Towards a digital model to edit the different paratextuality levels within a textual tradition

Paolo Monella,
University of Palermo
paolo.monella@gmx.net

Peer-Reviewed Article

Dedication

This article is published as part of *"Though much is taken, much abides": Recovering antiquity through innovative digital methodologies*, a special collaboration between Digital Classicist and the Digital Medievalist Journal presented in honor of Ross Scaife (1960-2008)

Abstract

In the textual tradition of a literary work, our sources (manuscripts, printed books etc.) commonly bear, together with the "main text", different kinds of "paratexts" commenting on it (including interlinear annotations, glosses, scholia, footnotes, modern scholarly introductions and commentaries, and many others). This article proposes a unified model for a document-based digital critical edition including both the main texts and the paratexts as they appear in different single sources. The problematic aspects of such an "enlarged" digital edition are discussed, including the relations between the different paratexts and the main text they refer to within each single textual source, as well as the "alignment" of different main texts and paratexts in different sources.

Keywords: Classics; electronic publication; electronic editing; scholia; gloss; paratext.

Contents

The critical edition as a representation of the textual variance

§ 1 Critical editions, i.e. editions of texts with a text-critical apparatus, respond to the necessity of representing one aspect of the complex reality of textual tradition: the textual variance. Their function is double: on the one hand, they present the different versions of a text within the context of the textual tradition; on the other hand, they try to 'extract', out of the different texts born by many carriers (manuscripts, incunabula, modern and contemporary print editions), a reconstructed Text, the closest possible to the 'original' one prior to its 'corruption' due to the very process of textual tradition, thus ideally recovering the *intentio auctoris*[1].

§ 2   Adopting the traditional opposition of 'document vs. text', we could say that the ideal movement of a critical edition is from the physical documents (the sources) to the abstract text[2].

§ 3   Within each document, the philologist selects the text to be included in his edition, and ignores whatever other texts may be carried by that source – such as different levels of glosses to the text, as well as other works by the same author, or belonging to the same literary genre, that happen to have been copied into the same codex, or published into the same print edition. The result of his selection is what we call, e.g., "codex A's text of Ovid's *Ars amatoria*". This, collated with the corresponding 'texts' of other sources, is the groundwork upon which the work of the editor is based[3].

§ 4   To sum up the process, we could say that a first selection takes place when the philologist extracts the 'texts' out of the documents. This selection happens while the philologist transcribes the textual variants carried by the primary sources but, just like the whole transcription process, it is rarely described[4]. This initial selection provides the groundwork for the second 'critical' phase of his activity, where he 'distills', from the many texts of the sources, the 'reconstructed' Text.

§ 5   The digital critical edition approximately aims to do the same work, but taking advantage of the flexibility offered by technology. In particular, it promises to accomplish better the goal of presenting in detail the textual variance (down to the detail level of the very sources of the text, both transcribed and reproduced with digital images). This provides two main advantages:

1. a digital critical edition allows the reader to verify and call into question the work of the editor
2. it builds up an 'open' model of the text, not implying that the text created by the editor is the text.

Maintext and paratexts

§ 6 However, in many of the primary sources that constitute the tradition of a text we can find around the 'main text' – so to speak – other texts, that point in different ways to it – e.g., a series of glosses or *scholia* written physically around a classical text in a manuscript, or a series of footnotes in a modern edition of the same text, commenting on the 'main text'. We shall name those many sorts of comments paratexts, as opposed to what we will call maintext, since the former comment on the latter[5].

§ 7   The overall content of a document may display a differentiated, and problematic, range of 'levels of paratextuality'. A complex document such as a modern print edition of a classical Latin work – not even a critical edition, but just a good 'commercial' paperback including text, translation and some notes – might provide, at the beginning of the book, together with the title and author of the work, a plethora of typographical and bibliographical information on the book (the document) itself, including the modern editor(s), the translator, the publication place, etc.; then, a preface by a well known scholar, commenting on the poetics of the author, its time, the sense of his whole literary production; then again, an introduction by the editor of the edition, commenting more closely on the work published. All sorts of biographical, metrical, textual-critical prefatory notes could follow, and our maintext would not yet have begun. Then, we can imagine the Latin maintext on the left pages, and its translation on the facing page. One also expects to find, in a really good paperback edition, at least some textual-critical notes below the Latin maintext informing the reader on the most meaningful variants, and some erudite explanatory footnotes below the modern translation, to assist the reader with the comprehension of the text. Those commentary footnotes, of course, could also be placed at the end of the maintext as endnotes, without changing the meaning of their relation to the maintext. After the whole text and commentary, a number of indices (*rerum*, *nominum*, *locorum*, etc.) are normally more than welcomed by the reader.

§ 8   The typical situation of a worthy manuscript is simpler in some ways, but not as much as one might think: other than the marginal commentary glosses, many codices present supralinear insertions, ranging from the typology of the *varia lectio* or textual correction[6], to glosses on difficult words, or commentaries on single *realien* (places, mythological characters etc.), or on the language or the style of the passage. Manuscripts often include as well a variety of notes (by different hands, composed over a period of time) that can appear in the margins of a handwritten page.

§ 9   Such a rich 'ground cover' of secondary texts (paratexts), 'growing' upon and around a primary text (our maintext), is an interesting textual phenomenon belonging to the complex reality of a literary tradition, and surely one that deserves to be represented.

§ 10   Among those texts, in particular the ancient *scholia* have gained attention in the philological tradition: we can think of the editions of whole corpora of *scholia*, in which the commentary notes found in the different manuscripts of a classical work are gathered, labeled with codes referring to the codex where the single *scholion* is found, and published separated from the text they comment – with pointers making clear which location in the maintext is being commented by each note, that is by each portion of that specific paratext.

The paratextuality levels

§ 11 My proposal is to elaborate a digital model to 'edit' (that is to give a representation of) the complex phenomenon of the relations between a maintext and each paratext commenting on it (or just 'pointing' to it) in the handwritten and print tradition of that text.

§ 12   Such a model should include both the maintexts and the paratexts of each source, expressing explicitly the relation between single portions of each paratext and the precise portions of maintext they refer to. This implies that, rather than a traditional edition of *scholia*, it would be both an edition of the text and of its ancient (and modern) commentaries – and the relationships between the text and its commentaries.

§ 13   Let us imagine now, within the textual tradition of a literary work, four sources: a (a manuscript); b (another manuscript); c (a modern print critical edition); and d (a modern commercial – not critical – print edition). If we agree to call maintext(a) the maintext of the source a, paratext(a)1 one of the paratexts attached to the maintext in the source a and so on, a plausible list of texts involved in the construction of our edition would include:[7]

- maintext(a)
- paratext(a)1 {*rubricae*}
- paratext(a)2 {explanatory glosses}

- maintext(b)
- paratext(b)1 {glosses}
- maintext(c)
- paratext(c)1 {philological introduction}
- paratext(c)2 {scholarly commentary to the text}
- maintext(d)
- paratext(d)1 {frontispiece etc.}
- paratext(d)2 {preface}
- paratext(d)3 {prefatory essay}
- paratext(d)4 {introduction}
- paratext(d)5 {commentary footnotes}

§ 14   What we mean to do is to instruct the computer to read the TEI-compliant transcriptions of the primary sources (transcription(a), transcription(b) etc.),[8] parse the TEI-XML markup and assign a 'paratextuality level' to each textual portion of the transcription. For example, when the computer reads the transcription of a source a and parses a code such as the following[9]:

```
<l n="1.200" xml:id="1.190">Theseus <note place
= "supralinear" type = "explanatory glossa">rex
Atheniensium</note> […] rapuit </l>
```

§ 15   it understands that "Theseus […] rapuit" belongs to the paratextuality level maintext, and is therefore a part of the object maintext(a), and that "rex Atheniensium" belongs to the paratextuality level 'explanatory glosses', and therefore to the object paratext(a)1,[10] pointing to the portion of maintext(a) whose id is "1.190".[11]

§ 16   The computation on the TEI-XML markup of transcriptions/descriptions of primary sources to deduce extensive information about the paratextuality levels may appear 'smooth', as far as we confine ourselves to simple examples such as the one above.

§ 17   But however refined the software we create might be, on many occasions it will find itself at pains to 'translate' the transcription markup into information about the paratextuality level, simply because the task of bearing information about our defined 'paratextuality levels' is not the purpose for which a standard TEI-compliant transcription of a primary source is written, and – I would add – for which the whole TEI-XML transcription markup was developed.[12]

Transcriptions of primary sources and paratextuality levels

§ 18 From a practical point of view, we can ask ourselves now whether any TEI-compliant transcription already available for the sources that constitute the textual tradition of a text would be suitable for the edition, or if we shall finally find ourselves compelled to create and use exclusively our own "project-oriented" transcriptions.

§ 19   The problem is that, in the latter case, we would create markup strongly oriented towards the needs of a specific research project, and such a practice would break a principle which, in my opinion, should inform any project elaborated in the Digital Humanities, especially in this still 'pioneering' stage of its development: any project should tend to the highest degree of standardization possible. The input-data themselves (in our case, the transcriptions) should be based on existing standards in order to allow the project to build on the work of other researchers, and to ensure that the output may be re-used by other projects. When there are relatively few people working in a research field, and when the paths followed by those different researchers diverge, standardization becomes a critical issue. By using standard technology, researchers ensure that their work will not become incomprehensible to others.

§ 20   But such practical considerations (about the possible need of 'purpose-created' transcriptions) invite us to raise a more theoretical issue – though bound to very practical ones – concerning the amount and nature of the information to be encoded in the transcriptions, and the responsibility of the transcriber and the editor to make their text-critical decisions,[13] as opposed to the liberty (and the specular responsibility) of the user[14] to apply his judgment on the choices of both the transcriber and the editor, and to actually make his own decisions on the text. I am thankful to Prof. Willard McCarty, who drew my attention on those issues during a discussion about the ideas presented in this article in the summer of 2006 at the Centre for Computing in the Humanities at King's College London.

§ 21   We can imagine a situation where the choice about the paratextuality level assigned to a portion of text has been made by the transcriber (the creator of a source's transcription), and explicitly 'encoded' into the transcription itself. An XML code like the following (an example we've already seen above):

```
<l n="1.200" xml:id="1.190">Theseus <note place
= "supralinear" type = "explanatory glossa">rex
Atheniensium</note> […] rapuit</l>
```

§ 22   bears unambiguous information about the assignment of a paratextuality level ('explanatory gloss') to "rex Atheniensium".

§ 23   As an example of the editor's choice (that is of the person

who cures the actual digital edition), we can imagine the case where an editor has decided to instruct the software to assign by default a certain paratextuality level to a certain markup pattern, for example deciding that all text encoded as follows (please note the lack of the 'interpretive' attribute @type):

```
<l n="1.200" xml:id="1.190">Theseus <note place
= "supralinear">rex Atheniensium</note> […]
rapuit</l>
```

§ 24   belongs to the paratextuality level 'explanatory gloss', and therefore, since it is carried by source a, to the object paratext(a)1.

§ 25   The point is that such decisions about paratextuality level assignments (both those encoded in the transcription by the transcriber and those made by the editor when creating or configuring the software) may often be problematic and questionable, due to the 'blurring' of the paratextuality levels into each other.

§ 26   In the last example above, for instance, the user could re-examine the textual situation (i.e. both the text and the assumed gloss), and also the digital image of the primary source, if available, and finally argue that the words written over the line ("rex Athenienium") constitute not a gloss, but part of the maintext.

§ 27   In a situation like the one we outlined above (i.e.: 'paratextuality level-neuter' transcription markup, paratextuality level-switches due to a different judgment on the paratextuality level a piece of text belongs to), Willard McCarty thinks that the software managing the digital edition should allow the user to change the paratextuality level-assignment to that portion of text.[15] In our example this means that the user of the digital edition should be able to change the paratextuality level from the 'default' paratext(a)1 (the editor's choice) to maintext(a) (the user's own choice).

§ 28   Naturally, such a 'flexible' software could allow also the editor, during the construction of the digital edition, to assign the paratextuality level-assignment case by case. Willard McCarty's suggestion is therefore to create (or re-use), 'paratextuality level-neuter' transcriptions, so to speak, which should not bear any explicit information about the assignment of paratextuality levels, and to transfer as many of the interpretative choices as possible to the software level, for the very simple reason that it would be difficult for the user to change the transcriptions, which are the 'input-data' of the system, whereas the software we create can be made flexible enough to allow for 'paratextuality level-shifts'.

§ 29   In general terms, I agree with this point of view, particularly when it comes to the need of giving the user ways to 'modify' the

edition itself in case of diverging opinions about certain editorial choices. In view of the realization of our project, though, doubt remains whether the encoder can create (and the software can work on) a transcription markup completely free of interpretive information about the paratextuality levels.

§ 30   But not only a single portion of text can have an ambiguous status: from a wider point of view, the interpretation of a whole paratextuality level (belonging to one of the many possible paratext categories) in terms of its relation with the maintext could be problematic. For example: manuscript c containing a collection of tales, in which each tale is preceded by a short summary of the story (in the transcription, something like <div type="summary">). A specific paratextuality level, called paratext(c)4, could be created to include all summaries within source c. But this is one of the cases where the user might want to make decisions about the 'role' of the paratextuality level. He could choose to 'include' those summaries in the text, and therefore have it displayed on screen, have text analysis software search through it together with the maintext etc. Or he might choose to 'separate' the paratext(c)4 from the maintext completely.

§ 31   This takes us to a model where the single paratextuality levels, in their turn, can be grouped into what we could call 'families of paratextuality levels'. An example, relative to a source d (a print edition) could be the following:

- Family "A" {the 'core': the text to be read sequentially}, including:

  - maintext(d)

  - paratext(d)1 {paragraph numbers}

  - paratext(d)2 {titles of paragraphs and chapters}

- Family "B" {in-line material not belonging to the 'core'}, including:

  - paratext(d)3 {'in-line' rubrics}

  - paratext(d)4 {'in-line' summaries}

- Family "C" {commentary}, including:

  - paratext(d)5 {footnotes}

  - paratext(d)6 {endnotes, printed after the maintext}

- Family "D" {prefatory material}, including:

  - paratext(d)7 {introduction}

  - paratext(d)8 {prefatory essay}

- paratext(d)8 {prefatory essay}
- Family "E" {bibliographical coordinates of the volume}, including:

  - paratext(d)9 {frontispiece}

  - paratext(d)10 {copyright information and warnings}

- The above list of paratextuality levels runs through an ideal range of 'paratextuality', from those texts ideally 'closer' to the maintext, up to paratext(d)10, which can be said to 'comment on' the maintext only in a very loose and general sense.[16] The grouping of the paratextuality levels into such families is surely one thing that the software should leave to the user's choice.

The general structure of the model, and the Alignment-Text

A document-oriented structure

§ 32 Let us delineate now the relational structure of our model.

§ 33 The whole project originates from a specific attention to the 'document' (i.e. the primary source for both the maintext and the paratext[17], so it is quite obvious that at the 'center' of the model itself we cannot put an abstract 'Text', a reconstructed text resulting from the philological work of an authoritative scholar (that is, predictably, straight from the pages of the most important critical edition of the work), like in the following structure:

Figure 1: This is a structure based on an abstract 'Text', which we don't want
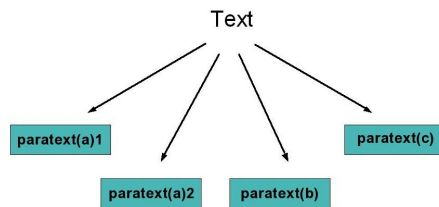
§ 34 On the contrary, keeping in mind the quite obvious consideration that any commentary, though aiming to be a commentary on the Text, is always necessarily a commentary on one text[18], we could imagine for our model a structure in which each paratext(x) is directly connected to its own maintext(x), that is to the maintext of the source x that bears both. The resulting structure would look like this:

Figure 2: This is the 'source/document-oriented' structure that we do want
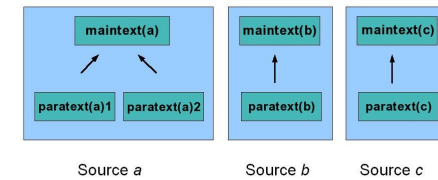
The alignment

§ 35 However, in the resulting model the 'alignment' among the paratexts (and the maintexts) carried by different sources becomes an issue: we need to put, within each transcription, some 'milestones' to create the cross-references between corresponding portions of different maintexts, and between those portions and the parts of the different paratexts commenting on them[19].

§ 36 An issue like the alignment of the different versions of the text and between text and *scholia* (or modern commentary notes) is hardly taken into account, either in the traditional work on print critical editions, nor in editing a scholastic tradition, for a number of reasons:

1. in 'classical', well attested literary texts, the discrepancies between the texts of the different sources (including *lacunae*, verse order alterations etc.) are normally too slight to constitute a serious problem for the traditional alignment practices in non-electronic editions;

2. in poetic texts, the progressive numbering of verses provides a good, almost 'natural' means to 'partition' the text[20];

3. for prose texts, the principle of the "authoritative edition" (see, e.g., Plato or Aristotle's editions), that is partitioning the text after the page and row numberings of a well-known past edition, is considered efficient enough for the formalization standards required by non-electronic processing of texts;

4. this takes us to the fundamental point: whenever a text is edited in order to be read and analysed by the reader, the partitioning strategies take into account the obvious fact that ultimately it will be the reader himself, with the help of synoptic tables of concordances[21], who will 'align' quite easily different text versions and commentaries with each other.

§ 37   But since the intelligence of a human being is (in most cases today) much more flexible than that of a computer, when we create electronic texts – meant to be processed automatically by computers – the text segmentation system must be formalized in a much more rigorous way. The easiest way is inserting such 'milestones' in the tagging of each transcription (both of the maintexts and of the paratexts), but even then a number of issues arise.

§ 38   Confining ourselves to verse texts, for which partitioning and numbering problems are reasonably easy to approach, we must consider that many phenomena quite common in the manuscripts – but also present in print editions – can alter in a single source the numbering of verses (and poems). These phenomena include interpolations (portions of text born by a source, but supposedly absent in the 'original'); entire verses missing, intentionally expunged, or accidentally lost in the process of transcription; transposition of verses[22]; discrepancies in the separations between poems (very frequent, e.g., in the second book of Propertius' elegies); and even concurrent divisions between poetic books (e.g. the third/fourth book of the *Corpus Tibullianum*).

§ 39   In all those cases, the scholars normally anchor the numbering to an abstract model of the text. For example, in the case of a generalized *lacuna*, a verse or group of verses lacking in all our sources, which for some reason must have existed in the 'original' text, a progressive verse number is often assigned even to the 'phantasm verses' that exist only in our abstract (not connected with any source) reconstruction of the text. We could cite the following passage, from the tenth poem of the first book of Tibullus' elegies:

At nobis aerata, Lares, depellite tela,	25	
· ·		(26)
· ·		(27)
Hostiaque e plena rustica porcus hara.	26	(28)
Hanc pura cum veste sequar myrtoque canistra	27	(29)
Vincta geram, myrto vinctus et ipse caput.	28	(30)

§ 40   After verse 25, a *lacuna* of at least two verses is almost certain, for quite obvious reasons of sense. Aside the lines, I copied the two concurrent numberings: the one including, and the other not including the two lost verses (but were they two, or more?).

§ 41   In the opposite case, represented by interpolations, a portion of text actually born by at least one source is commonly not assigned any numbering, because it is commonly judged by the scholars as extrinsic to the 'original' (i.e. reconstructed) text.

The Alignment-Text

§ 42 Keeping in mind that the transcriptions are the groundwork of our project, and that we need to partition and 'label' the text as it appears in each single source, the issue of aligning texts between sources becomes central.

§ 43   If we confine ourselves to creating a different, 'idiosyncratic' numbering system for every single source, only taking into account its actual, peculiar sequence of verses (i.e. all and only the verses which that source contains), we would face a situation where, for instance, line 200 of manuscript a (full of interpolations) could well correspond to line 190 of manuscript b, because b does not include a's interpolations (though it may include others), and has a long *lacuna* at a certain point of the text before that verse. The computer would have no way to know that line 200 in manuscript a corresponds to (i.e. is a different version of) line 190 in b, and therefore to understand that the gloss of paratext(b)1 commenting on line 190 of maintext(b) can also be seen, in a broader sense, as a comment on line 200 of maintext(a).

§ 44   The only way to overcome this problem seems to be the one that Prof. Tito Orlandi suggested to me during an interesting discussion in Rome: we need to develop a unifying numbering system (not mirroring any specific source), an abstract 'Alignment-Text', which we can imagine as a 'pure structure', a 'blank' sequence of place holders (in TEI-XML terms, a simple sequence of blank elements marked by non-ambiguous @xml:id attributes), each identifying unambiguously a textual *locus* (i.e. a part of the maintext), attested at least in one source (but obviously recurring, in most cases, in all sources). Consequently, even the most evident interpolation in the most complex manuscripts, or print editions, would be included, but the 'phantasm verses' like those whose existence we reconstructed for Tibullus' elegy above shouldn't.

§ 45   In the case of a verse text, in each single TEI-XML transcription a verse (an <l> element), or any other portion of the maintext, could be identified by both an @n attribute, referring to whatever numbering system we want[23], and by an @xml:id attribute, unambiguous within each transcription, i.e. not repeated within that file. Therefore if each transcription is an XML file, two elements will never have the same @xml:id attribute. In the example above, therefore, line 200 in the transcription of manuscript a could be tagged as follows:

```
<l n = "1.200" xml:id="1.190">
```

§ 46   and the corresponding line 190 in the transcription of manuscript b:

```
<l n = "1.190" xml:id="1.190">
```

§ 47   What the Alignment-Text file (A-Txt.xml) should actually look like is not a point that I will discuss in any detail in this paper. In any case, I lean towards using another XML file, containing a simple sequence of void elements to provide the software with a 'map' of all the possible maintext portions to be found in at least one carrier, and on their sequence. A chunk of this file could look like this:

```
<l xml:id="1.187"></l>

<l xml:id="1.188"></l>

<l xml:id="1.188a"></l>

<l xml:id="1.188b"></l>

<l xml:id="1.189"></l>

<l xml:id="1.190"></l>
```

§ 48   According to our conventions, this XML code tells us that there is one (or more) source(s) having, after line "1.188", two lines absent in other sources ("1.188a" and "1.188b"). But from the point of view of the computer, the code just says that a line with id "1.188a" exists in some manuscript, and that in the 'abstract' structure of our maintext (i.e. in our Alignment-Text) it comes after <l xml:id="1.188"></l> and before <l xml:id="1.188b"></l>.

§ 49   We could represent the new structure of our model with the following scheme:

Figure 3: Alignment-Text
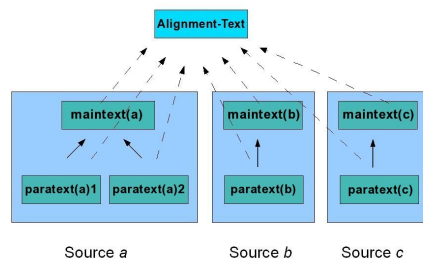
Linking strategies

§ 50 As to the paratext/maintext alignment, a first strategy to formalize the relation between a gloss and the portion of the maintext it refers to has been already discussed above: if, in the

transcription of manuscript a, the element <note place="supralinear" type="explanatory glossa"> is a child of an element <l n="1.190" xml:id="1.190">, the system will easily deduce that the former comments on the latter.

§ 51   The software could be instructed to store the information about this link in an XML file (external to the transcriptions) called, let us say, paratext(a)1.xml, through the use of one of the TEI XPointer Schemes, such as XPath1(), as recommended by the TEI P5 guidelines[24].

§ 52   Let us imagine that the transcription of manuscript a (file: transcription_a.xml) includes the following portion of code (that we already know pretty well):

```
<l n="1.200" xml:id="1.190">Theseus <note place
= "supralinear" type = "explanatory glossa">rex
Atheniensium</note> […] rapuit</l>

<l n="1.201" xml:id="1.191">Aegaeis<note place =
"supralinear" type = "explanatory
glossa">Aegaeus est pater Thesei</note>
aquis</l>
```

§ 53   And the Alignment-Text file A-Txt.xml includes a row reporting the existence of this verse (in the maintext of at least one witness) as follows:

```
<l xml:id="1.190"></l>
```

§ 54   The software should generate the following code into the paratext(a)1.xml file (corresponding to the 'explanatory glosses'paratextuality level within source a):

```
<link evaluate="all" targets="A-
Txt.xml#xpath1(//l[@id='1.190'])
transcription_a.xml#xpath1(//l[@id='1.190']/note[1])"/>
```

§ 55   In the preceding case, we can expect a software to 'understand' that any <note> element child of a <l> element comments on it, and create automatically the appropriate code in the paratext(a)1 file.[25] In many cases, though, the portion of text a note refers to must be encoded explicitly by the transcriber of the primary source. In particular, this will be necessary every time that a note whatsoever (summary, marginal annotation, footnote, prefatory essay, etc.) comments on wider portions of text.

§ 56   For instance, when a footnote of a modern print edition (that we will call source c) comments on a whole poem (and not only on a verse of it), or – even better – when it comments to a portion of the

poem (e.g. its introductory section, from line 1 to 3), the transcriber needs to use the XML linking markup to create 'by hand' an explicit link connecting the <note> element with all the elements it comments on (that is with all <l> elements whose @xml:ids span from "3.1" to "3.3"). Differently from the preceding example, the following link is supposed to be inserted by the transcriber into the transcription file. In this specific case, the simplest solution, according to the TEI P5 guidelines, would be the use of the @target and @targetEnd attributes in the <note> element. The following XML code could therefore be included in the transcription file for source c (transcription_c.xml):[26]

```
<l n="1" xml:id="3.1"> [...] </l>

<l n="2" xml:id="3.2"> [...] </l>

<l n="3" xml:id="3.3"> [...] </l>

[...]

<l n="12" xml:id="3.12"> [...] </l>

</div>

<note type="footnote" xml:id="fnote_3.1-3.3"
target="#3.1" targetEnd "#3.12"> [...] </note>
```

§ 57   When the software parses the transcription file of this source (transcription_c.xml) to store the maintext/paratext linking information into the paratextuality level files, it should transform the preceding code to generate (and write to the appropriate paratext file, let us say paratext(c)2.xml) the following rows, including an intermediate pointer <ptr>, with an @xml:id attribute (in this case "ATxt_3.1_3.2_3.3") automatically generated by the system:

```
<ptr xml:id="ATxt_3.1_3.2_3.3" targets="A-
Txt.xml#xpath1(//l[@id='3.1']) A-
Txt.xml#xpath1(//l[@id='3.2']) A-
Txt.xml#xpath1(//l[@id='3.3'])">

<link evaluate="all" targets="#ATxt_3.1_3.2_3.3
transcription_c.xml#xpath1(//note[@id='fnote_3.1-
3.3']")
```

Conclusion

§ 58 To sum up, the process of editing we have been outlining should include the following phases:

1. The transcriber creates the transcriptions of the primary sources

1. either confining himself to encode information neutral with regards to the paratextuality levels (not adding to elements such as <note> any @type attribute directly pointing to a precise paratextuality level)

2. or appending to any element of the like an 'interpretive' @type attribute[27]

2. The editor, working interactively with a specific software:

1. assigns a paratextuality level to any pertinent portion of the transcription[28]

2. generates the Alignment-Text by gathering all the xml:ids of the maintext in the transcriptions

3. stores the linking information – necessary for the alignment between the maintext Alignment-Text and the different paratexts – in the appropriate 'paratextuality level-files' (like maintext(a), paratext(b)1, paratext(c)2 etc.)[29]

3. From this point on – when the objects constituting the structure of our edition (the Alignment-Text and the 'paratextuality level-files') have been generated – the work on the maintext-files is perfectly analogous to what we would make to use the transcriptions of the primary sources in order to build a digital critical edition. As to the paratext s, the next phase is creating a software (or different modules of the same software) which, working on the objects mentioned above – and in particular on the linking information stored in the paratextuality level-files referring to different paratexts – performs at least presentational solutions to offer the user

1. dynamic and customizable access to both the literary work (the maintext) and the various forms of commentary grown around it within its textual tradition (the paratexts), and

2. flexible procedures to change the editor's choices in the ways discussed above, thus making our 'extended' digital edition dynamic and interactive enough to realize the main task of a digital scholarly edition: allowing the user, as I said above, to verify and call into question the editor's work, end eventually to intervene actively in the editorial process.

Notes:

1. Mordenti 2001, pp. 47-52 has an interesting discussion on the many different functions of a 'traditional' critical edition.

2. On this dichotomy, and on the digital models of text and of the

2. On this dichotomy, and on the digital models of text and of the document, see Ciotti 1994, pp. 220-224.

3. My personal research background relates principally to classical Latin literature, and to Classics in general. Not only this will affect the choice of examples throughout this article, but, as the reader will easily note, the bulk of my reflection originates from the specific task of editing classical literary texts that tend to have a long and complex tradition, in which the numerous (handwritten and printed) testimonies of the literary text are often accompanied by a complex corpus of different kinds of *glossae* and commentaries. This does not mean, of course, that I don't envisage a possible further development of the model I am proposing, to fit the specific issues associated with the editing of other textual forms.

4. The transcription-encoding process has been at the center of the theoretical reflection on the digital critical edition: see Mordenti 2001, pp. 53-82 and Adamo 1987.

5. I am grateful to Dr. Patrick Sahle, of the University of Cologne, who, in his review of the paper before its publication, encouraged me to switch from the original term I had adopted, i.e. "metatext", to "paratext" (within the theoretical frame given by Genette 1987). The first term was meant to draw attention particularly to the most explicit forms of 'commentary' on the text (such as glosses and modern foot- or endnotes), which used to be the original main focus of my reflections. But the term "paratext" has the indubitable advantages of relying on a terminology well-established in the studies on textuality, and of including a broader range of textual objects that my model will take into account, such as handwritten rubrics, chapter numbering, copyright information at the beginning of print editions etc. The use of the plural (paratexts) aims to highlight the plurality and diversity of such 'secondary texts' within a textual tradition as well as within a single witness of the text.

6. Do these kinds of annotations belong to the maintext, or to a very 'close' level of paratexuality? – we will address this issue later.

7. Obviously this list does not pretend to be exhaustive in any way: its only sense is to give an idea of what I mean for 'paratextuality levels'.

8. My model does not require necessarily that the TEI-XML markup conventions be adopted, yet the TEI is, so far, the framework within which I imagine this project to be developed, and the TEI P5 Guidelines will be the main reference for the encoding conventions in the examples – see in particular chapter 11, *Representation of Primary Sources* (http://www.tei-c.org/release/doc/tei-p5-doc/en/html/PH.html) and chapter 16, *Linking, Segmentation, and Alignment* (http://www.tei-c.org/release/doc/tei-p5-doc/en/html/SA.html) – all Internet addresses quoted in this paper

doc/en/html/SA.html) – all Internet addresses quoted in this paper are valid up to January 2008). For one of the most important papers about the text-critical and transcriptional markup in the TEI, see Cover and Robinson 1995.

9. One could argue that this gloss actually comments on the single word "Theseus", and not on the entire line. We could easily account for this more detailed information, if we structured the transcriptions markup – or at least some portions of it – at a word (not line) level by using <w>, not <l> elements. All textual examples in this papers will be taken from Latin classical texts, but modified in some instances to illustrate the argument being made.

10. I shall define better later the way these 'objects' (that I will also call 'paratextuality level-files') could be realized, keeping in mind that their only function is to store the linking information about a particular paratextuality level in a specific source.

11. The linking strategies to be adopted are discussed below. Surely the more efficient way to identify at least the portions of maintext within a transcription will be a consistent system of @xml:id attributes.

12. The TEI-XML transcription of a manuscript, in particular, tends to be focused primarily on the description of the physical disposition of the text on the page, even while combining such information with other informational levels (for example, the 'abstract' internal structure of the text itself).

13. 'Transcriber' refers here to the creator of the transcription, whereas 'editor' refers to the philologist who creates the whole edition. The two may be the same individual, different individuals, or even the same, different, or overlapping groups of people.

14. 'User' refers to the end-user of the edition.

15. This suggestion is part of the vey useful feedback that I received from Prof. McCarty during the discussion the discussion at the Centre for Computing in the Humanities I mentioned above

16. To be precise, one could argue that it comments on the book as a physical document, than on the 'abstract text' represented by the maintext.

17. I owe many suggestions to Thaller 2004, especially about the need to base the digital scholarly editions on a standardized, wide base of digital reproductions and transcriptions of textual primary sources. I also agree very much, as it is already clear, with his idea of the digital edition as a process where many actors (from the transcriber, to the scholarly editor, including what I call the 'user') play their role. In the same volume, another paper I owe much to is

Huitfeld 2004.

18. That is, a commentary on a certain version of the maintext.

19. We could call such portions of the paratexts 'annotations', but this would be an incomplete definition as an introductory essay, for instance, could be considered a paratext (commenting on the whole maintext).

20. Yet, a first example of exception to this (only apparent) ease is given by ancient Greek lyric texts, with no certain distinctions between the verses.

21. Those are required, e.g., when concurrent numbering systems are proposed in different editions (let us think of the editions of Aesop's fables). Many print editions of prose texts solve those issues by showing parallel numberings in-line or in the margins, and sometimes a differentiation of the formatting conventions is required to distinguish one numbering system from the other. The alignment between the *scholia* and their 'target' in the text in print scholastic editions is often reached recalling the most diffused segmentation system of the maintext, and through the use (which in some cases turns out to be very useful) of *lemmata*, repeating in the paratext the precise portion of the maintext being commented. The need to use the *lemmata* shows in itself, I think, the arbitrary nature and the not complete efficiency of text partitioning conventions that can vary together with the change of the 'reference' editions.

22. This case is very common in modern print editions, even though in those cases the verse numbering 'of the manuscripts' is often preserved. But we can have manuscripts, or non-critical print editions, where such transpositions are not highlighted by the use of the old numbering, because of a lack of cure by the editor of that source, or simply because the transposition was not intentional.

23. One could use an 'idiosyncratic' numbering system counting the actual verses that appear in that version of the text; or a 'conventional' numbering including also the 'phantasm verses' of the lacunae; or whatever else be useful for the visualization of that text on screen.

24. See http://www.tei-c.org/release/doc/tei-p5-doc/en/html/SA.html#SATSXP.

25. The consistency of the unique @xml:id attributes in the Alignment-Text file A-Txt.xml and in the single transcription files would be essential for the software to perform this task, that is to link a note belonging to a transcription file to a 'placeholder' <l> element belonging to the Alignment-Text file.

26. Although different solutions would be required in more complex cases, for example when a note refers to a non-sequential portion of

the text. See http://www.tei-c.org/release/doc/tei-p5-doc/en/html/SA.html#SAPTIP for a discussion of the possible options, that could include the use of <ptr> elements (as intermediate pointers) as well.

27. In any case, as we saw above, the transcriber must always encode explicitly all information about the linking between text and commentary within a transcription any time that this cannot be 'deduced' automatically by the software, like in the case when a single 'note' (a gloss, a summary, a little introduction to a whole poem in a poetic collection, a prefatory essay etc.) comments on wider portions of text.

28. Those processes cannot be completely automatized, because the intervention of the editor will be required for the paratextuality level assignment in the most problematical cases, and, even more, because he must decide in the first place which parts of the transcription are 'pertinent'. With this term I mean, for example, that within the transcription of a miscellaneous manuscript (or print edition) containing many different literary works, only the part concerning our work must be parsed by the system.

29. If the Alignment-Text simply works as an ordered sequence of xml:ids, conceived to build up a uniform alignment system, the paratextuality level-files, in their turn, contain simply void elements expressing linking information. In other words, they inform the software about which portions of paratext in each source comment on which portions of maintext.

Works Cited

G. Adamo, La codifica come rappresentazione, in *Studi di codifica e trattamento automatico di testi* (ed. by G. Gigliozzi), Bulzoni Editore, Roma 1987.↵

F. Ciotti, Il testo elettronico: memorizzazione, codifica ed edizione, in *Macchine per leggere. Tradizioni e nuove tecnologie per comprendere i testi* (edited by C. Leonardi, M. Morelli and F. Santi), Centro italiano di studi sull'alto Medioevo di Spoleto, Firenze 1994.↵

C. Cover and P.M.W. Robinson, Encoding Textual Criticism, *Computer and the Humanities* 1995, 29, 123-136.↵

G. Genette, *Seuils*, Éditions du Seuil, Paris 1987.↵

C. Huitfeld, Text technology and textual criticism, in *Digital technology and philological disciplines* (edited by A. Bozzi, L. Cignoli, J.-L. Lebrave), Istituti editoriali e poligrafici internazionali, Pisa-Roma 2004 (= *Linguistica Computazionale* 1994, vol. XX-XXI),

Pisa-Roma 2004 (= *Linguistica Computazionale* 2004, vol. XXXXII),
259-275.↵

R. Mordenti, *Informatica e critica dei testi*, Bulzoni editore, Roma
2001.↵

M. Thaller, Digital manuscripts as base line for dynamic editions, in
Digital technology and philological disciplines (edited by A. Bozzi,
L. Cignoli, J.-L. Lebrave), Istituti editoriali e poligrafici internazionali,
Pisa-Roma 2004 (= *Linguistica Computazionale* 1994, vol. XX-XXI),
489-511.↵

XML