



Open Library of Humanities



Part of the Ubiquity  
Partner Network

Digital Studies /  
Le champ numérique

---

## Method

**How to Cite:** Vazquez, Adam. 2021. "Transcribing and Collating for Digital Stemmatology. The Case of *Troilus and Criseyde*." *Digital Studies/Le champ numérique* 11(1): 7, pp. 1–34. DOI: <https://doi.org/10.16995/dscn.372>

**Published:** 18 October 2021

---

## Peer Review:

This is a peer-reviewed article in *Digital Studies/Le champ numérique*, 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 Studies/Le champ numérique* 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.

---

## METHOD

# Transcribing and Collating for Digital Stemmatology. The Case of *Troilus and Criseyde*

Adam Vazquez

University of Saskatchewan, CA  
[adam.vazquez@usask.ca](mailto:adam.vazquez@usask.ca)

---

This paper explores the decisions involved in a digital stemmatology project. Traditionally, transcription, collation, and the creation of a stemma have been processes linked to the edition of texts. However, stemmatology on its own can provide valuable insight to further the understanding of literary works since it sheds light on their production process. The research on the textual tradition of *Troilus and Criseyde* offers the possibility to reconsider what are the difficulties and implications of a digital project of this sort. I focus particularly on transcription and collation. Transcription can be affected by aspects such as the availability of high-quality reproductions of manuscripts and early printed editions. Then, in order to produce useful transcriptions that serve the purpose of the project, the level of detail has to be established. It is important to find a balance between the overwhelmingly detailed and scarcity that could result in unfruitful transcripts. In regard to collation, which is the process of identifying variants, I mention the characteristics and purpose of a base-text. It is essential to understand what is a variant and how to work with them so that reliable stemmata can be produced. Thus, I examine the case of two readings present in my research. By the end, I provide a brief example of how a phylogenetic tree can help us understand the distribution of variants and the relationships that the witnesses of *Troilus and Criseyde* bear. With that, I also hope that the usefulness of digital stemmatology is made evident.

---

**Keywords:** transcription; collation; stemmatology; phylogenetics; Chaucer; *Troilus*

---

Cet article explore les décisions prises dans un projet de stemmatologie numérique. Traditionnellement, la transcription, la collation et la création d'un stemma étaient des processus liés à l'édition de textes. Cependant, la stemmatologie elle-même peut fournir des informations précieuses qui contribuent à une meilleure compréhension d'œuvres littéraires, puisqu'elle éclaire leur processus de production. La recherche sur la tradition textuelle

de *Troilus et Criseyde* offre la possibilité de réexaminer les difficultés et les implications d'un projet numérique de ce genre. Je me concentre en particulier sur la transcription et la collation. La transcription peut être affectée par des aspects tels que la disponibilité de reproductions de haute qualité de manuscrits et d'éditions imprimées anciennes. Ensuite, pour produire des transcriptions utiles qui atteignent le but du projet, il faut établir le niveau de détails. Il est important de trouver l'équilibre entre des transcriptions massivement détaillées et celles qui sont peu détaillées pour éviter des résultats infructueux. En ce qui concerne la collation, le processus d'identifier des variantes, je mentionne les caractéristiques et buts d'un texte de base. Il est essentiel de comprendre ce qui c'est une variante, ainsi que comprendre la façon dont il faut la traiter pour que des stemmata fiables puissent être produits. Par conséquent, j'examine le cas de deux lectures présentes dans ma recherche. Je fournis un exemple de la façon dont un arbre phylogénétique peut nous aider à comprendre la distribution de variantes et les relations que témoignent *Troilus et Criseyde*. Dans cet esprit, j'espère aussi rendre évidente l'utilité de la stemmatologie numérique.

---

**Mots-clés:** transcription; collation; stemmatologie; phylogénétique; Chaucer; *Troilus*

---

## Introduction

Transcribing and collating are chores that generally precede the making of a critical edition. A critical edition should represent a theory about a work, as argued by many textual scholars (Pugliatti 1998, 163; Cerquiglini 1999, 80). However, in the digital age of editing, differing understandings of how best to take advantage of the tools that are in reach, as well as the general state of textual scholarship, have opened diverse options about how best to further our understanding of works, texts, and documents. Should one take advantage of every single aspect of them? In my research on the textual tradition of *Troilus and Criseyde*, transcription and collation have been critical aspects. The end goal, however, is not to produce an edition, but to add to the scholarship we have on the relationship that the eighteen witnesses of Geoffrey Chaucer's work bear through digital stemmatology. This does not mean that an edition could not take place, but the purpose of this paper is to reflect and comment on how this particular goal models the processes of transcription and collation. In that sense, this text serves both as a rationale and a place to meditate on the implications that these activities bring to a project of digital recension. *Recensio* specifically refers to the analysis of the variants of all the extant manuscripts as well as the affiliation of the later ones (Blecu 1983, 31).

*Troilus and Criseyde* was finished around 1386. This situates the poem in a crucial time in Chaucer's life, since as Paul Strohm states: "Chaucer was a man of literary accomplishment, standing on the brink of his decision to write the *Canterbury Tales*" (Strohm 2016, 184). Strohm refers to *Troilus and Criseyde* as his "first undoubted masterpiece" (Strohm 2016, 186) and believes that the body of work he had produced by then, at the age of 43 was "more than sufficient to establish him as the most eminent English writer before Shakespeare" (Strohm 2016, 186). Let us acknowledge that "Chaucer had a typically medieval attitude toward completion, in which works were valued even in fragmentary form, but this is the major poem that he *did* complete, and triumphantly so" (Strohm 2016, 189). *Troilus and Criseyde* is Chaucer's most important complete poem. The circumstances around him were also changing, as new patterns of manuscript production and of circulation of vernacular texts emerged, plus "Chaucer's

literary contemporaries were beginning to think differently about claiming credit for their work" (Strohm 2016, 185). Regarding Chaucer, Strohm believes that although he was fame worthy, by 1386 he was "certainly not famous yet" (Strohm 2016, 189). This was about to change. The final section of *Troilus and Criseyde* "registers the first signs of a shift; that poem's conclusion expresses a welter of confused and disturbing new literary ambitions and desires" (Strohm 2016, 185). Chaucer worries about the reproduction and performance of his poem because he wants to establish himself as an author (*autoritas*) among "Uirgile, Ouide, Omer, Lucan and Stace" (Chaucer 1984, V, 1792). It is because of all of these reasons, that it is of the utmost importance to approach this poem with new eyes and new goals. To work with Chaucer's *Troilus and Criseyde* and digital tools offers the opportunity to create a stemma or stemmata for different sections of the work. This is of great significance since as of today, there is no stemma for the witnesses of Chaucer's poem. To survey the text of the manuscripts and early printed editions with digital tools and to create a stemma is a definite step forward in the understanding of the textual tradition of *Troilus and Criseyde*, of Chaucer's literary production, and opens the possibilities for further projects.

As for the rationale of this work and for its position within the context of other Middle English editing projects: it is close to *The Canterbury Tales Project* on which I worked, thus learned from it to work with large textual traditions. The strategies for transcription, collation and the use of phylogenetic analysis are heavily inspired by that project. In that sense, it also agrees in a broader sense with Manly and Rickert's efforts to study the witnesses of *The Canterbury Tales* as they state in the second volume of their edition:

The purely mechanical procedure of collation will have resulted in groupings of MSS according to their readings without reference to whether the readings are correct or incorrect. These are all, *prima facie*, merely variational groups. Many of them are also in fact genetic groups [...] The test of a genetic group, as distinguished from a mere accidental grouping is that the same sigils [these represent manuscripts in their cards] should appear together persistently and consistently. (Manly et al. 1940, 20)

The collation and analysis of *Troilus* will reveal groups and their reoccurrence will reveal patterns that suggest whether these groups have a genealogical relation. Following this train of thought, this project does not share George Kane's conclusions on recension and its uselessness. Kane and Donaldson (1988) argued in the introduction of the B version of *Piers Plowman* the following:

In this situation lodges the ultimate absurdity of recension as an editorial method: to employ it the editor must have a stemma; to draw the stemma he must first edit his text by other methods. If he has not done this efficiently his stemma will be inaccurate or obscure, and his results correspondingly deficient; if he has been a successful editor he does not need a stemma, or recension, for his editing. (Kane and Donaldson 1988, 17, note 10)

Since the aim of this project is not to produce a critical text, but to shed light to the relationship that the witnesses bear, this criticism of the base text does not apply necessarily to this project. However, the methodological basis of the work I present here fundamentally disagrees with Kane's statement. Kane and Donaldson qualify recension as absurd because they portray the process as circular. They remind us that in order to collate the witnesses, a base text must be used. If the base text has scribal variation, the agreements in original readings would not be genetic. To avoid that, a base text with exclusively original readings is needed, then all the variants would be scribal. When several witnesses attest the same variant, presumptions of genealogy can be done. To obtain this ideal base text, the editor would have to create one by choosing the best manuscript and then proceed to remove the unoriginal readings with the variants offered by the rest of the manuscripts relying on his knowledge of the *usus scribendi* of both author and scribes. If this was done successfully, the base text would be the final text and the *stemma* would no longer be necessary as an editorial tool.

Kane and Donaldson's argument against recension is that they believe that originality must be detected in the beginning of the process. This is not necessarily true since agreement in variance can show allegiance and statements of originality can

be made after the examination of attestation. The process could be done manually, but technology has allowed for progress. With the use of digital tools there is no need for any *a priori* judgment of originality, and as it will be seen further, the *stemma* itself does not need a root or a center to show how the witnesses relate to each other.

Although the base text does not need to be free of scribal variation as Kane suggests, the role it plays is critical. It should not be confused with a best text in the tradition of Bédier (1928), nor with a copy-text in Greg's tradition (1950). The base text should allow for comparison between the witnesses. Let us keep in mind that *Troilus and Criseyde* is a poetic work organized in books, stanzas, and lines. Therefore, the base text must include all the possible books, stanzas, and lines that the text of all the witnesses present, or at least of the excerpts that are being analyzed. The base text is used as a reference and will make it possible to show the witnesses' agreements in variation. Its other purpose is to guide the transcription of the texts.

In this particular case I was fortunate to find an xml document with Barry Windeatt's text for his Longman edition of Chaucer's work thanks to the Oxford Text Archive (OTA) (Chaucer 1984). The xml document had numbered lines, stanza division, book division, and indications of *incipits* and *explicitis* which saved a considerable amount of time since I did not have to number the lines manually and structure the text according to levels of the work: book, stanza, line. However, the OTA xml text did not have numbered stanzas, page breaks were relative to the printed version and some characters like yoghs ("□", "□") were not recorded. Regardless, it was a good start for a base text since it could be compared with the witnesses. Accordingly, it could be modified to fit the number of stanzas or lines per folio and then guide the transcription. Although it is very important to talk about the base text in regard to collation, I will first explore the process of transcription and then explore collation in the second part of this text.

## Transcription

The process of acquiring images of the manuscripts and printed editions requires a separate discussion. For now, let it be said that one would like to think that we are in an era in which manuscripts are digitised and made available for everyone. But as we know, as much as this is the direction in which libraries are going, it is not applicable

to every particular case. Even when manuscripts have been digitised, their accessibility falls under a diverse range of possibilities. For example, both printed versions of *Troilus* were accessible through Early English Books Online (EEBO) from the moment my research began, most likely a long time before that. On the other hand, out of the 16 manuscripts that have the text of Chaucer's poem, only the very beautiful Corpus Christi College Parker Library MS 061 and HM 114 from the Huntington Library give access through the International Image Interoperability Framework (see IIF Consortium 2021) that delivers high quality images and allows for software like (Robinson 2021) to use those images without the need to host them (Huntington Library 2020). During the process of my research, some manuscripts have been made available. These are the ones that belong to the Harley collection housed at the British Library (H1 2280, H2 3943, H3 1239, H4 2392, H5 4912). The images of these manuscripts can exclusively be seen on the websites of the institution that host them. The rest of the manuscripts remain unavailable online. It is possible to photograph the ones that can be found at the Bodleian Library as long as the pictures are for research purposes only. For some manuscripts such as Cambridge University Library Gg. 4.27, it is possible to look at a facsimile; for others such as Durham University Library Cosin MS V.ii.13, the only possible way to get images is to pay for the digitization of a microfilm reel from the Hill Museum & Manuscript Library. Needless to say, it is not easy for a graduate student to get hold of all of these, but it should also be noted that it is getting easier.

The fact that images come together with transcription has, at the very least, two implications that should be considered. The first one is that the quality of the images has a direct impact on the transcription. While obvious, the quality of an image can hide valuable information. From now on, I will refer to witnesses according to their sigils (see appendix). An example of this is 793 and 795 of Book II in H4 (for all sigla, see Appendix 1, List of Witnesses). The *Riverside Chaucer* reads:

How ofte tyme hath it yknowen be  
The tresoun that to wommen hath ben do!  
To what fyn is swich love I kan nat see,  
Or wher bycometh it, whan that is ago. (Chaucer 2008, 484, 792–795)

While I was using images from a digitized microfilm, my transcription of lines 793 and 795 read “The treson̄/that to women ay is do [...] Or where becomyth it/whan it is go.” At first glance, that is not inaccurate and as far as I can tell, it represents the scribe’s intention, but it is hard to tell from that image that the scribe wrote something different and then corrected it. It is not hard to tell however if one looks at the high resolution image on The British Library Digitised Manuscripts (BLDM 2020a, H5). Initially, the scribe had written *doon* and *goon*, but it was later erased. While it could be stemmatically irrelevant, this piece of information can affect our perception of scribes and their rigor, or at least of this particular one. The previous images did not represent this aspect of the text in enough detail, and it was easily missed.

Another example of how better images enable the work to improve is line 1127 of Book II in H5. The *Riverside Chaucer* reads “He may nat longe liven for his peyne” (Chaucer 2008, 504). The line in H5 is “He may not longe lyue <sup>in<sup>bs</sup></sup> langou<sup>r</sup> for his payne”; however, this is only noticeable in the latest images uploaded to the British Library (BLDM 2020b, “Harley MS 4912,” 31r). The reproductions that I was using previously, which came from a microfilm, made such subtle traces as to be unreadable. Therefore, the superscript *bs* was omitted from my transcription and *his* was not struck out. It is not possible for me to reproduce the aforementioned images so that the readers can reach their own conclusions. However, I can cite Windeatt’s critical apparatus, which coincidentally did not take into account the same corrections that I overlooked.

I cannot assume that he was working with microfilm images instead of the manuscript itself, but his notation for line 1127 is interesting: “may] ne may Gg/for his] in thys Cx/his] the J/lyuen] lyve in langour H5” (Chaucer 1984, 213). According to Windeatt’s edition, H5 reads “He may not longe lyve un langour for his peyne.” This misses out completely on the interlinear corrections that coincide with Caxton’s *in thys*. It also ignores that *his* should be omitted. Various authorities suggest that only one hand is involved in the inscription of this manuscript. Failure to recognize this variant means that information with possible stemmatological relevance is lost. Only Caxton’s printed version and H5 include that reading. It is clear that filiation cannot depend on a single variant and that polygenesis could also explain the evidence. Regardless, poor quality images should not determine our understanding

of the tradition. (The British Library website states that: “Occasional marginal notes in English and Latin are in the hand of the scribe.” The Late Medieval English Scribes website only lists one hand as Unknown for this manuscript. In the Marginal Headings section, we can read: “Some small marginal names of characters in ink of text and hand of scribe.” Finally, Windeatt in his description of the manuscript says: “Written in one fifteenth century hand with quite frequent corrections, apparently by the same hand” [Chaucer 1984, 72]).

The second implication that we should reflect on is on the meaning of text. In her article “The Texts We See and the Works We Imagine,” Bárbara Bordalejo defines text as “all the meaningful marks on the page made by someone with the intention of communicating something” (Bordalejo 2013, 67). As part of the “text of the document,” she includes “any indications as to which text might be considered erased or what needs to be included, marks that suggest a change in order or any other meaningful signs on the page” (Bordalejo 2013, 67). I agree with this definition, since it is broad enough that it could include layout and gives place to think about variant states of the text, which is the issue Bordalejo explores in this article. Yet one needs to ask as a transcriber: what is meaningful? What would then be meaningless? Both interpretation and purpose play a critical role in responding to these questions.

The purpose of the editorial work needs to be kept in mind. In my particular case, the readers who will have access to my transcription will also be able to see a visual reproduction of the folio that has been transcribed. This is not new by any means. On the left side of the screen, one would find an image of a document (or part of it), and on the right side, the transcription of the text contained in that document would be found. In his book *La edición de textos*, Miguel Ángel Pérez Priego states that, for a diplomatic edition, “it is mandatory to keep graphic signs such as long s, sigma, the tironian note, abbreviations (they can be expanded as long as they are in italics or in square brackets), even punctuation as long as it reveals a particular and interesting practice” (Pérez Priego 1997, 43). Nevertheless, in this context, my transcription does not intend to fill the requirements of an ultra-diplomatic edition. This is not due to the unfaithfulness of my transcription or due to a lack of rigor; instead, my transcription is meant to help the reader make sense of the inscriptions that they can

see on the reproduction of the document. In that sense, the transcription does not need to reflect every possible detail.

A dialectical tension between transcription and the spelling of medieval works arises. This tension is not necessarily due to the graphical differences between them because readers understand that scribes and editors deal with different material supports which imply different capabilities. The tension has more to do with the logic behind each; it has to do with how we choose to represent letters and words, and it has to do with how we encode text. As literate individuals from the 21<sup>st</sup> century, one of the conventions that informs our habits of reading and writing is that of standardized spelling. This means that we all operate on the assumption that there is one correct spelling for each word, which turns all other variants into incorrect or misspelled words. This impulse is inherited from the tendency of print culture to standardize spelling. As Marshall McLuhan in *The Gutenberg Galaxy* states: “Manuscript culture had no power to fix language or to transform a vernacular into a mass medium of national unification” (McLuhan 1962, 229), there are “two questions directly related to the printed form of any language at all, namely the drive for fixity of spelling and grammar” (McLuhan 1962, 229). The material conditions in which scribes and typesetters work are completely different. The typesetters, in their task to represent what they see on the page, were limited to a finite amount of characters in a context in which, as Frederick H. Brengelman states, “no entirely redundant characters should be tolerated” (Brengelman 1980, 345). In contrast, scribes had an array of strikes to represent entities that today we represent in one particular way.

There is a possible analogy to be made with phonemic and phonetic transcriptions. A phonemic transcription is an abstraction that represents, regardless of the actual utterance of any particular speaker and redundant as it may be, a succession of phonemes, that are “the smallest sound unit in a language. Meaningless in themselves, phonemes are the building-blocks of language. Changing one for another changes the meaning of a word, as with /p/ and /b/ in *pat* and *bat*” (Chandler and Munday 2020). In this sense, a change in a phonemic transcription could cause the representation of different items; similarly, in print, a spelling mistake might misrepresent what was intended or might represent something entirely different.

Meanwhile, a narrow phonetic transcription will “show whatever differences between sounds can be perceived, regardless of whether they are distinctive in the language represented” (Matthews 2014). If two speakers utter the same phrase, it is likely that a narrow phonetic transcription would show different signs if the speakers used allophones which are “a difference in sound within a language that does not produce a difference in meaning” (Calhoun 2002) but the phrase would still have the same meaning. If both transcriptions were translated to a phonemic one, they would be identical. The same can be said of two scribes who copy the same line but transcribe it in different ways. It may have the same meaning and they might transcribe the same words but with different characters because they might represent the phonetic choices of their context, among other reasons. According to Febvre, an aspect of spelling in the transition between manuscript and print is that it “came to correspond less and less with pronunciation” (Febvre and Martin 1976, 319). It stopped representing dialectically specific uses gradually in favor of consistency.

To compliment these ideas, it is also useful to reflect on what Robinson and Solopova (1993) refer to as the level of transcription. In his article “Graphemic Analysis of Late Middle English Manuscripts,” W. Nelson Francis characterizes his work as graphetic in relation to graphemic transcriptions in the following way: “This is a *graphetic*, not a *graphemic*, transcription; that is, it records every detail whether or not it has linguistic significance. It corresponds methodologically to the linguist’s narrow phonetic transcription” (Francis 1962, 36). In this specific kind of graphetic transcription, each graph-type is assigned a number. In the case of groups, essentially graph-type variants, they “use the same number and add distinguishing letters” (Francis 1962, 36).<sup>1</sup> Regardless of representation, this definition matches Robinson’s and Solopova’s (1993): “graphetic: every distinct letter-type is distinguished (as: r “short” is transcribed apart from r “round” and r “long descender”, etc.)” Francis also indicates where the graphemes of a graphemic transcription come from: “A finite and smaller

---

<sup>1</sup> To go into more detail seems unnecessary, but it is interesting to see that a graphetic transcription of the kind that W. Nelson Francis refers to looks like this: #23b|10 6a||23a|28|23a||9b 19||9b 9c 4||16a 4 9c 9 14c 27||9d 9d 9d||6^7e||9c 19c|| 12| 4 | 5a| 16a 2b||10 6c||14c | 13 14b#. The transcribed line is “þat þ̄ is no louer in th is world at ese.”

repertory of graphemes is deduced [from the graphetic transcriptions], and each graph-type is assigned as an allograph of one of these on the basis of its distribution and linguistic reference" (Francis 1962, 43). Their relation is equivalent to phonetic and phonemic transcription. Instead of allophones that fall under the same phoneme, various allographs correspond to the same grapheme. This is also in concordance with Robinson's and Solopova's position on what constitutes a graphemic transcription:

Every manuscript spelling is preserved (as: "she", "sche") without distinction of separate letter forms as in a graphetic transcription. Diplomatic transcripts, for example those of Ruggiers for the Hengwrt manuscript and Furnivall for the Chaucer Society, are centred on a graphemic reproduction. (1993)

This analysis of spelling and levels of transcription is necessary because one needs to know the aspects of what one is dealing with to decide how detailed the transcription will be. Every transcription implies a hierarchy of what is interpreted to be meaningful, but there are no easy or strict general rules. One would like to keep as much information as possible without being redundant or being so specific that the data is overwhelming. My transcription can be characterized as graphemic with some elements of graphetic. Thus, spelling variants are transcribed because they reflect the dialectal variety, as well as the diversity of the *usus scribendi* of *scriptoria* and/or scribes. However, when it comes to letterforms, opposed to what Pérez Priego argues, I do not differentiate between long and short s. It is no longer due to the limitations of type because, as we know, TEI allows for different letterforms to be recorded with the <glyph> tag ("5 Characters, Glyphs, and Writing Modes" [TEI 2020]). Nonetheless, a line must be established for what is relevant to the project. Because an image of the manuscript is next to the transcription, I inform the reader that there are several medieval letterforms that correspond to modern characters. In that way, the reader can appreciate the differences between modern and medieval writing.

The transcription of abbreviations should also be considered in this discussion. If one chooses to expand them, one needs to interpret what an abbreviation mark is and what it represents. For example, we can compare lines 225 to 227 of Book 1 in H5 and the *Riverside Chaucer*:

So ferd it by þis fers prowde <i>knygth</i>	So ferde it by this fierse and proude <i>knyght</i> :
Thov he aworthy kynges soon̄ were	Though he a worthy kynges sone were,
And wende no thyng hadde had suche a <i>myth</i>	And wende nothing hadde had swich <i>myght</i>
H5	<i>Riverside Chaucer</i> (Chaucer 2008, 476)

According to the rhyming scheme of the royal stanza, the first and the third lines should rhyme. We notice that this is not the case with H5 (“Harley MS 4912,” 4r). In the manuscript, the final *h* has a stroke on top that crosses the ascender. Should that be considered to be a mark that means nothing in line 225 but restores the *g* in 227? This same line has no shortage of *h* (thyng, hadde, had, suche), but none of them have a crossed ascender. Most likely, the crossed *h* means nothing, and the scribe made a mistake. Therefore, I transcribed all the occurrences of *h* as a simple *h*. The Late Medieval English Scribes Project has a section dedicated to *h* in their description of this hand, and their analysis agrees with mine. There is no inclusion of this particular form of *h*, which may mean that the team did not consider it worthy of any clarification, ignored it because their instinct was not to consider it significant, or just ignored it as an honest mistake.

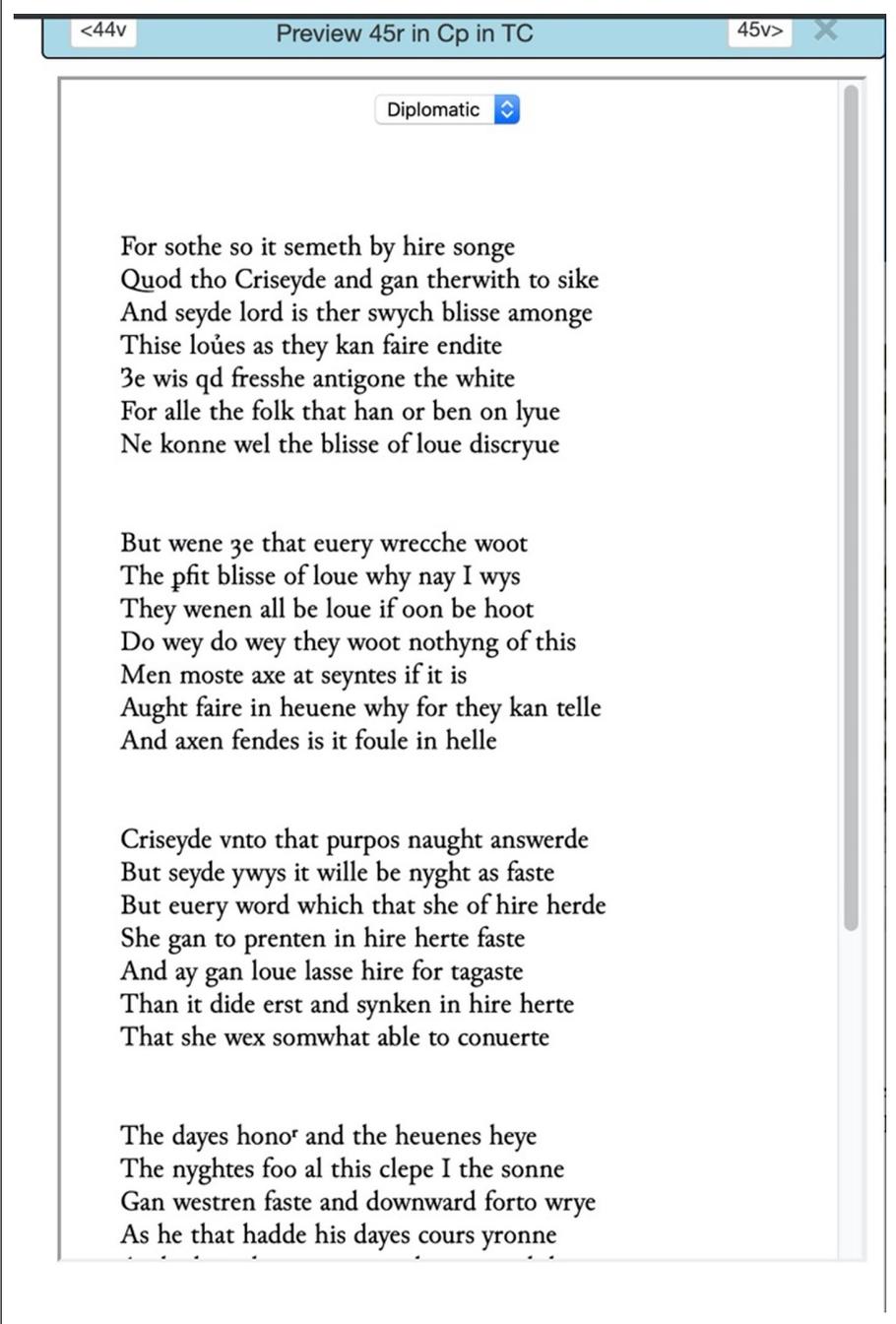
Is it then safe to state that within the textual tradition of *Troilus and Criseyde*, any *h* with a stroke across the ascender should be considered as an ordinary *h*? Not according to H4 (“Harley MS 2392,” 30r), R (there is no digital reproduction of Rawlinson Poet. 163), and Ph (“MssHM 114,” 217r). In line 235 of Book II, H4 and Ph read “That to myn h̄tis [...],” and the rest of the manuscripts do not abbreviate and have a version of *hertes*. It is obvious that the stroke in that specific context is abbreviating “er.” Thus, an exception has to be made in order to record this use of the *h* with a stroke across the ascender. The transcriber always has to be vigilant; any stroke is potentially meaningful.

Not every case of potential abbreviation is as clear. As anyone who has transcribed knows, dealing with macrons and minims is problematic if one chooses to expand abbreviations. Words like *wōman* could be expanded to *womman*, but is that necessary? Here, the decision to expand the macron is the responsibility of the transcriber, but transcribers, like scribes, change over time. I have experienced moments during my transcription where I have thought that if a macron is present, it must mean something. There have been other times in which I feel like a particular scribe adds macrons on top of nasal consonants as part of his *usus scribendi*. There

are cases in which it is obvious that a macron is meaningless, such as in line 184 of Book I, where the final word is *down*, and H2, H5, R, and S1 read *douñ/down̄*. There are other cases in which a stroke should be expanded, like the one just examined above. Sometimes, it is difficult to know what to do with cases like *wōman* or with a word that ends with *r* and then has a hook or a flourish that could mean final *-e*, since it is spelled out in other manuscripts. There is no easy answer.

A desire to fight redundancy should not turn into a habit of oversimplification. In the name of modernizing spelling, valuable information could be lost. I have a hypothesis for what happens in line 1153 that has to do with the scribe's craft, which would be lost if thorns were transcribed as *th*. The line in the *Riverside Chaucer* reads: "But for al that that ever I may deserve" (Chaucer 2008, 505). It is interesting to see the different ways scribes have written this line. For example, A, J, S1, read *that þat*, while H5 and Ph read *þat that*. Why would scribes choose to copy the same word with different spellings? Not all of them did. Cx, Dg, Gg, R, and S2 do not register the word twice, which means that either by mistake or by active choice one of the words was removed because it was judged it to be redundant. Cl, D, H1, and H2 read *þat þat* and Cp, H3, and H4 read *that that*, which suggests that the scribes were not concerned by the repetition. However, five witnesses show variant spellings of the same word next to each other. My guess is that it is done to emphasize that the word should be there twice. It is not a repetition due to distraction. By having two variant spellings, it is also easier for subsequent scribes to copy it and be sure that this is no mistake. Significance does not only operate on a discursive level; we can learn about the context of the production of medieval works by allowing these nuances to reveal their purpose.

How should we decide what kind of transcription to generate? Does one need to decide between providing an insufficient transcription that does not help the reader to understand the marks on the document or a transcription that provides expanded abbreviations at the risk as mischaracterizing every stroke on the page? Fortunately, in the age of digital editions, platforms like Textual Communities (Robinson 2021) allow the transcriber some flexibility. It is possible to produce different types of transcriptions. A "diplomatic" transcription seen in **Figure 1** will show the text with no expansions, while an "edited" version in **Figure 2** shows the expanded



The image shows a digital manuscript viewer interface. At the top, there is a header bar with the text "<44v" on the left, "Preview 45r in Cp in TC" in the center, and "45v>" on the right with a close button. Below the header, there is a dropdown menu set to "Diplomatic". The main content area displays a diplomatic transcription of text from a manuscript, organized into four distinct paragraphs. The text is in a black serif font and is centered on the page. The first paragraph begins with "For sothe so it semeth by hire songe". The second paragraph begins with "But wene 3e that euery wrecche woot". The third paragraph begins with "Criseyde vnto that purpos naught answerde". The fourth paragraph begins with "The dayes hono<sup>r</sup> and the heuenes heye".

<44v Preview 45r in Cp in TC 45v>

Diplomatic

For sothe so it semeth by hire songe  
Quod tho Criseyde and gan therwith to sike  
And seyde lord is ther swych blisse amonge  
Thise loües as they kan faire endite  
3e wis qd fresshe antigone the white  
For alle the folk that han or ben on lyue  
Ne konne wel the blisse of loue discryue

But wene 3e that euery wrecche woot  
The pfit blisse of loue why nay I wys  
They wenen all be loue if oon be hoot  
Do wey do wey they woot nothyng of this  
Men moste axe at seyntes if it is  
Aught faire in heuene why for they kan telle  
And axen fendes is it foule in helle

Criseyde vnto that purpos naught answerde  
But seyde ywys it wille be nyght as faste  
But euery word which that she of hire herde  
She gan to prenten in hire herte faste  
And ay gan loue lasse hire for tagaste  
Than it dide erst and synken in hire herte  
That she wex somewhat able to conuerte

The dayes hono<sup>r</sup> and the heuenes heye  
The nyghtes foo al this clepe I the sonne  
Gan westren faste and downward forto wrye  
As he that hadde his dayes cours yronne

**Figure 1:** Diplomatic transcription.  
Visualization created on Textual Communities.

<44v Preview 45r in Cp in TC 45v>

Edited

For sothe so it semeth by hire songe  
 Quod tho Criseyde and gan therwith to sike  
 And seyde lord is ther swych blisse amonge  
 These *loueres* as they kan faire endite  
 3e wis *quod* fresshe antigone the white  
 For alle the folk that han or ben on lyue  
 Ne konne wel the blisse of loue discryue

But wene 3e that euery wrecche woot  
 The *parfit* blisse of loue why nay I wys  
 They wenen all be loue if oon be hoot  
 Do wey do wey they woot nothyng of this  
 Men moste axe at seyntes if it is  
 Aught faire in heuene why for they kan telle  
 And axen fendes is it foule in helle

Criseyde vnto that purpos naught answerde  
 But seyde ywys it wille be nyght as faste  
 But euery word which that she of hire herde  
 She gan to prenten in hire herte faste  
 And ay gan loue lasse hire for tagaste  
 Than it dide erst and synken in hire herte  
 That she wex somewhat able to conuerte

The dayes *honour* and the heuenes heye  
 The nyghtes foo al this clepe I the sonne  
 Gan westren faste and downward forto wrye

**Figure 2:** Edited version.

Visualization created on Textual Communities.

abbreviations. The future reader can decide how to engage with the work and is always welcome to agree or disagree with the editorial choices. The process is visible and open for debate, which is one of the richest aspects of scholarship.

Transcription is an informed suggestion of what is significant based on the experience of looking at manuscripts over and over. A platonic fetishist, which I might as well be making up for the sake of the argument, could consider that every step from the original to the edited transcription leads to a loss in fidelity because each mediation is further removed from the original, from the Truth. I argue that every step is also the product of agents engaging with the artifact to facilitate its understanding. Thus, while each mediation is in no way a substitute for the manuscript, each has its own benefits. The adequacy of transcriptions will change over time depending on the needs of the ever-developing interpretative communities. For now, let us hope that our work is useful.

## **Collation**

For this project, it is necessary to keep in mind that transcription is a step towards collation. In order to compare the texts of witnesses, it is important to regularize orthographic variants in order to obtain substantive variants instead of accidentals. These are the terms coined by W.W. Greg:

A distinction between the significant, or as I shall call them 'substantive', readings of the text, those namely that affect the author's meaning or the essence of his expression, and the others, such in general as spelling, punctuation, word-division, and the like, affecting mainly its formal presentation, which may be regarded as the accidents, or as I shall call them 'accidentals' of the text." (Greg 1950, 21)

Here is where a note about the base text is required. The base text has to be modified so that it can always be compared to whatever evidence the witnesses offer. Its purpose is to show the relationships between the witnesses. Thus, when witnesses add a line or a stanza that is not present anywhere else and might not be authorial, the base text needs to account for that evidence so that it can show the agreement of those two witnesses. As stated before, the OTA xml text of Windeatt's edition was

extremely useful since it provided all the possible lines any of the witnesses could include, plus it had line numbers allowing the collation software to retrieve them in order to compare every lines word by word, the tags had to be complimented by adding stanza numbers, page breaks according to each witness and marks of book division where needed. Altering the base-text is common practice, as Robinson did with his work with Old Norse *Svipdagsmál*:

I used a purely arbitrary master text as the base of the collation, with the single aim of discovering exactly and efficiently just what manuscripts agree in what readings. I altered the master over and over, rerunning the collation each time, simply to maximize the detail about agreement between manuscripts. Then I turned to preliminary analysis of these agreements, examining the patterns of distribution across the manuscripts and searching for evidence of direction of variation in the variants themselves and in their distribution. At this point the first groupings of manuscripts began to appear, and the first certain judgements about the originality of particular readings could be made. (Robinson 1994, 97)

The fact that he uses a “purely arbitrary” text as base also coincides with J. Froger’s (1970) notion that any text could serve as a false original to show the relations between witnesses. What seemed to be wishful thinking by Froger (1970) was achieved by Robinson some twenty years after.

Since the final purpose is to generate *stemmata*, it is also convenient to make a final note about transcription. If a word has an abbreviation mark that was not expanded, that will not affect the final stemma. The example of line 184 of Book 1, where the final word is *down*, H2, H5, R, and S1 read *doun̄/down̄*, and the final word in H4 is encoded as works to illustrate this.

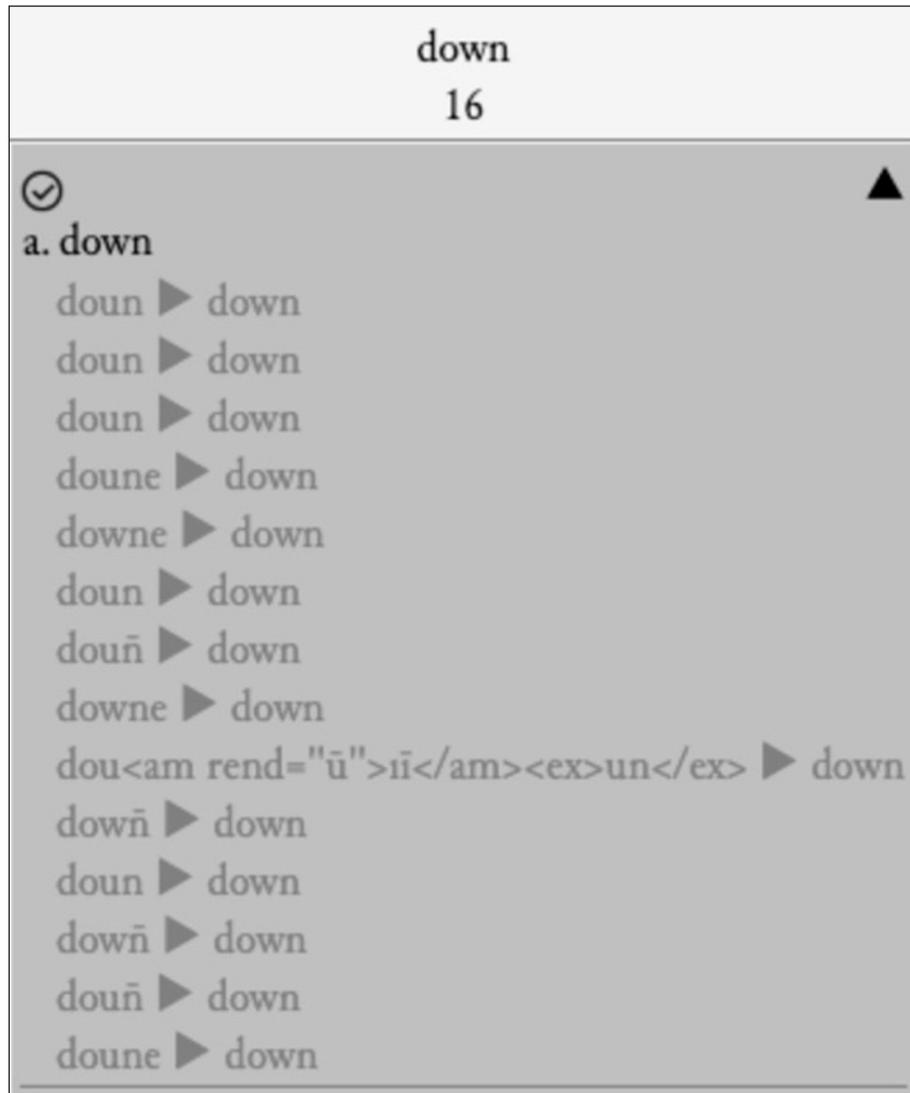
On **Figure 3**, we can see the apparatus for the phrase. Since every spelling variant of *down* has been regularized, as seen in **Figure 4**, they will all count as the same reading. That will allow the collation to focus on substantive variants. To perform this task I used the ITSEE- Birmingham Collation Editor (Smith 2019) that “provides

His	3	ong	knyghtes,	lad	hem	vp	and	down
2	4	6	8	10	12	14	16	
▲					▲			
a. his					a. lad hem			
+ b. hire					+ b. lad thame			
+ c. he					+ c. lad			
					+ d. hasladde hem			
					+ e. and lad thame			

**Figure 3:** The witnesses present spelling variants that have been regularized, thus the apparatus does not show variation. There is no genealogical information.

an interface to the CollateX engine developed by the INTEREDITION project, the successor to the COLLATE program by Peter Robinson” (Houghton, Sievers, and Smith 2014). It was refined to work with Textual Communities (Robinson 2021) to interface with the online transcriptions generated on the platform. These transcriptions are stored in a JSON database, rather than in files.

Other tools, such as the ones discussed by Jänicke and Wisley’s “Visualizing Mouvance: Toward a Visual Analysis of Variant Medieval Text Traditions,” (2017) propose very interesting ways of exploring medieval texts and textual traditions but may not be productive for this project. The textual tradition of *Troilus* is larger than the ones they work with (“the thirteenth century *Du chevalier qui fit les cons parler*, extant in seven manuscripts, six in continental French and one in the French of England” (Jänicke and Wisley 2017, 108), “the *Chanson de Roland*, known to be transmitted in six major versions” (Jänicke and Wisley 2017, 108). The degree of textual instability, in the terms of Paul Zumthor that the authors quote: “an ‘interplay between variant readings and reworkings’, balancing both the textual, literary elements of written works with oral, performative ones” (Jänicke and Wisley 2017, 106) is not of such a degree in the textual tradition of *Troilus* that could give place to the very interesting results of the analysis in that article. Plus, to visualize



**Figure 4:** Regularization. The various spelling variants of the witnesses This prevents the spelling variants from creating noise when building a *stemma*.

some variance in the 18 witnesses would be difficult since the “design relies upon a larger screen to view the entire tradition” (Jänicke and Wrisley 2017, 113) as they acknowledge when working with seven witnesses. Variation has different faces in medieval literature, and they lead to different questions that have to be tackled

with specific tools: the interplay between the reworkings of literary texts and their oral performance and dissemination is palpable in many works, while in others the scribal intervention and copying dynamics play a preponderant role compared to the oral delivery. The later seems to be the case of *Troilus* or at least it is that aspect on which this project focusses.

There is a debate regarding what exactly constitutes a variant. For the purposes of this project, the useful variants are what Bordalejo calls “stemmatically significant variant” (Bordalejo 2002, 96). In order to understand what that means, it is important to know some characteristics of both significant and non-significant variants. Benjamin Salemans lists some characteristics of non-significant variants in his *Building Stemmas with the Computer in a Cladistic, Neo-Lachmannian, Way* (2000):

- a. Differences in use of capitals and small letters ('Karel de Grote' vs. 'karel de grote').
- b. Differences in spelling ('roesen' vs. 'roisen').
- c. Differences in dialect and language ('brood' vs. 'bread').
- d. Differences in use of punctuation marks ('oh! oh!' vs. 'oh, oh').
- e. Differences in boundaries of words ('metten' vs. 'met den').
- f. Differences in clause headers (or incorrect placement of or clear absence of clause headers), if the same clause of a play is spoken by different people (in some *Lanseloet* texts clauses occur with incorrect clause headers; a copyist familiar with the text could detect these false headers and simply correct them).
- g. 'Ungrammaticalities' (ungrammatical sentences can often be easily corrected).
- h. 'Nonsense Readings' [...]
- i. Copy mistakes ('Karel de Grote' vs. 'Krl de Grote' [...]).
- j. Names (...). In the *Lanseloet* text versions reference is made to 'sint Jan' and 'sint iohan'. This is a clear example of a ungenealogical variant (...): the names of 'Sint Jan' and 'Sint Johannes' are still used for indicating the same saint. Different names in text versions, which refer to the same per-

son, are only genealogically relevant (...) if the names concern unknown persons, who normally do not play an important role in the story or 'the world'. [...]

- k. Archaic words. Many copyists will use a more contemporary word when confronted with an archaic word. Therefore, there is considerable chance that copyists working with minimally related exemplars introduce the same more modern word in their copies. The occurrence of the same more contemporary word in the text versions is not due to equal descent, but by diachronic change of language: they are not genealogical variants but parallelisms.
- l. Frequently used words, which are usually not kinship-revealing (or text-genealogical) and, therefore, must be treated with the highest caution. (Salemans 2000, 67–68)

Salemans also adds “*Synonymous parallelism* ('white'  $\longleftrightarrow$  'pale')” (2000, 68), and “*Inflectional parallelism* ('is'  $\longleftrightarrow$  'was')” (2000, 70). According to Paolo Trovato, Salemans lists the following as significant variants: variation in word order, rhyming conventions in verse should be followed, addition or omission of words when they are not small or very common (Trovato 2014, 111). In general, there is an agreement with Bordalejo's list: “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).

Learning how to distinguish between significant and non-significant variants will affect the collation. A big concern with collation is that no one wants to alter the process so that it reflects the editor's assumptions. Yet, as much as one would like to draw hard lines to guide the task, it is frequent that variants will also push our limits. Let us consider the next couple of examples.

In line 768 of Book II, the *Riverside Chaucer* reads: “A cloudy thought gan thorough hire soule pace” (Chaucer 2008, 499). Some witnesses, like A, Cl, Cp, D, Gg, H1, H5, J, and S1, indeed read *soule*. Some others, like Cx, H2, H3, H4, R, Ph, and S2, read *heart*, and only Dg reads *thought*. At first glance, these variants are substantive, but they fit

into Salemans' synonymous parallelism. A modern reader might say that I am terribly confused since soul, heart, and thought are not synonyms. However, a medieval scribe could think so. Let us remember that in a medieval context, the soul is responsible for intellectual tasks. If we follow Stephen Batman's 1582 translation of Bartolomeus Anglicus' *De Proprietatibus Rerum*, we find that the human soul had three parts: "one is called naturall, and is in the ly[u]er, the other is called vitall, or spiritall, & hath place in the heart, the third is called Animal, & hath place in [the] brayn" (Anglicus 1584, iii, 14). This is not new; these are Aristotelian notions of the soul. Accordingly, Batholomeus explains that: "the vertue vitall, that giueth lyfe to the bodye, whose foundation or proper place is the heart" (Anglicus 1584, iii, 15). Ioan P. Culianu states that:

The doctrines expounded by Bartholomaeus were based on the idea prevalent in Arab medicine that the heart is the unique generator of the vital spirit which, once it has reached the brain, is called sensitive. The messages of the five "external" senses are transported by the spirit to the brain, where the inner or common sense resides. (Culianu 1987, 11)

And finally, according to Mary Carruthers in her *Book of Memory*:

[E]ven though the physiology of consciousness was known to occur entirely in the brain, the metaphoric use of heart for memory persisted [...] *The Middle English Dictionary* records an early twelfth-century example of *herte* to mean "memory"; there is an Old English use of *heorte* to mean "the place where thoughts occur," *cogitationes*." (Carruthers 2008, 59)

Memory is an intellectual process. An oversimplified version of the procedure is that the soul moves from the heart to the brain to produce thoughts. This brief *excursus* through medieval notions of anatomy sufficiently explain why in the line "A cloudy thought gan thorough hire soule pace," *heart* and *thought* could be synonyms of *soule*. To Salemans' credit, the fact that *thought* is present in only one manuscript could indicate that it is the product of scribal resolution and, regardless of its presence in six witnesses, the same applies to *heart*. However, it is hard to believe that even if scribal

substitution explains the origin of the variants, that necessarily means there is no vertical transmission in the case of *heart* and *soule*. I consider these two readings as stemmatically significant variants. With this in mind, let us examine the next example.

Just some lines ahead, we can read in 771 “That thought was this: “Allas! Syn I am free.” There are two readings present in the witnesses that at first glance are accidentals: *syn* and *sith*. It turns out that A, Cl, Cp, Dg, H1, H5, R, and S2 read *syn*, while Cx, Gg, H2, H3, H4, J, Ph, and S1 read *sith*. This point is better illustrated by **Table 1**. *Syn* was written in five cases where *soule* was also present, as many as *sith* is with *heart*. Is this any indication of the distribution of variants through the witnesses, that is, of the place the witnesses would occupy in a stemma? Is it possible to consider that A, Cl, Cp, H1, and H5 share a common ancestor that has both *soule* in ii, l. 768 and *syn* in ii, l. 771? Does the same apply for *heart*, *sith* and Cx, H2, H3, H4, and Ph?

I am aware of how misleading these questions are. Variant distribution and filiation cannot and should not depend on four readings. But let us imagine the moment where I have just collated line 768, and then I notice that a spelling variant resembles the attestation of *heart* and *soule*. It is tempting to feed this information to the collation software since I can argue that I am trying to strengthen the previous information I had. However, the reality is that I would be using an accidental variant to partially reinforce the idea that *heart* and *soule* are indeed stemmatically significant, instead of a case of Salemans' synonymous parallelism. It would be very attractive to say that the textual tradition of Chaucer's *Troilus and Criseyde* is one that hinges upon the distinction of heart and soul. However, under no other circumstances would I consider that *syn* and *sith* are stemmatically significant. Therefore, as appealing as it could be not to, I regularized *sith* to *syn*.

**Table 1:** Variants and agreements. The column on the left contains different readings and the column on the right shows the attestation of each variant. If compared, the sigils that are highlighted suggest certain consistency.

B. II, l. 768	<i>soule</i>	<b>A, Cl, Cp, H1, H5</b> , D, Gg, J, S1
	<i>heart</i>	<b>Cx, H2, H3, H4, Ph</b> , S2, R
B. II, l. 771	<i>syn</i>	<b>A, Cl, Cp, H1, H5</b> , Dg, R, S2
	<i>sith</i>	<b>Cx, H2, H3, H4, Ph</b> , S1, Gg, J

After the full transcription and collation of the excerpts of *Troilus* that my research tackles, the data will be analyzed by phylogenetic software to further the results of this work. In “The *Canterbury Tales* and other Medieval Texts,” Robinson states:

we were able to show that phylogenetic software developed for biological sciences gave useful results when applied to manuscript traditions. That is: we can turn our lists of agreements and disagreements among the manuscripts into a form which can be input into a program used by biologists to hypothesize trees of descent among species; we can then use these programs to hypothesize trees of descent among the manuscripts. (Robinson 2006)

Many years after the creation of the *Canterbury Tales Project* and of Textual Communities, it is possible for me to abuse the multi-cited phrase of the dwarf standing on giants' shoulders and apply similar methods to my research on *Troilus and Criseyde*. Examination with Phylogenetic Analysis Using PAUP (PAUP\*) has produced unrooted trees that reflect the relationship that the witnesses bear. According to Prue Shaw, editor of Dante's *Commedia*, phylogenetic analysis produces stemmata that are represented:

as ‘unrooted phylograms’. The ‘unrooted’ view means that the branching appears to occur as an organic growth, from a relatively central point [...] This may free the reader from an over-simple view of the tradition, presented as series of vertical straight lines running down from the ancestor signifying cumulative corruption over time. One striking advantage of the ‘unrooted phylogram’ display compared with a traditional geometric representation lies in the correlation between the length of the branches and the degree of divergence from other witnesses. (Shaw and Robinson 2010)

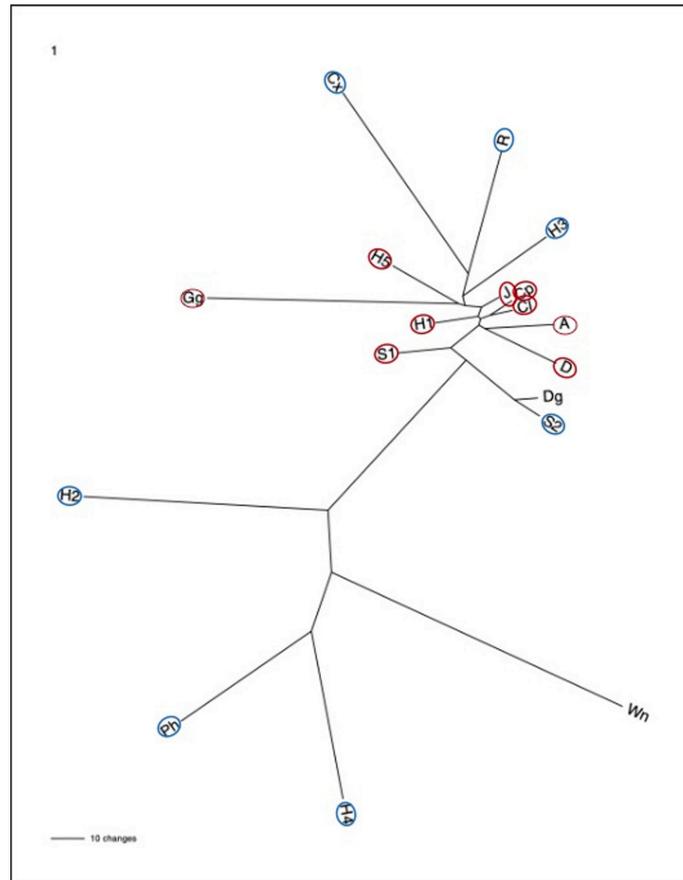
What this means is that we do not need to identify the original reading before carrying out the analysis as traditional stemmatics does. As Shaw and Robinson did before, I follow their example by using “the default criterion, maximum parsimony,

to search for optimal trees" (Swofford 2003) which "considers all possible bifurcating trees and identifies the one that requires the smallest number of changes (i.e. is most parsimonious). The various species, or manuscripts, are in this way grouped according to their shared derived characters" (Windram et al. 2008, 445) I will perform a heuristic search since the exhaustive search supports a maximum of 12 taxa and I am dealing with 18 witnesses. Let it show that the base text is not present in the diagram since PAUP\* allows us to remove taxa, so the fact that Windeatt heavily based his edition on Cp does not bear any significance.

The phylogenetic analysis of the first hundred lines of Book I seen in **Figure 5** shows that those witnesses that read *soule*, (in a red circle) with the possible exception of H5 and Gg, are closely related. On the other hand, the ones that read *heart* (in a blue circle) present a more diverse position in the unrooted tree: H2, H4, and Ph are closely related, the same goes for H3, R, and Cx, but S2 does not fit. Two things must be considered. The first that *heart* could be a coincidence in S2, but it is indicative of vertical transmission regarding H2, H4, and Ph on the one side, as well as for H3, R, and Cx on the other. The second is that analysis of Book II needs to be done to assess these particular variants. There might be changes of filiation, and what stands for the first hundred lines of Book I could have changed. But the full transcription and collation of the excerpts will allow me to test any particular variant and try to make sense of the conditions that gave place to variation.

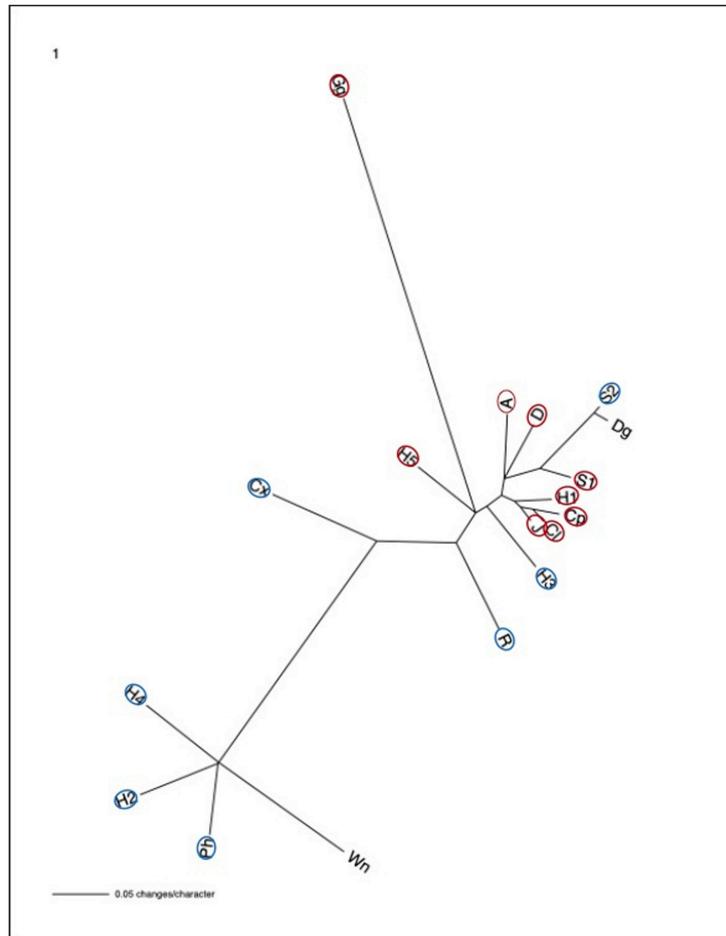
Very similar results can be seen when using the Likelihood criterion (see **Figure 6**) with a neighbor joining analysis which:

constructs a tree by sequentially finding pairs of neighbors, which are the pairs of OTUs [taxa] connected by a single interior node [...] The NJ algorithm starts by assuming a star-like tree that has no internal branches. In the first step, it introduces the first internal branch and calculates the length of the resulting tree. The algorithm sequentially connects every possible OTU pair and finally joins the OTU pair that yields the shortest tree. (Salemi, Vandamme, and Lemey 2009, 26)



**Figure 5:** Phylogenetic tree using the parsimony criterion and a heuristic search of the first 100 lines of *Troilus and Criseyde* with 18 witnesses. The witnesses that read *soule* in a red circle, the witnesses that read *heart* in a blue circle.

It is noticeable that the witnesses that read *soule* are closely related and their relationships are similar: A and D remain linked by node, the same can be said for H5 and Gg, S1 remains close to A, D, and H1. As for the other group, Ph, H2 and H4 are closely related and apart from the rest, while Cx, R, and H3 are not extremely distant from each other and S2 continues to be an outlier. Both diagrams confirm the general assumptions of witness relations and show encouraging results that will be later refined when more data is fed into the software so that precise information can be produced and we as editors or readers can make sense of it.



**Figure 6:** Phylogenetic tree using the likelihood criterion and the neighbor joining method of the first 100 lines of *Troilus and Criseyde* with 18 witnesses. The witnesses that read *soule* in a red circle, the witnesses that read *heart* in a blue circle.

I stated in the beginning of this text that the end goal of this project was not to make an edition, but to understand the relationship that the 18 witnesses to *Troilus and Criseyde* bear. The project is still at its early stages, but it shows potential. I find myself in agreement with Robinson (2006), who elegantly expressed why this kind of project is important: “Like the stemmatics of the last century, its aim is to illuminate the history of the text. Unlike the stemmatics of the old century, its aim is not a well-made edition, but a well-informed reader.”

## Competing interests

The author has no competing interests to declare.

## Editorial contributions

**Recommending Editor:** Luis Meneses, University of Victoria

**Section Editor:** Nathir Haimoun, University of Lethbridge Journal Incubator

**Lead Copy Editor:** Darcy Tamayose, University of Lethbridge Journal Incubator

**Lead Bibliography Editor:** Shahina Parvin, University of Lethbridge Journal Incubator

## Appendix 1

List of witnesses.

- A** Additional 12044, British Library
- Cl** M 817, Pierpont Morgan Library
- Cp** Corpus Christi College, Cambridge, 61
- D** Cosin MS V.ii.13 University Library, Durham
- Dg** Digby 181, Bodleian Library, Oxford
- Gg** Gg.4.27 Cambridge University Library
- H1** Harley 2280, British Library
- H2** Harley 3943, British Library
- H3** Harley 1239, British Library
- H4** Harley 2392, British Library
- H5** Harley 4912, British Library
- J** Cambridge, L.1, St. John's College
- Ph** HM 114, Huntington Library
- R** Rawlinson Poet. 163, Bodleian Library
- S1** Arch. Selden B.24, Bodleian Library, Oxford
- S2** Arch. Selden Supra 56, Bodleian Library, Oxford
- Cx** Caxton 1483
- W** Wynkyn de Worde 1517

## References

- Anglicus, Bartholomaeus.** 1584. *Batman Vppon Bartholome His Booke De Proprietatibus Rerum, Newly Corrected, Enlarged and Amended: With Such Additions as Are Requisite, Vnto Euery Seuerall Booke: Taken Foorth of the Most Approued Authors, the like Heretofore Not Translated in English. Profitable for All Estates, as Well for the Benefite of the Mind as the Bodie.* Translated by Stephen Batman. London: Thomas East. Accessed December 6, 2020. <https://quod.lib.umich.edu/e/eebo/A05237.0001.001/1:22?rgn=div1;view=toc>.
- Bédier, Joseph.** 1928. "La Tradition Manuscrite Du 'Lai de l'ombre': Réflexions Sur l'art d'éditer Les Anciens Textes." *Romania* 54: 161–196. Accessed February 2, 2021. [https://www.persee.fr/doc/roma\\_0035-8029\\_1928\\_num\\_54\\_214\\_4345](https://www.persee.fr/doc/roma_0035-8029_1928_num_54_214_4345). 321–356. DOI: <https://doi.org/10.3406/roma.1928.4345>
- Blecua, Alberto.** 1983. *Manual de crítica textual.* 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." PhD dissertation, De Montfort University. Accessed January 7, 2021. [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).
- . 2013. "The Texts We See and the Works We Imagine: The Shift of Focus of Textual Scholarship in the Digital Age." *Ecdotica: Rivista di Studi Testuali* 10: 64–76.
- Brengelman, Frederick H.** 1980. "Orthoepists, Printers, and the Rationalization of English Spelling." *The Journal of English and Germanic Philology* 79. Accessed February 2, 2021. <https://www.jstor.org/stable/27708682?seq=1> 3: 332–354.
- BLDM (British Library).** 2020a. "Harley MS 2392, f.34v." The British Library Digitised Manuscripts. Accessed July 6. [http://www.bl.uk/manuscripts/Viewer.aspx?ref=harley\\_ms\\_2392\\_f034v](http://www.bl.uk/manuscripts/Viewer.aspx?ref=harley_ms_2392_f034v).
- . 2020b. "Harley MS 4912, f.31r." The British Library Digitised Manuscripts. Accessed July 6. [http://www.bl.uk/manuscripts/Viewer.aspx?ref=harley\\_ms\\_4912\\_f031r](http://www.bl.uk/manuscripts/Viewer.aspx?ref=harley_ms_4912_f031r).

- Calhoun, Craig.** 2002. "Allophone." In *Dictionary of the Social Sciences*. Oxford University Press. Accessed December 6, 2020. <http://www.oxfordreference.com/view/10.1093/acref/9780195123715.001.0001/acref-9780195123715-e-51>.
- Carruthers, Mary J.** 2008. *The Book of Memory: A Study of Memory in Medieval Culture*. 2nd ed. Cambridge Studies in Medieval Literature. Cambridge, UK; New York: Cambridge University Press.
- Cerquiglini, Bernard.** 1999. *In Praise of the Variant: A Critical History of Philology*. Baltimore: Johns Hopkins University Press.
- Chandler, Daniel, and Rod Munday.** 2020. "Phoneme." *A Dictionary of Media and Communication* (3ed.). Oxford University Press. Accessed December 6. <https://www.oxfordreference.com/view/10.1093/acref/9780198841838.001.0001/acref-9780198841838-e-2034>. DOI: <https://doi.org/10.1093/acref/9780198841838.001.0001>
- Chaucer, Geoffrey.** 1984. *Troilus & Criseyde: A New Edition of Chaucers The Book of Troilus*, edited by B. A. Windeatt. London: Longman.
- . 2008. *The Riverside Chaucer*. 3rd ed. Edited by Larry Dean Benson. Oxford; New York: Oxford University Press.
- Culianu, Ioan P.** 1987. *Eros and Magic in the Renaissance*. Chicago: The University of Chicago Press.
- Febvre, Lucien, and Henri-Jean Martin.** 1976. *The Coming of the Book: The Impact of Printing 1450–1800*. New ed. Foundations of History Library. London: N.L.B.
- Francis, W. Nelson.** 1962. "Graphemic Analysis of Late Middle English Manuscripts." *Speculum* 37 (1): 32–47. DOI: <https://doi.org/10.2307/2850597>
- Froger, Dom Jacques.** 1970. "La Critique Des Textes Et L'ordinateur." *Vigiliae Christianae* 24 (3): 210–217. <https://doi.org/10.1163/157007270X00056>
- Greg, Walter W.** 1950. "The Rationale of Copy-Text." *Studies in Bibliography* 3: 19–36. Accessed December 6, 2020. [https://christopherohge.com/greg\\_rationale\\_copy-text.pdf](https://christopherohge.com/greg_rationale_copy-text.pdf).
- Houghton, Hugh, Martin Sievers, and Catherine Smith.** 2014. "The Workspace for Collaborative Editing." *Lausanne*. Accessed December 6, 2020. <http://dharchive.org/paper/DH2014/Paper-224.xml>.

- Huntington Library, Art Museum, and Botanical Gardens.** 2020. "MssHM 114, f. 217r." The Huntington Digital Library. Accessed July 6. <https://hdl.huntington.org/digital/collection/p15150coll7/id/30971>.
- IIIF Consortium.** 2021. International Image Interoperability Framework (IIIF). Accessed February 24. <https://iiif.io/>.
- Jänicke, Stefan, and David J. Wrisley.** 2017. "Visualizing Mouvance: Toward a Visual Analysis of Variant Medieval Text Traditions." *Digital Scholarship in the Humanities* 32 (suppl\_2): ii106–23. DOI: <https://doi.org/10.1093/llc/fqx033>
- Kane, George, and E. Talbot Donaldson.** 1988. "Classification of the Manuscripts." In *Will's Visions of Piers Plowman, Do-Well, Do-Better, and Do-Best*, by William Langland, Revised ed. Piers Plowman—the Three Versions 2. London, England: Berkeley, California, U.S.A: Athlone Press; University of California Press.
- Manly, John Matthews, Edith Rickert, Mabel Dean, Helen McIntosh, Margaret Josephine Rickert, and Geoffrey Chaucer.** 1940. *The Text of the Canterbury Tales, Studied on the Basis of All Known Manuscripts*. Vol. 2. Chicago: The University of Chicago Press.
- Matthews, P. H.** 2014. "Narrow Transcription." In *The Concise Oxford Dictionary of Linguistics*. Oxford: Oxford University Press. Accessed January 7, 2021. <https://www.oxfordreference.com/view/10.1093/acref/9780199675128.001.0001/acref-9780199675128-e-2537>
- McLuhan, Marshall.** 1962. *The Gutenberg Galaxy: The Making of Typographic Man*. Reprinted. Toronto: University of Toronto Press.
- Pérez Priego, Miguel Angel.** 1997. *La Edición de Textos*. Teoría de La Literatura y Literatura Comparada 20. Madrid: Editorial Síntesis.
- Pugliatti, Paola.** 1998. "Textual Perspectives in Italy: From Pasquali's Historicism to the Challenge of 'Variantistica' (and Beyond)." *Text* 11 (1998): 155–88.
- Robinson, Peter M. W.** 1994. "Collation, Textual Criticism, Publication, and the Computer." *Text* 7: 77–94. <https://www.jstor.org/stable/30227694>.
- . 2006. "The Canterbury Tales and Other Medieval Texts." In *Electronic Textual Editing*, edited by John Unsworth, Katherine O'Brien, and Lou Burnard. <https://tei-c.org/Vault/ETE/>.

———. 2021. Textual Communities (version 2.0). Accessed January 11. <https://textualcommunities.org/app>.

**Robinson, Peter M. W., and Elizabeth Solopova.** 1993. "Guidelines for Transcription of the Manuscripts of the Wife of Bath's Prologue." In *The Canterbury Tales Project Occasional Papers*, edited by Norman Francis Blake and P.M.W. Robinson, 1:19–52. Oxford, UK: Office for Humanities Communications. <https://wiki.usask.ca/display/CTP2/Original+Transcription+Guidelines>.

**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 University Press. Accessed January 13, 2021. [https://www.dbnl.org/arch/sale003buil01\\_01/pag/sale003buil01\\_01.pdf](https://www.dbnl.org/arch/sale003buil01_01/pag/sale003buil01_01.pdf).

**Salemi, Marco, Anne-Mieke Vandamme, and Philippe Lemey,** eds. 2009. *The Phylogenetic Handbook: A Practical Approach to Phylogenetic Analysis and Hypothesis Testing*. 2nd ed. Cambridge, UK; New York: Cambridge University Press.

**Shaw, Prue, and Peter M. W. Robinson,** eds. 2010. "VI. The Phylogenetic Analysis." In *Commedia*, by Dante Alighieri. Birmingham, UK: Scholarly Digital Editions. <http://sd-editions.com/AnaAdditional/commediaonline/home.html>.

**Smith, Catherine.** 2019. itsee-birmingham/collation\_editor\_core 1.0.4 (Version 1.0.4). *Zenodo*. Accessed June 25, 2020. DOI: <https://doi.org/10.5281/zenodo.3539578>

**Strohm, Paul.** 2016. *The Poet's Tale: Chaucer and the Year That Made the Canterbury Tales*. UK: Profile Books.

**Swofford, D. L.** 2003. "Quick Start." *PAUP\* Phylogenetic Analysis Using Parsimony (\*and Other Methods)* (version 4.0a [build 167]). Sinauer Associates. Accessed June 25, 2020. <https://paup.phylosolutions.com/tutorials/quick-start/>

**TEI (The Text Encoding Initiative).** 2020. "5 Characters, Glyphs, and Writing Modes - The TEI Guidelines." Accessed July 6. <https://www.tei-c.org/release/doc/tei-p5-doc/en/html/WD.html#D25-30>

**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. Translated by Michael D. Reeve. Storie e linguagie. Padova, Italy: Libreriauniversitaria.it Ed.*

**Windram, Heather F., Prue Shaw, Peter Robinson, and Christopher J. Howe.**

2008. "Dante's Monarchia as a Test Case for the Use of Phylogenetic Methods in Stemmatic Analysis." *Literary and Linguistic Computing* 23 (4): 443–63. DOI: <https://doi.org/10.1093/lc/fqn023>

**How to cite this article:** Vazquez, Adam. 2021. "Transcribing and Collating for Digital Stemmatology. The Case of *Troilus and Criseyde*." *Digital Studies/Le champ numérique* 11(1): 7, pp. 1–34. DOI: <https://doi.org/10.16995/dscn.372>

**Submitted:** 06 December 2019 **Accepted:** 12 June 2020 **Published:** 18 October 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 Studies/Le champ numérique* is a peer-reviewed open access journal published by Open Library of Humanities.

**OPEN ACCESS** The Open Access icon, a stylized 'O' with a person inside.