

Textual Variants

Marjorie Burghart

In this chapter you will learn:

- how to understand a traditional printed edition and its conventions;
- how to encode textual variants according to different methods and styles;
- how to solve several special cases common in critical editions.

When a text is transmitted through more than one witness, a critical edition will generally take a strong interest in recording the variant readings of some or all of those manuscripts or editions. There are many schools of textual criticism, more or less prevalent depending on the geographic area, field (literary studies, history, ...), and period of the original text. The history of (and relationships between) currents of textual criticism goes far beyond the scope of this chapter, and deserves a book of its own - if not a full library. The readers wishing to familiarise themselves with the various approaches could turn to Greetham 2013 for a short summary. But whatever the school or current of textual criticism, the TEI offers the same mechanism to record textual variance, described in the [Critical Apparatus chapter of the Guidelines](#).¹

For some readers, the concept of critical edition may be new. We will therefore open this chapter with a brief introduction to textual variants and the traditional way to record them, before moving to a survey of the TEI methods available for digital editions. We will give an overview of the different possible styles of apparatus, within the frame of the TEI. Finally, we will see some examples of particular cases commonly occurring in critical editions.

1. Understanding textual variants in a critical edition

The presence of variants and errors in the various witnesses of a text (manuscripts and editions) is particularly important in manuscript cultures, especially for classical and medieval texts. As texts were copied from manuscript to manuscript, the scribes introduced in their copies readings that differed from the model (or models) from which they were copying. A common distinction is to call “errors” the differences that are mere mistakes (typo, forgotten or repeated word, etc.), and “variants” the ones willingly introduced by a scribe, showing some creativity. However, it is sometimes difficult to distinguish between the two, and besides the concept of “error” implies the existence of a correct, reference text, which can be seen as a view too partial to particular currents of textual scholarship. We will therefore use only the more neutral term of “variant” or

1. <http://www.tei-c.org/Vault/P5/3.0.0/doc/tei-p5-doc/en/html/TC.html>

“reading” here. For editions of texts published in the print era, the proportion of variants is naturally much less significant, but nevertheless variant readings exist between editions and must be taken into account.

Critical editors need to render account of their work on the tradition and the choices they made, in a way that would allow their readers to verify their work. In a traditional printed edition, scholars have to conciliate, in the two-dimensional space of the page, the legibility of the text they propose with the informativity of the annotations representing their exploration of the tradition, and their choices (if any) among the variant readings. More than a century of modern scholarship has shaped the traditional critical edition page: it is composed of a main text, which is the version of a text proposed by the editor, to which refer one or more groups or layers of footnotes (one for textual variants, one for source identifications, one for the reference of biblical citations, one for historical notes, etc.). To avoid the multiplication of footnote markers in the text, which impairs its legibility, it is common in printed editions to link the notes to the text by the means of line numbers: the lines of the text are numbered, and in the footnotes a reference is made to a line number, or the numbers of a range of lines.

The [following illustration](#) is an example of such traditional critical edition layout. It has three distinct layers of footnotes. The first two (textual variants, and biblical quotations) refer to line numbers in the main text, while the last (source identifications) has footnote markers in the text. The editor has proposed a text from the evidence found in four manuscripts, A, S, T and H. Using the textual variants footnotes, the reader must be able to deduct the readings of each witness. Each footnote about a textual variant has the same structure: following the line number is the “lemma,” that is the word or group of words from the main text for which there is a variant reading. The lemma is followed by a description of what changes (is this word omitted? is it replaced with another word? etc.), and in which manuscript(s). For instance, if we look at the footnotes referring to line 44 in the text, we can reconstruct the text of each manuscript:

- A has the same sentence as the one presented in the main text: “scilicet malis que obuenerunt ei,” which translates as “that is to say the misfortunes that hit him”;
- S has one omission, “scilicet” is missing: “malis que obuenerunt ei”;
- T has two omissions, “scilicet” is missing at the beginning, and “ei” at the end: “malis que obuenerunt”;
- H has one variant reading “euenerunt” instead of “obuenerunt”: “scilicet malis que euenerunt ei”

You will notice that many abbreviations are used in those footnotes, in order to save space: *om.* for an omission, *add.* for an addition to the text of the edition, etc.

omnibus hiis, scilicet malis que obuenerunt ei, non peccauit
 45 *Iob labiis suis*, nec contra Deum, nec contra proximum, nec
 contra uxorem. Item sufferunt omnia patienter, confortati
 exemplo eiusdem Iob iuxta illud Iacobi ultimo: *Sufferentiam*
Iob audistis. Et bene debet tale exemplum animare. Si enim
 50 patienter portauit amissionem tantorum bonorum et proprio-
 rum filiorum, et tam grauem et tam uilem infirmitatem a *plan-*
ta pedis usque ad uerticem, quanto magis alii in quibus desunt
 conditiones predictae debent esse patientes in minoribus!

Preterea notandum quod amor presentis uite occasio est
 55 omnium malorum sicut ostendit beatus Sebastianus in sermo-
 ne quem fecit Marco et Marcelliano, ubi, enumeratis multis
 et grauibus malis, concludit dicens: « Numquid non presentis
 uite amore ducti ista committunt? »¹. Sed huiusmodi infirmi-
 tas acerba et desperata et abhominabilis amaricat omnes io-
 60 cunditates huius uite, et ideo retrahere debet ab amore huius
 uite, sicut amaritudo circumposita uberibus retrahit puerum
 ab amore uberum. Item amaritudo huius uite debet nos mouere
 ad desiderandum aliam uitam beatam que est sine amaritudi-
 ne. Augustinus: « Cum molestie in huius uite fragilitate cre-
 65 brescunt, eternam requiem nos desiderare compellunt »². Et
 ideo in huiusmodi infirmitatis constituti amaritudine debent
 ad illam beatam dulcedinem uehementer suspirare. Et sic fe-
 cit quidam religiosus leprosus qui sanctissime se habens in

⁴⁴ scilicet *om.* S T obuenerunt / euenerunt H ei *om.* T ⁴⁶ sufferunt / su-
 ferant S T ⁴⁷ ultimo / 5 S T sufferentiam / sufficientiam S T ⁴⁹ quasi *om.*
 H S T tam / et *praem.* S T ⁵¹ filiorum *om.* A ⁵² desunt *om.* S
 T ⁵⁴ preterea / postea H ⁵⁶ enumeratis / et innumeratis S ⁵⁷ et graui-
 bus *om.* H numquid / numquam S T non *om.* S T ⁵⁸ committunt / alias di-
 mittunt *add.* H ⁶⁰ et *om.* A ⁶⁰⁻⁶¹ et ideo... uite *hom.om.* A ⁶¹ uberibus
 / ubique S T ⁶³ desiderandum/considerandum A ⁶⁴ in *om.* A fragilitate /
 flagelitate S T, fragilitates H ⁶⁵ requiem *om.* S T ⁶⁷ uehementer *om.*
 A ⁶⁸ se habens / sequens S T

⁴⁷ Iac 5, 11 ⁵² Iob 2, 7

¹ Cf. *Passio sancti Sebastiani*, PL 17, 1025D-1026A.

² AUGUSTIN, *Ep.* 145, 2, PL 33, 983.

Figure 1: An example of traditional critical edition in print (from Nicole Bériou et François-Olivier Touati, *Voluntate Dei leprosus: les lépreux entre conversion et exclusion aux XIIIe et XIIIe siècles*, Spoleto: Centro Italiano di Studi sull'Alto Medioevo, 1991, 162)

Now that we have deciphered a print edition, we can work in the opposite direction, and see how we would build such an edition from manuscript evidence. Let us imagine a simple case, where our text is transmitted through three surviving manuscripts that we will call A, B and C. These witnesses of the text form its “tradition,” that is all manuscripts and editions which have born the text. In these three manuscripts, the text is almost the same, but in manuscript B, we have emphasised a word that differs from the readings of the two other manuscripts:

A	B	C
... ut dum spiritus reficitur dolores corporis leniantur...	... ut dum spiritus reficitur dolores corporis <i>laniantur</i> ut dum spiritus reficitur dolores corporis leniantur...

The Latin sentence born by manuscripts A and C, “... ut dum spiritus reficitur dolores corporis leniantur...,” means “so that, while the spirit is restored, the sufferings of the body are soothed.” It is grammatically correct and makes perfect sense. The sentence born by manuscript B, “... ut dum spiritus reficitur dolores corporis *laniantur*...,” has only one word that changes (actually, only one letter), but its meaning is totally different: it translates as “so that, while the spirit is restored, the sufferings of the body are *torn*.” This is not a correct and meaningful sentence, and a critical editors will deduce that the scribe copying manuscript B made a mistake while copying “leniantur,” or copied from a manuscript (now lost) already bearing “laniantur” instead of “leniantur.” If we were to produce a traditional printed edition, here is how we would build our critical apparatus note:

- *Text:*
ut dum spiritus reficitur dolores corporis leniantur^(a)
- *Footnote:*
a) leniantur] laniantur *B*


Figure 2: Structure of the information contained in a textual variant footnote

Digital editions, unlike printed ones, are not limited to “the two-dimensional space of the ‘page’ and to typographic means of information representation” (Sahle 2008). Yet, it is very important for digital Humanists to understand the traditional semiology of critical editions and the decades of scholarship that produced it. Conversely, it is capital for scholars familiar with this traditional semiology to be able to think beyond it, in order to create truly digital scholarly editions which are free from the constraints of the printed page and convey to the reader the same information, and even more, through different means. This is what we will learn in the following section, explaining how textual variants are represented in a TEI XML edition.

2. Encoding textual variants in TEI

To record textual variants, the TEI uses the `<app>` (apparatus entry) element, which “contains one entry in a critical apparatus, with an optional lemma and usually one or more readings or notes on the relevant passage.”² An apparatus entry can be critical, i.e., express a choice made by the editor to favour one reading over the others. In this case, it will contain two types of elements:

- a single `<lem>` element (lemma), which records the reading chosen by the editor. In a printed edition, this is the text we would find in the main text of the edition;
- one or more `<rdg>` (reading) elements, which record the rejected readings. In a printed edition, this is the information we would find in a critical footnote, saying that instead of the lemma which manuscript has which variant instead of the lemma.

Or the apparatus entry can be neutral and simply record how the various witnesses vary one from another at a given point of the text, without expressly favouring a reading. In this case it will only contain `<rdg>` elements.

The critical or neutral style of apparatus entries is chosen according to the school of textual criticism in which belongs the editor. But even editors who want critical apparatus entries in their final edition might want to use neutral entries, simply recording variants, while they are still collating manuscripts and have not yet decided which reading will be the lemma. It is perfectly possible to start working on an edition with neutral entries, then choose a lemma for each entry in the later stages of the work.

For `<lem>` as well as for `<rdg>`, the `@wit` (witness) attribute is used to materialise the link between the given text and the manuscript(s) bearing it: `@wit` may contain one or more reference(s) to the unique `@xml:id` of witnesses described in a list of witnesses (`<listWit>`), typically placed in the `<sourceDesc>` section of the header or in the `<front>` part of the `<text>`. This list contains several `<witness>` elements identified by their siglum (`@xml:id`). The first step must therefore be the creation of this list of witnesses, to be able to use the references to the witness sigla in the apparatus entries. A minimalist version of this list could be encoded like this:

```
<listWit>
  <witness xml:id="A">Short description of witness A (city, library,
shelfmark)</witness>
  <witness xml:id="B">Short description of witness B (city, library,
shelfmark)</witness>
  <witness xml:id="C">Short description of witness C (city, library,
shelfmark)</witness>
</listWit>
```

2. <http://www.tei-c.org/Vault/P5/3.0.0/doc/tei-p5-doc/en/html/ref-app.html>

To learn more about the encoding of witness descriptions, which you can make as complete and precise as you wish, please read the chapter Manuscript Description.

The `<lem>` and `<rdg>` elements wrapped in an `<app>` are the basic tools to represent textual variants, but the TEI offers three different methods to combine them, each with its pros and cons. Even though we generally recommend to use the Parallel segmentation method, the following review of the three methods might help you make well-informed choices. This is of course just an overview, and we refer you to the [Guidelines](#) for a full description of the methods and their options. Let us examine them individually, and see how our latest example, where we have one variant among three manuscripts, would be encoded with each of these methods.

A	B	C
... ut dum spiritus reficitur dolores corporis leniantur...	... ut dum spiritus reficitur dolores corporis <i>laniantur</i> ut dum spiritus reficitur dolores corporis leniantur...

2.1. Location-referenced method

In the Location-referenced method, we need to have a base text to attach apparatus entries to it. It means that it is not suitable if we want to make an edition without using the concept of base text. The apparatus entries are not linked to a precise spot in the text, but to a block of text: we indicate to our readers that there is an apparatus entry linked to a particular paragraph, or line of verse, etc. Human readers will be able to understand the meaning of the entry and its relation to the base text, but a script cannot process the information automatically beyond simply displaying an apparatus note linked to the relevant block of text. It means that this method is not suitable either if you wish to be able to virtually reconstruct the individual text of each witness, for instance.

One of its advantages though is that, since the apparatus entries are linked to the text by a simple system of references, they will not interfere with other hierarchies, i.e., the `<app>` elements will not overlap with other elements, like the ones used to encode citations, text structures, etc.³ The apparatus entries may be “internal” (stored in the same file as the edition) or “external” (stored in a different file).

This method can be useful when converting printed edition, if you simply want to display notes loosely attached to the text, without the need to process the result in elaborate fashions. Its compatibility with overlapping hierarchies is also an argument in its favour, but the other methods also have solutions. However, we would not generally recommend using the Location-referenced method, since it is just as long to encode as with other methods, but offer less possibilities for processing.

Here is how our example would be encoded with the Location-referenced method. In the base text we would have a block of text identifiable by some sort of reference (here, paragraph “1” in the division “sermo1”:

3. For a discussion of the methods available, see section Handling overlap in chapter Citations and references.

```
<div n="sermo1">
  <p n="1">...ut dum spiritus reficitur dolores corporis leniantur...</p>
</div>
```

Somewhere else, either in the same file as the edition (in the **<back>** for instance) or in another related file, we would have the apparatus entries. The **@loc** attribute of **<app>** is used to represent the reference to the relevant block of text in the base text:

```
<app loc="sermo1 1">
  <lem>leniantur</lem>
  <rdg wit="#B">laniantur</rdg>
</app>
```

2.2. Double-end-point-attached method

The Double-end-point-attached method, just like the previous one, implies the existence of a base text - with the same consequences. But this time, the apparatus entries are linked to a precise span of the base text, using the empty tags **<anchor>** if no other element is already present: one to mark up the beginning of the lemma, and another to mark up the end. This means that, beyond generating notes at the relevant places, you will be able to process your edition in more elaborate ways than with the previous method. Since it uses empty tags, marking up just points in the text, this method is also convenient if you have to deal with overlapping hierarchies.

This method is intellectually satisfying: it allows us to link your apparatus precisely to the base text, leaving all the possible processing options, while allowing potentially overlapping hierarchies to coexist without any problem. Its major drawback is practical: first, you will have to place the elements or anchors marking the limits of the lemmata in the base text, giving each an **@xml:id**, and then you will need to be very careful when linking the **<app>** to these identifiers, since a mistake will mean that your apparatus entry will be linked to a wrong span of the base text. In the absence of a specialised XML editor that would take care of these delicate steps, it is a tedious and error-prone task. Second, and perhaps more importantly, at the time of writing there are no available tools offering support to display or process critical editions encoded following the Double-end-point-attached method. This means that you will have to develop your own scripts and applications to be able to display and/or process an edition encoded with this method.

Our example could be encoded as follows with the Double-end-point-attached method. Anchors would mark the beginning and end of the lemma in the base text:

```
<div n="1">
  <p n="1">... ut dum spiritus reficitur dolores corporis <anchor
xml:id="applbegin"/>leniantur<anchor xml:id="applend"/> ... </p>
</div>
```

Somewhere else, either in the same file as the edition (in the **<back>** for instance) we would have the apparatus entries. The **@from** and **@to** attributes of **<app>** contain a reference to the **@xml:id** of the tags marking up the beginning and end of the lemma in the base text:

```
<app from="#applbegin" to="#applend">
  <rdg wit="#B">laniantur</rdg>
</app>
```

2.3. Parallel segmentation method

The parallel segmentation method does not necessarily imply a base text: we may have one, if we use **<lem>** in our apparatus entries, or we may not, if we only use **<rdg>**. With this method, the apparatus entries are placed “inline,” instead of being placed somewhere else in the edition file or in a related XML file. It means that the apparatus does not need to be linked to the text through references, since it occurs in the flow of the edition’s text. Each time witnesses have different readings, an **<app>** element is inserted, containing either a lemma and one or more readings, or a simple list of variant readings. These **<app>** are of course at risk of overlapping with other elements, but should it occur there are mechanisms allowing you to handle them.⁴

Relatively easy to encode manually, giving the option of critical or neutral apparatus entries, and supported by several third-party tools,⁵ the Parallel segmentation method is the most widely used. It is the method we recommend to use, unless you have a particular reason to prefer another. All the further examples of apparatus entries will be encoded using this method.

We might encode the example as follows - we could also add a **@wit** attribute to **<lem>**, but it is optional as, by default, all the witnesses not mentioned in a **<rdg>** are supposed to bear the text of the lemma:

```
<div n="1">
  <p n="1">... ut dum spiritus reficitur dolores corporis <app>
    <lem>leniantur</lem>
    <rdg wit="#B">laniantur</rdg>
  </app> ... </p>
</div>
```

3. Different styles of apparatus

Besides the encoding methods, editors must opt for a style of apparatus. The choice of a style of apparatus has nothing to do with technical considerations: it is a methodological issue, linked to the type of edition you intend to produce. Some schools of textual criticism will recommend the neutral recording of the variants, other will consider a lemma absolutely necessary. However, even for the latter type of editions, we recommend a mixed approach: a neutral recording of variants during the collation stage, then a conversion to critical apparatus entries by choosing a lemma.

4. See section Handling overlap in chapter Citations and References. The mechanisms are explained with the example of quotations, but they may be applied to any other element.

5. At the time of writing, the Parallel segmentation method is the only one supported by tools dedicated to TEI critical editions, like [Edition Visualization Technology](#), [Stemmaweb](#), the [TEI Critical Apparatus Toolbox](#) or the [Versioning Machine](#)

3.1. Neutral recording of the variants

When opting for a neutral recording of the variants, the editor lists the varying readings of the witnesses, without choosing a lemma. Editors who do not want to privilege the readings of a witness over the others may choose this style of apparatus.

It is worth mentioning that even for editors wishing to produce an edition using lemmata, this style of apparatus may be a valuable first step: when the editor has not yet collated all the witnesses, it may be difficult to decide which reading is the lemma. In this case, a common strategy is to encode first the apparatus in a neutral style, then, once the collation stage of the edition is over, make an editorial decision for each `<app>` and choose which `<rdg>` will become the `<lem>`.

With this style of apparatus, the same example would be encoded as follows:

```
ut dum spiritus reficitur dolores corporis <app>
  <rdg wit="#A #C">leniantur</rdg>
  <rdg wit="#B">laniantur</rdg>
</app> ...
```

3.2. Negative critical apparatus

When editors wish to produce an edition using lemmata, critical apparatus entries are required, but they may belong in two different styles: negative, or positive. In a negative critical apparatus entry, the witnesses are listed in `@wit` only for the variant readings (`<rdg>`). By default, all the witnesses which are not listed in a `@wit` attribute of a `<rdg>` of the current `<app>` are supposed to bear the text of the `<lem>`.

Practically, a negative critical apparatus is quicker to encode, especially when there is a long list of witnesses: we do not need to add a `@wit` attribute, and insert references to the witnesses in its value. On the other hand, processing it may be more complicated, depending on your goal, since the scripts will need to reconstruct for themselves the list of witnesses bearing the lemma. Another argument against a negative critical apparatus is that it is more difficult to verify: during the collation, it is only human to make mistakes - forgetting to add the reference to a new witness bearing the same variant as an existing `<rdg>` for instance, or making a typo which will result in the wrong manuscript being mentioned twice in the `@wit` value or a reading, etc. If this is a concern, we recommend choosing the positive style. It is worth noting that, in terms of display, it is perfectly possible to display or print in a negative style an apparatus that has been encoded as positive.

Example with a negative critical apparatus:

```
<app>
  <lem>leniantur</lem>
  <rdg wit="#B">laniantur</rdg>
</app>
```

3.3. Positive critical apparatus

In a positive critical apparatus entry, the witnesses are listed for the lemma as well as for the other readings. This is the method we generally recommend for editions using the concept of lemmata. This style is slightly longer to encode than a negative apparatus, but has some advantages: by forcing the encoder to explicitly state which witnesses bear the lemma and which have variants, a positive apparatus allows automatic verifications of the consistency of your encoding. Scripts will be able, for instance, to check whether each apparatus entry mentions a reference to all the witnesses - if one does not, it means that a mistake has been made and we must verify this apparatus entry. Another advantage is that scripts will generally have less difficulties processing your edition, since less operations are required to reconstruct the text of individual witnesses. Finally, as we noted above, encoding a positive apparatus does not imply you will have to display a positive apparatus: those are two different things, and the display scripts may perfectly be configured to display one style or the other, since in both cases the same information is recorded - implicitly in the first case, explicitly in the latter.

Example with a positive critical apparatus:

```
<app>
  <lem wit="#A #C">leniantur</lem>
  <rdg wit="#B">laniantur</rdg>
</app>
```

4. Going further with the Parallel segmentation method

We have seen the basic principle of the Parallel segmentation method, in various styles, and how they apply to one type of variant, the “substitutions” - when a word or phrase is replaced with another word or phrase. In this section we will first examine the other three main types of variants: “omissions,” “additions,” and “transposition” and suggest efficient ways to encode them.⁶ Then we will turn to some examples of the variety of particular that may arise.

4.1. Omissions, additions and transpositions

4.1.1. Omissions

There is an omission when a witness has nothing where the exemplar from which it was copied has a lemma. In this case, you can simply leave the `<rdg>` element empty where the witnesses have nothing. For instance, if we consider the following example, manuscript B has an omission: the word “corporis,” which is present in the other manuscripts and makes perfect sense in the sentence, is absent.

A	B	C
... ut dum spiritus reficitur dolores corporis leniantur...	... ut dum spiritus reficitur dolores _____ leniantur...	... ut dum spiritus reficitur dolores corporis leniantur...

6. The use cases presented here are derived for the most part from Burghart 2011.

- *Text:*
ut dum spiritus reficitur dolores^(a) corporis leniantur
 - *Footnote:*
a) dolores *om.* B
-
- lemma | editorial explanation | witness

Figure 3: Representing this omission in a printed apparatus

We may encode this variant in the following way:

```
<app> ... ut dum spiritus reficitur dolores <lem wit="#A
#C">corporis</lem>
<rdg wit="#B"/>
</app>
```

4.1.2. Additions

Conversely, there is an addition when one or more of the witnesses has some text where the exemplar from which it was copied has nothing. Additions may stem from innovation (a conscious intervention of the scribe) or mere errors (copying the same sentence twice, for instance), and if we wish to make this distinction we may use the **@type** attribute on **<rdg>** to categorise them.

In case of an addition, it is the **<lem>** element that we will leave empty, since there is no text at this place in the text of our edition. For instance, if we consider the following example, manuscript B has an addition: the word “corporis” has been written twice by the scribe - a type of scribal error known as “dittography.”

A	B	C
... ut dum spiritus reficitur dolores corporis leniantur...	... ut dum spiritus reficitur dolores corporis <i>corporis</i> leniantur...	... ut dum spiritus reficitur dolores corporis leniantur...

- *Text:*
ut dum spiritus reficitur dolores corporis^(a) leniantur
 - *Footnote:*
a) corporis] corporis add. B
- lemma | rejected reading | ed. explanation | witness

Figure 4: Representing this addition in a printed apparatus

We may encode this variant in the following way; here, since it is a particular type of addition, we have chosen to use the optional @cause attribute to indicate the type of phenomenon occurring:

```
... ut dum spiritus reficitur dolores corporis <app>
  <lem wit="#A #C" />
  <rdg wit="#B" cause="dittography">corporis</rdg>
</app> leniantur...
```

Additions raise a particular issue when it comes to displaying your edition. Reconstructing the individual text of each witness is not a problem, but if we wish to display a main text with notes, in the fashion of traditional printed apparatus, we will need to display, as lemma, the word preceding the addition in the main text, so our readers may know where the addition occurs. Since the <lem> is empty, this word is the one immediately preceding the <app> in the main text, and it is not encoded as a lemma. There are various strategies: one consists in handling everything at the processing stage, typically by writing an elaborate XPath selecting this word in the XSLT - a perfectly reasonable solution, but requiring a good level of XPath knowledge. Another strategy is to record the desired lemma, as you wish it to appear in the future notes, in an unused attribute of <lem> - for instance, @n. Our encoding would in this case be the following, after documenting our particular usage of @n:

```
... ut dum spiritus reficitur dolores corporis <app>
  <lem wit="#A #C" n="corporis" />
  <rdg wit="#B" cause="dittography">corporis</rdg>
</app> leniantur...
```

4.1.3. Transpositions

There is a transposition in a witness when the order of the words, or sentences, or paragraphs from the exemplar is changed. In this example, for instance, manuscript B has “corporis dolores” where the other manuscripts have “dolores corporis”:

A	B	C
... ut dum spiritus reficitur dolores corporis leniantur...	... ut dum spiritus reficitur <i>corporis dolores</i> leniantur...	... ut dum spiritus reficitur dolores corporis leniantur...

In a traditional printed edition, transpositions are typically represented by giving the full phrase or sentence in the lemma, and only the first letter(s) of the transposed words in the reading part of the note, to save space on the page:

dolores corporis] c. d. *B*

Transpositions are the most difficult type of variants to encode, at least with the Parallel Segmentation method. In simple cases, like the above, the encoding is rather simple:

```
... ut dum spiritus reficitur <app>
  <lem wit="#A #C">dolores corporis</lem>
  <rdg wit="#B" type="transposition">corporis dolores</rdg>
</app> leniantur...
```

When, however, a passage is moved to a different place in the text, for instance when the third paragraph of a text is moved to the first place in a witness, the encoding is a little more complex. To achieve this, we are going to use a linking mechanism available in the TEI, and more precisely the **@copyOf** attribute.

Let's consider two hypothetical witnesses of a nursery rhyme. In manuscript B, although the text is the same, the usual order of the verses has been changed: the third and fourth verses are swapped.

A	B
Hey diddle diddle, the cat and the fiddle The cow jumped over the moon The little dog laughed to see such sport And the dish ran away with the spoon	Hey diddle diddle, the cat and the fiddle The cow jumped over the moon And the dish ran away with the spoon The little dog laughed to see such sport

The first step is to give an **@xml:id** to each of the elements which are transposed. Here, we are working on lines of verse which can be encoded with the **<l>** element. The mechanism would be identical if we were dealing with words (**<w>** element), sentences (**<s>**), paragraphs (**<p>**) or even divisions (**<div>**). Of course it would be perfectly acceptable to give an **@xml:id** to each **<l>**, but it is not necessary for our purpose, so in this example we will do that only for the verses swapped in manuscript B:

```
<l>Hey diddle diddle, the cat and the fiddle</l>
<l>The cow jumped over the moon</l>
<l xml:id="verse3">The little dog laughed to see such sport</l>
<l xml:id="verse4">And the dish ran away with the spoon</l>
```

Now that the transposed elements have an **@xml:id**, we can add the markup for an apparatus entry. In the lemma, we simply put the two lines of verse swapped in manuscript B. In a **<rdg>**,

we represent the transposition by adding two `<l>` elements which are virtual copies of verse 3 and verse 4 respectively. This is done with the help of the `@copyOf` attribute, which is used to point to an element of which the current element is a copy.

```
<l>Hey diddle diddle, the cat and the fiddle</l>
<l>The cow jumped over the moon</l>
<app>
  <lem wit="#A">
    <l xml:id="verse3">The little dog laughed to see such sport</l>
    <l xml:id="verse4">And the dish ran away with the spoon</l>
  </lem>
  <rdg wit="#B">
    <l copyOf="#verse4"/>
    <l copyOf="#verse3"/>
  </rdg>
</app>
```

4.1.4. Nested apparatus entries

It is also worth mentioning that variants may be “nested.” Let us consider the following example: manuscript B has a substitution (“laniantur” instead of “leniantur”), but manuscript C has an omission spanning over the same passage (“dolores corporis leniantur” is omitted)

A	B	C
... ut dum spiritus reficitur dolores corporis leniantur...	... ut dum spiritus reficitur dolores corporis <i>laniantur</i> ut dum spiritus reficitur _____ ...

To encode this, we simply nest the shortest variant inside the lemma of the longest:

```
... ut dum spiritus reficitur <app>
  <lem wit="#A #B">dolores corporis <app>
    <lem wit="#A">leniantur</lem>
    <rdg wit="#B">laniantur</rdg></app>
  </lem>
  <rdg wit="#C"/>
</app>...
```

4.1.5. Overlap

As a final note, it may happen that overlaps occur between variants. The risk of having to deal with such phenomena increases greatly with the number of manuscripts used in an edition. Let us consider the following example: manuscript B has a substitution (“mentis laniantur” instead of “corporis leniantur”), but manuscript C has an omission overlapping the same passage (“reficitur dolores corporis” is omitted)

A	B	C
... ut dum spiritus reficitur dolores corporis leniantur...	... ut dum spiritus reficitur dolores <i>mentis laniantur</i> ut dum spiritus _____ leniantur ...

We will discuss in details the various methods available to handle overlaps at the end of the chapters dedicated to *Citation and References*. Here we will only consider encoding the example according to the Fragmentation and Reconstitution of Virtual Elements method, but you are invited to consider other methods. Here we may split in two the substitution in B, which will lead us to have a nested variant:

```
... ut dum spiritus <app>
  <lem wit="#A #B">reficitur dolores <app>
    <lem wit="#A">corporis</lem>
    <rdg wit="#B">mentis</rdg>
  </app>
</lem>
<rdg wit="#C" /></app>
<app>
  <lem wit="#A #C">leniantur</lem>
  <rdg wit="#B">laniantur</rdg>
</app> ...
```

4.2. Some special cases

In some cases, the apparatus entry may be complicated by other factors: the necessity to add precisions about how some text has been corrected (or at least changed) in a witness, where, and by which hand; or to point out (and maybe correct) inaccurate text; or to supply missing letters or words.

In those cases, the encoding of an apparatus entry must be combined with elements used to represent the more documentary aspects of a primary source. We give here an overview of some of the most common cases.

4.2.1. Word or phrases corrected by the scribe

Scribes make mistakes, which are sometimes corrected in the manuscript, either by the scribes themselves or by later readers. To encode this phenomenon, we must combine the apparatus entry with elements used in transcriptions. In this case, we can use the **<subst>** mechanism, used to encode a “substitution,” which in TEI “groups one or more deletions with one or more additions when the combination is to be regarded as a single intervention in the text”⁷ In other words, a substitution happens when the scribe or a medieval reader deletes a mistake (by scraping the parchment, crossing over the wrong word, or expunctuating it for instance), and in the same movement adds what he considered the correct word.

If we had the following situation:

7. See section Substitutions in the chapter dedicated to Transcription. Note that what the TEI calls a “substitution” is different from what philologists call a “substitution.”

4.2.3. *Sic*: pointing out inaccurate or incorrect text

The element `<sic>` in TEI “(Latin for thus or so) contains text reproduced although apparently incorrect or inaccurate,”⁹ which corresponds to the definition used by philologists, who add the mention *sic* after an unexpected reading. In the example below, for instance, manuscript B has “carporis” instead of “corporis,” which is a spelling mistake.

A	B	C
... ut dum spiritus reficitur dolores corporis leniantur...	... ut dum spiritus reficitur dolores <i>carporis</i> leniantur...	... ut dum spiritus reficitur dolores corporis leniantur...

In a traditional printed edition, we could represent this information as follows:

corporis] carporis *sic* B

This is how we could encode this phenomenon:

```
... ut dum spiritus reficitur dolores <app>  
  <lem wit="#A #C">corporis</lem>  
  <rdg wit="#B"><sic>carporis</sic></rdg>  
</app> leniantur...
```

In this case, it is obvious enough that the inaccurate word is a misspelt version of “corporis,” although in some cases it is worth mentioning explicitly what we think the correct form should be. In a traditional printed edition, we could represent this as follows:

corporis] carporis *sic pro* corporis B

The corresponding encoding would be:

```
... ut dum spiritus reficitur dolores <app>  
  <lem wit="#A #C"/>  
  <rdg wit="#B">  
    <choice>  
      <sic>carporis</sic>  
      <corr>corporis</corr>  
    </choice>  
  </rdg>  
</app> leniantur...
```

4.2.4. *Suppleamus*: when the editor supplies missing text

When the text of a witness is damaged or unclear, the editors may have to supply their own conjecture. Let us imagine for instance that, in yet another variant of our example, manuscript B has an extra word added after “corporis,” which has been made illegible by some later damage to the manuscript (for instance a stain):

9. <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-sic.html>

A	B	C
... ut dum spiritus reficitur dolores corporis leniantur...	... ut dum spiritus reficitur dolores corporis leniantur...	... ut dum spiritus reficitur dolores corporis leniantur...

If conjecture as to the nature of this word is impossible, we may simply consider that it is a “locus desperatus” (or “hopeless passage” in Latin), and content ourselves with marking up the words as a damaged part of the manuscript, using the **<unclear>** element, which “contains a word, phrase, or passage which cannot be transcribed with certainty because it is illegible or inaudible in the source.”¹⁰ This element has various useful attributes, among which **@reason** that lets us specify what makes the passage unclear, and **@extent** that lets us give a full-text estimate of the number of letters or words of the unclear passage:

```
... ut dum spiritus reficitur dolores corporis <app>
  <lem wit="#A #C"/>
  <rdg wit=" #B"/>
</app> leniantur...
```

When a conjecture is possible, the editors will indicate which word(s) should be in the text according to them, using the **<supplied>** element, which “signifies text supplied by the transcriber or editor for any reason; for example because the original cannot be read due to physical damage, or because of an obvious omission by the author or scribe.”¹¹ Among the attributes available for **<supplied>**, **@reason**, which may be used to explain why this word had to be supplied by the editors. Here, for instance, the editors think that the illegible word is “uestri,” meaning “your”

```
... ut dum spiritus reficitur dolores corporis <app>
  <lem wit="#A #C"/>
  <rdg wit=" #B"><supplied reason="damage">uestri</supplied></rdg>
</app> leniantur...
```

In some cases, there might be various hypothesis regarding the elucidation of the illegible passage. Here for instance, the word added in manuscript B could be read as “uestri” (“your”), or maybe as “nostri” (“our”). To express those two possible readings of the illegible word, we could use the **<choice>** element to wrap several possible **<supplied>** elements. The optional **@cert** (certainty) attribute on **<supplied>** could be used to indicate our level of certainty regarding each supplied possibility. Here for instance, if we are fairly certain that the word is “uestri,” but still think there is a possibility it might be “nostri,” we could encode the phenomenon like this:

10. <http://www.tei-c.org/Vault/P5/3.0.0/doc/tei-p5-doc/en/html/ref-unclear.html>

11. <http://www.tei-c.org/Vault/P5/3.0.0/doc/tei-p5-doc/en/html/ref-supplied.html>

```

<app>
  <lem wit="#A #C"/>
  <rdg wit=" #B"><choice>
    <supplied cert="high">uestri</supplied>
    <supplied cert="low">nostri</supplied>
  </choice></rdg>
</app>

```

When different editors have different opinion on the right solution, we can use the `@resp` attribute to indicate who suggests to supply what:

```

<choice>
  <supplied cert="high" resp="#MB">uestri</supplied>
  <supplied cert="low" resp="#PJ">nostri</supplied>
</choice>

```

4.2.5. Displaying long lemmata

To finish up this long and rather dense chapter, let us look at a practical issue: we might have in our edition long lemmata, spanning over several lines of text, if not several paragraphs (for a long omission, for instance). This is not an issue in itself, but it may pose a problem if we wish to display our edition in a traditional fashion, in print or online, with footnotes giving some information to the readers. When a lemma is longer than a couple of words, the traditional convention is to simply give its first and last word separated by an ellipsis, instead of repeating the whole passage (the first and last few significant words may be given when the passage is long, or to avoid confusion).

Let us imagine for instance that a manuscript omits these two long sentences:

```

<app>
  <lem wit="#A #C">Amicitie primum, ut mihi uidetur, ipsa natura
humanis mentibus impressit affectum, deinde experientia auxit,
postremo legis auctoritas ordinavit. Deus enim summe potens, et summe
bonus, sibi est ipsi sufficiens, qui bonorum nostrorum non eget;
uoluit ut omnes creaturas suas pax componeret, et uniret
societas.</lem>
  <rdg wit=" #B"/>
</app>

```

In a traditional printed critical edition, the footnote about this phenomenon would have a shortened version of the lemma, like this:

amicitie primum ... uniret societas] *om. B*

If we wish to display the same sort of notes for some version of our digital edition, we have two solutions:

- Leave our encoding as above, and process the lemma (typically in XSLT) to transform its text into the shortened version before displaying it. It is a good solution for people who

are familiar enough with a suitable processing language, and also if the encoding is not too complicated by nested variants, which is likely to happen in a very long lemma.

- Add an element to our encoding, to record how we wish the lemma to be displayed in notes. There is no standard way of doing this, but we could use for instance the generic `<note>` element, with a specific `@type` attribute value (for instance `"altLem"` for “alternative to the lemma”):

```
<app>
  <lem wit="#A #C">Amicitie primum, ut mihi uidetur, ipsa natura
humanis mentibus impressit affectum, deinde experientia auxit,
postremo legis auctoritas ordinavit. Deus enim summe potens, et
summe bonus, sibi est ipsi sufficiens, qui bonorum nostrorum non
eget; uoluit ut omnes creaturas suas pax componeret, et uniret
societas.</lem>
  <rdg wit=" #B"/>
  <note type="altLem">amicitie primum ... uniret societas</note>
</app>
```

With this encoding, it is much easier to display a shortened version of the lemma in notes when we need to, as an alternative to the full text.

Bibliography

- Marjorie Burghart, TEI: Critical Apparatus Cheatsheet, 2011, <http://marjorie.burghart.online.fr/?q=en/content/tei-critical-apparatus-cheatsheet>
- David C. Greetham, “A History of Textual Scholarship,” in Neil Fraistat and Julia Flanders (eds), *The Cambridge Companion to Textual Scholarship*, Cambridge:Cambridge University Press, 2013, 16–41.
- Patrick Sahle, *A Catalog of Digital Scholarly Editions*, 2008, <http://www.digitale-edition.de/vlet-about.html>