Exercises in Modelling: Textual Variants

Abstract

The article presents a model for annotating textual variants. The annotations made can be queried in order to analyse and find patterns in textual variation. The model is flexible, allowing scholars to set the boundaries of the readings, to nest or concatenate variation sites, and to annotate each pair of readings; furthermore, it organizes the characteristics of the variants in features of the readings and features of the variation. After presenting the conceptual model and its applications in a number of case studies, this article introduces two implementations in logical models: namely, a relational database schema and an OWL 2 ontology. While the scope of this article is a specific issue in textual criticism, its broader focus is on how data is structured and visualized in digital scholarly editing.

En forçant un peu, on pourrait imaginer que si quelqu'un trouvait un manuscrit des *Exercices de style* il se demanderait s'il ne s'agit pas d'une collection de variantes, trace d'une hésitation de Queneau entre diverses manières de raconter son histoire.

D. Ferrer, *Logique du brouillon*, Seuil 2001, p. 133

Introduction

Textual variation is a central object of study for textual criticism, *philologie*, scholarly editing.

The variation takes place when there are competing readings of a portion of a work. It might take different shapes: it occurs inside the same document (striking out, additions, etc.) or between documents (witnesses of the same work). The nature of the variation is also variegated: the difference among readings might concern formal or substantive text features, where—generally and traditionally—the first relate to orthography (spelling, punctuation, etc.) and the second to all other linguistic categories (morphology, syntax, lexis).

Finding patterns in the moving universe of textual variation is one of the scholar's goals. A writer might consistently remove references to his private daily life, moving from a note in a diary to a draft of a chapter[i]. A copyist might rewrite an entire text, according to changed orthography conventions[ii]. These kinds of patterns indicate the direction of changes, tracing precious paths for exploring the work and its *mouvance*[iii]; they help making sense out of a shapeless set of variants and shed light on textual dynamics. In stemmatics, patterns of substantive variants and, in particular, errors are also used to infer relationships among the witnesses and for drawing a stemma that accounts for the textual transmission.

This article introduces a model for annotating textual variants. Querying the annotations made, allows us to find patterns in textual variations. Instead of looking at a variation site as a single entity, the model attempts to decompose it and to explore its constituent parts: the readings and their relationships. For doing so, the model proposes to use a set of common general categories and other optional specific categories. These categories describe the features of the readings and those of the variation between them.

The model aims to be generic and applicable to a wide range of works. Nevertheless, the specific categories to be used for annotating the texts might vary greatly, depending on the texts themselves and on the scientific approach[iv]. For example, a relevant category for studying the transmission of a medieval text might be the *saut du même au même*: it proves the tight relation among the witnesses because it is an error which hardly occurs by chance at the same point in unrelated witnesses. When studying modern manuscripts, a relevant category might be that of instant rewriting,[v] which is the opposite to later rewriting. Often, the same phenomenon can be covered with different approaches: in the example of the removal of references to private life in an author's papers, above, an *ad hoc* category could be created, to annotate every relevant passage; another approach would be to decompose the phenomenon into smaller ones, and use multiple categories, such as the replacement of proper nouns with common ones,[vi] the removal of dates, etc., all leading to the removal of private-life references.

Modelling, in this article, refers to the "heuristic process of constructing and manipulating models" [McCarty, 2004],[vii] and, in particular, data models. A data model is a formalization of the understanding and interpretation of an object, which should be consistent, coherent and explicit; these characteristics allow to move from a conceptual model to a logical model, that is a computable object to be implemented in one or more physical models (Flanders and Jannidis, 2015: 11; Flanders and Jannidis, 2016).[viii] The conceptual model is here introduced using an entity-relationship diagram, while the logical view is presented in two schemas (relational tables and

OWL ontology). A number of case studies where the conceptual model is implemented are also presented.

Conceptual model

The model covers textual variants, that is, competing readings, and does not take into account the rest of the text. This means that it does not allows to reconstruct the entire text of each witness or stage; on the contrary, it only represents what is traditionally gathered in the critical apparatus.[ix]

A reading is the atomic unit of the model. A reading is a string of characters in plain text, with no typographical, structural or semantic markup; it is composed by one or more letters, or one or more words. The scholar is at liberty to choose the boundaries for each reading, following strategies that might differ from case to case, also within the same text. Because the model does not represent the rest of the text, the reading might include some non-variant words, in order to better contextualize the variant reading. This is what happens in a traditional critical apparatus, where the non-variants words are often abbreviated, while the variant words are spelled in full, as in the following example:

Critical text: *Il se vantoit de folie*
Apparatus: *Il se vantoit*] A, *qui se v.*[x]

The model describes two main aspects of the elements involved in the variation: the features of each single reading and those of the variation between them [Illustration 1]. This distinction is a fundamental characteristic of the model.

Illustration 1: In blue, the space for the features of each reading and in red the space for the features of the variation between them. This distinction is a fundamental characteristic of the model.

Features of the reading

For each single reading, two general features must be set: the witness to which the reading belongs, and the location of the reading in the witness[xi]; optionally, the location of the reading in the work might be added [Illustration 2].

Illustration 2: General features of each reading.

Each single reading can also be annotated using customized categories, which might vary greatly. A relevant feature recorded in a category might be the writing tool associated with the reading, mostly in the case of modern manuscripts. Another category can be set to record erroneous reading, for instance bringing to a metric violation when too short or too long, or repeating erroneously a word remained in the memory of the scribe. These *ad hoc* categories are to be added to the general ones [Illustration 3].

Illustration 3: Example of general and specific features of each reading.

Features of the variation

The features of the variation express what kind of difference exists between the competing readings. Two categories are used to record the general features of the variation: the category of change and, in the case of substitution, the linguistic aspect involved [Illustration 4].

The categories of change are addition, deletion, substitution and transposition. These four classes, referred to as *quadripartita ratio* (*adiectio, detractio, immutatio, transmutatio*) are defined as the categories of mutation by stoic philosophers and used by classical and modern rhetoricians. They correspond to the operations used for calculating the difference between two strings in computer science, known as edit distance[xii], and have been used in Textual Criticism for classifying variants (Stussi, 2011: 182). A substitution includes everything that is not only an addition, a deletion or a transposition: it might contain them, but not be limited to it.

The linguistic category defines which aspect of the language is involved in the variation: orthography, morphology, syntax, lexis.

Illustration 4: General features of the variation.

An example for the use of such general categories is the following: 'I still had one bad leg' vs 'I had still one bad leg' (O'Reilly, Van Hulle, Verhulst & Neyt, 2016)[xiii], which can be annotated as a transposition (category of change). Another case might be: 'Et lors parla mestre Helie di Tolose' vs 'Et lors parla maistre Helie di Tolose' (Micha, 1978-83, IV), where 'mestre' vs 'maistre' is a substitution (category of change) concerning orthography (linguistic category).

Specific categories can also be used to describe precise features of the variation. A relevant one might be the direction of the relation, that is from reading A to reading B, or the contrary. A specific category can be used, for instance, to record the type of intervention occurring: in the case of a substitution, reading A might be crossed out and reading B written above, below, after, etc. (Italia, 2010, 64).

These specific categories for describing the variation between the readings are to be added to the general ones (Illustration 5).

Illustration 5: Example of general and specific features of the variation.

The features of the readings coexist with the features of the variation [Illustration 6].

Illustration 6: Example of features of the readings and of the variation.

Variation site: pairs of readings

When a variation site involves more than two readings, a number of phenomena take place at once, and describing them might require complex annotations. This is particularly relevant when no direction of change has been set in advance, that is when the relations between the readings are not known. In most of the case in medieval textual transmissions, for instance, at first the scholar might want to compare all the readings, without setting, more or less arbitrarily, a base text (Spadini 2016).

A simple example of variation site involving four readings is the following[xiv]:

BnF fr. 1466 (*A*): *totes bontez pardue*
BnF fr. 1430 (*B*): *totes hennors pardues*
BnF fr. 118 (*C*): *toutez honneurs perdues et toutes ioyes*
BnF fr. 751 (*D*): *totes honors perdus et totes lois*.

As said above, the boundaries of each reading can be decided freely. In this case, the texts might be divided in various ways: for example, aligning word by word, considering the entire sentence at once, or separating the sentence in two at the conjunction "et". The latter scenario gives:

(1)
A: *totes bontez pardue*
B: *totes hennors pardues*
C: *toutez honneurs perdues*
D: *totes honors perdus*
(2)
A: /
B: /
C: *et toutes ioyes*
D: *et totes lois*

In (1), "bontez" (*A*) is different from "hennors" (and its orthographic variants, *BCD*).

In (2), *A* and *B* are null, while *C* and *D* have readings which are close at the paleographical level, but whose meanings are far ("ioyes" vs "lois").

Using the model (only the general features of the variation, that is category of change and linguistic aspect), they can be described as follows:

(1) *A* vs *BCD* substitution lexis orthography; *B* vs *C* vs *D* substitution orthography.
(2) *AB* vs *CD* addition/deletion; *C* vs *D* substitution lexis orthography.

Given that the combinations of readings may change for each variation site (*A* vs *BCD, B* vs *C* vs *D, AB* vs *CD, C* vs *D*), the more consistent way to pursue the variation is to examine the witnesses in pairs[xv], which produces:

(1) A vs B substitution lexis orthography; A vs C substitution lexis orthography; A vs D substitution lexis orthography; B vs C substitution orthography; B vs D substitution orthography; C vs D substitution orthography.

(2) A vs C addition/deletion; A vs D addition/deletion; B vs C addition/deletion; B vs D addition/deletion; C vs D substitution lexis orthography.

From this complete description, it is possible to obtain other, less redundant, ones, combining the readings as above.

In principle, the model could accept more than two readings for each variation, and use the same features of the variation to describe the differences between all of them. One of the main characteristic of the model, however, is to break up the variation in its constituent parts, in order to achieve the maximum of expressiveness[xvi].

This description only covers the features of the variations between the readings. Each reading *per se* can also be annotated with specific categories; here an appropriate category would be 'error', since "pardue" (A) is erroneous because singular and "perdus" (D) is erroneous because masculine.

All the selected features of the variation site can be represented together [Illustration 7].

Illustration 7: A variation site with multiple readings.

Boundaries of the readings, nested variants and concatenation

Setting the correct reading boundaries is not the only way to manage the variation extent. A variation site might also be contained by another variation site. This is the case, in particular, for variations of smaller size (for number of characters involved) inside a variation, to be called nested variants; and for recording the evolution of a reading in a variation site, to be called concatenated variants. It is important to remember that the sub-reading inherits the features of the reading it is part of.

An example of the first type—variation of smaller size inside a variation—is *A* "La luna o la Ricordanza" vs *B* "La Ricordanza" [Italia, 2010, 68-71]. *A* vs *B* might be described as an addition/deletion; inside it, there is an orthographic substitution, opposing "la" to "La" [Illustration 8]. In this case, the two sub-readings are parts of two different readings.

In the second case – recording the evolution of a reading– a sub-reading is involved in another variation site, tracking previous alternatives. An example from the same poem is at v. 8: *A* "il tuo viso apparia, perché dolente" → *B* "al mio sguardo apparia, perché dolente" → *C* "il tuo volto apparia; chè travagliosa". A part of reading *C* is the result of the change from *Ca* ", che" to *Cb* "; chè": *Ca* is thus a sub-reading of one reading only, that is *C*, and it is involved in a variation site with *Cb* [Illustration 9].

Illustration 8: Example of nested variants.

Illustration 9: Example of concatenated variants.

Model outline

The model outlined here allows:
- to distinguish between the features of the reading and those of the variation between the readings;
- to append more than one feature to each reading and variation;
- not to set a base witness to orient the variation;
- to annotate each pair of witnesses or a combination of them for each variation site;
- to nest and concatenate variation sites.

Case studies

The model has been used in the web-application *La* Commedia *di Boccaccio* (Spadini and Tempestini, 2018). Here, other case studies in the form of graphics are presented to test its applicability [Illustrations 10, 11, 12]
In the first three examples, specific categories are employed to annotate common types of morphological variation, in addition to the general categories. The text in the examples is that of an Old-French *pastourelle, "Par un matinet l'autrier"* (Rivière, 1964, III, n° LXXVI);[xvii] the distinction of types of morphological variations is relevant here, because certain types of them recur often, i.e. the alternation between present and past tense, while others are rare. Note that the combination of witnesses changes for each variation site.

Illustration 10: Case study 1.

Illustration 11: Case study 1.

A more complex example [Illustration 13], where three alternative readings are involved, is taken from Giacomo Leopardi's *La ricordanza*, mentioned above. Its manuscripts are conserved at the National Library in Naples[xviii], and an edition of the poem is provided by Italia (2010:68-71).

In the methodological chapter of the same volume [*ibid:* 64], Italia introduces a list of types of interventions occurring in a draft. The list includes: *corretto in* (reading A is corrected into reading B), *soprascritto* (reading B is overwritten on reading A which is crossed-out in the line), *sottoscritto* (reading B is underwritten to reading A which is crossed-out in the line), *inserito* (reading B is inserted), *prima* (reading B is preceded by reading A crossed-out in the line), *dopo* (reading B is followed by reading A crossed-out in line and then abandoned). In the model, it is possible to create a specific category of variation to record this information, here called intervention; in the example [Illustration 13], values for this category are 'overwritten' (as in *soprascritto.*) and 'corrected in' (as in *corretto in*). Furthermore, the relation between the readings has a direction, expressed with an arrow replacing the line. The readings also have a specific category, indicating the writing tool in use for each of them. A comment is attached to the third reading.

Illustration 13: Case study 2.

Logical model

The model can be implemented in different data structures: an OWL ontology and a relational database schema will be presented in this section[xix].
A comparable XML/TEI solution will not be pursued here. This is because overlapping annotations are constituent of the model (e.g., the relation between A vs B and B vs C); therefore, a XML solution would be possible, but requires some workarounds. Nevertheless, a TEI compliant result can be achieved using the Feature Structures module or stand-off mechanisms.

Relational tables

A schema for a relational database, only covering the general features of the reading and of the variation, is presented below [Illustration 14]. Specific categories can be added by means of new tables, connected to the Variation table.

Illustration 14: Relational DB schema representing the model.

OWL ontology

The model can be implemented in the following OWL 2 ontology, formulated in Turtle syntax[xx] and visualized below[xxi] [Illustration 15]. Here too, only the general, and not the specific, features of the reading and of the variation are represented.

The choice of an OWL ontology is dictated by the fact that it is a standard data-model, part of the architectural formalisms of the Semantic Web[xxii]. Note, however, that using a labeled property graph, such as Neo4j, the Variation class would not be needed because the information it carries could be stored as properties of the edge between the Readings.

```
@prefix : <http://example.org/TextualVariation> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix dc: <http://purl.org/dc/elements/1.1/> .


############
##
## ONTOLOGY
##
############


:       a owl:Ontology ;
        dc:title "Ontology for textual variation"@en ;
```

dc:description "This ontology is an implementation of the model for textual variants proposed in this article" ;
dc:modified "2017-01-01"^^<http://www.w3.org/2001/XMLSchema#date> ;
dc:rights "This work is distributed under a Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/)."@en ;
dc:creator "Elena Spadini" ;
owl:versionInfo "1.0" .

\#\#\#\#\#\#\#\#\#\#\#\#
\#\#
\#\# CLASSES
\#\#
\#\#\#\#\#\#\#\#\#\#\#\#

:Reading   a owl:Class ;
rdfs:label "reading"@en ;
rdfs:comment "A reading is a portion of a text and it is composed by a string of characters (one or more letters or words). Competing readings are different readings of a portion of a work; they might belong to the same witness (i.e. alternative readings in the margins) or to different witnesses (i.e. various copies of the same work)."@en .

:Variation   a owl:Class ;
rdfs:label "variation"@en ;
rdfs:comment "The variation expresses what kind of difference exists between the competing readings."@en .

:Witness   a owl:Class ;
rdfs:label "witness"@en ;
rdfs:comment "A witness is a copy of a work. It is a phisical object, containing the text of a work, in manuscript, print or digital form."@en .

:LocationInWitness   a owl:Class ;
rdfs:label "location in witness"@en ;
rdfs:comment "The location in the witness indicates where the reading is located."@en .

\#\#\#\#\#\#\#\#\#\#\#\#\#\#
\#\#
\#\# PROPERTIES
\#\#
\#\#\#\#\#\#\#\#\#\#\#\#\#\#

:hasText   a owl:DatatypeProperty ;
rdfs:label "has text"@en ;

```
                    rdfs:domain :Reading ;
                    rdfs:range rdfs:Literal .

:hasLocationInWitness a owl:ObjectProperty ;
                    rdfs:label "has location in witness"@en ;
                    rdfs:domain :Reading ;
                    rdfs:range :LocationInWitness .

:hasLocationInWork a owl:DatatypeProperty ;
                    rdfs:label "has location in work"@en ;
                    rdfs:domain :Reading ;
                    rdfs:range rdfs:Literal .

:hasWitness    a owl:ObjectProperty ;
               rdfs:label "has witness"@en ;
               rdfs:domain :LocationInWitness ;
               rdfs:range :Witness .

:hasPageNumber          a owl:DatatypeProperty ;
               rdfs:label "has page number" ;
               rdfs:domain :LocationInWitness ;
               rdfs:range rdfs:Literal .

:hasSigil      a owl:DatatypeProperty ;
               rdfs:label "has sigil"@en ;
               rdfs:domain :Witness ;
               rdfs:range rdfs:Literal .

:isInArchive   a owl:DatatypeProperty ;
               rdfs:label "is in archive"@en ;
               rdfs:domain :Witness ;
               rdfs:range rdfs:Literal .

:hasShelfMark        a owl:DatatypeProperty ;
               rdfs:label "has shelfmark"@en ;
               rdfs:domain :Witness ;
               rdfs:range rdfs:Literal .

:consistsOf    a owl:ObjectProperty ;
               rdfs:label "consists of"@en ;
               rdfs:domain :Variation ;
               rdfs:range :Reading .

:hasChangeCategory        a owl:DatatypeProperty ;
               rdfs:label "has category of change"@en ;
               rdfs:domain :Variation ;
               rdfs:range rdfs:Literal .

:hasLinguisticCategory        a owl:DatatypeProperty ;
               rdfs:label "has linguistic category"@en ;
               rdfs:domain :Variation ;
               rdfs:range rdfs:Literal .
```

:isPartOf   a owl:ObjectProperty ;
                 rdfs:label "is part of"@en ;
                 rdfs:domain :Reading ;
                 rdfs:range :Reading .

:isPartOf   a owl:ObjectProperty ;
                 rdfs:label "is part of"@en ;
                 rdfs:domain :Reading ;
                 rdfs:range :Reading .

Illustration 15: Visualization of the OWL 2 ontology representing the model.

Conclusions

This article presents a model for annotating textual variants. Once the annotations are made and conveniently stored, they can be queried, in order to find patterns and analyse the *mouvance* of the work. Possible queries depend on the categories of reading and variation in use. The distinction between features of the readings and features of the variations is fundamental to the organization of the categories. In addition to the general categories (additions, deletions, substitution, transposition; orthography, morphology, syntax, lexis), the annotations might cover: verbal tenses, paleographical variations, errors of different types (*coniunctivus*, *separativus*), dialectal forms, synonyms; over selected sections of the work and selected witnesses or stages. Specific queries can be performed in order to isolate, for studying of removing the noise of, the phenomena covered by the annotations: all the changes of verbal tense in section A, all the deletions between witness/stage A and witness/stage B, all the instant rewriting, etc. The model is flexible, as much as it ensures freedom to the scholar in choosing the categories and setting the boundaries of the readings; the length of the readings, in particular, might vary in the annotations of the same text.

Adopting the model is cumbersome work. On the other hand, it provides detailed and organized information, which is fundamental for certain projects of scholarly editing. Asking precise questions to a machine often requires this kind of thorough work: eventually, we can only ask what we previously gave it.[xxiii] Annotating variations following the model could benefit from a dedicated GUI. In addition, some of the categories might be identified automatically[xxiv].

The implementation in different data structures proves that the relational DB schema and the OWL ontology have the same expressiveness: namely, in articulate relationships. XML, on the contrary, is less suitable for conveying the information gathered using the model, even if XML solutions can

eventually be implemented. This conclusion should be evaluated taking into account that the model covers a textual phenomenon, that of variation; even if, in the model, this phenomenon is detached from the rest of the text, it should be possible to expand the model in order to include the contexts, or, better, the co-texts. Now, in digital scholarly editing the *de-facto* standard data structure for text is XML. This is of course related to the adoption of the TEI Guidelines, but also, more generally, to the fact that digital scholarly editing often results in digital publishing, and the language of the web is XML, in the form of HTML. Comparing relational databases and graphs with XML, we note that from the first is less intuitive to retrieve a stream—which is a fundamental quality for working with texts—, and the second lacks of tools for handling entire texts to be published digitally. In short, they are commonly used for data which are much more structured and fragmented than texts.

Ongoing experiences, however, prove that there is an interest in the digital scholarly editing community to explore solutions other than the tree formalism of XML. In particular, the graph structure is emerging, as a conceptual model to be implemented in different ways[xxv]. The adoption of graphs raises a number of technical and theoretical challenges. Among the technical ones, there might be the need to integrate the information stored in graphs within the XML (or HTML) representation of the text: the discussion on the TEI List about the integration of RDF annotations in a TEI document shows that the discussion is open-ended;[xxvi] stand-off solutions can peer out here, for overcoming the limitation of XML and for filling the gap with other data structures. Among the theoretical challenges, on the other hand, there is the possibility to call into question the way texts are employed and consumed, which is not unrelated to the way they are visualized. This means, for instance, that scholarly editing can produce various outputs: diplomatic or critical texts; but also SVG objects and, more in general, graphics and dynamic visualizations results of analysis, which might represent some of the features of the texts better than typographical devices reproduced by HTML (Andrews and van Zundert 2016; Cummings, Hadley and Noble 2017). The terms visualization and analysis recall that what is represented is data, and not only words or sentences. In this scenario, it is easier to take advantage of data structures such as graphs or relational tables.

The exercise in modelling presented in this article is intended as a minor contribution to the broad discussion briefly addressed here above, but primary as a way to explore how computational methods may contribute to the old issue of handling textual variation. Applying it to other case studies will prove its usefulness and versatility.

References

Andrews, T. L. (2016). Analysis of Variation Significance in Artificial Traditions Using Stemmaweb. *Digital Scholarship in the Humanities*, *31*(3), 523-539. DOI: https://doi.org/10.1093/llc/fqu072

Andrews, T. & van Zundert, J. (2016). Apparatus vs. Graph: New Models and Interfaces for Text. *Interface Critique* (pp. 183-206). (Kaleidogramme; Vol. 139). Berlin: Kulturverlag Kadmos.

Brandoli, Cristina. "Due Canoni a Confronto: I Luoghi Di Barbi E Lo Scrutinio Di Petrocchi." In Paolo Trovato (Ed.), *Nuove Prospettive Sulla Tradizione Della Commedia. Una Guida Filologico Linguistica Al Poema Dantesco*, (99-214). Firenze: Cesati, 2007.

Camps, Jean-Baptiste. (2012) Louis Havet, Cesare Segre, critique verbale et diasystème."Blogpost. *Sacré Gr@@l* (blog). Accessed March 8, 2018: https://graal.hypotheses.org/550.

Ciotti, F., & Tomasi, F. (2016). Formal Ontologies, Linked Data, and TEI Semantics. *Journal of the Text Encoding Initiative*, 9. DOI: https://doi.org/10.4000/jtei.1480

Ciula, A., & Eide, Ø. (2017). Modelling in digital humanities: Signs in context. *Digital Scholarship in the Humanities*, *32*(1), 33-46. DOI: https://doi.org/10.1093/llc/fqw045

Colwell, E. C., & Tune, E. W. (1964). Variant Readings: Classification and Use. *Journal of Biblical Literature*, *83*(3), 253–261. DOI: https://doi.org/10.2307/3264283

Cummings, J., Hadley, M., & Noble, H. (2017). It has moving parts! Interactive visualisations in digital publications. Presented at the DiXiT Workshop The Educational and Social Impact of Digital Scholarly Editions, Cologne, Germany. Retrieved from http://dixit.uni-koeln.de/programme/materials/#aiucd2017

Eide, Ø. (2014). Ontologies, Data Modelling, and TEI. *Journal of the Text Encoding Initiative*, 8. Retrieved from https://jtei.revues.org/1191

Haentjens Dekker, R., & Birnbaum, D. J. (2017). It's more than just overlap: Text As Graph. In *Proceedings of Balisage: The Markup Conference 2017*. DOI: https://doi.org/10.4242/balisagevol19.dekker01

Italia, P. (2010). *Che cosa è la filologia d'autore*. Roma: Carocci.

Italia, P., Vitali, F., & Di Iorio, A. (2015). Variants and Versioning Between Textual Bibliography and Computer Science. In *Proceedings of the Third AIUCD Annual Conference on Humanities and Their Methods in the Digital Ecosystem* (2:1–2:5). New York, NY, USA: ACM. https://doi.org/10.1145/2802612.2802614

McCarty, W. (2004). Modeling: A Study in Words and Meanings. In S. Schreibman, R. Siemens, & Unsworth, J. (Eds.), *A Companion to Digital Humanities*. Oxford: Blackwell. Retrieved from: http://www.digitalhumanities.org/companion/.

Micha, A. (1978-1983). *Lancelot: roman en prose du XIIIe siècle.* Genève: Droz.

Pierazzo, E. (2015). *Digital scholarly editing: theories, models and methods*. Ashgate. Basingstoke.

O'Reilly, M., Van Hulle, D., Verhulst, P. & Neyt, V. (2016). *Samuel Beckett Digital Manuscript Project*. Retrieved from http://www.beckettarchive.org.

Rivière, J.C. (1974). *Pastourelles.* Genève: Droz.

Schauweker, Y. (2013). Variantes « significatives » et variantes « récurrentes ». Repenser l'appareil critique. In *Actes du XXVIIe Congrès international de linguistique et de philologie romanes*. Nancy, 15-20, July 2013. ATILF.

Spadini, E. (2017). The role of the base manuscript in the collation of medieval texts. In Boot, P., Cappellotto, A., Dillen, W., Fischer, F., Kelly, A., Mertgnes A., Sichani, A. M., Spadini, E. & Van Hulle, D. (Eds.), *Advances in Digital Scholarly Editing. Papers presented at the DiXiT Conferences in The Hague, Cologne, and Antwerp* (pp. 345–350). Leiden: Sidestone Press.

Spadini, E. & Tempestini, S. (2018). *La Commedia di Boccaccio. Un apparato in movimento*. Retrieved from: http://boccacciocommedia.it.

Stussi, A. (2011). *Introduzione agli studi di filologia italiana*. Bologna: Il Mulino.

Tomasi, F., Daquino, M., & Giovannetti, F. (2018). Linked data ed edizioni scientifiche digitali. Esperimenti di trasformazione di un Quaderno di appunti. Presented at the 7th AIUCD Conference. Cultural Heritage in the Digital Age, Bari, Italy. Retrieved from http://www.aiucd2018.uniba.it

Unsworth, J. (2002). What is Humanities Computing and What is not? *Jahrbuch Für Computerphilologie*, 4, 71-84.

Vanhoutte, E. (2007). Traditional editorial standards and the digital edition. In Stronks, E. & Boot, P. (Eds.), *Learned Love* (pp. 157-174). The Hague: DANS.

i   The example is taken from Gustave Roud's *œuvre*: his writing is rooted in diary's notes taken during ramblings in the Vaud region; the notes are elaborated for articles published in literary magazines and then assembled in collections of short pieces. A project of edition of the complete works of Gustave Roud is ongoing at the Université de Lausanne, under the direction of Daniel Maggetti: Gustave Roud, *« Œuvres completes »* <http://unil.ch/crlr/home/menuinst/projets-de-recherche/gustave-roud-oeuvres-completes.html>.

ii   It happens, for instance, for every literary work whose textual transmission spans various centuries.

iii   While Zumthor's term *mouvance* is related to anonymity and textual variations in medieval manuscripts, his definition of 'moving work' might be valid also for modern literature: 'l'unité complexe, mais aisément reconaisable, que constitue la collectivité des versions en manifestant la matérialité […]. L'oeuvre est fondamentalement mouvante' (Zumthor 1972: 73).

iv   The literature on the topic is vast and specific to literary periods and languages; most of the analysis are disseminated in editions and studies of specific authors or works. Some inspiring contributions are Colwell and Tune (1964), Brandoli (2007), Camps, (2012), Schauweker, (2013), Italia *et al.,* (2015), Andrews (2016).

v   *Variante d'écriture* (Grésillon 1994: 246); *varianti immediate* (Italia and Raboni 2010: 54). The definitions are gathered under the entry 'Instant rewriting' in *Lexicon of Scholarly Editing* <http://uahost.uantwerpen.be/lse/index.php/lexicon/instant-rewriting/>.

vi   This example springs again from the analysis of Roud's papers. A first examination of the drafts connected to *Petit traité de la marche en plaine* (Mermod 1932) suggests that proper nouns are replaced by generic characters.

vii   For what concerns Textual Criticism, particular attention is devoted to modelling in Unsworth (2002) and Pierazzo (2015).

viii   The aim here is the creation of a 'model for production' (Eide, 2014:15), and the model in use is a 'metaphor-like model' (Ciula and Eide, 2017).

ix   The model, highly interpretative, can be used with profit together with facsimiles of the images, more and more common in the digital panorama, or might be expanded to take into account the context (or, better, the co-text) of each reading. See Buzzetti (2002, 62): 'the diacritical signs or the forms of markup are no longer conceived as an aid in visibly reconstructing an absent document, but rather as a means of "modelling" the physical and textual information contained in the original for the purpose of further processing', and '[a]n adequate digital text representation must therefore be compatible with the application of the formal procedures of information processing which give algorithmic form to current methods and practices of textual criticism and interpretation.'.

x   (Rivière 1974), vol. III, pièce n° LXXVI.

xi   Formalization of how to point to the location of a reading in the physical object and in the literary work is beyond the scope of this contribution.

xii   The edit distance between two strings is based on the number of operations required to transform the first string into the second one. The edit distance calculated using all four operations is the Damerau-Levenshtein distance.

xiii   *Molloy* module, <http://www.beckettarchive.org/molloy/collatex/1606?lang=EN>.

xiv   *Lancelot*, in four manuscripts of the Bibliothèque nationale de France. Cf. (Micha, 1978-83, III:, § XXVI).

xv   See (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'. A location variant corresponds to a reading. In line with Vanhoutte study, the model analyses the variation in pairs of readings. This is not only the most consistent way to do it, but also the most thorough, because most of the time it would not be possible to summarize in one single annotation all the differences between all the readings.

xvi	It should also be remembered that the model proposes one precise interpretation of the phenomenon at stake; a different interpretation would lead to a different model. Thus the model might not be suitable for all editorial projects.

xvii	The critical text of Rivière's edition is: 'Par un matiner l'autrier | oï chanter un fou berchier ; | s'en sui esmeü , | qu'il se vantoit qu'il ot geü | tout nu | entre les deux bras s'amie. | Il se vantoit de folie, | car cele amour est vilaine, | més j'aim certes plus loiaument que nus ; | puis que bele dame m'aime | je ne demant plus.' The text is present in four manuscripts, indicated here with the corresponding sigils.

xviii	Digital facsimiles are available on the library website at <http://digitale.bnnonline.it/index.php?it/119/giacomo-leopardi-canti>.

xix	Some details of the schema and the ontology are omitted, such as data-types and cardinality.

xx	To enhance readability, subjects are in bold and predicates are underlined.

xxi	The visualization is obtained with WebVOWL 1.0.6, available at <http://visualdataweb.de/webvowl/>.

xxii	The mapping to Vocabularies used for Linked Open Data is beyond the scope of this article; for the Witness class, the FRBF model and FABIO, its OWL formalization, should be considered. See FRBR-Aligned Bibliographic Ontology (FABiO), <http://www.sparontologies.net/ontologies/fabio>. In (Flanders and Jannidis 2015: 9-10) ontologies "are restricted to the conceptual model"; it is important to distinguish between the conceptual ontology and its logical implementation in an OWL Ontology, in order to understand why RDF Schema is considered a logical model in the same article (*ibid* 11).

xxiii	Except for unsupervised machine learning.

xxiv	It is the case, at least, for additions and deletions, and for linguistic categories using NLP tools.

xxv	The graph structure is prominent in research connected to modelling text (Haentjens, Dekker & Birnbaum 2017), semantic editions (Eide 2014), (Ciotti and Tomasi 2016), (Tomasi, Daquino & Giovannetti 2018), software framework infrastructures based on graph solutions, such as Knora <http://www.knora.org/> and Alexandria Markup Text Repository (Haentjens, Dekker & Birnbaum 2017).

xxvi	The first mention of RDF in the TEI-List goes back to 1999, see <https://listserv.brown.edu/archives/cgi-bin/wa?A0=TEI-L>.