



Open Library of Humanities



Part of the Ubiquity  
Partner Network



---

## Research

**How to Cite:** Bordalejo, Barbara and Adam Vázquez. 2021. "You're Collating Just Fine and Other Lies You've Been Telling Yourself." *Digital Medievalist* 14(1): 6, pp. 1–46. DOI: <https://doi.org/10.16995/dm.8066>

**Published:** 15 December 2021

---

## Peer Review:

This is a peer-reviewed article in *Digital Medievalist*, a journal published by the Open Library of Humanities.

## Copyright:

© 2021 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

## Open Access:

*Digital Medievalist* is a peer-reviewed open access journal.

## Digital Preservation:

The Open Library of Humanities and all its journals are digitally preserved in the CLOCKSS scholarly archive service.

---



## RESEARCH

# You're Collating Just Fine and Other Lies You've Been Telling Yourself

Barbara Bordalejo and Adam Vázquez

University of Saskatchewan, CA

Corresponding author: Barbara Bordalejo ([barbara.bordalejo@usask.ca](mailto:barbara.bordalejo@usask.ca))

---

Although textual scholars agree that collation is a crucial component of the editing process, it often goes undefined and only briefly explained. This article defines the term, explains different kinds of collation, and explores some of its applications. We emphasize stemmatology and medieval textual traditions. By drawing from editorial examples and the theoretical frameworks of projects centred on works such as the *Canterbury Tales*, *Troilus and Criseyde*, Dante's *Commedia* and the Greek New Testament, the article seeks to compare manual and computer-assisted approaches to collation methods. We delineate the scope of this activity and argue that computer-assisted collation minimizes the risk of missing out on relevant data. We examine the advantages of full-text collation over sample collation and conclude that no decisions about stemmatically significant variation can be made a priori and that variant distribution is the major factor weighing on significance.

---

**Keywords:** Collation; Stemmatology; Variation; Textual Scholarship; Chaucer; Computer-Assisted Methods

---

## 1. Introduction

§1 Although most theorists and textual scholars refer to collation in one way or another, they do so in passing, as if anyone coming upon their texts would understand unequivocally their meaning, and no one should ever require further explanations. This article examines the concepts behind the word collation, with a focus on textual collation, stressing the fundamental considerations for the optimization of computer-assisted collation. Because our research interest resides in the investigation of textual filiation, this article emphasizes the stemmatological purposes of our collations, describing and questioning some procedures. To conclude,

we restate that editions depend on a limited number of variants. The methods used for identification and selection of variants through collation affects the researcher's understanding of textual relationships and her perception of the text.

§2 In researching this article, we found that a significant number of scholars write the word *collation* and emphasize the importance of the process without going into further details about the concept. We are not the first ones to point this out. In a piece published in 2017, Elena Spadini states: "If the reasons why we collate are well known, the way we do it, especially when we do it manually, is less documented: handbooks and essays seem to take for granted this delicate task or summarize it in a couple of sentences." (Spadini 2017, 245). Although Spadini is not strictly right in this assertion (see our discussion of David C. Parker's advice for manual collation below), the spirit of what she writes resonates with anyone researching collation theory.

§3 In 2016, Tara Andrews presented, as part of the preliminary DiXiT activities before the conference of the European Society for Textual Scholarship in Antwerp, a paper which was eventually printed (in the same volume where the article by Spadini can be found) under the title "What we talk about when we talk about collation" (Andrews 2017). With its four printed pages, this is one of the most substantial considerations surrounding collation that we have been able to find. Andrews relies on the entry for "Collation," as found in the *Lexicon of Scholarly Editing* (Nury, 2017), which partly explains the heterogeneity of her references. The authors of record are as diverse as Grésillon, Hockey, and Plachta. They include Biblical scholars (Colwell and Tune), Victorianists (Shillingsburg), Modernists (Eggert), and Armeniologists (Andrews herself). And yet, as Andrews seems to acknowledge, the mere mention of the word collation does not imply an explicit definition of the concept but rather the implicit understanding that scholars know what it is and how to carry it out. Indeed, more often than not, authors mention the process without pausing to reflect on the meaning of collation in the context of their own work. It is left to others to extrapolate what is being said and what is its context. Like the authors of the *Lexicon*, we have also found references to collation in the context of textual scholarship (Blecuca 1983; Gabler 2007; Parker 2008; Waltz 2013; Trovato 2014; Bordalejo et al.

2014; Bordalejo 2014; Driscoll and Pierazzo 2016; Bordalejo 2018; Fischer 2019), the vast majority of which live up to Spadini's description of articles that make a note of collation but never elaborate.

## 2. What is collation?

§4 To collate is to compare by close examination. The word, however, has different specialized meanings, even among textual scholars. One can collate books or collate texts; one may speak of horizontal or vertical collation (Williams and Abbott 2009, 92). It is also possible to collate documents. Although these terms are not in any way obscure for specialists, it is reasonable to include them briefly as part of this article for two reasons: first, they could serve as a reference point for the textually curious; and second, they clarify exactly where the emphasis of our argument lies.

§5 Some textual scholars are mainly bibliographers, i.e. their main interest resides in studying the physical characteristics of the book. Like book historians, they seek to learn about the book as a material object. When bibliographers use the term collation, they are talking about bibliographical description. In the context of analytical and descriptive bibliography, collation refers to the accounting for the quire structure that constitutes the physical form of the codex (McKerrow 1927; Greetham 1994; Bowers 1995; Gaskell 2000). This is expressed in a collation formula, "a shorthand note of all the gatherings, individual leaves, and cancels as they occur in the ideal copy" (Gaskell 2000, 328). Bibliographical collation, as fascinating as it is, does not concern us for the purposes of this article.

§6 In textual criticism, collation refers to the systematic comparison of two or more texts with the aid of a base-text (a text specifically used as a referencing system) or without one. Those texts could have come to be by different means: copied by scribes, printed in a manual press, or reproduced by a mechanical press. There are two types of text collation with a very different focus: vertical and horizontal.

§7 Horizontal collation occurs while one is comparing different instances of the same print, a technique used by bibliographers to compare different copies of printed books, often using optical collators but now also replicating the optical

processes by digital means. The objective is to detect differences that might shed light on the material history of the print to detect stop-press variants or resettings. The variants uncovered by this type of collation are not *transmissional* (as is the case with manuscripts copied by scribes) but *revisional* if errors are detected during the printing process, or *remedial*, when they are the result of a forme resetting due to accidents occurred during the book production process (Plachta 1995, 504–506).

§8 An example of revisional variation can be found in Mari Agata's research. Agata detected stop-press variants in the Gutenberg Bible from which she inferred that the paper print predated the vellum print (Agata 2006; 2011). It makes sense, as the less expensive material was used as a sort of preliminary state before the more expensive production was undertaken. For her research, Agata used digital methods, which included semi-transparent images in Photoshop and fast animation alternating images in Macromedia Director. The first method emulated what purely optical collators (McLeod's or Hailey's collators) would do, while the latter produced results similar to the optomechanical alternating lights of the Hinman collator (Hinman 1955).

§9 Vertical collation investigates successive textual stages, usually within manuscript traditions or pre-print documents, instances likely to present a less than clear chronology to guide research. There are different reasons to carry out vertical collations, which can be related to linguistic or textual matters. This article focuses on vertical collation and considers some of its potential applications with a particular emphasis on collation for stemmatological and textual analysis. Nearly thirty years of research by the Canterbury Tales Project seek to achieve a thorough understanding of its textual transmission. This complex task requires many steps of which transcription and collation are fundamental.

### 3. Defining variation

§10 In order to compare texts effectively, we have to define what is a variant. According to Vinaver's analysis of the "movements" that constitute the act of copying, scribal variants come to be if any of the following steps is ineffective: "(a) the reading of the text; (b) the passage of the eye from the text to the copy; (c) the

writing of the copy; and (d) the passage of the eye from the copy back to the text" (Vinaver 1976, 142; Blecua 1983, 16–17). Then, an initial categorization of the nature of variants resides in the now-classic division of substantive and accidental. W.W. Greg coined the terms in his well-known essay, "The Rationale of Copy-Text" (Greg 1950, 21). According to Greg, substantive variants are ones that "affect the author's meaning," this division responds to the way in which scribes or compositors "may in general be expected to react" (Greg 1950, 21). Greg argues scribes and typesetters reproduce substantive readings, while they "will normally follow their own habits or inclination" (Greg 1950, 22) in reference to accidentals. João A. Hansen in Volume 5 of his edition to Gregório de Matos e Guerra's poetry, expands on Greg's thoughts and hypothesizes that accidentals are more likely to be changed in the copying process because they are not considered to be as related to the author's intention as substantives (Hansen and Moreira 2013). Daniel O'Donnell adapts Greg's for his edition of *Cædmon's Hymn*, where he divides them in orthographic, substantive, and (potentially) stemmatically significant (O'Donnell [2005] 2018, §7.6 to §7.9). Textual scholarship emphasizes the distinction between these three categories, yet when it comes to editing projects, the definition of a variant changes.

§11 Greg presents a definition *a priori*, a model that distinguishes between relevant and irrelevant variants. This is also followed by other critics such as Ben Salemans. By making value decisions before hand, they are trying to establish principles allowing anyone to understand what a variant is and apply that notion of variant to any text. This process of defining variation *a priori* results either in a series of vague suggestions or in a prescription which cannot be applied to every textual tradition.

§12 Salemans lists characteristics of non-significant variants (Salemans 2000, 68–70). Significance in this specific context refers to the potential stemmatical relevance these variants may carry in so far as their usefulness in establishing a genealogy of witnesses within a textual tradition. The following, according to Salemans, are non-significant variants: differences in capitalization, orthographic variants, dialectological variants, punctuation, word division, difference in clause

headers, ungrammatical sentences, nonsense readings, evident copy mistakes (Karel de Grote vs Krl de Grote), names, archaisms, frequently used word, synonymous parallelism, and inflectional parallelism. Paolo Trovato appears to agree with Salemans' attempt to distinguish between *variants*, that are numerous, polygenetic, and irrelevant, and *significant errors*, which, according to him, are as a rule few, can derive from previous copies, thus, are useful for the construction of a *stemma* (Trovato 2014, 110). After analyzing Salemans' list, Trovato considers the following as significant variants: variation in word order, following rhyming conventions in verse, addition or omission of words when they are not small or very common (Trovato 2014, 111). Although O'Donnell makes a similar distinction, when he describes significant variants:

[The] apparatus entries include only those forms that might be understood as involving a change in metrical, lexical or syntactic "significance" from the editorial lemma: i.e. variation involving the substitution of one lexical form, metrical pattern, or syntactic construction for another, or the irreversible destruction of sense, metre, or syntax. Such substitutions include variation between contextually appropriate alternatives[...] and contextually inappropriate alternatives and nonsense forms that cannot easily be restored to the archetypal form. (O'Donnell [2005] 2018, §7.8)

By referring to "potentially" significant variants, O'Donnell suggests that the judgement on significance will be made at a later point; otherwise, there would be no need for the qualifier. Trovato follows Caterina Brandoli's description of polygenetic variants: those independently produced by different scribes, not transmitted from a common ancestor (Brandoli 2007). Manly and Rickert (1940) refer to this phenomenon as agreement by coincidence. Since Brandoli's results largely agree with Saleman's (Trovato 2014, 220), Trovato's conclusions aspire to be comprehensive and applicable to every textual tradition. It is an argument that favours the judgement of variants *a priori*: the editor decides what will be relevant for stemmatological purposes ahead of the actual collation.

§13 In her doctoral dissertation, Bordalejo outlines the criteria she used for collating the texts of Caxton's printed editions of the *Canterbury Tales*: "I have considered as significant all additions, deletions and substitutions, all the changes in word-order, all substantive variants [as opposed to Greg's accidental variants]" (Bordalejo 2002, 104). She agrees with some of Salemans' categories, but at no point does Bordalejo indicate that these criteria are applicable to any other textual tradition or even to other aspects of the study of the *Canterbury Tales*. It is more productive to describe the type of variation taken into account during the collation process and how the results of the collation were employed in the creation of the apparatus. In this sense, it is preferable to treat cautiously and as a suggestion any list that distinguishes *a priori* between stemmatically significant and non-significant variants. Vázquez, in his article "Transcribing and Collating for Digital Stemmatics" (Vázquez Forthcoming), shows examples of how some of the items in Salemans' list can be questioned. Furthermore, Bordalejo's approach aims to postpone the judgment of variants since the advances on stemmatics, and the use of digital tools do not require the *a priori* judgment of readings that would likely belong to the archetype (Bordalejo 2002, 98). This emphasis on "potentially" stemmatically significant variation is similar in approach, if not in substance, to what O'Donnell proposes. Although the judgment of variants is postponed, it must be emphasized that it is not abandoned; but only after and through the lens of the analysis of all the variants is it possible to make legitimate claims about the genealogy of the witnesses of a textual tradition (Bordalejo 2002, 99). Hence, the difference between prescriptive and descriptive collation for stemmatics: the distinction between assuming that one knows what should be taken into account versus allowing the textual evidence to speak for itself while also being aware of the objectives of an individual project. In the case of the Canterbury Tales Project, research concentrates on stemmatically significant variation containing genetic information.

#### **4. The purposes of collation**

§14 In the context of textual criticism focused on medieval materials, collation can have one or more end goals. Scholars might be concerned with the range of variation present in a series of texts, they might want to achieve a better understanding of



the relationships between different instances of a text, to explain the historical or linguistic circumstances surrounding a textual tradition, to understand the development of a text, or they might be seeking to isolate readings to include in their editions. These are just a few examples of what can be achieved by collation. Each scholar can choose to focus on some aspects more than others, or she can use multiple simultaneous approaches. Although Andrews states that “[t]he comparison may be done at the word level, at the character level, or at another unspecified syntactic or semantic level, according to the sensibilities of the editor,” (Andrews 2017, 232) we maintain that the choice has little to do with “sensibility” and everything to do with knowledge and the nature of the extant primary documents.

### **a) Apparatus**

§15 A common goal of collation is the production of a critical apparatus for inclusion in an edition. The type of edition (historico-critical, genetic, reader's edition) is not as relevant as it is the dialectic relationship between apparatus and text. Marina Buzzoni makes this distinction when she states:

The apparatus is indeed different from the descriptive *lectio variorum* one can get by, say, applying any collation software to the transcription of the witnesses. The apparatus is critical—i.e. interpretative—in that it accommodates certain variant readings, and excludes some others, according to the editorial principles to which the philologist conforms. (Buzzoni 2016, 76)

She goes on to explain that while stemmatologists might only record those variants that preserve genetic information (excluding singletons, for example), anyone interested in the linguistic features or language evolution, might include the *lectiones singulares* as a record of a philologically significant moment. This was particularly pertinent in light of the discussions of a possible change of name for the TEI Critical Apparatus group. Suggestions like “textual variance” or “textual variants” do not carry the same weight or implications of “critical apparatus.” The relationship of the latter with the text it informs, differs from the output of a collation process (Buzzoni 2016, 76).

### **b) Stemmatology**

§16 For stemmatological purposes, transcription and collation form a solid base from which scholars can investigate the relationships between different witnesses within a textual tradition. To research textual filiation, scholars must first decide which variants are going to be deemed stemmatologically significant. And yet, every editorial decision made before that point affects the potential results.

§17 A thorough understanding of matters related to the textual tradition builds a trustworthy foundation from which collation decisions can be made. A dependable collation should, in turn, provide the basis for further research. This can be carried out by hand or, as is the case of the Canterbury Tales Project with a combination of phylogenetic analysis and advanced database searches (see Bordalejo 2020). Below, we refer to various projects using collation for stemmatological purposes: the *Commedia*, the *Canterbury Tales*, *Troilus and Criseyde*, and the Greek New Testament. (A collation for stemmatological purposes requires, at least in the case of medieval texts, a process of regularization and alignment. [See below the section on Chaucer's *Canterbury Tales*.])

### **c) Popular lyrics**

§18 Margit Frenk's *Nuevo Corpus de la antigua lírica popular hispánica* is, as the title indicates, a *corpus* of traditional lyrics from the 15th to the 17th century. The purpose of the *corpus* is to register traditional lyrics. Given the traditional nature of these poetic compositions, there is no author. As the compiler explains in her book *Entre la voz y el silencio*:

The collectivity possesses a poetic-musical tradition, a limited amount of melodic and rhythmic resources, of literary motives and topics, of metrical and stylistic modes (limited amount of resources does not mean small or reduced). The author must move within these limited resources so that his composition is accepted as well as the innumerable individuals that will retouch it and transform it as time goes by. There is a very small margin left for originality and innovation, although these are not excluded. (Frenk 1971, 11)

The authorial figure and the uniqueness of his or her expression and the intention of his or her artistic endeavor is not the focus in this quotation. Since this is a *corpus* of traditional lyrics, every intervention is equally important. It belongs to the community. The nature of these texts guides the goal that the editorial criteria obey.

§19 Frenk states that this work is “something like a critical edition” (Frenk 2003, 23). The main concerns are the “fidelity to the original sources, the scrupulous attestation of variants, and the coherence of the editing criteria” (Frenk 2003, 23). For Frenk, a *version* of a lyric is “every occurrence of a lyric, either in different sources, in the same source, whether it presents variation in regard to the “base text,” or if it is identical to it” (Frenk 2003, 24 n23). The focus is on the recurrence of the lyric, not on originality. Frenk discusses the selection of a base text by stating, “given that this is an oral tradition, there is no version that can be considered either the first or the best –initially, they all carry the same value–” (Frenk 2003, 25–26). Therefore, the criteria for selecting a base text has to do with commonality: “thus, the preferred feature has been to choose the version that has the most features in common in respect to the others, so that it reflects what may have been the preponderant version through the xvi and xvii centuries as well as to reduce the apparatus” (Frenk 2003, 26). Frenk collates the text of the different versions to meticulously register the variants in the apparatus, but not to reconstruct an original because that notion is alien to this type of literature (see also Bordalejo 2020). Aside from registering the variants in the apparatus so that the reader can reconstruct the text of the witnesses, Frenk explains the apparatus has a chronological direction so that it is possible “to build an idea of the temporal trajectory of the changes” (Frenk 2003, 26). The *corpus* registers more than 2000 traditional lyrics in a two-volume edition that have around 1000 pages each. Needless to say that the physical format makes it a difficult *corpus* to work with because of the size, plus one needs to get acquainted with the relation that the base text and the apparatus hold. If this project had started in an era where digital publication was available, it is possible that the *Nuevo Corpus* would have the format and the appearance of a digital variorum edition. In essence, it is a variorum since it attempts to represent as many versions of a text as possible to enable the reader to navigate the dynamics and results of traditional lyrics. It is a celebration

of difference, rather than a judgement of what belongs to the critical text and what must be relegated to the apparatus. Frenk uses the tools from textual criticism to present lyrics in a universe where the recurrence of the poems and chronological order are more important than authors and originality. The purpose and the nature of texts guide how and why to collate.

## **5. Different approaches to collation**

§20 The importance of manual collation methods resides in the fact that, conceptually, they also form a natural basis for our understanding of computer collation. For example, the use of a base-text as a referencing system that allows us to record the diverging variation behind some types of automatic collation, while pair-comparison of texts forms the basis of other systems. Collation is an error-prone activity, manual collation is more so. Blecua suggests that collation should be done by more than one individual (Blecua 1983, 45). Full-text collations based on complete transcriptions minimize the risk of ignoring relevant evidence. We examine a few examples of manual collation to show how they relate to computer-assisted methods and how they compare to them.

§21 We also compare sampling versus the collation of complete transcriptions. The general procedure for sampling collation is to make a selection of a few significant witnesses, which might then be transcribed and collated in full. Nury argues that “[i]n practice, when dealing with a large manuscript tradition, only a few witnesses will be fully collated, i.e., the ones selected as the most relevant to edit the text” (Nury 2018, 68). However, there are instances in which it is possible to find examples in which the collation of complete transcriptions has produced different results than the ones achieved by sampling collation.

### ***a. Manual collation technique explained***

§22 David C. Parker's *An Introduction to the Greek New Testament Manuscripts and their Texts* (Parker 2008) presents a detailed description of the process of manual collation, as there might be cases in which a scholar is circumstantially forced to carry out the process without the help of a computer (Parker 2008, 95–100). Although Parker focuses on the *Greek New Testament* and so some of his references

are only relevant for the edition of said text, most of the described techniques should be useful for anyone attempting to produce hand-written collations of manuscripts or printed editions. Parker makes an important distinction between collation and transcription, where he states:

The general rule is that you are recording data that will be of value in establishing the relationship between the manuscripts, not all the data about the manuscript you are collating. If you wanted to do the latter, you would make a transcription. (Parker 2008, 97)

This separation of transcription and collation, so clearly drawn for the purposes of a hand-written collation, is also essential for our understanding of how a digital collation is carried out. In Parker's view, a collation of the *Greek New Testament* should not include punctuation, accentuation, italicisms, spelling variations, or abbreviations (Parker 2008, 97). Although it might appear as if Parker were being prescriptive, the reality is that he has been working on the manuscripts of the *Greek New Testament* since he began his career, and with computer-assisted collation for some 25 years. Parker is describing the collations produced within his John 18 project and, in fact, describing how his digital transcripts are regularized when they are processed using collation software, translating his procedures into a possible manual collation. Moreover, he does not suggest that his exclusions apply to other texts beyond the one he is studying.

***b. Sampling vs complete text collation. Loci and deteriores***

§23 In order to decide what to collate, long before digital collation, there were two routes that can be illustrated with Barbi's *loci* and Petrocci's edition, that is, limiting the variational places subjected to analysis or limiting the number of witnesses due to extratextual criteria. According to Blecua, when there is no textual tradition, or the previous analyses are not trustworthy, the collation of test passages should be done in order to choose a text to serve as the base for collation. He adds that when the textual tradition is abundant, and a complete collation is close to impossible, the *loci critici* method can be applied. This approach examines passages that have been

considered problematic. Blecua further states that the purpose of this method is to eliminate *codices descripti* or *deteriores*, but he warns that although the method is fast, it can be dangerous (Blecua 1983, 44).

§24 In 1891, Michele Barbi published in the *Bulletino della Società Dantesca Italiana* his 400 *loci critici* (actually 396) to further the analysis of the textual tradition of Dante's masterpiece. The efficacy of Barbi's *loci* has been called into question. Trovato warns his readers about the transitory purpose of the *loci*, thus states that:

Barbi was fully aware that this was not the only necessary collation; it was merely a preliminary operation, but one that would be sufficient for picking out from the vastness of the tradition, on the one hand, all mss. belonging to the most numerous families, on the other, the more promising isolated witnesses—independent of or predating the formation of the vulgate texts—which could then be subjected to a more “accurate examination.” (Trovato 2014, 305)

§25 Trovato refers to Barbi when he stated that nothing certain could be concluded after a partial examination given the “oscillation” he saw (Trovato 2014, 305). Doubts about Barbi's *loci* accuracy are not new. Robinson reports that Petrocchi “warned explicitly against the use of the Barbi *loci* as a base for the textual reconstruction of the whole tradition” (Robinson 2012, 24) because any partial collation can only offer “a preliminary and general sense of direction within the thicket of relationships among the manuscripts” (Robinson 2012, 25), something Barbi had expressed.

§26 Sebastiano Timpanaro argued against the *collazioni saltuarie* in favour of systematic collation (Timpanaro 1985, 6, 11,30–31). He considered it a mistake to only collate predetermined places of variation while he thought that the collation of complete texts would yield more solid results. Greetham describes the process as a “trial collation” (Greetham 1994, 365) used to establish an initial base-text. Such process entails a collation in two stages: first, to establish the base-text; second, to establish filiation. We refer to this first process as sampling collation (what Timpanaro calls *collazioni saltuarie*) in recognition that many scholars do this

without the intention of identifying a base-text or ever moving towards the collation of complete texts.

§27 Giorgio Petrocchi offered a different solution. He did not aim to provide the authorial text of the *Commedia*. His edition *La commedia secondo l'antica vulgata* was intended to represent "the standard edition of the 1330s" (Cherchi 2008, 413). Petrocchi never called it a critical edition since the text was based on 27 manuscripts (there are more than 800 extant witnesses for the *Commedia*). Petrocchi acknowledged that "...the ways and the velocity in which the *Commedia* was circulated in the first thirty years [after Dante's death] compromised the faithfulness of the copies compared to the original text" (Vergani 1969, 192). Boccaccio's intervention in the *Commedia*'s textual tradition produced "untrustworthy editions of the poem around the middle of the Trecento" (Cherchi 2008, 413). Boccaccio created his own recensions borrowing variants from different manuscripts and performing *emmendatio ope ingeni* when deemed preferable (Petrocchi 1966, 9). For Petrocchi, "Boccaccio contaminates more than he corrects" (Petrocchi 1966, 45). According to Petrocchi, everything after 1355 was at risk of contamination. With this rationale, Petrocchi created a *stemma codicum* which guided his editorial decisions and his edition for the *Società Dantesca Italiana* that "for decades has been the 'critical edition' of the poem, superseding Vandelli's edition of 1921" (Cherchi 2008, 413).

§28 Trovato and Angelo E. Mecca are skeptical of setting Boccaccio's recensions as a chronological limit. Mecca concludes in his "Il canone editoriale dell'antica vulgata di Giorgio Petrocchi e le edizioni dantesche del Boccaccio" that: "even though Giovanni Boccaccio is famous as a great author and a lover of Dante, he did not have equal success in the manuscript tradition of Dante's text" (Mecca 2013, 181–182). Thus, he does not think the textual tradition should be divided as *before* and *after* the author of *Filostrato*. Trovato, on the other hand, states that Petrocchi's 27 have been mistaken to represent the whole pre-Boccaccio tradition, "of which it is merely a random sample (the surviving mss. that meet Petrocchi's definition of 'antica vulgata' are close to one hundred)" (Trovato 2014, 311). It does not seem to be a mere random selection, as in that same text, Trovato wrote just before that a lot of Petrocchi's conclusions regarding the tradition, such as the importance of Urb "still hold water

today" (Trovato 2014, 311). Vatican Library ms. Urbinate latino 366 has become a controversial witness in the scholarship of the *Commedia's* textual tradition.

### ***c. The Commedia, Sanguineti and Shaw***

§29 Peter Robinson, in "The Textual Tradition of Dante's *Commedia* and the Barbi 'loci,'" reviews the differences between the treatment of variants in Shaw's and Federico Sanguineti's editions of the *Commedia*. According to Robinson:

Sanguineti declared that not only could traditional stemmatics be applied to the whole tradition, but he had done it. He had looked at all the Barbi *loci* in every one of the 800 manuscripts (so achieving on his own, with virtually no support, what scholars had failed to achieve in over a hundred years), and from analysis of the readings at these *loci* he had created a comprehensive account of the whole tradition, and isolated just seven manuscripts as necessary and sufficient for the creation of a critical text. (Robinson 2012, 5)

Shaw and Robinson dubbed these witnesses the Sanguineti seven. Then, Shaw's digital edition (Dante 2010), which had begun as a collaboration with Sanguineti, "became a test of Sanguineti's arguments about the relationships among these seven manuscripts" (Robinson 2012, 6). Sanguineti argues that out of all the witnesses that have been divided into *tradizione*  $\alpha$  and *tradizione*  $\beta$ , only the text of Vatican Library ms. Urbinate latino 366 (Urb) was a good representative of *tradizione*  $\beta$ . On the other hand, Prue Shaw's edition of the *Commedia* argues that there is a common ancestor between Urb and Ms. Riccardiano 1005 (Roddewig n. 302) (Rb). The reason for the discrepancy is that a high proportion of the variants considered by Shaw as demonstrative of genealogical relevance "would not satisfy Barbi's criteria" (Robinson 2012, 28). It is not the *a priori* appearance of these variants what supports their argument, but the consistency of the agreements between these witnesses which reveals a pattern, or as Petrocchi would put it and Shaw quotes, the "foltezza di statistica" (Robinson 2012, 29). The discrepancy is the result of the collation of complete texts versus the use of *loci*. This reveals the conflicting assumptions underlying Barbi and Shaw: that the editor can predetermine what is



stematologically relevant or that the textual evidence will show the relationships borne by the witnesses. The definition of what is a stemmatically significant variant evinces differences in method and attitude towards textual criticism. Shaw, with the help of a team led by Robinson, carried on a computer-assisted collation (which should not be confused with automatic collation). Each word was individually regularized, according to Shaw's direction, and later aligned with the help of Jennifer Marshall. Both regularization and alignment are under editorial control to ensure the collation is optimal for stemmatological purposes.

§30 Robinson presents five readings that are "likely to have been introduced by the common ancestor of Urb/Rb; none of these five lines appear among the Barbi *loci*" (Robinson 2012, 27). These are the variants.

Inf. i 89: aiutami da lei, famoso saggio,	
famoso e saggio	LauSC Rb Urb FS
famoso saggio	Ash Ham Mart Triv PET

Inf. ii 71: vegno del loco ove tornar disio;	
di	LauSC Rb Urb FS
del	Ash Mart Triv PET
dal	Ham

Inf. ii 110: a far lor pro o a fuggir lor danno,	
pro e a	Mart-orig Rb Urb FS
pro ne a	Ash LauSC-c2 Triv
prode et a	Ham
pro [...] a	LauSC-orig
pro o a	Mart-c2 PET

Inf. iii 3: per me si va tra la perduta gente.	
ne la	Rb Urb FS
tra la	Ash Ham LauSC Mart Triv-c1 PET
tra	Triv-orig

Inf. iii 22: Quivi sospiri, pianti e alti guai

altri            Ash-orig Rb Urb

alti             Ash-c2 Ham LauSC Mart Triv FS PET

(Robinson 2012, 27)

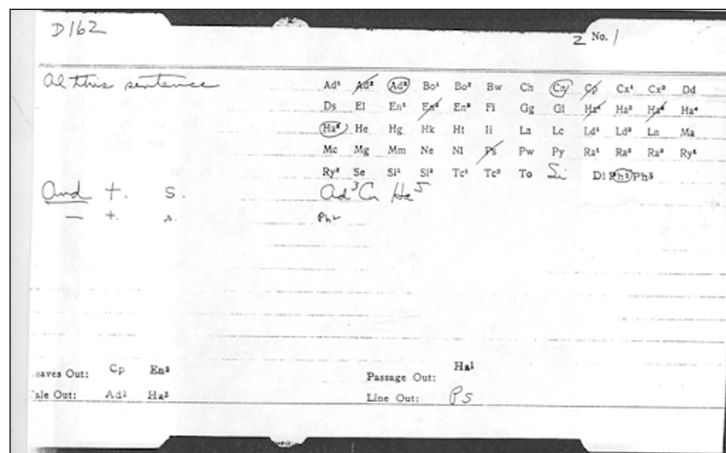
Robinson understands why these variants would not catch Barbi's attention: they might have appeared as polygenetic readings to him, errors "liable to arise independently in independent manuscripts" (Robinson 2012, 28). His work demonstrates *a priori* judgments should not be made; the weight should be placed on the variant distribution across texts in a tradition. Assumptions about which variation is important and how to recognize and isolate that importance are central for the contrasting approaches between Barbi and Petrocchi, but also between Shaw and Sanguineti. Brandoli, examines the passages that Barbi and Petrocchi considered useful to constitute a *stemma* through the lens of various definitions of what constitutes polygenetic and monogenetic readings. It is an attempt to systematize what is genealogically relevant *a priori*. She concludes that out of the 396 Barbi *loci*, 366 fit her definition of monogenetic variants, thus are stemmatologically relevant (Brandoli 2007, 113). As for Petrocchi's key passages, she states that 282 out of 477 seem to be monogenetic, which is still a majority but not as decisive as Barbi's. Then she points out that out of those 282, 132 were already considered by Barbi (Brandoli 2007, 122). According to Brandoli, Petrocchi identified 150 relevant readings. The rest were polygenetic variants that should be of no interest and would compromise his results. This difference extends to the discrepancy between Sanguineti's edition of the *Commedia* and Shaw's and corresponds to the distinction between prescriptive and descriptive collation for stemmatology. The textual analysis conducted by Shaw's edition rests on full-text transcriptions and computer-assisted full-text collation, which minimizes the risk of missing any data by trying to analyze the tradition with manual methods and a limited set of variants. Shaw demonstrates that there is no *a priori* valid method to judge variants, makes the process as transparent as possible, and reports the ties that the textual evidence shows, instead of the assumptions of the editor in detriment of what he did not judge to be relevant.

**d. Chaucer's *The Canterbury Tales***

§31 For their collation of the *Canterbury Tales*, John Manly and Edith Rickert (Manly and Rickert 1940) recorded information in some 50.000 cards. Each card registers a number (D162) on the top left corner, corresponding to individual lines of the *Canterbury Tales* (in this case, "Al this sentence/me liketh eury del"). However, the indication, in the right-hand corner, that this card is one of two explains why we are only dealing with the first half of the line. The bottom of the card explains why some witnesses have left this particular line out, i.e. we find a record of whether the line, the passage or the tale are not present. The sigils of the witnesses with variants (Cn, Ha5, Ad3 and Ph2) are encircled with a pen, and the variation is recorded in **Figure 1**.

§32 It is a deceptively simple system, so effective, that it allowed the editors a degree of accuracy that has not been sufficiently recognized (the *Canterbury Tales* Project preliminary research suggests a punctilious precision beyond what one might expect from the tools available to the collators). Those of us who have carried out computer-assisted collations of the *Tales* can attest to the level of work that can be found in the Manly and Rickert volumes. The exactness of this collation illustrates the importance of thoughtful consideration of these most basic matters.

§33 The *Canterbury Tales* Project (along with any other projects its leaders have been involved in) produces complete encoded transcriptions of each witness of the *Tales*. The project attempts to improve over the manual work carried out by



**Figure 1:** Collation card by Manly and Rickert.

Manly and Rickert in the 1920s and 1930s, not because their edition is imprecise, but because their interpretation of the *Canterbury Tales'* vast variants corpus created problems that muddled their understanding of the textual filiation in a sea of variation.

§34 The main goal of the Canterbury Tales Project is to understand the textual history of the *Tales* as fully as possible with the information currently available. Under Bordalejo's leadership, the project is pressing on with the production of edited texts which will be included in our publications (the first instance of an edited text appearing in a project publication was Bordalejo's General Prologue reading text in the *CantApp* [Chaucer 2020]), but this does not alter the original goal of understanding the relationships between the different texts in the extant fifteenth-century witnesses of the *Canterbury Tales*.

§35 From the beginning, the Canterbury Tales Project's transcriptions' were conceived to retain information which would be useful for stemmatic analysis, even though the project has always made an emphasis on retaining scribal spellings which must be regularized during collation in order to produce meaningful stemmatological results (see Bittner and Dase 2021; Bordalejo 2016, 2020). The project's original transcription guidelines (Robinson and Solopova 2020) retained tails and flourishes that might have stood in place of final *e*. As the project carried on, it became clear that these flourishes were merely ornamental. So, they were finally excluded from the transcriptions. Moreover, the implementation of separate encoding based on the one used for the *Divine Comedy* (Bordalejo 2010) to account for the text of the document and the variant states of the text (Bordalejo 2016) also meant a modification of the transcription guidelines which culminated in its current version (Bordalejo and Robinson 2018). In this way, the project produces rich and detailed transcriptions that record places of variation within each document (with the use of the apparatus element) and which can be published alongside the images, thus allowing readers access to the same resources we use in our analysis.

§36 Our work is designed for generosity. We want others to benefit from the many hours spent on the creation of individual transcripts, which is why we make those available for reuse. Since a significant portion of our research was funded with

money from various governmental sources, we have the duty to make them available for others to use in their research. Perhaps other scholars are interested in analyzing idiosyncratic peculiarities or quirky spellings. Someone interested, for example, in toponymy, could encode all the place names and repurpose our transcriptions to build maps; or they could encode historical figures or characters for a different type of study.

§37 For our own purposes, if our initial transcription fails to record an individual place of variation, it can be easily modified and processed again. Because we have complete transcriptions, we can choose to publish editions of individual witnesses (Chaucer 2000; Bordalejo 2003) or of multiple ones (Chaucer 2004; 2006; 2020). We can rebuild our transcriptions for different purposes and select what version of the text and in which format will be displayed (see Bittner and Dase's description of the multiple encoding of our apparatus tag).

§38 If there is a downside to the full-text transcription is the trap of the illusion of control. The transcription of complete witnesses postpones the detection of variants since it does not require to make decisions *a priori*. Moreover, the researcher might feel some relief as she waits for the eventual results of the computer-assisted collation. There is safety in not having to jump in a vast sea of variation with decisions being made at the moment. Postponing judgment offers the opportunity to study and analyse variation without fear of having left anything important out.

§39 As we collate the *Canterbury Tales*, we carry out two further processes: regularization and alignment. The aim of these processes is two-fold. On the one hand, we seek to flatten spelling, (which we have found to be not significant for stemmatological purposes) information in order to analyze the collation results with the help of evolutionary biology software (see Bordalejo 2021). As we have pointed out before, the project seeks to do a genealogical analysis of the textual tradition, and spelling variation obscures the filiation of the witnesses. On the other hand, Bordalejo and Robinson intend to produce a readable apparatus that can be easily understood by human reading. In the same way that the project's transcription guidelines have evolved, so have our regularization ones. The original guidelines for regularization indicated that the different witnesses should be regularized to a very lightly edited version of Hengwrt (National Library of Wales, Peniarth, 392 D), likely to be the oldest

manuscript of the *Tales* and witness to one of the best texts in the tradition. But the words “lightly edited” mean nothing unless further specified. By lightly edited, it is meant that all medieval characters, no longer in use in contemporary English, are substituted by modern forms, abbreviations are expanded, and all lines found in any witness not present in the base-text added. This created the Canterbury Tales Project's base-text for collation, which Collate used for comparison purposes. Later, as the editions are built, the base-text for collation is discarded.

§40 Although the base-text is of little interest otherwise, it is useful because it anchors the spellings used for regularization purposes. The Canterbury Tales Project's rule for this is to regularize to the most common spelling of Hengwrt. There is also a list of preferred witnesses in case this turns out not to be possible (in cases in which Hengwrt does not include a particular term, for example). Elsewhere Bordalejo and Robinson explain the concept that underlines their view on variation:

It is our core conviction, based on decades of work with digital tools, that “significant variants” are defined entirely by how the variants are distributed across the whole tradition. That is: if we find a number of variants which are present, over and over again, in the same distinctive pattern of witnesses, then those variants are significant. (Bordalejo and Robinson 2019, 37)

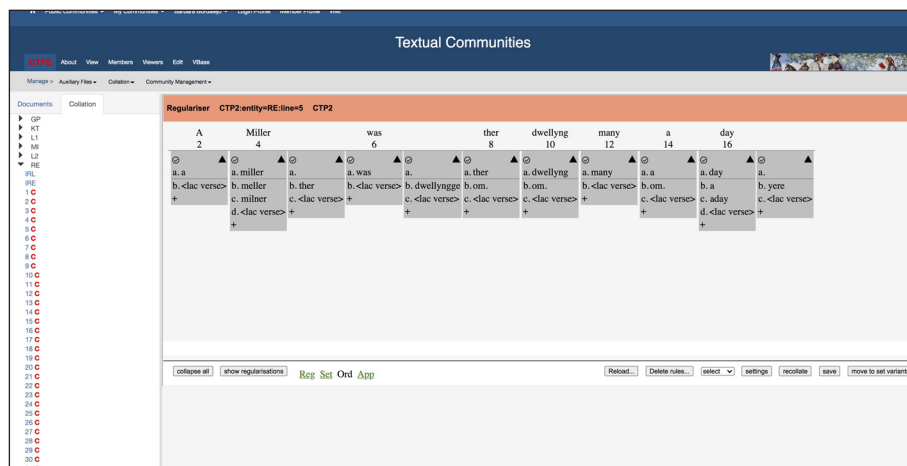
For the purposes of the Canterbury Tales Project, whether variants are polygenetic or monogenetic is not crucial. Instead, the project considers variant distribution, which can only be known *a posteriori*, the essential factor which determines stemmatically significant variation.

§41 The collations, carried out with the Textual Communities modification of Catherine Smith's Collation Editor, itself based on CollateX (Smith 2019), present similar challenges. The collations identify stemmatically significant variation which includes all text present in some witnesses and not others, changes in order, substitutions, and all substantive variation. From our collations, we seek to both produce files that can be used with evolutionary biology (such as PAUP [Swofford 2003]) or stemmatological software (RHM [T. Roos and Heikkila 2009] or SemStem [Teemu Roos and Zou 2011]).

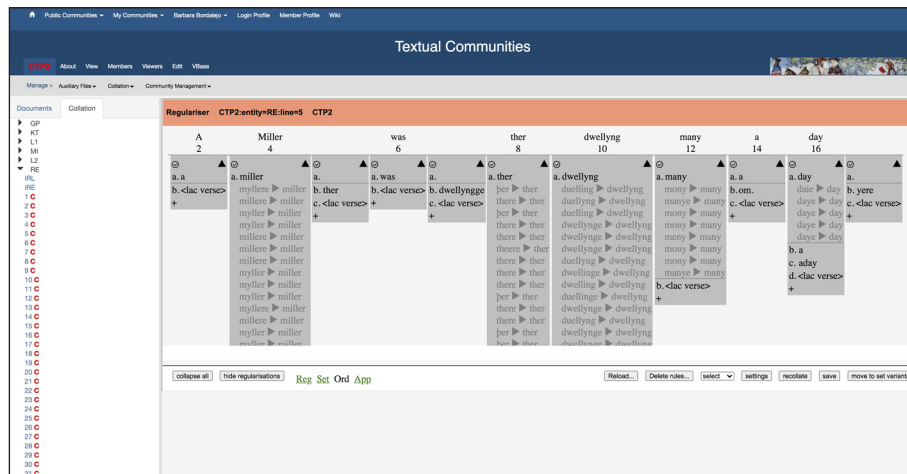
§42 For the purposes of the Canterbury Tales Project, a fundamental step during the collation process is the regularization and alignment of variants. Variants are regularized and aligned to produce a better apparatus: one that is both more readable, presenting the information as clearly as possible, and gives the most effective representation of variation at each point for phylogenetic analysis. Let us consider, for example, line 5 of the Reeve's Tale (recently collated by Thomas Farrell, see Chaucer 2021).

§43 Here we have the collated text, with place of variation 4 showing the readings *miller/meller/milner* in **Figure 2**, which are of dialectological importance and preserved by Farrell in the current collation. These, however, do not represent spelling variation which, in accordance with the Canterbury Tales Project's collation guidelines, are levelled during the collation process.

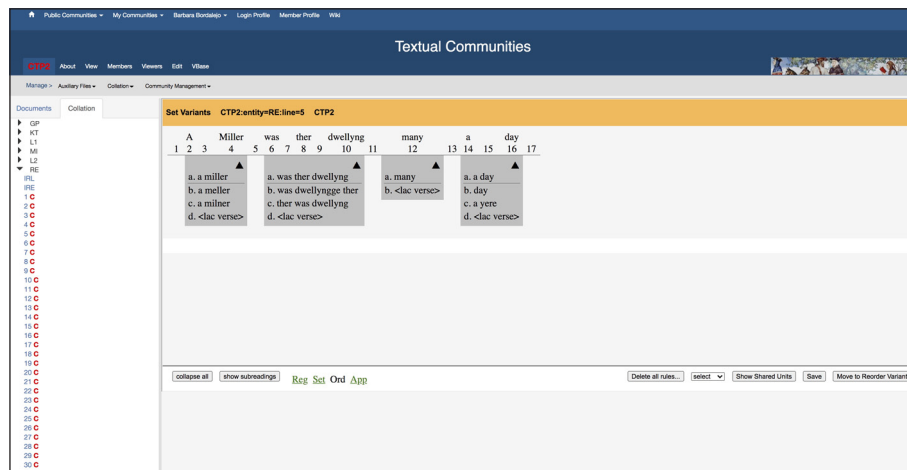
§44 The multiple spellings of each word are of no significance from a stemmatological perspective, so they are regularized to a lemma (generally the form of the base-text, although the project has specific protocols for when this is not feasible) (see **Figure 3**). After the process of regularization is complete, it is followed by variant alignment (**Figure 4**), during which the editor optimizes the apparatus for use with phylogenetic software as well as for ease of readability. In place of variation 2, we find that *a* presents no stemmatically significant variation. For this



**Figure 2:** Textual Communities' integration of the Collation Editor and CollateX showing a regularized version of the line, The Reeve's Tale Line 5, as collated by Farrell (Chaucer 2021).



**Figure 3:** Textual Communities' integration of the Collation Editor and CollateX showing spellings which have been regularized, The Reeve's Tale Line 5, as collated by Farrell (Chaucer 2021).

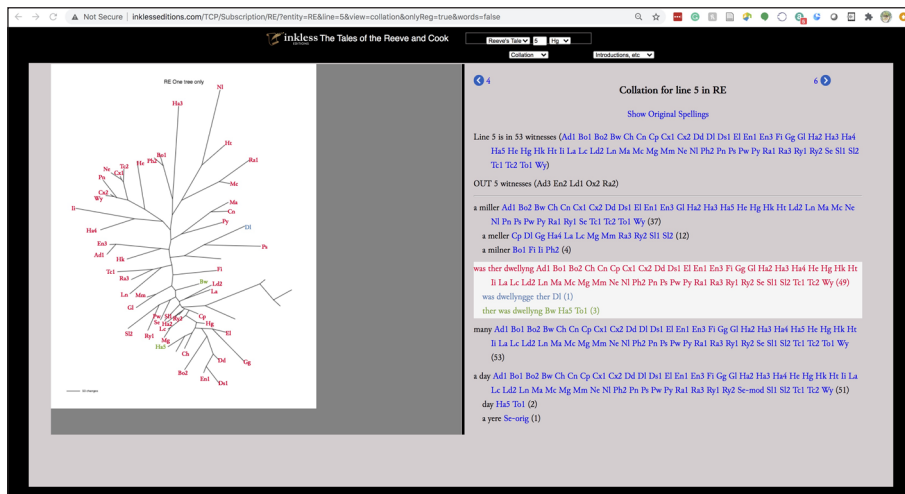


**Figure 4:** Textual Communities' integration of the Collation Editor and CollateX showing the setting of variants, The Reeve's Tale Line 5, as collated by Farrell (Chaucer 2021).

reason, the editor has aligned it with *miller*, so the apparatus shows three variants for the combination of places of variation 2 and 4:

- a miller
- a meller
- a milner





**Figure 5:** The Tales of the Reeve and the Cook, Variant map and apparatus for The Reeve’s Tale Line 5 (Chaucer 2021).

These decisions affect the way in which the apparatus will be displayed as part of the editions, and how it will be processed in phylogenetic analysis. Consider the example in **Figure 5**, which shows both the Reeve’s stemma as well as the corresponding apparatus. The decision to display *a* and *millar* as a phrase-variant makes the apparatus clear and does not obscure variation. The following variant phrase, *was ther dwelling*, collates against both *was dwelling ther* and *ther was dwelling*. The transposition of *was* and *ther* in D1 (Takamiya MS 24, ex Delamere) encourages the longer phrase, which becomes both a more readable alternative while presenting all of the information and a more efficient representation of the variation for stemmatic analysis.

§45 The process of collation and alignment relies on a series of decisions based on internal project protocols: they are labour intensive and intellectually demanding. This is the reason why we insist on referring to computer-assisted collation, rather than expressing ourselves in ways that might suggest that the process is automatized. Just as for the Canterbury Tales Project processes of transcription and encoding, the project leaders have spent countless hours developing the collation protocols used by the Canterbury Tales Project. We should also emphasize that, once the regularization and alignment are complete, both require revisions and adjustments before they can be used in formal publication.

### ***e. Chaucer's Troilus and Criseyde***

§46 There is no stemma for the textual tradition of *Troilus and Criseyde*. Previous editions and analysis (Chaucer 1926; 1984; 2008) have stated that it is not possible to conduct traditional recension due to the constant changes in filiation of the witnesses. Thus, this is an opportunity for analysis with the aid of digital tools. For the initial stages of the recension of the textual tradition of Chaucer's *Troilus and Criseyde*, Adam Vázquez decided to transcribe and collate the first 546 lines of Book One, as well as lines 764–833 Book One, and 490–1225 Book Two, from the 16 manuscripts and two early printed editions. The purpose of the project is to conduct phylogenetic analysis. The first 546 lines work as a point of reference since we know thanks to the work of past scholars that, from line 547 to the end of the poem. Wynkyn de Worde follows Caxton's edition but not for lines 1–546, that is, there is a change in filiation. Then the analysis of lines 764–833 show Wynkyn changing place in the phylogram. That proves the efficacy of the method. Then, the analysis of lines II. 490–1225 attempts to shed light on this problematic excerpt of Book two, given the changes in filiation of some witnesses that have been investigated previously by various scholars (Root 1926; Hanna 1996).

§47 This method postpones judgment. The project does not seek to make general claims on the textual tradition based on the analysis of 1350 lines. The aim is to analyze the selected excerpts rigorously. The transcription of the excerpts and their complete collation enables the textual critic to get acquainted with the material. The stemmatological aspect of the project does not rest on the pre-selection of a few variants. As seen with the discrepancies between Barbi and Petrocchi, or Shaw and Sanguineti, variants that would escape the attention of the manual collator and alter the direction of genealogy, are less likely to be missed by a textual critic that uses digital tools. The *Troilus* project is far from complete, but it shows consistent results so far (Vázquez 2020).

§48 By comparing the critical apparatus that Barry Windeatt and Robert K. Root provide in their editions with a computer-assisted collation, we can see that there is a small omission as early as line four. The line reads "fro wo to wele, and after out of ioie" (Chaucer 1984, 84). All the witnesses but one agree on "out of." Rawlinson

reads “on to.” Neither apparatus registers this variant. It is a small variant, it is also not relevant for genealogical purposes since it is not present in any other witness, yet one may argue that since both their editions walk away from making genealogical statements, the apparatus is there to inform the reader of what is present in the textual tradition. On another occasion, Windeatt’s apparatus draws attention to the fact that the text in the Corpus manuscript presents a peculiar spelling in the third line, thus: “*auentures] auentuirs* Corpus (3 minims after *t*)” (Chaucer 1984, 85). Thus, a lack of interest in small detail is not to blame for Windeatt’s disregard for the Rawlinson variant: the most likely explanation for this omission is that it is easy to miss when doing manual collation. A digital collation tool that draws variants from full-text transcriptions cannot fail to bring this reading to the collator’s attention.

#### ***f. The Greek New Testament, Nestle-Aland, and the Editio Critica Maior***

§49 The Institute for New Testament Textual Research has collected the necessary materials of the textual tradition of the *Greek New Testament*. It was then necessary to filter the material so that “new views about important manuscripts find their way into the minor editions of the institute, and finally to present an *Editio Critica Maior*” (Mink 2004, 17). In order to access the relevance of the text of manuscripts, a sampling method was devised and the results published “in the five volumes of the *Text und Textwert der Griechischen Handschriften des Neuen Testaments*” (Morrill 2012, 7). It meant to separate the manuscripts that contained “the relatively uniform text which was standard at the end of the Byzantine tradition from the still large number of manuscripts which must be considered relevant on account of their deviations from the majority text” (Mink 2004, 17). The results made it possible to select manuscripts that do not contain the uniform text from the end of the tradition.

§50 The Claremont Profile Method that was used for the classification of 1385 manuscripts of the Gospel According to St. Luke (Morrill 2012, 23) is another example of sampling collation. The test passages were picked out after “the complete collation of 282 manuscripts. A sample of three chapters, 1, 10, and 20, were selected, and all variation in these chapters was evaluated” (Morrill 2012, 21). After careful consideration, 196 were used to analyze 816 more manuscripts “for a total of 1385

manuscripts in 196 passages" (Morrill 2012, 23), and fourteen groups of witnesses were created after the analysis. The purpose of this collation was not to create an apparatus, not an edition. The reasoning was that the "critical apparatus would adequately represent the manuscript tradition if it included representatives of the groups, plus those manuscripts that did not fall into definable groups" (Morrill 2012, 19). It cannot be denied that the Claremont Profile Method achieved remarkable results, yet it will always be preferable to rely on full-text transcriptions and full-text collations.

§51 The Institute for New Testament Textual Research has used computer-assisted collation tools for more than 20 years. First Collate and now CollateX have played a significant role in the development of their editions. Both the Nestle-Aland *Greek New Testament* and the *Editio Critica Maior* built their research and apparatuses with the help of semi-automated software after carrying out complete transcriptions of witnesses (Houghton et al. 2020).

§52 Bordalejo has before described the *Editio Critica Maior* as a born-digital printed edition, in reference to the techniques used in its construction and as an example of how similar digital textual critical tools are to their analogue counterparts, despite their speed and higher accuracy (Bordalejo 2013, 65n). What is remarkable about this edition is how its apparatus, based on a computer-assisted collation, differs from that of the Nestle-Aland edition. Despite the same tools being employed by both, the distinct purposes of the editions are made evident in the apparatuses generated from very different collations.

## 6. Digital collation tools

§53 There are several tools that can be used in order to conduct a computer-assisted collation. In this section, we examine some of the tools with collation features, their characteristics, and their potential uses.

### *i. TUSTEP*

§54 TUSTEP (TUebingen System of TExt processing Programs) is a toolbox for scholarly processing textual data. According to Gabler, the "algorithm was devised 30 years ago and is still among the most powerful and efficient of collation" (Gabler 2007, 4). A demonstration of TXSTEP in 2015 shows the algorithm at work. It shows

```

71 <compare-units mode="word"><!-- normalize: ignore upper and lower case; ß = ss = f = s, c = k, i = ie = y, th
71 = 1, ignore punctuation marks -->
72 > <normalize><exchange-table><exception-string>eie</exception-string><string-pair><!-- would else become ei
72 -->
    
```

Figure 6: Normalize function in TXSTEP.

3 Varianten in 11 Textzeugen	
Und leider auch Theologie!	
Und leider auch <u>die</u> Theologie	1 H.5
Und leider auch Theologie	S
Und leider auch Theologie!	A, D.1
Und leider auch Theologie.	T 5 H.1
Und leider auch Theologie!	B, D.2, B.a, 1 H.1, C.1 12, C.3 12, Text

Figure 7: Line with three variants from the Faustedition.

versions of the text, but it seems like one has to program subroutines so that the collation shows significant variants instead of different spellings (like the German long s) or punctuation marks (see Figure 6).

§55 While this can work for texts that have a consistent orthography, it is not likely that a scholar dealing with non-fluid spelling systems could foresee a set of exceptions that would satisfy the needs of the collation. Nevertheless, TUSTEP has been fundamental for the creation of the FaustEdition (2020). Here a small snippet of the variation that a reader can see when reading the text (See Figure 7).

**ii. Juxta**

§56 A similar thing could be said of Juxta or JuxtaCommons, the online accessible version of Juxta. While it may be easier to use and the interface is friendlier, Juxta will be handy for people that are interested in collating texts with unified orthographies. Juxta, as their example shows, is an efficient way to show differences between documents (See Figure 8).

§57 Juxta highlights the variation place in question. It is not perfect, but the accuracy of the alignment of variants is remarkable. It also calculates the difference between two or more witnesses, it encodes the witnesses using TEI tags, and that information can be visualized in the experimental Edition Starter or in their experimental Versioning Machine. Unlike TUSTEP, there does not seem to be an

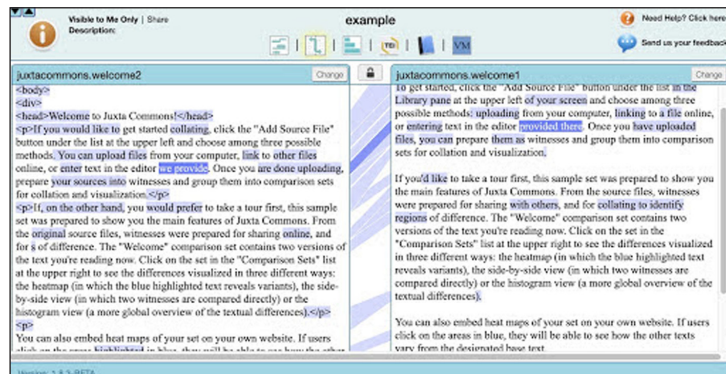


Figure 8: Side-by-side view in Juxta.

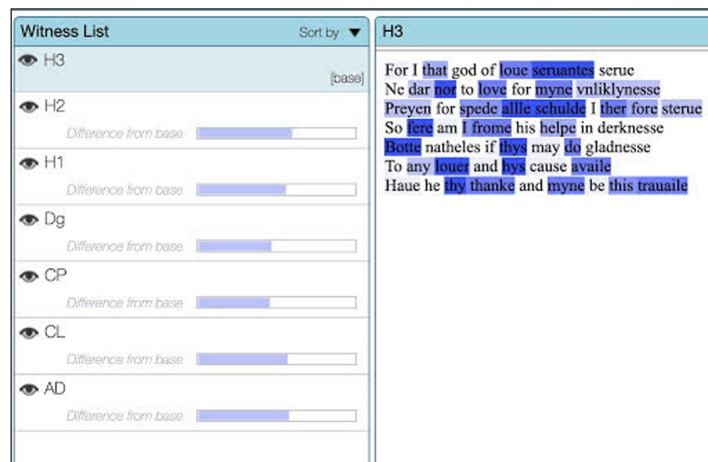


Figure 9: Heat map in Juxta.

option for personalized rules. I used Juxta to collate the third stanza of *Troilus and Criseyde*, and I selected eight witnesses for this exercise (See Figure 9).

§58 Juxta highlights the places of variation and differentiates degrees of variation by using different hues of blue. It is not possible to see more than two witnesses in the side by side view. The TEI tagging process took a long time, and even though these are eight witnesses, it should also be said that this is just one stanza. It would be next to impossible to use this feature with the 18 witnesses or with more than seven lines. But the main problem is that for a textual scholar that studies texts and deals with non-fluid spelling systems, this tool is almost unusable. It is good at showing difference, which is not the same as variation, or not the variation that

might be interesting for a textual scholar. In other words, although the alignment is accurate, variation cannot be properly displayed because there is no option for regularization. Differences in orthography are important, especially if they are indicative of dialectological information, or if one is interested in studying particular manifestations of different *usus scribendi*, but it is not useful for the creation of a critical edition that researches the genealogy of the witnesses.

§59 The variation of line seven in **Figure 10** is meaningful since pronouns are dissimilar in the witnesses, and it changes the sense of the stanza. Still, it is not apparent in this collation since orthographic and substantive variants are treated equally. One can manually get rid of the insignificant variants, but one wonders if this is an ideal way to collate and edit.

§60 Juxta also includes an experimental Versioning Machine that is reminiscent of the project with the same name by Susan Schreibman. The Versioning Machine 5.0 was released in 2016. The website has sample texts so that one can see the capabilities of the software. The reader can compare as many witnesses as their screen can fit. By hovering the cursor on top of a line, or a word, the Versioning Machine highlights the equivalent word, phrase, or line in other witnesses. It is at its core, a collation visualizer, in the sense that it aids in the comparison of places of variation the editor has defined manually as such. It provides valuable information if one is interested in authorial states of composition. It is not intended as a collation tool, since the editorial collation precedes what the Versioning Machine visualizes. This is why it can work in tandem with Juxta since the program provides the encoding and the Versioning Machine the visualization.

TC	
For I that god of loues seruauntz serue	7 Haue he] W1, W3, W4, W5, W6, W7; <i>not in</i> W2
Ne dar to loue for myn vnlikynesse	7 my] W2, W3, W5, W6, W7; chy W1; the W4
Preyen for speed al shoalde I ther fore sterue	7 thonk] W5; thank W4, W7; thanke W1; þank W6; þonk W3; lover W2
So fer am I from his help in derknesse	7 thonk] W5; lover or my love auaille / Haue he thanke and W2
But natheles if this may don gladnesse	7 myn] W2, W3, W5, W6, W7; myne W4, W1
To any louere and his cause auaille	7 he] W1, W2, W4, W5, W6, W7; by W3
Haue he my thonk and myn be this trauaille	7 this] W1, W4, W5; his W6, W7; the W2; þis W3
	7 trauaille] W3, W5, W6; trauaile W4, W1; trauaile W2; trauayle W7

**Figure 10:** Variants in Juxta.

### ***iii. Collate, CollateX, and the Collation Editor***

§61 Up to 2010, the Canterbury Tales Project used the Collate software tool (Robinson, 2000) developed by Peter Robinson. The experience of using Collate was a window into the mind of its developer and evinced Robinson's understanding of textual criticism and scholarly editing.

§62 Collate is complex because the tasks editors require the software to perform are complex. Preparing the files for use with Collate was a necessary prerequisite if one intended to get the best results and, although the software handled ASCII files, the text was processed more precisely if a light encoding was included to mark textual sections (Robinson 1994, 32). For most of the *Canterbury Tales*, this means to add line numbers within different poetry sections (Blake 1996). Collate, even in its second iteration, Collate 2 (Robinson 2000), was not optimized to compare witnesses differing in large sections of text and was clumsy when dealing with transpositions. By 2005, when Apple announced they were discontinuing support of OS9 because of the move to Intel processors, it became clear that Collate needed to be replaced.

§63 Robinson reports that it was around that time that he started talking to Fotis Jannidis and Joris Van Zundert, who would under the InterEdition Cost Action eventually develop Collate's successor under the leadership of Ronald Dekker (Robinson 2014). Robinson offers a detailed explanation of the desirable characteristics of Collate's replacement. The outlined requirements, some dating as far back as February 2007, made their way into CollateX.

§64 In that same document, Robinson painstakingly describes the series of tests for alignment undertaken by Collate in order to show its results (Robinson 2014). Further, Robinson emphasizes the distinction between alignment and variant identification. This distinction is crucial because although the software is able to align the text, only an editor can make the decisions that lead to variant identification. Again, this is the reason why our theoretical understanding of variation is fundamental: does capitalization count as a variant for the purposes of our collation? Does punctuation? The decisions made, at every previous point,



particularly during the transcription process, of what to record have a profound impact on alignment, while the theoretical framework of the edition and the textual context colour editorial decisions.

§65 The official development of CollateX is said to have started in 2010, despite being part of the InterEdition agenda for three years before that. However, it is possible that they are just referring to the specific timeline for the development of the code. We learn that:

CollateX was planned as a complete rewrite of Collate that was primarily addressing the architectural challenges of its predecessor. Over the years though and with more and more participants contributing their requirements and ideas, it developed a different agenda. On the one hand, Collate is a complete solution for producing a critical apparatus, with features ranging from its very own algorithm for comparing versions of a text to a powerful graphical user interface that lets the user control the collation process. On the other hand, CollateX has become a software component which can be embedded into other software or be made a part of a software system. Its goal is the provision and advancement of current research in the field of computer-supported collation involving natural language texts. (The Interedition Development Group 2020)

A fundamental difference between Collate and CollateX is that the former was a standalone programme, while the latter is a software component. Users can choose between embedding CollateX within other systems, as Smith did for the Collation Editor, which was in turn modified and embedded in Textual Communities (Robinson, 2000; Smith 2019), or download the stand-alone version-controlled from a command line. Today, CollateX is the most widely used specialized tool for textual scholars engaged in the collation of large textual traditions. For medievalists and other scholars working on pre-modern texts, CollateX presents another challenge as the software is not optimized to deal with erratic spelling variations and the peculiar word divisions of languages in development. The regularization and alignment the

Canterbury Tales Project relies upon are only possible thanks to the integration of the modified Collation Editor, a tool that supplies features that were present in Collate and Collate 2 but are not part of CollateX.

## **7. Conclusions: Learning from collation**

§66 Collation is not an isolated process that happens in a vacuum, but a practice occurring within the wider context of textual-critical research and which, at times, leads to the production of an edition. It is the process of variant identification that allows scholars to further their research agenda. However, editions depend on a limited number of variants. Even in the case of digital editions, where every piece of variation can be included, we are forced to reckon with the reality represented by our limited number of texts. Human intellect is also limited, particularly when it refers to a large number of items and, for this reason, we rely on other systems to help us process the vast number of variants we detect during a regular collation process.

§67 We have shown that, although the collation process relies on the same principles, whether it is carried out manually or with the aid of computers, there are significant advantages in using computer-assisted methods over manual ones. Although the preparation of files for computer collation requires a significant investment of time and effort, by creating full-text transcriptions and making them publicly available, we ensure that our work can be evaluated by other scholars and reused in future research. These advantages are more evident in the context of research on large textual traditions, but they do not disappear in reference to briefer or less distributed texts.

§68 The success of a critical edition relies on its ability to connect a system of data. With computer-assisted collation methods and complete text transcriptions, the process that leads to a critical text becomes comprehensive, thorough, and more transparent to the reader. In consequence, the critical text turns into a window through which we can observe the circumstances and the intervention of many of the agents that made it possible for us to engage with the text of the documents.

### ***Exemplum: The Miller's Tale: Manual vs computer-assisted collation***

-Señor conde Lucanor -dixo Patronio-, mucho me plaze desto que dezides,  
et para que vós mejor lo podades fazer, plazerme ya que sopiédeses  
lo que consteçió a un muy grand philósopho et mucho ançiano.  
(Don Juan Manuel, *El Conde Lucanor*)

§69 This research was carried out in the years after the publication of *The Miller's Tale on CD-ROM*. It is offered here as an example of how much more accurate the computer is. This, naturally, could be an isolated instance of carelessness. Bordalejo is aware that this text exposes her and her research in ways that are not often privy to others.

§70 During the preparation of Bordalejo's De Montfort University Ph.D. thesis, she used Collate 2 to compare the encoded transcriptions of Caxton's first (Cx1) and second (Cx2) editions of the *Canterbury Tales*.

§71 Both of the Caxton transcriptions used in Bordalejo's 2003 study attempted to represent spelling as accurately as possible but ignored form distinctions between ragged and Roman r or long and round s. The system overlooked other distinguishable sorts, including tailed d and ligatures, because it was primarily developed for the transcription of manuscript materials and not for incunables.

§72 Because Bordalejo's work focused on isolating potentially stemmatically significant variants, it required her to separate those from the orthographic and graphetic variation included in the transcriptions by default. The Canterbury Tales Project single-tale editions use software record and save (in a separate file as did Collate or in a database as is the case of CollateX) regularizations of orthographic variants (see Farrell's article on his work on dialectological variation in *The Reeve's Tale*). This process ensures that the transcriptions remain a close reflection of what is found in the source document.

§73 At the time, Bordalejo did not carry out a complete regularization of Caxton's editions. The regularization process would have allowed to hide all accidental variation (such as spelling and punctuation) and would have shown only

variation at word level, including modifications in word-order and other substantive changes. She collated Caxton editions using the raw transcriptions (without the help of a regularization file). The result of this process was an extremely long list of differences between Cx1 and Cx2. The lists were printed out and read before deciding whether further analysis of the variants would be required. In **Figure 11**, one can see an example of a collation of both editions, using Cx1 as a base, of the first lines of the General Prologue.

§74 The Collate output shows all the differences between the two editions, even those that only represent a difference in the type (“breth/breth; with/wyth”) or encoding (“[3orncp]W[/3orncp]Han/[4orncp]W[/4orncp]Han”). These variants were eliminated from the analysis, as they were of no use to trace the affiliations

```
<L 1>
[3orncp]W[/3orncp]Han ] [4orncp]W[/4orncp]Han Cx2
Aprill ] Apryll Cx2
with ] wyth Cx2
his ] hys Cx2
<L 2>
And the ] The Cx2
pcid þe ] percyd the Cx2
<L 3>
badid ] bathyd Cx2
licoue ] lycour Cx2
<L 4>
whiche ] whyche Cx2
vertu ] vertue Cx2
engendrid ] engendryd Cx2
<L 5>
zepherus ] Zepherus Cx2
with his ] wyth hys Cx2
breth ] breth Cx2
```

**Figure 11:** Caxton's second edition of the Canterbury Tales (Cx2) compared against the first edition serving as base-text for collation. Originally produced by Collate 2 and edited by Bordalejo to remove variants which were not potentially stemmatically significant.

of the source for the corrections of Cx2. A great deal of attention was required to detect those variants that could have the potential to be stemmatically significant. In the previous example, such variants are represented by “And the/The” in line 2 and “licoue/lycour” in line 3. Bordalejo read all the variants between the editions attempting to isolate those with potential stemmatic significance. Although the aim was to record the majority of the variants, and great care was devoted to the data gathering, human error remained a consideration. However, the proportion of the potential differences between the electronic and manual data gathering were not clear, not until the regularization process had been completed for the *Miller's Tale on CD-ROM*, published in 2003. This publication allowed users to retrieve the information about the differences between both of Caxton's editions in seconds. The non-computerized search, s 'manual' collation, used to gather data for Bordalejo Ph.D. yielded 79 stemmatically significant variants in *The Miller's Tale*. This number did not take into account the addition, deletion and substitution of lines, as this was being dealt with in a different section of her work where she established that there were three line substitutions, and seven lines were added in Cx2. No lines were deleted that were on Cx1 without being replaced with an alternative line. Although the data about addition, deletion and substitution of lines were easily retrieved using Collate, individual (word) variants detected were the result of reading and marking an unregularized printed collation. These accounted for 33 variants that Bordalejo was considering separately (reference to thesis) and which included lines 47, 47a, 28, 416–1, 465–1, 466, 577 to 584, and 585.

§75 By subtracting 33 from the 150 variants of the computer collation, there 117 variants of which 38 were not accounted for in the manual collation. There were several reasons which could have explained these discrepancies:

1. Transcription mistakes
2. Different segmentation
3. Incorrect regularization
4. Different understanding of variation
5. Human error

§76 The latest can be further divided into errors helped by the display in Collate, which was not optimized to be read but rather to be processed, and inaccuracy or oversight.

1. Transcription mistakes could have occurred at any point. These could have been input after Bordalejo's final collation, but before the Miller's Tale on CD-ROM was finished or they could have been corrected for the publication, in which case the mistake would have remained part of the final collation.
2. The Canterbury Tales Project uses parallel segmentation. This ensures that variant phrases are reduced to their minimum variant expression (sometimes two words but in a few cases six or seven). However, Collate just shows sequences of differences. When evaluating the Cx1/Cx2 collation, variant phrases were counted as a single variant (even in cases in which these could have been separated into more than one).
3. Because the regularization process is carried out by humans, it is prone to error. Occasionally, there has been overregularization (MI 322, MI 564, MI 619) or under regularization of the variants. There are also simple cases of misregularization when a word goes to the incorrect lemma (MI 64)
4. In very few cases, the regularization policy employed by a particular editor might not be the same that was used when detecting the variants for Bordalejo's thesis. She might have made a different regularization choice in lines MI 16, MI 37, MI 153.

§77 Analyzing these factors it is possible to see that of the 150 original variants which result of the CD-ROM's search:

- 33 are line variants considered in a different section of Bordalejo's thesis.
- 7 are the result of different segmentation (-7)
- 4 are the result of incorrect regularization (-4)
- 3 are the result of a different understanding of variation (-3)

This gives a total of 14 variants which we need to subtract to obtain 103 net variants of which 79 had been correctly isolated. This translates into 24 variants which were not detected in the original Caxton collation when reading the Collate output. Of those 24 overlooked variants:

3 are related to the Collate display

2 (MI 208 and MI 210) were errors in judgment due to lack of experience.

§78 The other 19 even though present in the original collation between the Caxton editions, were not detected. Four of those occurred in lines that have some other kind of variation, and it might be possible that their closeness to other variants might explain why they were overlooked. This leaves a total of fifteen missed variants for which we cannot find any explanation. Those fifteen variants translate into around 20% of the variation. Projecting such a number to the rest of the text would be a total of some 600 undetected variants if one were to carry out a manual collation with a steady error rate.

§79 Even with the risks and decisions made during the regularization process, its accuracy rate is much higher. Since its publication in 2004, only three regularization mistakes have been found in "The Miller's Tale." Not taking into account disagreements with the editor or parallel segmentation matters, then the machine aided collation has an inaccuracy rate of only 2%—a much better ratio than the one achieved by Bordalejo.

§80 If we assume that these numbers are representative, the case for computer-assisted collation is solid. Although it is possible to produce more accurate manual collations if researchers can dedicate time and effort to their work, it seems evident that the computer's capacity to count and classify should produce better results when a similar timeframe is allowed and provided that the data is relatively free of errors. Although computer-assisted collation is time-consuming in its data preparation, it is a superior system when limited time is available.

## **Competing Interests**

The authors have no competing interests to declare.

## **Author Contributions**

### ***Editorial***

#### **Recommending editor:**

Franz Fischer, Università Ca' Foscari Venezia, Italy

#### **Recommending referees:**

Elena Spadini, University of Lausanne, Switzerland

Klaus Wachtel, University of Munster, Germany

#### **Section/Copy/Layout editor:**

Shahina Parvin, The Journal Incubator, University of Lethbridge, Canada

### ***Authorial***

Authorship is alphabetical after the drafting author and principal technical lead. Author contributions, described using the CASRAI Credit typology, are as follows:

Barbara Bordalejo: BB

Adam Vazquez: AV

**Authors are listed in descending order by significance of contribution. The corresponding author is BB:**

Conceptualization: BB, AV

Methodology: BB, AV

Formal Analysis: BB, AV

Investigation: BB, AV

Writing – Original Draft Preparation: BB, AV

Writing – Review & Editing: BB, AV

## **References**

- Agata, Mari.** 2006. "Stop-Press Variants in the Gutenberg Bible." Graduate thesis, Tokyo: Keio University.
- . 2011. "Improvements, Corrections, and Changes in the Gutenberg Bible." In *Scribes, Printers, and the Accidentals of Their Texts*, edited by Jacob Thaisen and Hanna Rutkowska, 135–55. Frankfurt am Main; New York: Peter Lang.



- Andrews, Tara.** 2017. "What We Talk about When We Talk about Collation." In *Advances in Digital Scholarly Editing*, edited by Peter Boot, Anna Cappellotto, Wout Dillen, Franz Fischer, Aodhán Kelly, Andreas Mertgens, Anna-Maria Sichani, Elena Spadini, and Dirk Van Hulle, 233–36. Leiden: Sidestone Press. <https://www.sidestone.com/books/advances-in-digital-scholarly-editing>.
- Blake, Norman Francis.** 1996. "The New Lineation System." Edited by Norman Francis Blake and Peter M. W. Robinson. *The Canterbury Tales Project: Occasional Papers II*.
- Blecua, Alberto.** 1983. *Manual de Crítica Textual*. Literatura y Sociedad 33. Madrid: Castalia.
- Bordalejo, Barbara.** 2002. "The Manuscript Source of Caxton's Second Edition of The Canterbury Tales and Its Place in the Textual Tradition of the Tales." De Montfort University. Accessed November 12, 2020. [https://www.academia.edu/2987324/THE\\_MANUSCRIPT\\_SOURCE\\_OF\\_CAXTONS\\_SECOND\\_EDITION\\_OF\\_THE\\_CANTERBURY\\_TALES\\_AND\\_ITS\\_PLACE\\_IN\\_THE\\_TEXTUAL\\_TRADITION\\_OF\\_THE\\_TALES](https://www.academia.edu/2987324/THE_MANUSCRIPT_SOURCE_OF_CAXTONS_SECOND_EDITION_OF_THE_CANTERBURY_TALES_AND_ITS_PLACE_IN_THE_TEXTUAL_TRADITION_OF_THE_TALES).
- . ed. 2003. *Caxton's Canterbury Tales: The British Library Copies*. Leicester: Scholarly Digital Editions.
- . 2010. "The Encoding System." In *Dante Alighieri. Commedia: A Digital Edition*, edited by Prue Shaw. Birmingham and Florence: Scholarly Digital Editions and Sismel.
- . 2013. "The Texts We See and the Works We Imagine: The Shift of Focus of Textual Scholarship in the Digital Age." *Ecdotica* 10: 64–76.
- . 2014. "Caxton's Editing of the Canterbury Tales." *Papers of the Bibliographical Society of America* 108(1): 41–60. DOI: <https://doi.org/10.1086/680833>
- . 2016. "The Genealogy of Texts: Manuscript Traditions and Textual Traditions." *Digital Scholarship in the Humanities* 31(3): 563–577. DOI: <https://doi.org/10.1093/lc/fqv038>
- . 2018. "Digital versus Analogue Textual Scholarship or The Revolution Is Just in the Title." *Digital Philology: A Journal of Medieval Cultures* 7(1): 7–28. DOI: <https://doi.org/10.1353/dph.2018.0001>

- . 2021. "Well-Behaved Variants Seldom Make the Apparatus: Stemmata and Apparatus in Digital Research." *Digital Medievalist*. DOI: <https://doi.org/10.16995/dm.8065>
- Bordalejo, Barbara, Fiona Maguire, Manuel Moreno, Peter Robinson, and Dorothy Severin.** 2014. "An Electronic Corpus of Fifteenth-Century Castilian Cancionero Manuscripts." *Digital Philology: A Journal of Medieval Cultures* 3(1): 11–23. DOI: <https://doi.org/10.1353/dph.2014.0003>
- Bordalejo, Barbara, and Peter Robinson.** 2018. "New Transcription Guidelines – Wiki – Textual Community." 2018. Accessed November 12, 2020. <http://www.textualcommunities.usask.ca/web/canterbury-tales/wiki/-/wiki/Main/New+Transcription+Guidelines>.
- . 2019. "Manuscripts with Few Significant Introduced Variants." *Ecdotica* 15(March): 37–65. DOI: <https://doi.org/10.5281/zenodo.3344880>.
- Bowers, Fredson.** 1995. *Principles of Bibliographical Description*. New Castle, DE: Oak Knoll Press.
- Brandoli, Caterina.** 2007. "Due Canoni a Confronto: I Luoghi Di Barbi e Lo Scrutinio Di Petrocchi." In *Nuove Prospettive Sulla Tradizione Della Commedia*, edited by Paolo Trovato, 99–148. Florence: Casa Editrice Leo S. Olschki s.r.l.
- Buzzoni, Marina.** 2016. "4. A Protocol for Scholarly Digital Editions? The Italian Point of View." In *Digital Scholarly Editing: Theories and Practices*, edited by Matthew James Driscoll and Elena Pierazzo, 59–82. Cambridge, UK: Open Book Publishers. DOI: <https://doi.org/10.11647/OBP.0095.04>
- Chaucer, Geoffrey.** 1926. *The Book of Troilus and Criseyde*. Edited by Robert K Root. Princeton: Princeton University Press.
- . 1984. *Troilus & Criseyde: A New Edition of "The Book of Troilus."* Edited by Barry A. Windeatt. London; New York: Longman.
- . 2000. *The Hengwrt Chaucer Digital Facsimile*. Edited by Estelle Stubbs. Leicester: Scholarly Digital Editions.
- . 2004. *The Miller's Tale on CD-ROM*. Edited by Peter Robinson. Canterbury Tales Project. Leicester: Scholarly Digital Editions. Accessed November 12, 2020. <http://www.sd-editions.com/AnaAdditional/miller/images/millerhome.html>.

- . 2006. *The Nun's Priest's Tale on CD-ROM*. Edited by Paul Thomas. Leicester: Scholarly Digital Editions. <http://www.sd-editions.com/NP/index.html>.
- . 2008. *The Riverside Chaucer*. Edited by Larry Benson. 3rd ed. Oxford New York: Oxford University Press.
- . 2021. *The Tales of the Reeve and the Cook*. Edited by Thomas Farrell. Saskatoon: Inkless Editions. <http://www.inklesseditions.com/TCP/Subscription/RE/?wit=Hg&page=50r&view=Image/Text>.
- Cherchi, Paolo.** 2008. "Review Article. Nuove Prospettive Sulla Tradizione Della Commedia. Una Guida Filologico Linguistica Al Poema Dantesco." *Annali d'Italianistica* 26: 413–22.
- Dante.** 2010. *Commedia*. Edited by Prue Shaw. Birmingham: Scholarly Digital Editions. Accessed December 14, 2020. <http://sd-editions.com/AnaAdditional/commediaonline/home.html>
- Driscoll, Matthew James, and Elena Pierazzo.** eds. 2016. *Digital Scholarly Editing: Theories and Practices. Digital Humanities Series*, v. 4. Cambridge: Open Book Publishers. DOI: <https://doi.org/10.11647/OBP.0095>
- FaustEdition.** 2020. "Johann Wolfgang Goethe: Faust." Accessed November 11. <http://www.faustedition.net/>.
- Fischer, Franz.** 2019. "Digital Classical Philology and the Critical Apparatus." In *Digital Classical Philology: Ancient Greek and Latin in the Digital Revolution*, edited by Monica Berti. De Gruyter Saur. Accessed December 14, 2020. <https://www.degruyter.com/view/title/537705>.
- Frenk, Margit.** 1971. *Entre Folklore y Literatura*. México, DF: El Colegio de México.
- . 2003. *Nuevo corpus de la antigua lírica popular hispánica (siglos XV a XVII) Vol. I*. Vol. 1. 2 vols. México, DF: Facultad de Filosofía y Letras, UNAM; Colegio de México; Fondo de Cultura Económica.
- Gaskell, Philip.** 2000. *A New Introduction to Bibliography*. New Castle, Delaware: Oak Knoll Press.
- Greetham, David C.** 1994. *Textual Scholarship: An Introduction*. London, UK: Routledge.

- Greg, Walter W.** 1950. "The Rationale of Copy-Text." *Studies in Bibliography* 3: 19–36.
- Hanna, Ralph.** 1996. *Pursuing History: Middle English Manuscripts and Their Texts*. Stanford California: Stanford University Press.
- Hansen, João A., and Marcello Moreira.** 2013. *Para que todos entendias. Poesia atribuída a Gregorio de Matos e Guerra*. Belo Horizonte: Autentica Editora.
- Hinman, Charlton.** 1955. "Mechanized Collation at the Houghton Library." *Harvard Library Bulletin* 9: 132–34.
- Houghton, H. A. G., David C. Parker, Peter Robinson, and Klaus Wachtel.** 2020. "The Editio Critica Maior of the Greek New Testament." *Early Christianity* 11(1): 97. DOI: <https://doi.org/10.1628/ec-2020-0009>
- Manly, John M., and Edith Rickert.** 1940. *The Text of the Canterbury Tales Studied on the Basis of All Known Manuscripts, Manly, John M.; Rickert, Edith; with the Aid of Mabel Dean e. a.; with a Chapter on Illuminations by Margaret Rickert*. Chicago (Ill.): University of Chicago press.
- McKerrow, Ronald B.** 1927. *An Introduction to Bibliography for Literary Students*. 1960 ed. Oxford: Oxford University Press/Clarendon Press.
- Mecca, Angelo E.** 2013. "Il canone editoriale dell'antica vulgata di Giorgio Petrocchi e le edizioni Dantesche del Boccaccio." In *Nuove Prospettive Sulla Tradizione Della Commedia: Seconda serie, 2008–2013*, a c. di E. Tonello, P. Trovato, Monterotondo (RM). Padova: Libreriauniversitaria.
- Mink, Gerd.** 2004. "Problems of a Highly Contaminated Tradition: The New Testament." In *Studies in Stemmatology II*, edited by Pieter van Reenen, August den Hollander, and Margot van Mulken, 13–85. Amsterdam, Netherlands: The John Benjamins Publishing Company. Accessed November 11, 2020. <http://ebookcentral.proquest.com/lib/usask/detail.action?docID=623163>.
- Morrill, Bruce.** 2012. "A Complete Collation and Analysis of All Greek Manuscripts of John 18." PhD thesis. Birmingham: University of Birmingham.
- Nury, Elisa.** 2017. "Collation." *Lexicon of Scholarly Editing: A Multilingual for a Multilingual Discipline*. <https://lexiconse.uantwerpen.be/index.php/lexicon/collation/>.

- . 2018. "Automated Collation and Digital Editions: From Theory to Practice. Classical Studies." London: King's College London. Accessed November 18, 2020. <https://hal.archives-ouvertes.fr/tel-02493805/document>.
- O'Donnell, Daniel Paul**, ed. (2005) 2018. *Cædmon's Hymn: A Multimedia Study, Edition and Archive*. 1.1 Internet Reprint. Vol. A.8. SEENET. Accessed July 2, 2020. <https://caedmon.seenet.org/htm/frontmatter/prefaceToInternetIssue.html#howtoCite>
- Parker, D. C.** 2008. *An Introduction to the New Testament Manuscripts and Their Texts*. 1 edition. Cambridge, UK; New York: Cambridge University Press.
- Petrocchi, Giorgio.** 1966. "Criteri fondamentali dell'edizione." In *La Commedia secondo l'antica vulgata/Dante Alighieri; a cura di Giorgio Petrocchi*, by Dante Alighieri, 1a ed. Vol. 1. Milano: Mondadori.
- Plachta, Bodo.** 1995. "German Literature." In *Scholarly Editing: A Guide to Research*, edited by David. C. Greetham, 504–29. New York: Modern Language Association of America.
- Robinson, Peter.** 1994. *Collate 2: A User Guide*. Oxford: The Computers and Variant Texts Project.
- . 2012. "The Textual Tradition of Dante's *Commedia* and the Barbi's Loci." *Ecdotica* 9: 1–32.
- . 2014. "Collate 2, and the Design for Its Successor: CollateXML (Now, CollateX)." *Zenodo*. DOI: <https://doi.org/10.5281/zenodo.3902430>
- . 2020. *Textual Communities*. Accessed November 12. <https://textualcommunities.org>.
- . 2000. *Collate* (version 2). Oxford: Oxford University Press.
- Robinson, Peter Max Wilton**, and **Elizabeth Solopova.** 2020. "Original Transcription Guidelines – Canterbury Tales Project 2 – Wiki." Accessed June 30. <https://wiki.usask.ca/display>.
- Roos, Teemu**, and **Tuomas Heikkilä.** 2009. "Evaluating Methods for Computer-Assisted Stemmataology Using Artificial Benchmark Data Sets." *Literary and Linguistic Computing* 24(4): 417–33. DOI: <https://doi.org/10.1093/lc/fqp002>

- Root, Robert K.** 1926. "Introduction." In *The Book of Troilus and Criseyde*, by Geoffrey Chaucer, xi–li. Princeton: Princeton University Press.
- Salemans, Ben J. P.** 2000. *Building Stemmas with the Computer in a Cladistic, Neo-Lachmannian, Way: The Case of Fourteen Text Versions of Lanseloet van Denemerken*. Nijmegen: Nijmegen Univ. Press.
- Smith, Catherine.** 2019. Itsee-Birmingham/Collation\_editor\_core 1.0.4. *Zenodo*. DOI: <https://doi.org/10.5281/zenodo.3539578>
- Spadini, Elena.** 2017. "The Role of the Base Manuscript in the Collation of Medieval Texts." In *Advances in Digital Scholarly Editing*, edited by Peter Boot, Anna Cappellotto, Wout Dillen, Franz Fischer, Aodhán Kelly, Andreas Mertgens, Anna-Maria Sichani, Elena Spadini, and Dirk Van Hulle, 345–49. Leiden: Sidestone Press. <https://www.sidestone.com/books/advances-in-digital-scholarly-editing>.
- Swofford, David L.** 2003. *PAUP\* Phylogenetic Analysis Using Parsimony (\*and Other Methods)* (version 4.0a [build 167]). Sunderland, Massachusetts: Sinauer Associates.
- The Interedition Development Group.** 2020. "About the Project." *CollateX*. Accessed June 23. <https://collatex.net/about/>.
- Timpanaro, Sebastiano.** 1985. *La Genesi Del Metodo Del Lachmann*. 1a ristampa corr. con alcune aggiunte. Saggi 5. Padova: Liviana.
- Trovato, Paolo.** 2014. *Everything You Always Wanted to know about Lachmann's Method: A Non-standard Handbook of Genealogical Textual Criticism in the Age of Post-structuralism, Gladistics, and Copy-text*. Storie e linguaggie. Padova, Italy: Libreriauniversitaria.it Ed.
- Vázquez, Adam.** 2020. "Transcribing and Collating for Digital Stemmatology." *Digital Studies/Le Champ Numérique*.
- Vergani, Luisa.** 1969. "Review Article. La Commedia, Secondo l'antica Vulgata." *Italica* 46(2): 191–93. DOI: <https://doi.org/10.2307/477953>
- Vinaver, Eugène.** 1976. "Principles of Textual Emendation." In *Medieval Manuscripts and Textual Criticism*, edited by Christopher Kleinhenz, 139–66. Berkeley, California: University of North Carolina Press.


**Waltz, Robert B.** 2013. *The Encyclopedia of New Testament Textual Criticism*. Last Preliminary Edition. London and New York: Routledge.

**Williams, William Proctor, and Craig S. Abbott+.** 2009. *An Introduction to Bibliographical and Textual Studies*. 4th ed. New York: Modern Language Association of America.

**How to cite this article:** Bordalejo, Barbara and Adam Vázquez. 2021. "You're Collating Just Fine and Other Lies You've Been Telling Yourself." *Digital Medievalist* 14(1): 6, pp. 1-46. DOI: <https://doi.org/10.16995/dm.8066>

**Published:** 15 December 2021

**Copyright:** © 2021 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

 *Digital Medievalist* is a peer-reviewed open access journal published by Open Library of Humanities.

**OPEN ACCESS** 