The Genealogy of Texts: Manuscript Traditions and Textual Traditions

Barbara Bordalejo
KU Leuven, Faculteit Letteren, Belgium

Abstract

For some time, scholars have been using computer-assisted methods to produce graphic representations of the relationships between witnesses within a textual tradition.[1] The use of methods originally developed by evolutionary biologists has been called into question on account of the perceived lack of identity between two different disciplines. This view arises from a misunderstanding about how the methods work in relation to texts and how the resulting stemmata should be interpreted. This article refines textual critical terminology, particularly the distinction between textual traditions and manuscript traditions, in the context of the use of computer-assisted stemmatological methods to further our understanding of how these fit within the wider theoretical framework of textual criticism and scholarly editing, and makes explicit the way in which stemmata produced by using evolutionary biology software should be read.

Correspondence: Barbara Bordalejo, Faculty of Arts, Blijde Imkomststraat 21, Faculteit Letteren, KU Leuven, Leuven 3000, Belgium.
E-mail: barbara.bordalejo@kuleuven.be

1 Introduction

The distinction between the concepts of textual tradition and manuscript tradition is central to our understanding of how stemmatological software works and what its limitations are. The history of stemmatology is the story of a series of exercises in the use of quantitative methods, sometimes paired with qualitative methods, to analyse textual traditions. To study manuscript traditions requires more than the quantitative analysis of variants, as I explain below, and relies on extra-textual elements that add a further dimension to our understanding of the text.

It can be argued that computer-assisted methods are just the latest chapter in the history of stemmatology. However, there have been strong objections to the use of computer-assisted stemmatic analysis. Some of these objections arise from a misapprehension about how the software works, while others flow from a fundamental fallacy which assumes that traditional 'hand-made' stemmata are closer to the historical reality involving the transmissions of texts in manuscripts. The root of this misunderstanding is the failure to distinguish textual traditions and manuscript traditions. This failure gives rise to an expectation that an analysis solely based on textual data (the textual tradition) might yield results that are equivalent to an analysis that includes textual and extra-textual data (the manuscript tradition). In this article, I answer both kinds of objections and, with examples from three different textual traditions, explain how to read and interpret computer-generated stemmata before drawing conclusions.

In order to take advantage of computer-assisted stemmatic analysis, one first must understand how it works and why and in what it differs from non-computational approaches. As a further step, one must be able to correctly read and interpret the computer-generated stemmata.

2 A Brief History of the Stemmatic Approach

Before the invention of the movable type printing press, texts were transmitted either by copying from

doi:10.1093/llc/fqv038

dictation or from a previously written manuscript. Although scribes were making a conscious effort to be as accurate as possible during the copying process, they made unintentional mistakes and introduced intentional changes to the text. Over a period of time, many different copies could have been made from any given manuscript. At each copying stage, the text was liable to be changed, and those changes took it further from the version that originated the tradition (the archetype).

2.1 Traditional stemmatic analysis

Traditional stemmatic analysis is attributed to the German scholar Karl Lachmann who presented his ideas about editing in several of his works, notably in his 1850 edition of Lucretius. Lachmann proposed to group manuscripts belonging to a particular tradition according to their errors. In this way, 'families' of manuscripts could be established and their relationships made clearer. His impact was so great that often this approach is referred to as the Lachmann method (Timpanaro, 1985 [1963]). This method was based on the belief that agreement in error could indicate identity in origin; hence, the method is referred to as 'the common error method'. In this way, two witnesses presenting shared errors could have the same ancestor (Maas, 1958).

2.2 Questioning the Lachmann method

Henry Quentin (1926) started to question several aspects of the Lachmann method. Firstly, he thought that the attempt to reconstruct the author's original (the autograph) was a serious mistake and that the best that editors could hope for was the production of an archetype. Secondly, he articulated for the first time the concept that the archetype could possibly contain errors and so all variants, not just the so-called common errors, should be taken into account to establish relationships between different witnesses. In this, Quentin was making a case that is still relevant today since the expectation that a text will be completely devoid of errors is unrealistic. Even editors who believe that authorial intention can be recovered accept that the physical manifestation of a text is likely to introduce mistakes that could not have been intended (see Section 2.3 below).

Later, the French editor, Joseph Bédier, after having edited *Le Lai de l'Ombre* using the stemmatic method (Bédier, 1913) became suspicious of it and wrote an essay denouncing its shortcomings (Bédier, 1929). Bédier became the first proponent of what became known as best-text editing. Bédier's best-text editing, in contrast with the Lachmann method, generated what is now perceived as a classical antinomy in textual criticism, but particularly within medieval literary studies. On the one hand, it might appear as if the Lachmann method implies necessarily the reconstruction of the archetype. On the other hand, best-text editing only requires the editor to select the best extant witness of a text and present it, perhaps slightly edited, to the readers.

2.3 Anglo-American textual criticism

Anglo-American textual criticism, known for its bias towards intentionalist editing,[2] makes a useful distinction between document and work, where document is any material container of text, and the work is an immaterial or ideal form of a text (Tanselle, 1989). The distinction between work and text, and the conceptualization of each of those terms, are fundamental to the way we understand textual criticism and its products. The documents that are the physical containers of text also hold extra-textual materials. There are illuminated capitals, textual corrections, marginal notes, or visible drypoint and many other marks that, although unrelated to the text, might shed light on the history of its production.

The difference between a work and the documents that witness it, at least in literary studies, can be put in terms of Bateson's (1961) question 'If the *Mona Lisa* is in the Louvre, where are *Hamlet* and *Lycidas*?'. The answer to this is that *Hamlet* is in every copy of the text that we recognize as *Hamlet*. In practical terms, a scholar faced with a document containing a text generally recognizes it as an instance of a particular work (if that work is known), even though it might present variation that deviates quite sharply from other instances of the work itself.[3] The importance of the concept of the work (which Paul Eggert has called a 'regulative concept' [2009, 2013]) lies in our realization that in the world of textual transmission, our acknowledgement of a document being a witness to a particular work is what defines whether that

document belongs to a manuscript tradition or not. For stemmatology, this is a fundamental step. There would be very little point comparing texts that are completely unrelated, so one must determine whether a text is an instance of a work (possibly one of many instances) and so belongs in a textual tradition, or not.

2.4 The text of the document and the variant states of the text

The distinction between work and document leads to a subtle distinction between the text of the document (Bordalejo, 2010, 2013), a series of physical marks on a page or a sequence of sounds in a recording, for example, and the text of the work, a conceptual entity. Elsewhere, when I deal with the notion of 'the text of the work', I conclude by referring simply to the work, as a referential object (cf. Bordalejo, 2013).[4] Much more important to my theoretical framework are the concepts of 'the text of the document' and 'the variant states of the text'. The text of the document can be a series of meaningful marks on a writing surface (Bordalejo, 2013), but it could also be, for example, the totality of an audio recording, including its pauses, repetitions, and other peculiarities it might present. The text of the document might include what I refer to as 'variant states of the text', which are versions of the text as interpreted by a reader from corrections made into a written document, for example. The variant states of the text are constructions of meaning or meanings based on textual and extra-textual indicators present on the physical medium (Bordalejo, 2010, 2013).

This means, in stemmatics, that a single source document can carry multiple versions of the same text. Depending on how these versions came to be, they might have an immediate bearing in the genealogical analysis of texts or, if they are copied and passed on, might have an impact on the sections of a tradition that derive from them. This is illustrated by my examples of the *Divine Commedy*, the Luca Martini collation, and the Trivulziano manuscript (see Section 5.4 below).

2.5 Computer-assisted stemmatology

Computer-assisted stemmatological analysis was first used in 1977 (Platnick and Cameron), but it became a relatively common tool in the 1990s (O'Hara and Robinson, 1992, 1993; Robinson and O'Hara, 1996; Robinson, 1991, 1996, 1997, 2000a,b, 2001, Salemans 1996, 2000; Spencer *et al.* 2003a,b, 2004). It has been used with artificial traditions (Spencer *et al.*, 2004a, Baret *et al.*, 2006, Roos and Heikkilä 2009) with Chaucer's *Cantebury Tales*, Dante's *Monarchia*, and the Greek New Testament.

Through my involvement in the Canterbury Tales Project, I have employed phylogenetic inference software with different methods such as split decomposition, parsimony, and distance to produce working hypotheses about textual relations. Other methods have been employed, most notably clustering (Evert, 1996) and the coherence method developed by Gerd Mink (2004). More recently, software specifically designed to work with texts has been developed. A method based on compression algorithms, RHM, described by Roos and Heikkilä (2009) and Semstem, which solves both the problem of bifurcating trees and allows for the placement of witnesses as internal nodes (Roos and Zuo, 2011). I have not experimented with these methods while they were in development. For now, my experience is confined to the use of SplitsTree and Phylogenetic Analysis Using Parsimony and Other Methods (PAUP).

2.6 Phylogenetic software

Phylogenetic software was originally developed to create hypotheses of the possible relationships between living organisms. For this, it uses DNA sequences and compares them before constructing a hypothesis based on a particular evolutionary model. Phylogenetic software works by assessing different DNA sequences. The four bases (adenine, cytosine, guanine, and thymine) combine to form three-letter sequences to form specific amino acids. Mutation alters the sequencing; if the change is successful, it will be copied and become a regular feature (Bordalejo, 2003). The similarity of this process with those involved in the copying of manuscripts is remarkable and has been discussed before (Howe *et al.*, 2001). The difference resides in the length of the sequences since human language does not follow the three-letter system which is the norm for the coding proteins. Despite that, the rest of the

relationship is one of identity, where contamination can be related to recombination and lateral gene transfer, and agreement by coincidence related to convergent evolution (Howe *et al.*, 2004).

2.7 Objections to the use of phylogenetics and other methods to study texts

Despite the discussions on how phylogenetic analysis can be used for the study of large textual traditions and its success with important literary, religious and historical texts, serious objections are still being raised about the suitability of these methods or of stemmatology as a valid textual critical approach (Hanna, 1996, 2000; Cartlidge, 2001; Robins, n.d, 2007b). Notably, neither Hanna nor Cartlidge has seriously attempted to use computer-assisted stemmatic analysis, and it is possible that this lack of experience in using these methods accounts for their misunderstanding of how the different methods work. Robins seems to completely dismiss the use of computer-assisted stemmatic methods when he states that '...[i]f cladistic [sic] approaches are used only to replicate traditional stemmatics—to distinguish between authorial originality and scribal error—then they are almost certain to be dismissed by many literary scholars. Few accept the premise that the significance of a text is controlled by authorial intentions, even more so in cases where no authorial documents survive' (Robins, 2007a, p. 114).[5] The problem with this description of stemmatics is that it overlooks the fact that, for some time now, it has been made clear that this is not how contemporary stemmatology works (Robinson, 1991, p. 153; Salemans, 1996, p. 22; Bordalejo, 2003, p. 60). Instead, what we have is a basic misinterpretation of stemmatics arising from historical misconceptions about how the method has been implemented in individual analysis, or by the way in which it has been described in print (Maas, 1958).

3 What is a Stemma?

In the past, stemmata have been thought to represent the relationships between the different witnesses of a text. It is likely that Bédier was criticizing the apparent claim that stemmata were literal representations of reality. However, it might be more productive to approach stemmata not as a literal reflection of history and more as an educated hypothesis created following a specific model.[6]

The use of phylogenetic software (or of other software specifically designed to deal with the relationships between texts) helps us to understand various aspects of stemmatology that might not otherwise be as clear. For example, unless specific historical data is introduced, any software used for stemmatology deals only with text and textual features. Unregularized collations, which preserve medieval spelling, word division, and capitalization, produce inconclusive results because of the amount of noise that these types of variations generate. A regularized collation, in which only substantive variation is taken into account, produces much more meaningful results (cf. Solopova, 2000; Robinson, 2004; Thomas, 2006).

One of the most contentious areas for textual scholars concerns the tendency of phylogenetic methods to give trees in which almost all relationships between manuscripts are in the form of bifid branching: that is, with almost every node giving rise to two branches.[7] In manuscript traditions, we know many cases where three or more manuscripts are copied from a single exemplar. But in the representations of textual traditions, created by standard phylogenetic methods, it will be very rare to see three descendants of a single node. This is a result of the way phylogenetics, and particularly parsimony, works. In phylogenetic methods, there are only two possible states: change or no change. No change is represented by 0, while there can be one or more changes at any single point. In order for standard phylogenetic software to come up with a tripartite tree for even a single node, with three branches descending from that node, it would be required that at least three witnesses presented equivalent variation at exactly the same points and that this variation should occur consistently throughout. In quantitative terms, this would mean that the three must have exactly the same number of divergences from the common exemplar, with no two of the three having any more variants in common against the third than any other pair. Coincident variation

and simple chance make this extremely unlikely: in all the phylogenetic reconstructions of textual traditions we have analysed, instances of this have never been recorded. If we accept this, we are also likely to understand that this software can only model our data according to its own assumptions and methods, and thus the hypothesis of possible relationships between texts will reflect these assumptions and methods. In other words, a hypothetical representation using phylogenetic software models the relationships between the text found in different witnesses based on the data it is given and the assumptions and methods governing the software. The representation it creates, of what I call the 'textual tradition', is valid in and for itself, but cannot be taken of itself as a full representation of the 'manuscript tradition'.

A possible objection to this type of representation, for example, is that we might know with certainty that three or more witnesses are descended from another one. We could know this from extra-textual evidence, as in the case of the *Svipdagsmál* manuscripts (Robinson, 1991, O'Hara and Robinson, 1993) or of the artificial textual traditions (Spencer *et al.*, 2004a, Baret *et al.*, 2006, Roos and Heikkilä, 2009). This objection can be answered by separating textual traditions from manuscript traditions.[8]

4 Manuscript Traditions and Textual Traditions

4.1 The textual tradition

Computer-assisted stemmatology bases its results only on textual data, unless non-textual data are deliberately included in the data set. A variant matrix produced using collation software can only take into account textual places of variation and the presence or absence of said data. In this sense, to expect that stemmata generated from such data should give exact representations of a known or suspected historical reality is not reasonable. At most, phylogenetic software can offer a hypothesis in reference to the analysed data, and to expect something different from that will only lead to puzzlement and disappointment. If we accept this fact, we will find that it becomes less pressing to produce a stemma that can serve as a true reflection of a historical reality. Instead, the results can be approached with a more objective perspective devoid of false expectations. This allows us to produce further analysis of the variation that generates a particular set of results in the knowledge that the hypothesis presented by the software is just a starting point, rather than the final objective.

4.2 The manuscript tradition[9]

In order to generate a stemma that takes into account other historical features, more than just textual data has to be introduced in the data set. Any information that is external to the text and that can be used to establish genetic relationships, time of copying, relationship to other witnesses, can be included into the data set so as to allow for a modification of the behaviour of the software that will now take into account more than just the textual variants. The result of this would be a graph or stemma that is the stemma of the manuscript tradition (or the document tradition, as the case may be), that is, one that makes an advised attempt to include information that, although available, is external to the text in order to move one step closer to a historical reality. Notice that I consider this just to be one step closer, rather than to be one with it.

As researchers, it is our duty to maintain a healthy scepticism that allows us to remain objective in our work. This applies to the use of any computer methods of investigation, as well as to the conclusions that we draw from them, but it should also apply to any preconceived notions that we might carry from our own history.

4.3 Robinson, O'Hara, and the textual tradition of *Svipdagsmál*

The concept of textual tradition and manuscript tradition can be illustrated by the work on the Old Norse tradition *Svipdagsmál* carried out by Robinson and O'Hara (1996).

Robinson had created a hand-drawn stemma that showed, by using arrows, which lines of descent were supported by both textual and extra-textual data (Fig. 1).

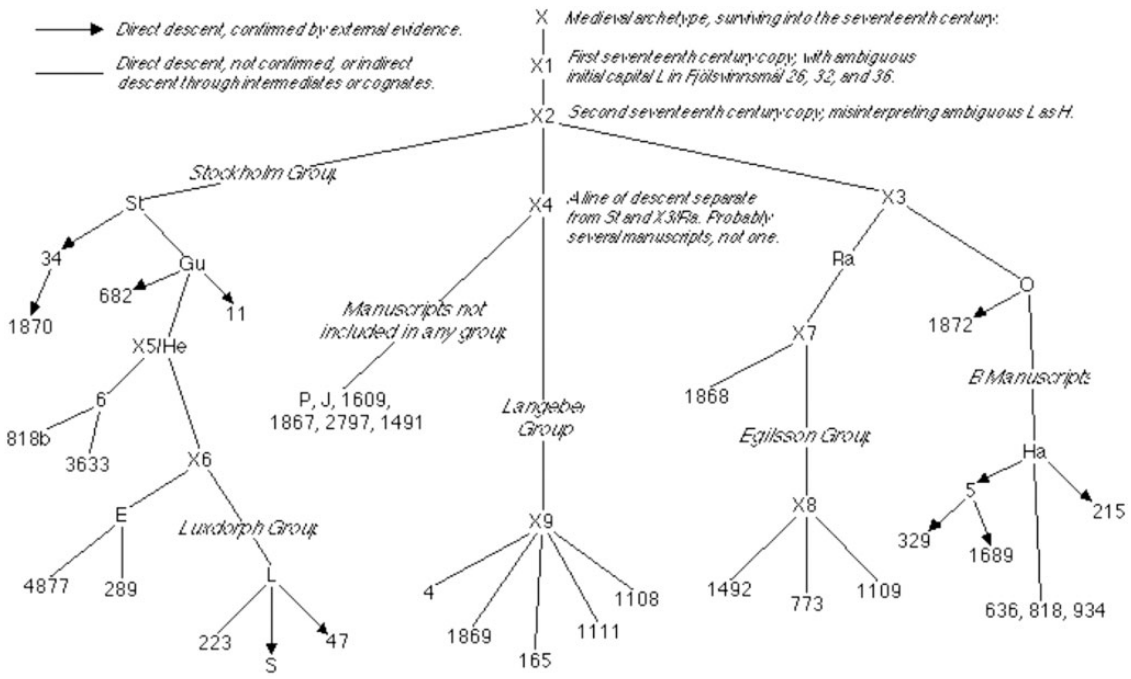

Fig. 1 Stemma of the manuscripts of *Svipdagsmál*. Relationships of the *Svipdagsmál* manuscripts, after Robinson (1991). Branch lengths and branching angles are arbitrary, and branches may be rotated about nodes arbitrarily. Arrows indicate relationships confirmed by external evidence. X-X9 are hypothetical ancestors. Ra may be identical with X3 rather than copy of it, and He may be either a copy of X5 or identical with X5

This is what I refer to as a stemma of a manuscript tradition since it takes into account not only the variation at word level, but also data derived from annotations in the manuscripts stating dates of copying, as well as occasionally, the names of scribes and the source of the text.

Robinson and O'Hara never made this distinction because, at the time, their main concern was to highlight the cases in which PAUP showed the same results as the traditionally produced stemma. However, the distinction might be fundamental to explain not the similarities, but the differences between this and the stemma of *Svipdagsmál* produced by PAUP (Fig. 2).

Indeed, when comparing the two graphs, O'Hara and Robinson (1996) expressed their idea that further similarities between the stemmata could have been achieved if, for example, the direction of variation could have been included within the data. From my perspective, it is much more interesting to see some of the differences between the two graphs. For example, the stemma produced using traditional methods, i.e. without the help of computers, shows that 34 was copied from St and that 1870 was copied from 34. In the PAUP stemma of the textual tradition, these appear next to each other, in adjacent nodes. This makes perfect sense, the software, making use only of textual data, shows a very close textual relation between those witnesses without specifying exactly what that relationship is. Only with the introduction of the extra-textual data could Robinson understand the exact nature of the relationship between those three manuscripts.

5 How to Read a Computer-Generated Stemma?

5.1 SplitsTree stemma of the textual tradition of the *Canterbury Tales* General Prologue

It is useful to consider some examples of computer-generated stemmata and how these should be read.

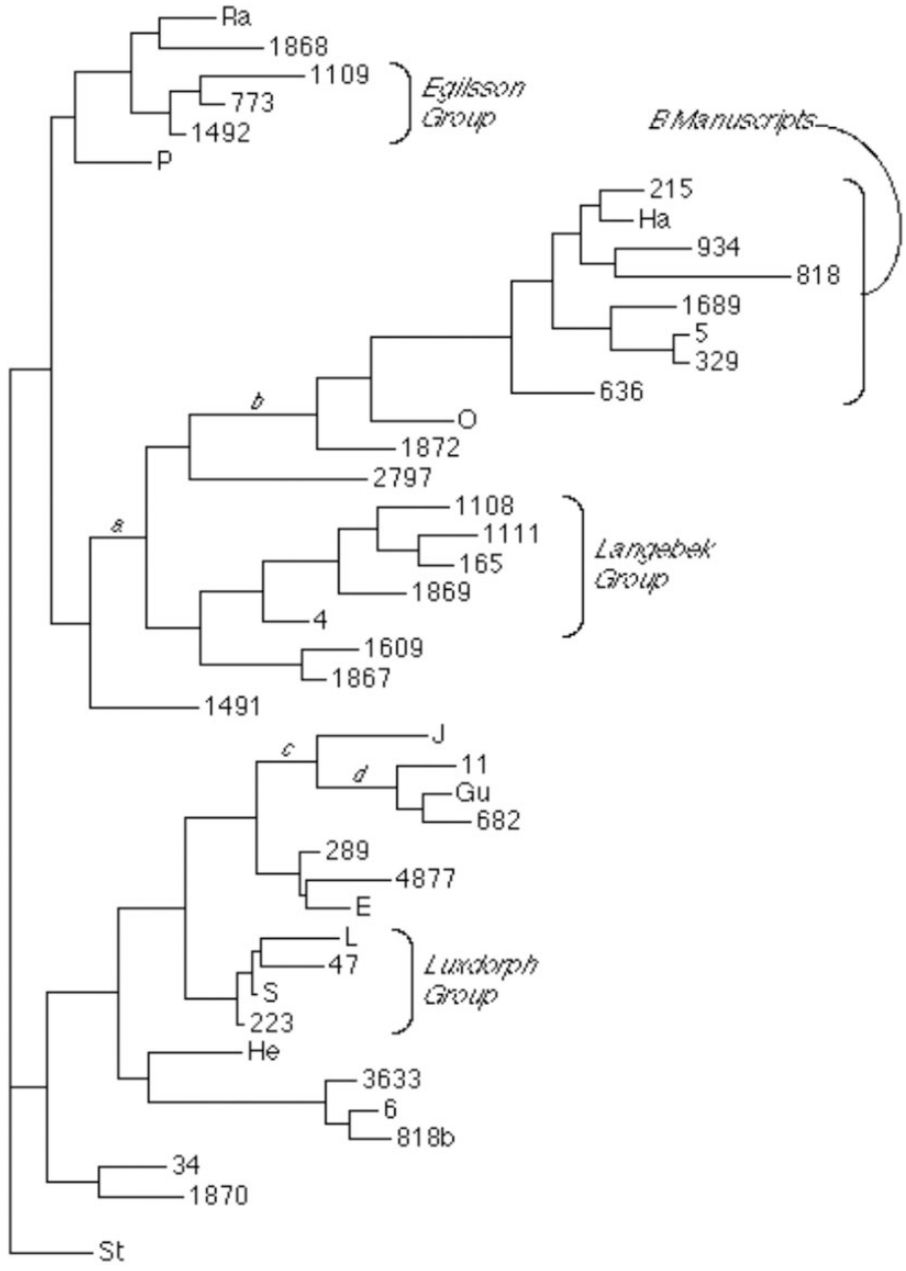

Fig. 2 Cladogram of the manuscripts of *Svipdagsmál*. Estimate of the history of the *Svipdagsmál* manuscripts generated by the cladistics program PAUP. Some of the major grouping of manuscripts common to this tree and to Robinson's stemma (Fig. 1) are indicated. Horizontal branch lengths are proportional to the number of character state changes along each branch. Vertical branch lengths are arbitrary, and branches may be rotated about nodes arbitrarily. See note 7 for additional details

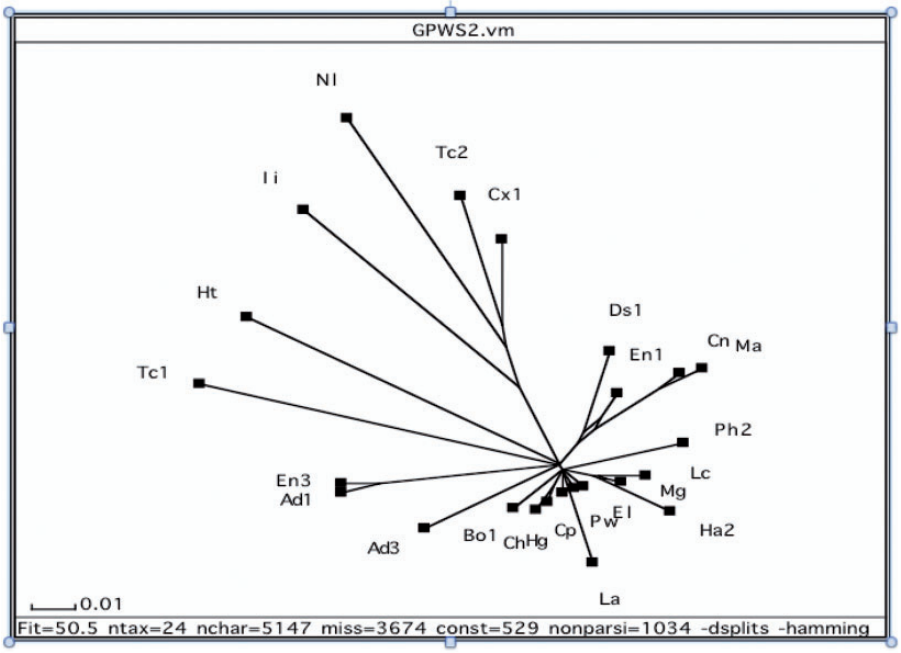

Fig. 3 Stemma of the General Prologue produced using SplitsTree. Originally published in Solopova 2000

This is particularly important because, as I show below, an understanding of the textual tradition as well as tools for further investigation are required to productively and accurately interpret these stemmata (Fig. 3).

This stemma of the manuscript tradition of the *Canterbury Tales* was published as part of the *General Prologue on CD-ROM* (Solopova, 2000). From the diagram alone, it not possible to ascertain the position of the archetype of the tradition. Instead, we can see how the texts of the different witnesses relate to each other. In order to read this stemma successfully, it is important to understand a multiplicity of aspects of the textual tradition. I here refer to Robinson's discussion of the results of the phylogenetic analysis of the General Prologue witnesses, published on this CD-ROM (2000a).

Before the Canterbury Tales Project, scholars had been hypothesising about the tales, and particularly John Manly and Edith Rickert (Manly and Rickert, 1940) had attempted to understand the relationships among all extant witnesses and established what computer analysis later confirmed were four distinct witness families, which they named a, b, c, and d. Even before them, Furnivall and Skeat had suggested that Hengwrt (National Library of Wales Peniarth 392d) and Ellesmere (Huntingdon Library, MS EL 26 C 9), represented in the stemma as Hg and El, were two of the most important textual witnesses of the *Tales.* Manly and Rickert agreed with this statement, as have various members of the Canterbury Tales Project. Notice, however, that in the General Prologue stemma, Hg and El do not appear very close together. Indeed, the closest manuscript to Hg is Ch (Christ Church, Oxford, MS 152) which is a manuscript generally ignored by textual critics on account of its late dating (around 1460), the fact that it contains the Tale of Gamelyn and the Plowman's Tale and, generally, because it is unremarkable in its making, lacking illustrations or any kind of illumination. Analysis of individual variants using the variant database, VBase,[10] shows that although Ch is a late manuscript of the *Tales*, probably copied around 1460, the text that it contains is close to the earliest versions of the text.

In this stemma, we can also see what appeared to Robinson to be an early split in the textual tradition

(2000a). This early split, which separates the top half of the stemma in which Ad1, En3, Tc1, Ht, Ii, Nl, Tc2, Cx1, Ds1, En1, Cn, and Ma appear, from the bottom half (including Bo1, Ch, Hg, Cp, La, Pw, El Ha2, Mg, Lc, and Ph2). The split suggests the existence of two different hyparchetypes. Robinson called these alpha and gamma. He thought that these were early copies of the archetype of the tradition. This might suggest that the archetype could have been somewhere close to those two, no longer extant, witnesses. In this case, and based on the analysis of individual variants to explain the relative positions of the different witnesses, we find that the archetype of the tradition was probably close to the two points in which the graph splits. Although we cannot be absolutely sure, a good case can be made for a point closer to Hg, El, and Ch to be the archetype of the tradition itself, since each of these witnesses represents an independent line of descent from that same archetype.

5.2 PAUP[11] stemma of the textual tradition of the Miller's tale

For some time now, Robinson and myself have been using PAUP to build stemmata of the textual tradition of the *Tales*. We have successfully analysed several sections of Chaucer's text, namely, the Miller's Tale, the Nun's Priest's Tale, the Merchant's Tale, and the Franklin's Tale, as well as those previously studied by Robinson, the Wife of Bath's Prologue and the General Prologue. I have also used it for the analysis of non-textual variation when I worked on tale-order (Bordalejo, 2003; Spencer *et al.*, 2003a 2003b).

This second example comes from the Miller's Tale on CD-ROM (Robinson, 2004) (Fig. 4).

This stemma shows Manly and Rickert's groups a, b, c, and d (divided in d1 and d2). There are also groups e and g, which have been found also in the other sections of the *Tales*. The witnesses labelled o are not a genetic group (Bordalejo, 2003); rather, they represent independent lines of descent from the archetype of the tradition.

This stemma shows more clearly than the one based on General Prologue data the early split into two different hyparchetypes, alpha and gamma. Robinson thought that this split in the General Prologue might suggest that the archetype could be located between those two hyparchetypes (around the visual centre of the graph), and analysis based only on the phylogenetic results, without any knowledge of the textual tradition and without means to access the data that generated it, might think this also true for the Miller's Tale (Robinson, 2004). However, analysis of the individual variants, drawing on scholarly judgement as to the likely direction of variation in each case, suggests that the archetype of the tradition is between the putative alpha and the branch in which Hg, Ch, and El are placed (see the discussions of the 'O variants' in Robinson 2004 and Thomas 2006). Again, we know this because the nature of the variants shared between Ch, Hg, and El show that, for this section of the text, these witnesses represent independent lines of descent from the archetype of the tradition (Robinson, 2006).

5.3 Extra-textual evidence in the manuscript tradition of the *Canterbury Tales*

Unlike the *Svipdagsmál* manuscripts, the *Canterbury Tales* does not present explicit information about the precedence of the text of the manuscripts. However, there are some extra-textual features in the witnesses that one might want to consider.

For example, Ch, dated around 1460, includes texts that were not written by Chaucer such as the Plowman's Tale and the Tale of Gamelyn. Without the help of the computer-generated stemmata it would be less obvious that this late manuscript is witness to a very early text. Ch places consistently with the best and earliest manuscripts of the *Canterbury Tales*, including Hg and El, and the shared readings are sometimes remarkable and unusual. This indicates that the textual information from this manuscript overrules the fact that it is a later copy. Indeed, just because a witness was produced at a later time it does not always follow that its text is a late text. If we were only to take into account the age of the manuscript or if we gave it an undue importance, we could risk overlooking the textual evidence that testifies to the closeness of the text of Ch to the archetype of the tradition.

Hg and El are also linked because they were copied by the same scribe, known as Hand b until

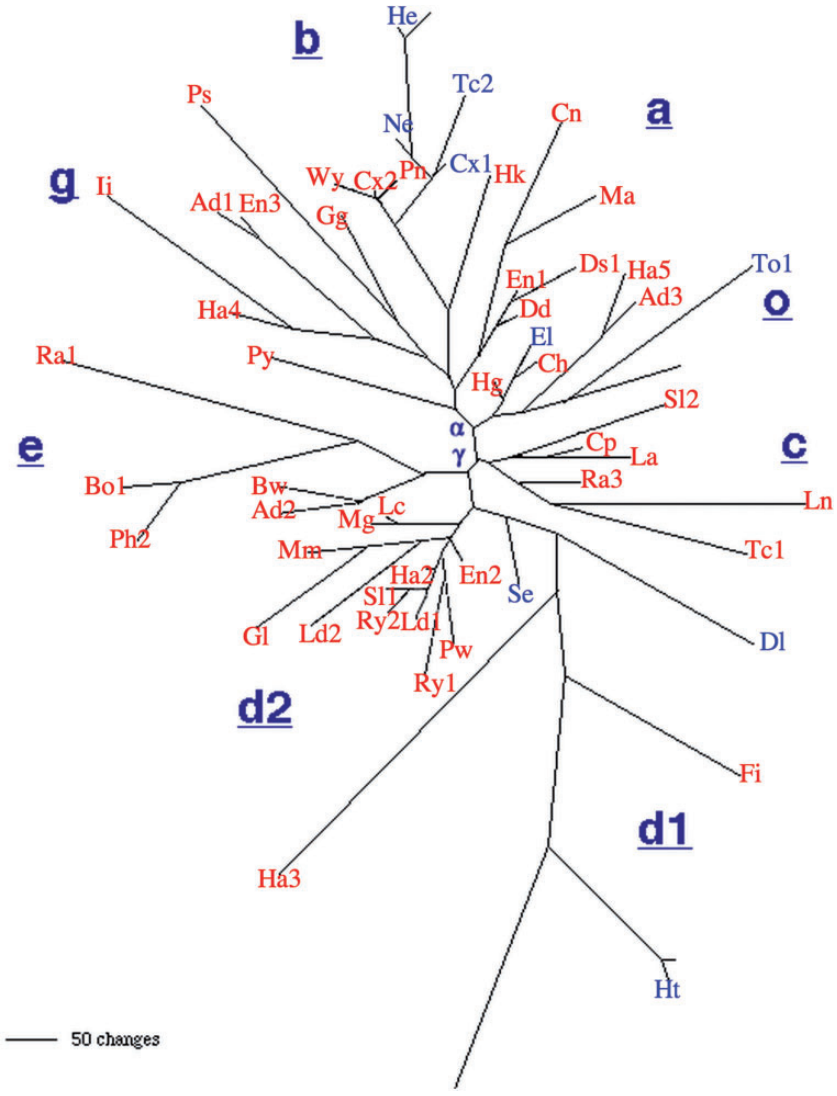

Fig. 4 Stemma of the Miller's Tale produced using PAUP. Originally published in Robinson 2004

Linne Mooney's identification (2006) which gave him a name, Adam Pinkhurst, and a story, as a scribe who might have been working closely with Chaucer over a period of time. We knew that Hg was an early manuscript and that El boasted a luxury production, together with a very good text. However, the identification of the scribe who wrote these manuscripts and the evidence that he was closely associated with Chaucer support statements about their importance, both textual and historical. I remark in the last section that scholarly judgement is essential for interpretation of the textual tradition; here, one must use scholarly judgement to determine which information sheds light on the transmission of the text and what obscures it, for example the late date of Ch.

5.4 Dante's *Divine Comedy*

The case of the *Divine Comedy* also shows some examples that can help us understand the layer of

meaning and interpretation afforded by the use of extra-textual information. This work, now published both online and in CD-ROM (Shaw, 2010), was born from a collaboration with Federico Sanguineti who proposed a new view of the textual tradition of the *Divine Comedy* in which only 7 of the more than 800 manuscripts and fragments were required to produce a reliable edition.

As the project progressed and started to show partial results it became evident that it would not support Sanguineti's conclusions. Shaw synthesizes this by stating that the position of Rb (Biblioteca Riccardiana, Ms. Riccardiano, 1005) within the stemma is so crucial to Sanguineti's view of the tradition that if it were to be shown to be different from what he suggested, his whole argument would collapse (Shaw, 2010). In Sanguineti's view, Rb is an alpha manuscript, which expressly contradicts what Petrocchi had stated in his edition, i.e. that it was a beta manuscript.[12]

Six of the manuscripts included as part of the Sanguineti seven predate 1355, which was the cut-off date for the Petrocchi edition. Petrocchi used 1355 as a date because this would ensure that the considered manuscripts predated the copies of the *Divine Comedy* produced by Boccaccio between 1373 and after which the tradition became widely contaminated with derivative readings.

The seventh manuscript included chosen by Sanguineti is LauSC (Biblioteca Medicea Laurenziana, Ms. Plut. sin 1). This manuscript was not included by Petrocchi precisely on account of its date. Sanguineti had considered it an important witness to be included and so it also became part of Shaw's edition.

LauSC presents many later corrections and, as part of the transcription programme for the edition, we decided to encode the corrections separately from the main text (Bordalejo, 2010). When the results of both the first text and the one including the later corrections are compared, we find that the branch for this witness becomes shorter (Fig. 5).

Robinson explains this, stating that the corrected version of LauSC (refered as 'c2') has more than 1,000 changes from the original form of the same witnesses and that, moreover, those changes made the text closer to the other witnesses, i.e. it shares more archetypal readings with them.

All of this can only be asserted because we have external evidence indicating that the manuscript is a later one, which initially had also a later text, that is, that having been copied after Boccaccio's intervention in the textual tradition its original, uncorrected text, had a text post 1357. The corrections in LauSC, however, must derive from an earlier text, one closer to the earlier witnesses, and this explains the difference in length in the compared branches.

The case of Mart (Biblioteca Nazionale Braidense, Aldina AP XVI 25) is revealing when considered in the light of extra-textual evidence. A copy of the 1515 Aldine edition was corrected by Luca Martini in 1548 against an early copy of the poem, using a manuscript that predated any surviving copy of the *Divine Commedy* and which is no longer extant (Shaw, 2010). Again, our encoding allows us to consider separately the Aldine text, which is not very informative because it belongs to the post-Boccaccio tradition, and the Martini collation. This separation of the texts revealed a close link between Martini's corrections in Mart and the Trivulziano manuscript of the *Divine Comedy* ('Triv': Biblioteca dell'Archivio Storico Civico e Trivulziana, Ms Trivulziano 1080), which supports the hypothesis that the manuscript source of Luca Martini's corrections was not only an early witness, but also textually close to Triv (Robinson and Shaw, 2010). Here, the extra-textual information, in the form of a note by Martini that he was using a manuscript dating from 1330 (and so earlier than any surviving manuscript of the *Divine Comedy*), is crucial to our understanding of the extraordinary closeness of the Martini corrections with Triv, written in 1337. Both the case of LauSC and that of Mart illustrate what I described in Section 2.4, that a single document may contain more than one variant state of the text. Our encoding system allows us to separate those variant states of the text so they can be considered and analysed separately from each other.

5.5 Using extra-textual evidence to inform stemmata

From these examples, one can infer that the use of extra-textual elements is an important tool in the

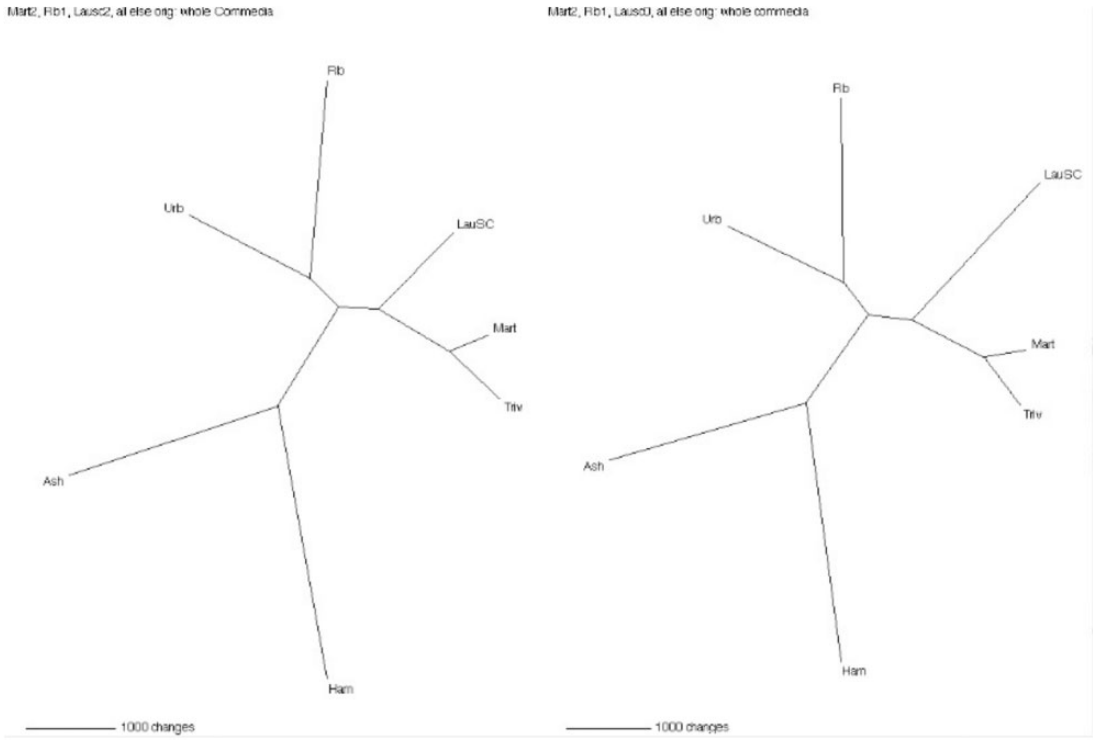

Fig. 5 Stemmata of the Divine Comedy produced using PAUP. Originally published in Shaw 2010

study and understanding of textual elements and fundamental for the construction of manuscript traditions. When the tradition is particularly explicit, as in the case of *Svipdagsmál*, we can compare the results of the software we use to build stemmata. In such cases, we might want to consider way in which both the data and the data processing can be improved to produce even more accurate results. Our understanding of particular aspects of the transmission can be enriched by our knowledge of particular witnesses, as is the case for the corrections to LauSC, a late manuscript containing corrections from an earlier text, or the Luca Martini collation of an early manuscript which he recorded in his Aldine edition of the *Divine Commedy*.

Also, where we find that our understanding of the textual tradition agrees with our extra-textual knowledge of the manuscript tradition, we can have greater confidence in the hypotheses offered by these methods concerning the textual tradition. Thus, both textual analysis and extra-textual evidence agree on the central importance of the Hengwrt and Ellesmere manuscripts of the *Canterbury Tales*. Indeed, the correspondence of textual and extra-textual evidence here is so strong that one may rely upon the textual evidence of the closeness of the Ch manuscript to the best extant manuscripts and its direct relationship with the archetype of the tradition, despite the extra-textual information suggesting its late production.

6 Conclusion

Phylogenetic analysis and other computer-assisted stemmatological approaches can be used productively when studying large textual traditions, despite the difficulties presented by contamination, changes in order, major alterations, and significant losses. The stemmata produced using computer-assisted methods are working hypotheses which serve as a starting point of investigation. These stemmata,

whether they correspond to a textual tradition or a manuscript tradition, are one of the tools that we can use to further our understanding of how texts are transmitted and how variants are inherited. What they do not do is to present us with a one-size-fits-all solution that could answer all of our queries. In the end, we are still subject to the remarks of A. E. Housman who said that knowledge and method were important, but that besides those a scholar was required to make use of her brain (Housman, 1921).

The interpretation of the stemmata generated by the use of phylogenetic software is fundamentally changed when we understand the difference between textual and manuscript traditions. Although the search for meaning in each of these follows a similar pattern, the recognition of the differences between the data sets will have an impact on our expectations.

A stemma, computer-generated or made by hand, is only a graphic representation of a hypothesis (machine or human or a combination of both) created following a specific model and has to be treated as such. The historical reality that underlies our hypotheses cannot be recovered in its totality, whether this reality corresponds with the textual tradition or with the manuscript tradition.[13] However, combining computer-assisted stemmatic analysis, database searches and historical knowledge of the production history of a particular text can help us build increasingly convincing hypotheses about it. Once we recognize this, we will be better equipped to use the tools at our disposal more efficiently and interpret the results of our research more accurately.

References

Bateson, F. W. (1961). Modern bibliography and the literary artifact. In Bonnard, G. A. (ed.), *English Studies Today*. Bern: Francke Verlag, pp. 67–77.

Baret, P. V., Macé, C. and Robinson P. (2006). Testing methods on an artificially created textual tradition. In Macé, C., Baret, P., Bozzi, A. and Cignoni, L. (eds), *The Evolution of Texts: Confronting Stemmatological and Genetical Methods*. Linguistica computazionale XXIV–XXV. Pisa, Roma: Istituti Editoriali e Poligrafici Internazionali. pp. 255–81.

Bédier, J. (1913). *Le lai de l'ombre*. Paris: Société des Anciens Textes Français.

Bédier, J. (1929). *La tradition manuscrite du Lai de l'ombre: réflexions sur l'art d'éditer les anciens textes*. Paris: Librairie Honoré Champion.

Bordalejo, B. (2003). *The Phylogeny of the Order in the Canterbury Tales*. Doctoral thesis, New York University.

Bordalejo, B. (2010). The encoding system. In Shaw, P. (ed.), *Dante Alighieri: Commedia. A Digital Edition*, Birmingham and Florence: Scholarly Digital Editions and SISMEL-Edizioni del Galuzzo.

Bordalejo, B. (2013). The texts we see and the works we imagine. *Ecdótica*, 10: 64–76.

Cartlidge, N. (2001). The *Canterbury Tales* and cladistics. *Neuphilologische Mitteilungen,* 102(2): 135–50.

Eggert, P. (2009). *Securing the Past*. Cambridge: Cambridge University Press.

Eggert, P. (2013). What we edit and how we edit. *Ecdótica*, 10: 50–63.

Evert, W. (1996). Clustering stemmatological trees. In van Reenen, P. and Van Mulken, M. (eds.), *Studies in Stemmatology*. Amsterdam: John Benjamins, pp. 123–34.

Hanna, R. (1996). *Pursuing History: Middle English Manuscripts and Their Texts*. Stanford, CA: Stanford UP.

Hanna, R. (2000). The application of thought to textual criticism in all modes - with apologies to A. E. Housman. *Studies in Bibliography*, 53: 163–72.

Housman, A. E. (1921). The application of thought to textual criticism. *Proceedings of the Classical Association*, 18: 67–84.

Howe, C., Barbrook, A. C., Spencer, M., *et al.* (2001). Manuscript evolution. *Trends in Genetics*, 17(3): 147–52.

Howe, C., Barbrook, A. C., Mooney, L. and Robinson, P. (2004). Parallels between systematics and phylogenetics. In van Reenen, P., den Hollander, A. and van Mulken, M. (eds), *Studies in Stemmatology II*. Amsterdam: John Benjamins, pp. 3–11.

Howe, C., Connolly, R. and Windram, H. (2012). Responding to criticisms of phylogenetic methods in stemmatology. *Studies in English Literature*, 52(1): 51–67.

Maas, P. (1958). *Textual Criticism*. Trans. Barbara Flower Oxford: Clarendon Press.

Mooney, L. (2006). Chaucer's scribe. *Speculum*, 81(1): 97–138.

McLaverty, J. (1984). The mode of existence of literary works of art: the case of the Dunciad Variorum. *Studies in Bibliography*, 37: 82–105.

Mink, G. (2004). Problems of a highly contaminated tradition: the New Testament. Stemmata of variants as a scource of a genealogy for wintesses. In van Reenen, P.,

den Hollander, A. and van Mulken, M. (eds), *Studies in Stemmatology II*. Amsterdam: John Benjamins, pp 13–85.

O'Hara, R. and Robinson, P. (1992). *Report on the Textual Criticism Challenge 1991*. http://rjohara.net/cv/1992-bmcr (accessed 31 March 2014).

O'Hara, R. J. and Robinson, P. M. W. (1993). Computer-assisted methods of stemmatic analysis. In Blake, N. and Robinson, P. (eds), *Occasional Papers of the Canterbury Tales Project*, vol. 1, Publication 5, Office for Humanities Communication, Oxford University, pp. 53–74. http://rjohara.net/cv/1993-ctp, (accessed 31 March 2014.)

Phillips-Rodriguez, W., Howe, C. and Windram, H. (2010). Some considerations about bifurcation in diagrams representing the written transmission of the Mahābhārata. *Vienna Journal of South-Asian Studies*, 52/53: 29–43.

Quentin, H. (1926). *Essais de Critique Textuelle*. Paris.

Robins, W. (2007a). Editing and evolution. *Literature Compass*, 4(1): 89–120.

Robins, W. (ed.) with **Motta Attilio**. (2007b). *Antonio Pucci: I cantari della Reina d' Oriente, edizioni critiche*. Bologna: Commissione per i Testi di Lingua.

Robins, W. (n.d.). Cladistics and Italian Philology. Abstract of Conference Presentation. http://folk.uib.no/hnooh/filologiadigitale/abstracts/Robins.pdf (accessed 18 January 2015).

Robinson, P. M. W. (1991). *The Textual Criticism Challenge*. Discussion List Archive. http://lists.village.virginia.edu/list_archive/Humanist/v05/0261.html (accessed 18 October 2001).

Robinson, P. M. W. (1996). Computer-assisted Stemmatic Analysis and 'Best-Text' Historical Editing. In van Reenen, P. and Mulken M. Van (eds), *Studies in Stemmatology*. Amsterdam: John Benjamins, pp. 71–103.

Robinson, P. M. W. (1997). A stemmatic analysis of the fifteenth-century witnesses to the Wife of Bath's Prologue. In Blake, N. F. and Robinson P. M. W. (eds), *The Canterbury Tales Project*. Occasional Papers, vol. 2. Oxford: Office for Humanities Communication, pp. 69–132.

Robinson, P. M. W. (2000a). Analysis workshop. In Solopova, E. (ed.), *The General Prologue on CD-ROM*. Cambridge: Cambridge University Press.

Robinson, P. M. W. (2000b). Stemmatic commentary. In Solopova, E. (ed.), *The General Prologue on CD-ROM*. Cambridge: Cambridge University Press.

Robinson, P. M. W. (ed.) with **Bordalejo B. and Da Rold O.** (Assistant Editors). (2004). *The Miller's Tale on CD-ROM*. Leicester: Scholarly Digital Editions.

Robinson, P. M. W. and O'Hara, R. J. (1996). Cladistic analysis of an old norse manuscript tradition. In Hockey, S. and Ide, N. (eds), *Research in Humanities Computing 4*. Oxford: Oxford University Press, pp. 115–37.

Robinson, P. and Shaw, P. (2010). The phylogenetic analysis. In Shaw, P. (ed.), *Dante Alighieri Commedia, A Digital Edition*. Birmingham: Scholarly Digital Editions.

Roos, T. and Heikkilä, T. (2009). Evaluating methods for computer-assisted stemmatology using artificial benchmark data sets. *Literary and Linguistic Computing*, 24(4): 417–33.

Roos, T. and Zuo, Y. (2011). Analysis of textual variation by latent tree structures. In Proceedings of the 2011 ICDM IEEE International Conference on Data Mining, IEEE Press, pp. 567–76.

Salemans, B. (1996). Cladistics or the resurrection of the method of Lachmann. Studies in Stemmatology. In van Reenen, P. and Van Mulken, M. (eds), *Studies in Stemmatology*. Amsterdam: John Benjamins, pp. 3–55.

Salemans, B. J. P. (2000). Building stemmas with the computer in a cladistic, neo-lachmannian, way. PhD thesis, Katholieke Universiteit Nijmegen.

Shaw, P. (2010). *Dante Alighieri Commedia, a Digital Edition*. Birmingham: Scholarly Digital Editions.

Solopova, E. (2000). *The General Prologue on CD-ROM*. Cambridge: Cambridge University Press.

Spencer, M., Bordalejo, B., Robinson, P. and Howe, C. (2003a). Analyzing the Order of Items in Manuscripts of *The Canterbury Tales*. *Computers and the Humanities*, 37: 97–109.

Spencer, M., Bordalejo, B., Wang, L., *et al*. (2003b). Analyzing the order of items in manuscripts of *The Canterbury Tales*. *Computers and the Humanities*, 37(1): 97–109.

Spencer, M., Davidson, E. A., Barbrook, A. C. and Howe, C. (2004a). Phylogenetics of artificial manuscripts. *Journal of Theoretical Biology*, 227(4): 503–11.

Spencer, M., Mooney, L. R., Barbrook, A. C., Bordalejo, B., Howe, C. J. and Robinson, P. (2004b). The effects of weighting kinds of variants. In van Reenen, P.; den Hollander, A. and van Mulken, M. (eds.), *Studies in Stemmatology II*. Amsterdam: John Benjamins.

Tanselle, G. T. (1989). *A Rationale of Textual Criticism*. Philadelphia, PA: University of Pennsylvania, pp. 14–15.

Thomas, P. (ed.) with **Bordalejo B.** (Assistant Editor). (2006). *The Nun's Priest's Tale on CD-ROM*. Birmingham: Scholarly Digital Editions.

Timpanaro, S. (1985 [1963]). *La genesi del metodo del Lachmann*. Florence: Le Monnier, 1963. Revised edition Padua: Liviana.

Trovato, P. (2014). *Everything You Always Wanted to Know about Lachmann's Method*. Padua: Libreriauniversitaria.it edizioni.

Notes

1 This essay was born of a series of workshops, known as Studia Stemmatologica, that took place between 2010 and 2012 and led by a three Finnish scholars, Tuomas Heikkilä, Teemu Roos, and Petri Myllymäki, who had been developing their own system to build stemmata. Before the first workshop, it seemed that a discussion of the future of stemmatics would be relatively easy as the participating scholars appeared to agree in both methodology and expectations. However, during the course of the first meeting, it became clear that some within the group had a very different view of the subject ranging from rejection of computerized tools to rejection of all the developments in the field since Karl Lachmann and his Anglo-Saxon voice, Paul Maas. It appeared that the disagreements started at the most basic level, with terminology, and continued to completely divergent theoretical understandings of how texts are transmitted and what approaches should be used to study their transmission.

2 W. W. Gregg, Fredson Bowers, and G. Thomas Tanselle have all been focused on the recovery of authorial intention. Jerome McGann initiated a counter-current in the 80s (along the lines of D. F. McKenzie) that attempted to bring forward the different aspects of book production into the business of editing.

3 For more on this subject see also McLaverty (1984).

4 In my own research, I avoid using the concept of 'the text of the work'. Instead, I refer to the work in a similar way as the one proposed by Paul Eggert in *Securing the Past*: we agree that it is a 'regulative concept' that allows us to classify and study a set of documents which we see as containing distinct texts of what we identify as a single work.

5 Robins (d'Oriente, 2007b) continuously refers to cladistic approaches, never making the distinction between cladistics and phylogenetics. In its original sense, cladistic methods were a part of phylogenetic systematics (also the title of the book by Willi Hennigs, first translated into English in 1966). In stemmatology, the confusion seems to have originated with O'Hara and Robinson (1992,1993) and perpetuated by Salemans (1996, 2000). An abstract of a conference presentation by Robins on 'Cladistics and Italian Philogology' (given at various venues in 2012 and 2014) suggests that his views are now close to the arguments presented in this essay: that one must supplement the 'quantitative' data drawn from analysis of the textual variation with 'qualitative' data drawn from study of the whole 'storia della tradizione'. On Robins' changing views, cf. Trovato 2014, p. 201.

6 This is true both of hand-made and computer-assisted stemmata since, no matter how a stemma is built, there are assumptions that underlie any approach.

7 For more details on the problem of bifurcation, see Phillips-Rodriguez *et al.* (2010) and for answers to general criticisms of the use of phylogenetics in stemmatology, see Howe *et al.* (2012).

8 Another possible solution would be to use methods that can handle multifurcating trees. One such method and which also can deal with latent tree structures (placing witnesses as internal nodes of the tree) is SemStem (Roos and Zuo 2011).

9 If we were dealing with other documents besides manuscript, like printed books, for example, it might be better to refer to a 'document tradition'.

10 VBase is available as part of all the Canterbury Tales Project's publications.

11 PAUP stands for Phylogenetic Analysis Using Parsimony and Other Methods and it is software developed by David Swofford. http://paup.csit.fsu.edu/

12 In 1923, Mario Casella produced an edition of the *Divine Comedy* in which he '. . . articulated the first tentative theory of manuscript relationships' (Shaw, Introduction, An Overview). The alpha family had three witnesses, Trivulziano, Martiano, and Laurenziano di Santa Croce. The beta family was formed by the rest of the manuscripts of the *Divine Comedy* (Shaw, Introduction, An Overview).

13 However, as textual critics, we might find ourselves in a better position to hypothesize about the missing aspects of a textual tradition, since this data set is, by definition, restricted to text only and to the accidents and happenings that can occur to it.