, behaviours, products and institutions*, ed. Willard McCarty (Cambridge, Eng., 2010), pp. 145–63.

Shillington, P., “How literary works exist: implied represented, and interpreted,” in *Text and genre in reconstruction. Effects of digitization on ideas, behaviours, products and institutions*, ed. Willard McCarty (Cambridge, Eng., 2010), pp. 165–82 (165).
Spencer, M., H.F. Windram, A.C. Barbrook, E.A. Davidson and C.J. Howe, “Phylogenetic analysis of written traditions,” in Phylogenetic methods and the prehistory of languages, eds. P. Forster and C. Renfrew (Cambridge, 2006), pp. 67–74.

Van Dalen-Oskam, K., “The secret life of scribes. Exploring fifteen manuscripts of Jacob van Maerlant’s Scolastica (1271),” in LLC: The Journal of Digital Scholarship in the Humanities 27(4) (2012), pp. 355–72. doi: 10.1093/llc/fqs034. First published online: August 8, 2012.

Van Dalen-Oskam, K. and K. Depuydt, “Lexicography and Philology,” in Dictionaries of Medieval Germanic Languages. A Survey of Current Lexicographical Projects, eds. K.H. van Dalen-Oskam, K.A.C. Depuydt, W.J.J. Pijnenburg and T.H. Schoonheim (Turnhout, 1997), pp. 189–97. (Selected Proceedings of the International Medieval Congress, University of Leeds, 4–7 July 1994)

Van Dalen-Oskam, K. and J. van Zundert, “Delta for Middle Dutch – Author and Copyist Distinction in Walewein,” in Literary and Linguistic Computing 22 (2007), pp. 345–62.

———, Digital editions with eLaborate: from practice to theory. Conference Abstract for Digital Humanities 2012 (Hamburg, 2012). http://www.dh2012.uni-hamburg.de/conference/programme/abstracts/digital-editions-with-elaborate-from-practice-to-theory/

Vanhoutte, E., “Defining electronic editions: a historical and functional perspective,” in Text and genre in reconstruction. Effects of digitization on ideas, behaviours, products and institutions, ed. Willard McCarty (Cambridge, Eng, 2010), pp. 119–44.

Van Zundert, J., “If you build it, will we come? Large scale digital infrastructures as a dead end for Digital Humanities,” in Controversies around the Digital Humanities. Special issue of Historical Social research / Historische Sozialforschung 37, ed. M. Thaller (2012), pp. 165–87.

Van Zundert, J. and P. Boot, “The Digital Edition 2.0 and the Digital Library: Services, not Resources,” in Digitale Edition und Forschungsbibliothek. (Beiträge der Fachtagung im Philosophicum der Universität Mainz am 13. und 14. Januar 2011). Special issue of Bibliothek und Wissenschaft, 44 (2011), pp. 141–52.

Windram, H.F., P. Shaw, P. Robinson and C.J. Howe, “Dante’s Monarchia as a test case for the use of phylogenetic methods in stemmatic analysis,” in Literary and Linguistic Computing 23 (2008), 443–63.

**Websites (All active 5 September 2013)**

Canterbury Tales Project: http://www.canterburytalesproject.org/index.html

eLaborate: www.e-laborate.nl/en/

Lodewijk van Deyssel, Menschen en bergen: http://menschenenbergen.huygens.knaw.nl/

Minitab: http://www.minitab.com/
Online Variorum of Darwin's *Origin of Species*: http://darwin-online.org.uk/Variorum/index.html

*Text Encoding Initiative*: http://www.tei-c.org/index.xml

*Vincent van Gogh: The Letters*: http://www.vangoghletters.org/vg/

*Walewein ende Keye*: http://www.waleweinendekeye.huygens.knaw.nl/