RESEARCH

Querying Variants: Boccaccio’s ‘Commedia’ and Data-Models

Sonia Tempestini¹ and Elena Spadini²

¹ Università della Svizzera Italiana, CH
² Université de Lausanne, CH
Corresponding author: Sonia Tempestini (tempestini.sonia@gmail.com)

This paper presents the methodology and the results of an analytical study of the three witnesses of Dante’s Commedia copied by Giovanni Boccaccio, focusing on the importance of their digital accessibility. These extraordinary materials allow us to further our knowledge of Boccaccio’s cultural trajectory as a scribe and as an author, and could be useful for the study of the textual tradition of Dante’s Commedia. In the first section of the paper, the manuscripts and their role in previous scholarship are introduced. A thorough analysis of a choice of variants is then offered, applying specific categories for organizing the varia lectio. This taxonomy shows how fundamental it is to combine the methodological tools for studying copies (as usual in medieval philology) and those for studying author’s manuscripts (as usual in modern philology) in dealing with the three manuscripts of Boccaccio’s Commedia: in fact, the comparative analysis of the three manuscripts has much to reveal not only of their genetic relationship but also of Boccaccio’s editorial practices. Furthermore, the analytic categories inform the computational model behind the web application ‘La Commedia di Boccaccio’, <http://boccacciocommedia.unil.ch/> created for accessing and querying the variants. The model, implemented in a relational database, allows for the systematic management of different features of textual variations, distinguishing readings and their relationships, without setting a base text. The paper closes on a view to repurposing the model for handling other textual transmissions, working at the intersection between textual criticism and information technology.

Keywords: digital philology; data model; Italian literature; database; textual variation
1. Introduction
§1 This paper presents the methodology and the results of an analytical study of the three witnesses of Dante’s Commedia copied by Giovanni Boccaccio, focusing on the importance of their digital accessibility. These materials allow us to further our knowledge of Boccaccio’s editorial practices as well as of his cultural trajectory as a scribe and as an author and could be useful for scholars who want to study the textual tradition of Dante’s Commedia.

§2 In sections 2 and 3, we summarize the analysis of the varia lectio of the text of Dante’s poem included in Boccaccio’s autograph manuscripts: Toledo, Archivo y Biblioteca Capitulares, Zelada 104 6 (To); Florence, Biblioteca Riccardiana, 1035 (Ri) (Figures 1–2); Vatican City, Vatican Apostolic Library, Chigiano L VI 213 (Chig). The witnesses have been entirely collated.¹

§3 In section 4, we introduce ‘La Commedia di Boccaccio’, <http://boccaccio­commedia.unil.ch/> , the web application created for accessing and querying the textual variants (cf. below, Figures 3–5). We focus on the conceptual model informing the database, with a view to its repurposing for managing other textual transmissions.

2. Giovanni Boccaccio, scribe and editor of Dante’s Commedia
§4 Giovanni Boccaccio’s works and autographs have seen a recent revival of popularity following the scholarly yield of the author’s 2013 centenary. Recent scholarly output spurred by novel discoveries of Boccaccio’s method has sought to elucidate one of his much neglected facet: his activity as editor and his concern for

¹ For sections 2 and 3, a more in depth study in Tempestini (2018a, 2018b); here they serve as a necessary but short introduction for the presentation of the web application and its subjacent model, in section 4. The complete transcription of To has not yet been published; the transcription of Ri, edited by Società dantesca italiana, is available at <http://www.danteonline.it/italiano/codici_frames/codici.asp?idcod=321>; the digital facsimile of Chig is available at <https://digi.vatlib.it/view/MSS_Chip.LVI.213> and its transcription is included in Tempestini (2018a). For the present study, transcriptions are diplomatic with a number of interventions: disambiguation of \(u/v\), expansion of the abbreviations (Tironian note \(? > et\); \(qua\to > quanta\)), separation of words in scriptio continua (\(chepiu > che pia\)); the scriptio continua is maintained in case of elision (\(chio > chio\)), apheresis of vowel-initial word (\(lonferno > lonferno\)), univerbation by syntactic doubling (\(allui > allui\)).
preserving and promoting the work of his masters (Cursi and Fiorilla 2013). Recent studies such as the discovery, by Sandro Bertelli and Marco Cursi, of a portrait of Homer signed in Greek by Boccaccio, on the reverse side of final sheet of the Toledano 104 6 manuscript, or of the recent identification, by Cursi himself, of Boccaccio's
handwriting on the margins of an Odyssey copy – ms. Gr. IX. 29 of the Marciana Library of Venice, that is Leonzio Pilato’s autograph manuscript annotated by Francesco Petrarca – have served to consolidate interest in Boccaccio’s cultural project and its design.\textsuperscript{2} For Boccaccio, Dante stands at once as the greatest representative of the recent literary past and the figurative heir to an unbroken tradition dating to the Greek Classics.

§5 The three copies of Dante’s poem transcribed by Boccaccio in order to broaden its circulation constitute a benchmark of his intellectual and editorial project. Boccaccio copied the \textit{Commedia} in the autograph manuscripts To, Ri and Chig; the latter was originally compiled in a single volume together with Chigiano L V 176.\textsuperscript{3} These manuscripts are anthologies of sorts: the first one, To, contains Boccaccio’s \textit{Trattatello in laude di Dante}, Dante’s \textit{Vita Nova}, Boccaccio’s \textit{Argomenti in terza rima}, Dante’s \textit{Commedia} and Dante’s fifteen canzoni “distese”; the second one, Ri, contains Boccaccio’s \textit{Brevi Raccoglimenti}, Dante’s \textit{Commedia} and fifteen canzoni “distese”; while the manuscripts known as Chigiano LVI 213 and Chigiano L V 176 are an anthology of Italian poetry, containing Boccaccio’s \textit{Brevi Raccoglimenti}, Dante’s \textit{Commedia}, Boccaccio’s \textit{Trattatello in laude di Dante} (in the second redaction), Boccaccio’s poem \textit{Ytalie iam certus honos}, Dante’s fifteen canzoni “distese”, Petrarca’s \textit{Rerum vulgarium fragmenta} and Cavalcanti’s \textit{Donna me prega} (Bologna 1994, 166–181). The three manuscripts are dated relatively close to each other: according to recent studies, To should be dated to the early 1350s and not later than 1355, Ri to 1360 and Chig between 1363 and 1366 (Cursi 2014). The sequence To — Ri — Chig, was already established by Barbi, then confirmed by Petrocchi and by recent scholarship (Alighieri 1907, CXCVI; Alighieri 1994, 40. Nevertheless Francesca

\textsuperscript{2} Cursi and Bertelli (2012, 2014) studied Homer’s portrait. Boccaccio’s signature is analysed in Petoletti and Martinelli Tempesta (2013). The annotations of Boccaccio and Petrarca are also studied respectively by Cursi (2015) and Pontani (2002–2003). For more about the role that Boccaccio played as a scribe in the birth of the Italian literary tradition, see Eisner (2013).

\textsuperscript{3} For more about To, see Cursi and Fiorilla (2013); Bertelli (2014); for Ri, see Bertelli (2014); for Chig, see Bertelli (2014), Cursi and Fiorilla (2013); for Città del Vaticano, Biblioteca Apostolica Vaticana, Chig. L V 176, see De Robertis (1974), Bertelli (2014, 4–6), Cursi and Fiorilla (2013, 48–49).
Malagnini [2006] provides a divergent opinion establishing the sequence To — Chig — Ri, on the basis of paratextual information.)

§6 In his edition of *La Commedia secondo l’antica vulgata*, Giorgio Petrocchi established that “il fondamento di To Ri e Chig è l’attuale Vat. lat. 3199 o un suo
Tempestini and Spadini: Querying Variants

section

Tempestini and Spadini: Querying Variants

Art. 1, page 6 of 28

"gemello" (Alighieri 1994, 42), referring to manuscript Vaticano Latino 3199 (Vat), probably gifted to Petrarch by Boccaccio (Traversari 1905; Billanovich 1947; Bertelli 2014, 35–38; Trovato et al. 2013). Although it was not Vat, as Petrocchi had already understood, the exemplar was certainly produced in the same workshop in Florence as Vat, now known as officina vaticana. Petrocchi identified with the workshop two manuscripts, Vat and Cha (Cha597 of Musée Condé de Chantilly), while Gabriella Pomaro later attributed other five manuscripts to the same family and to the same hand, excluding them as direct exemplars of Boccaccio’s copies (Pomaro 1986). Angelo Eugenio Mecca’s recent studies (2012, 2013, 2014) on Boccaccio’s manuscript rule out the possibility of their being directly derived from Vat (or its copy) and confirm their close proximity to the officina vaticana manuscripts, especially to manuscripts Laurenziano 40.13 and Marciano Zanetti 55. However it should be specified that Boccaccio’s editorial work on the Commedia did not rely on a single exemplar but was influenced by multiple sources.

§7 Finally, it is essential to recall the importance of the three manuscripts to the textual transmission of the Commedia as reconstructed by Giorgio Petrocchi: Boccaccio’s editorial activity would have increased contamination in the tradition and represented the limit of the antica vulgata, whose ante quem term would then be around 1355. Petrocchi’s edition is based on the complete collation of 27 manuscripts (referred to as antica vulgata, for the list see Alighieri 1994, 57–59) chosen among those dated, at the time of the edition, before 1355 – that is, before Boccaccio’s copies. However, Boschi Rotiroti’s recent codicological and palaeographical studies have shown that only 22 of these 27 manuscripts are actually ante 1355. Besides that, according to Boschi Rotiroti’s recensio codicum, a total of 85 manuscripts are datable within 1355 (Boschi Rotiroti 2004, 15–17).

§8 Today, the notion of an antica vulgata seems outdated (Tonello and Trovato 2013). Mecca’s studies, in particular, provide an incisive reflection on Boccaccio’s influence on the later tradition: “[...] tutto sembra indicare che da un punto di vista testuale non esiste uno sbarramento cronologico del Boccaccio nella tradizione manoscritta della Commedia” (Mecca 2013, 182).
§9 Concerning the relationship among Boccaccio’s manuscripts, in reference to To and Ri, Petrocchi concluded that, although they have been considered, “non s’è reso necessario l’analitico raffronto con Chig il quale si impone sugli altri con la qualifica di edizione ultima e definitiva del testo dantesco” (Alighieri 1994, 18–19).

3. A selection of variants

§10 In recent years, researchers have attempted to remedy the absence of analytical comparison between the three manuscripts, and a number of critical studies have sought to shed light on the significant textual differences among them (Mecca 2012, 2013, 2014; Breschi 2013; Bettarini Bruni et al. 2015).

§11 In this paper, we present a selection of data harvested from the results of the complete collation of Boccaccio’s autographs of the *Commedia*’s text, serving as case studies for the digital representation of the varia lectio discussed below.

§12 The varia lectio is conspicuous for the discrepancies it contains. There are more than 1500 places in the text where the three codices do not agree with each other and where they yield at least two different readings; there are over four hundred such instances in the first cantica, about five hundred in the second and almost six hundred in the third. Relevant linguistic variants were also considered in this calculation, while graphic variants are never included.

---

4 It is well known that different strategies have been used for editing the *Commedia*, concerning in particular the dichotomy between editions based on complete or partial collations of the witnesses, starting from Aldina edition by Pietro Bembo based on Vat (Alighieri 1502), and going through Carl Witte (1869–1879), Barbi’s *loci critici* (Bartoli, D’Ancona and Del Lungo 1891) and Vandelli’s edition (Alighieri 1927). If Petrocchi (Alighieri 1994) chose the complete collation of a small number of witnesses, in recent years Sanguineti (2001) and now Trovato (2007; Tonello-Trovato 2013) followed Barbi’s line using selective variant sites for building a *stemma codicum* (cf. the exhaustive paper by Viel (2011)).

5 The analysis (Tempestini 2018a) has been extended to include relationships with the *antica vulgata*, comparing the data emerging from the examination of the internal *varia lectio* with the *apparatus* of the critical edition by Giorgio Petrocchi. Comparisons have also been made with Mecca’s studies, particularly those concerning the manuscripts belonging to the officina vaticana, in order to distinguish the readings already present in the textual transmission from those only found in Boccaccio’s manuscripts. A new comparison should be made with the results of Tonello’s works (2018) on the Florentine tradition.
§13 Where there is a variation in the texts of the three cantiche, it is most often the Toledano manuscript that has a different reading, while Riccardiano and Chigiano manuscripts are typically in agreement. This is primarily due to the fact that most of the variants that were introduced by Boccaccio in Ri are replicated in Chig, while To is usually consistent with readings already present in the antica vulgata (as well as with the family of the exemplar of the manuscripts themselves, of which, as mentioned, is part Vat). The complete collation thus confirms Petrocchi’s argument concerning the three manuscripts’ progressive detachment from that family, as well as the chronological sequence To — Ri — Chig. It also appears possible to divide the codices into two groups: To on one side and the Ri — Chig couple on the other, since many of the innovations in Ri are retaken in Chig. The innovations in Ri and Chig add new and very significant lectiones singulares, a testament to Boccaccio’s continuous work on the text of the Commedia.

§14 Representing and organizing the varia lectio emerging from a complete collation is a complex and critical process, and the specific case is also unique for the notoriety of both the text and its scribe, as well as for the value of Boccaccio’s copies in the context of his broader cultural and editorial project. Boccaccio worked on the Commedia for over twenty years, editing a text which, as demonstrated by the three copies and the Esposizioni (Boccaccio 1965),6 would never acquire a stable form. In order to study the varia lectio, not with a view to settling on a definitive text of Dante’s Commedia, but rather in order to trace the evolution of Boccaccio’s Commedia and to try to glean something more concerning his editorial practices, we should reconsider the tools and categories through which we analyse textual variants. It is fundamental to establish and define a vantage point onto the matter of textual oscillation: given that our stress is on the modality of the transcription, rather than on the valeur ecdotique of the copied text, each reading should be examined

6 This is the text of the public lecture and commentary for which Boccaccio was commissioned and which he began on 23 October 1373 in Santo Stefano in Badia. The project was suspended in early 1374, likely due to Boccaccio’s illness. Boccaccio died in December of 1375; Esposizioni thus cuts off at the beginning of Inferno, XVII.
singly in order to theorise its origins and inner logic. Although some variants can easily be attributed to mechanical or palaeographic errors, the analysis of each reading will allow us to better appreciate the copyist’s method.

§15 The complete collation includes variants whose readings are absent in the *antica vulgata*, suggesting that they are Boccaccio’s own innovations: about 240 cases, from over 1500 variations sites emerging from the full collation. (The term “innovation” thus includes also erroneous readings, given that this is an analysis of places where the three manuscripts do not agree, which does not include innovations that passed on from To to Ri and to Chig.)

§16 In the attempt to analyse the qualitative aspects of the textual variants we have had to settle on a limited number of categories reflecting the most common forms of divergence in the innovations. We propose an articulation and differentiation, even if provisional, of the *varia lectio* that includes categories used by the *filologia d’autore* (Stussi 2011, 182–183). Indeed, as mentioned before, the study of the variants for themselves, of the text of Boccaccio’s *Commedia* in its oscillations and of Boccaccio’s editorial behaviours, requires a different point of view on the copyist, the author himself of that textual *mouvance*.

Five categories describe the most significant forms of divergence:

- inversion;
- addition/omission;
- inflection;
- echo/anticipation;
- lexical.

The first two categories are easy to identify: Inversions are readings where the order of two or more words is inverted; e.g.:

*Inf. XV, 87 mia lingua*] mia lingua To lingua mia Ri Chig; *Inf. XVI, 123 tuo viso*] tuo viso Ri Chig viso tuo To; *Purg. XX, 53 li regi antichi*] li regi antichi Ri Chig gli antichi regi To; *Par. IV, 47 Gabriel e Michel*] gabriel et michel To
michele et gabriel Ri Chig). (Reference readings before square brackets are taken from La Commedia secondo l’antica vulgata by Giorgio Petrocchi.)

By addition/omission we mean readings where words are added or omitted; these are often short words, mostly conjunctions, pronouns, and articles. Only in rare cases are they nouns or, in the case of scribal error, verses; e.g.:

Inf. I, 47 test’alta] testa alta To Ri; testa Chig; Inf. IV, 141 Tulio e Lino] tullio et lino To; tullio lino Ri Chig; Par. I, 53 imagine mia] ymagine mia To Ri ymagine Chig; II, 103 quanto tanto] quanto tanto To quanto il tanto Ri Chig).

The other three categories could be defined as three different forms of substitution or replacement, often with a fine line differentiating them. The first kind of substitution is related to readings that arise from a divergence in grammatical flexion: a change in number, gender, tense, verbal mode. For example:

Inf. II, 38 novi] nuovi To nuovo Ri Chig; Inf. XII, 57 andare] andare Chig andando To Ri; Inf. XVIII, 55 fui] fui To son Ri Chig; XXIX, 84 larghe] larghe To Ri largo Chig; Par. XXIV, 13 cerchi] cerchi To Ri cerchio Chig.

On occasion, they involve rhyme, as in the case of Inf. XXI, 129 credi] credi To creda Ri Chig, where the rhyme is: piedi/vedi/credi while creda is likely influenced by the rhyme in –ema at Inf. XIII, 128,130,132 scema/prema/gema or for Inf. XXII, 50 chiostri] chiostri To chiostro Ri Chig.

The fourth form of innovation is echo of a previous verse/anticipation: readings influenced by nearby passages in the text. This form of variant, sometimes attributable to scribal or mnemonic errors, is quite common. Among the most significant and broadly commented upon cases are Inf. V, 113 pensier] pensier To Ri sospiri Chig: Sospiri is influenced by Inf. V, 118: [...] doki sospir; Inf. VIII, 69 grande] grande To grave Ri Chig: Grave echoes the first hemistich of the verse (Coi gravi cittadin, col grande stuolo); Inf. XXV, 65 bruno To Ri nero (ad
marg. bruno) Chig: Nero: rhyme, written by mistake from the following verse: Inf. XXV, 65–66: per lo papiro suso, un color bruno/che non è nero ancora e ’l bianco more. Inf. XVIII, 99 piangendo] piagnendo Ri Chig correndo To: rhyme, scribal error from Purg. XVIII, 97 Tosto fur sovr’a noi, perché correndo; Par. IV, 36 eterno spiro] eterno spiro To dolce spiro Ri Chig: echo of verse 35 e differentemente han dolce vita.

§17 The fifth category is a proper replacement indicated as lexical variant: a substitution within the text, not attributable, or at least not clearly, to echo or anticipation even as it might generally be influenced by other distant passages in the text. At times synonyms, at others Boccaccio’s own comments on the text (often in conversation with commentaries to Dante’s Commedia pre-dating 1375), the most significant variants are nouns, verbs, and adjectives with substitutions of particles, pronouns and prepositions notably lacking. Those readings are among the most instrumental for understanding the degree of the copyist’s autonomy, given such variants cannot easily be identified as errors, and are more significant than a simple inversion; these variations were deliberately arbitrated by the copyist. We have analysed some of these readings in order to better explain their importance: Inf. I, 17 vestite] vestite To Ri; coperte (in marg. al. vestite) Chig: this reading (coperte) is likely derived from the Commedia’s oldest commentaries: see Jacopo Alighieri’s commentary (de raggi del sole coperta la vide) and Ottimo Commento (la vide coperta delli raggi del sole) (cited from the commentary to Inferno, I. 17 [Alighieri 1915; Di Fonzio 2008]. These resources are available in the Dartmouth Dante Project, https://Dante.Dartmouth.EDU.) Inf. XXXIII, 59 per voglia] per voglia To Ri per fame Chig: Boccaccio seems to analyse the text: per fame is synonymous to voglia di manicar. “ambo le man per lo dolor mi morsi; /ed ei pensando ch’io’l fessi per voglia/di manicar”. However, in the previous canto (when Ugolino bites Ruggieri’s head) we find per

---

7 The autographs’ marginal variants are noted by *in marg* and in round brackets. There are three kinds of marginal readings: readings not preceded by any sign, which correct copy errors, readings preceded by *al* (aliter) which indicate readings equally accepted (adifere) and variants preceded by *c* (corige o credo) which indicate corrections or conjecture (Feola 2008). For the different hypotheses regarding the meaning of *c* and for the relevant bibliography see also Finazzi (2013, 100–101).
fame followed by *si manduca*: Inf. XXXII, 127–129 e come 'l pan per fame *si manduca/* cosi 'l sovan li denti a l’altro pose/ là ve 'l cervel s’aggiugne con la. *Purg.* XXIII, 133

*lo vostro regno*] *lo vostro regno* To Ri; *il vostro monte* Chig *Purg.* XXIII, 131–133: e addita’lo; 'e quest’altro è quell’ombre/ per cui scosse dianzi ogne pendice/ *lo vostro regno, che da sé lo sgombra*” Dante pointed at Statius, for whom the mountain of Purgatory trembled: quand’io senti’, come cosa che cada, tremar lo monte, onde mi prese un gelo (*Purg.* XX, 127–128). In describing the earthquake, Boccaccio opted for *monte* over *regno* in ms. Chigiano. See also the commentary by Jacopo della Lana: E quest’altro, cioè Stazio, per lo quale tremò lo *monte*, come è detto nel XX capitolo del Purgatorio (cited from the commentary to *Purgatorio* XXIII, 131–133 (Scarabelli 1866–67). *Purg.* XXV, 13 *voglia*] *voglia* To Ri *voce* Chig: Tal era io con *voglia* accesa e spenta/*di dimandar*, venendo infino a l’atto/che fa colui ch’a dicer s’argomenta (Inf. XXV, 13–15). *Voce* is easily explained, this reading is like a comment: *voce* instead of *voglia* (*di dimandar*) and very close to the reading of the same manuscript at Inf. 33, 59, that is *fame* instead of *voglia* (*di manicar*).

§18 Complete transcriptions and collations of long textual works are, in general, not easy to produce and manage.8 Beside that, it is unusual to convey the complexity of the *varia lectio* not in a critical *apparatus* of a scholarly edition. Nevertheless, we consider fundamental to secure the availability of this data for scholars. On paper, the search for specific variants and categories, according to any classification, is certainly complex and demanding. A digital representation, on the

---

8 An important endeavour in this direction is Shaw’s digital edition of Dante’s poem (Shaw 2010), which presents full transcription and collation of the seven witnesses; in particular, the VBase system to access variants allows for fine-grained queries (see Spadini 2015). Nevertheless, the role of IT in the project discussed in the present article and in Shaw’s work are different, being far more complex in the second. In Shaw’s case, the collation and the analysis of the textual variants occupy a preliminary phase (*recensio*) in the editorial workflow and benefit from computer-assisted tools. Here, on the contrary, the collation has been performed manually (since the work was almost finished when the first attempts with automatic collation were made, and it would have been too time-consuming to redefine transcription and collation criteria for the software) and the analysis is not part of an editorial workflow; the IT contribution, here, is limited to the data-model informing the database schema and to the web application for querying the data.
contrary, might provide ready accessible and searchable materials: with this aim we devised ‘La Commedia di Boccaccio’ (Figures 3–5).

§ 19 This web application, further presented in the next section, allows for the visualization and querying of the variants between the three manuscripts, including Petrocchi’s critical text for reference. The unit, here, is the reading: the database

![Figure 3](image3.png)

Figure 3: Web application ‘La Commedia di Boccaccio’: home page with advanced query interface.

![Figure 4](image4.png)

Figure 4: Web application ‘La Commedia di Boccaccio’: results of a query selecting only variants in *Inferno*, V.
contains over 4500 readings from around 1500 variation sites. Each reading belongs to a witness (To or Ri or Chig) and corresponds to a textual location (cantica, canto, verse). Possible queries concern the single reading (present or absent in the antica vulgata), or the relationship between readings: the latter is described with the combination of witnesses involved (To vs Ri + Chig or Chig vs To + Ri etc.) and the category of change indicating the different kinds of variation (inversion; addition/omission; inflection; echo/anticipation; lexical). As mentioned above, only those readings that can possibly be considered as Boccaccio’s innovations, and not other variants, are marked with a category of change.

4. Textual variants and data-models

§20 As explained in the previous sections, scholars need to compare, classify and order textual variants in order to make sense of them. This is one of the areas where Textual Criticism meets Information Technology (IT). There are a number of benefits to the organization, storage and querying of the data in digital form, not least of which is the ease in retrieving, reusing and sharing them. In this section, we will demonstrate how a theoretical modelling on the data and the consequent creation of a data-model are pre-requisites to conceiving such classificatory systems, to be exploited by means of IT resources. In order to do so, we will further discuss
§21 Computers do not understand texts. Not only they are unable to comprehend the semantic information that they carry, but even their elements of structure, including the simple fact that texts are generally composed of words made up by characters. It falls to the user to tell the machine exactly what kind of data she is working with: whether it is a string (sequence of characters), a date, a decimal number, \textit{vel sim}. She also has to specify how to parse the data (tokenization) – into words, for example – and provide guidelines for that segmentation, specifying, for instance, that each section of string between two white spaces is a word. (As this task can be accomplished by scholars who are not strictly speaking programmers and to avoid to set a precise division of labour, which this kind of work challenges, we prefer the term ‘user’ taken in a broad sense.) Following this rule, two words separated by a hyphen or by an apostrophe will be considered a single unit. This specification may or may not be suitable for the end goal, be it the generation of a list of words or an analysis through complex algorithms. As this simple example demonstrates, the user has a fair degree of freedom; thus, it is crucial that she maintains a clear understanding of the data and of its eventual application. In the Humanities in general and, more relevant for our purposes, in Textual Criticism, understanding often requires some form of interpretation: from transcribing (Segre 1976), to translating, to judging variants, the significance of interpretation varies depending on the margin of consensus for a certain explanation. It is equally important to take the data's eventual application into account. Returning to the previous example, compound expressions can be considered multiple units when generating a frequency list and a single unit when conducting a semantic analysis; \textit{liez et dolenz}, for instance, functions as a single unit for a sentiment analysis of Old French Arthurian romances, in which this syntagma occurs often, and as three separate words in most other cases. To summarize: understanding the data hinges on interpretation and goals, or better, on the interpretation of the data itself and the interpretation of the goals. In order to fully exploit the realm of possibilities granted to the user, it is not only important
to understand the data, but also to convey this understanding to the machine. In other words, in order to control the machine and not to be controlled by it. For this purpose, knowledge must be organised into a model, in this case a data-model. A data-model is nothing more than a formalization of our understanding in ways sufficiently internally consistent, logically coherent and explicit to be applied in one or more computational implementations.\footnote{The implementation should be distinct from the conceptual modelling, but the stress here is on the consistency, coherence and explicitness required for the model to be computable. Note that the model outlined here is not represented in any formal language and can be considered a loose data-model. For a clear and synthetic approach to data-modelling in the fields of (Digital) Humanities, see Flanders and Jannidis (2016). Cf. materials available online from the workshop ‘Knowledge Organization and Data modelling in the Humanities’, organized by Flanders (Women Writers Project) and Jannidis (Center for Digital Editions at the University of Würzburg) in 2012, <http://www.wwp.northeastern.edu/outreach/conference/kodm2012/index.html>. The contributions of Tito Orlandi remain fundamental for understanding the practice of modelling related to systems theory in Digital Humanities (at least, Orlandi 1999). Willard McCarty’s analysis (2004, 2005). In the field of Textual Criticism, particular attention is devoted to modelling in Pierazzo (2015, 37–40 and elsewhere). Cf. Unsworth (2002, now in Terras et al. 2014).} In what follows, we present the model of textual variation conceived and implemented in ‘La Commedia di Boccaccio’. The source data, in this case, are the materials that emerged from the collation of the three witnesses: hundreds of words. (Note that only the variants, as emerging from the collation, are retained, i.e. the corpus is composed by all the variants, and not by the entire texts.) These words are not treated individually in themselves, but rather as elements of a reading. The reading, in this case, is the atomic unit composed of one or more words. (A first form of interpretation acts in the identification of the boundaries of the reading and the parsing of the variation sites, particularly in instances in which there are a number of possible alignments between witnesses.) Each reading has two features and can be analysed with reference to a series of categories. The features are the reading’s place in the work and the identity of the witness. For example, the reading ‘grameça’ is found in manuscript Chigiano in Inferno I, 50. The location is, in this case, expressed through three sub-features: cantica, canto and verse.
§22 In addition to these features, for which a low level of interpretation is required, readings can be analysed, classified, interpreted in a number of other ways. The aims and the criteria of the analysis may vary, heavily depending on the work in question and on the methodological approach.

§23 In the case of textual variation, an initial distinction can be made between categories that concern (A) a single reading and those that describe (B) the relation among the readings, i.e. the variation. For example, the notion of hypermeter or of error can only be applied to a single reading, and not to the relation among the readings, while an addition or deletion involves the presence of two or more readings. (An error can be spotted by comparing readings; nevertheless 'erroneous' is not a quality of the relation, but of the reading.) As far as concerns the relation among the readings, the model does not set a base text: thus, the readings are all on the same level, each of them being a variant to each other. Addition and deletion can be considered together because no orientation is set for interpreting a deletion as an addition or vice versa. As we shall see below, the absence of a base text is also essential to defining the entities to be compared.

§24 The taxonomy used in 'La *Commedia* di Boccaccio', presented in the previous section, is recalled here from the point of view of data-modelling. As mentioned above, different texts and methodologies require different categories to be enlisted. Here, for each individual reading (A), it is specified whether it is present in the textual transmission prior to Boccaccio (the *antica vulgata*). The relation between readings (B) might demand a generic classification, such as linguistic categories or categories of change (*adiectio*, *detractio*, *immutatio*, *transmutatio*). The taxonomy in use here includes some of them, while others have been expressly designed to address specific issues relevant to this corpus. Thus, the relation between the readings (B) is described in terms of addition/omission, inversion, morphological inflection, echo/anticipation and lexical change. (The 'echo/anticipation' category is here considered as a form of substitution, but could also be regarded as a feature of the single reading.) Furthermore, it is specified if the variation occurs in rhyme or not.
§25 As said above (section 3), the aim of this study is not the building of a *stemma* representing the textual transmission; rather, it is the analysis of the textual dynamics in Boccaccio’s copies: the methodology adopted is in fact closer to genetic criticism than to stemmatology. Therefore approaches such as those tested in Spencer, Mooney, Barbrook, Bordalejo, Howe and Robinson (2004) and Andrews (2016), exploring the use of weighted variants, have not been applied here.

§26 As far as concerns the relation between the readings, the combination of witnesses to analyse can vary: the three witnesses can be considered together, or in varying pairings.

Consider the following hypothetical variants:

(1) A: *cat* | B: *dog* | C: *cat* | D: *CATS* | E: *cat*;

it is impossible to give a unique and detailed description of the relations between all of them. These include, for instance, a lexical change, a morphological one and a typographical one. Further information is needed to specify among which witnesses these changes occur. A more thorough interpretation would be:

(2) ACE vs B, lexical change; ACE vs D, inflection change; ABCE vs D, typographical change.

Given that the combinations of witnesses may change for each variation site, the more consistent way to pursue the variation is to examine the witnesses in pairs.\(^\text{10}\)

(3) A vs B, lexical; A vs C, no variation; A vs D, inflection — typographical; A vs E, no variation; B vs C, lexical; B vs D, lexical — inflection — typographical; B vs E, lexical; C vs D, morphological — typographical; C vs E, no variation; D vs E, inflection — typographical.

---

\(^{10}\) Vanhoutte (2007): “Recording each class for each possible relationship each location variant can have with all corresponding location variants from the other witnesses is therefore the closest approximation to an explicit classification one can aim for.”
From this complete description (3), it is possible to return to the previous one (2), and vice versa. It is important to remember that no base text has been set.

§27 Given this theoretical framework, implementation may differ depending on the corpus. In the case of the Boccaccio’s copies of the *Commedia*, the pairs are replaced by combinations. This is because the variance never affects all three witnesses at once, but always opposes two witnesses to a third. Recording all the pairs would therefore be redundant.

To summarize, the model outlined here has a number of crucial characteristics:

- it distinguishes between the features of the reading and those of the relations between the readings;
- it allows us to append more than one feature to each reading and relation;
- it does not require a base witness to orient the variation;
- it permits to annotate each pair of witnesses or a combination of them for each variation site.

In the project ‘La *Commedia* di Boccaccio’, the model has been implemented into a relational database. Other possibilities such as XML (TEI) or a graph database might also have been suitable, given that they meet the requirements for the use of a widespread format that follows rules defined by standards: this is fundamental for facilitating preservation, data sharing and in order to make use of existing frameworks. As said, in this case data are organized in a relational database (Figure 6), which can be queried through SQL and published. Data are inserted into the database using spreadsheets automatically converted into SQL instructions.

§28 The two central tables, ‘reading’ and ‘annotation’, collect most of the others’ information. The table ‘reading’ carries the features of a single reading (in the taxonomy referred to as A), while the ‘annotation’ table contains descriptions of the relations between the readings (B) and the combination of witnesses involved.

---

11 A first attempt to encode the variants following the taxonomy just described with XML/TEI suggests the use of the elements <app>, <rdgGrp>, <rdg>, each of which may carry attributes such as @type and @ana. E.g.: <app><rdgGrp type="add_om"><rdg wit="RI">testa alta</rdg><rdg wit="CH" type="absent">testa</rdg></rdgGrp></app>

A standoff approach could be envisaged for storing information about all possible pairs of witnesses.
§29 As with any other row in a table, each reading has a unique identifier. While each reading has features such as location and witness belonging, they do not suffice for its identification, since two readings can belong to the same verse and to the same witness (e.g. ‘nella sua’ and ‘grameça’ in Chigiano at Inf. I, 50).

5. Conclusions

§30 In this study and the related web application, we aim at providing resources for a better understanding of the *Commedia* copied by Giovanni Boccaccio, his role as a copyist — he being an author —, and the relations between the three witnesses.

§31 As mentioned in the first part of the article, it is not our intention to sponsor a cult of textual variants. Rather, our aim is to give voice to the variants in the text in order to analyse the editorial practices of a *sui generis* scribe. The data collected highlights not only a much greater tendency for innovation in mss. Riccardiano and Chigiano, but also the considerable importance of some of these innovations. Ms. Chigiano in particular does not appear as Boccaccio’s final and definitive edition of Dante’s *Commedia* as Petrocchi suggests (Alighieri 1994, 19), but as an experimental text in which the copyist manifests his freedom; ms. Chigiano can thus be defined as Boccaccio’s own edition. This likely explains Boccaccio’s choice

---

**Figure 6**: Model of the relational database of ‘La *Commedia* di Boccaccio’.
of ms. Toledano, the least original and customized of the manuscripts, for reference in his *Esposizioni sopra la Commedia*, his public reading and commentary. Finally, in discussing variants, it is not entirely possible to judge whether some of them are due to mnemonic errors or to a conscious choice without entering into a discussion that goes beyond the limits of empirical data. Nevertheless, in analysing the trajectory from To to Chig, it seems evident that the development of the work did not aim at the most authoritative *ante litteram* edition of Dante’s *Commedia* but rather at a simplification of the text in which the boundary between exegete and scribe/copyist comes increasingly to be blurred.

§32 In the second part of the article, we propose a conceptual data-model for textual variants, which can be tested on other corpora beyond the project presented here and its specific taxonomy. A model is inherently selective in its features, and describing it requires providing an answer to the question, “what attributes of the original does it capture and make explicit, and which does it omit?” (Unsworth 2002, now in Terras et al. 2014). When considered as the result of selection and abstraction, the model resembles a map: as the expression has it, it cannot be the territory. Modelling is also a fundamental step for any attempt of automation. This is not to say that our long-term goal is to have machines that recognize and classify textual variants, but that structuring our understanding of these phenomena is a first, fundamental step towards leaving to computers some of the most time-consuming work and keeping for scholars the tasks that requires all their knowledge and experience: for instance, choosing the categories and making sense of the output.

§33 In theory, most of the categories used in the model could already be automatically identified. This is the case for additions, deletions and transpositions,

---

12 Petrocchi’s first hypothesis was that the reference code used for the text of *Esposizioni* was Vat (Alighieri 1994, 42). When he returned to the issue ten years later, he wrote, «Un’altra mia vecchia scheda mi chiede: quando scrive il Commento quale dei suoi testi ha dinanzi? Credo il Toledano; [...] anche se non mancano presenze d’altre tradizioni, ma così saltuarie e testualmente poco rilevanti da non consentire riferimenti perentori» (Petrocchi 1979, 20). Francesco Feola (2008) concurs with Petrocchi in a study based on the analysis of the marginal variants in To Ri and Chig. For a more comprehensive bibliography and discussion around this subject see also Tonello (2015) and Tempestini (2015). Those studies agree to identify To as the main reference for the text of *Esposizioni*. 
recognized by the most advanced software for automatic collation; for morphological variation, which can be spotted using Natural Language Processing tools; for the position of the variation site, in rhyme or not, which can be checked with a simple algorithm. The identification of lexical variations and echoes would require more complex algorithms, while ascertaining whether the reading exists in the *antica vulgata* ‘merely’ requires the transcriptions of all the witnesses of the *antica vulgata* to be available, with a suitable copyright, in digital form. But this is only a possible future. At present, the variants of Boccaccio’s *Commedia* are available and will hopefully stimulate and facilitate further studies.

**Contributors**

*Editorial*

**Recommending editors:** Franz Fischer, Università Ca’ Foscari Venezia

**Recommending referees:** Peter Robinson, University of Saskatchewan, Canada; Michelangelo Zaccarello, Università degli Studi di Pisa

*Authorial*

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

- Sonia Tempestini: st
- Elena Spadini: es

Authors are listed in descending order by significance of contribution. The corresponding author is es.

- Conceptualization: st, es
- Data curation: st
- Formal analysis: st, es
- Methodology: st, es
- Software: es
- Project administration: st, es
Writing: st, es

Competing Interests
The authors have no competing interests to declare.

References


Eisner, Martin. 2013. Boccaccio and the Invention of Italian Literature: Dante, Petrarch, Cavalcanti, and the Authority of the Vernacular. Cambridge University Press. DOI: https://doi.org/10.1017/CBO9781107300484


**Tempestini and Spadini: Querying Variants**

---


