TEI BY EXAMPLE



MODULE 6: PRIMARY SOURCES

Ron Van den Branden

Edward Vanhoutte

Melissa Terras

Centre for Scholarly Editing and Document Studies (CTB) , Royal Academy of Dutch Language and Literature, Belgium, Gent, 9 July 2010

Last updated September 2020

Licensed under a Creative Commons Attribution ShareAlike 3.0 License

TABLE OF CONTENTS

3. *Representing* Primary Source Phenomena 1

3.1 Additions and Deletions 1

3.1.1 Simple Additions and Deletions 1

3.1.2 Complex Additions and Deletions 7

3.2 Facsimiles 9

3.3 Damage 10

3. *Representing* Primary Source Phenomena

3.1 Additions and Deletions

In primary sources, prominent traces of the editing process are additions and deletions. Additions may be marked by differing positioning, shifts in hands, ink, or font, and can be explicitly indicated by all kinds of markers. Deletions are often visible as struck out text. Because they may shed light on the writing process of the text or hold alternative readings and interpretations, additions and deletions can be very valuable elements in an electronic transcription.

3.1.1 Simple Additions and Deletions

Our sample text contains a clear addition at the top of the second page. The phrase “the phone box” has probably been added afterwards, as it didn’t fit in the space available. This can be transcribed as an addition with the <add> element (from “addition”).

```
<p xmlns="http://www.tei-c.org/ns/1.0">There seems to be no-one around <add>the phone
 box</add>.</p>
```

Example 2. Encoding of an addition.

Similarly, deletions can be marked with a specific element: <del> (from “deletion”). If we look closely at the same sentence, we see how the author has corrected a writing mistake by striking through a letter. This can be transcribed as follows:

```
<p xmlns="http://www.tei-c.org/ns/1.0">There seem<del>e</del>s to be no-one
 around <add>the phone box</add>.</p>
```

Example 3. Encoding of a deletion.

Notice how additions, just like deletions, can be transcribed at character level, so that they enclose exactly those letters or phrases that have been deleted or added. We can be more precise about how additions and deletions are realised. Actual rendition information can be specified in the global @rend attribute. For example, we could use the @rend attribute to state that the deleted text has been crossed out, and the addition has occurred above

the line, possibly using some kind of formalised expression. For deletions, information like `"strikethrough"` or `"overwritten"` may be sufficient (you can craft your own typology). Additionally, the <add> element has a specific attribute to record the place where the text had been added: @place. It can combine keywords like `"below"` or `"above"` for text added below or above the line; `"bottom"`, `"margin"`, or `"top"` for text added at the bottom, margin, or top of the page; `"opposite"` or `"overleaf"` for additions on the opposite page or at the other side of the page. Our example could be extended as follows:

```
<p xmlns="http://www.tei-c.org/ns/1.0">There seem<del rend="crossout">e</del>s to be no
-one around <add place="above">the phone box</add>.</p>
```

Example 4. Encoding rendition details about additions and deletions.

Often, additions and deletions can be traced by shifts in ink or writing material. As this information can convey useful insights in the writing process, it can be useful to include it in the transcription. Of course, the @rend attribute could be used for this purpose, but <add> and <del> have a more sophisticated mechanism to record aspects of the hand in which they are written: the @hand attribute. This attribute in itself doesn't indicate any specific features directly, but rather holds a reference to a hand description elsewhere in the document. A hand can be defined in a <handNote> element in the <profileDesc> part of the TEI header, which groups different hand definitions inside a <handNotes> element. A <handNote> definition can contain a loose prose description of the hand inside paragraphs, as well as more formalised identifications of different aspects of the hand in specific attributes: @scribe (an name for the scribe), @script (the writing style or font of a hand), @medium (the type of ink), and @scope (the dominance of this hand in the document). In order to make references to such a hand definition elsewhere in the transcription, a unique @xml:id value must be provided. Inside "transcriptional" elements such as <add> and <del>, reference to a hand definition can be made with the @hand attribute. As with all references in TEI, this takes the form of an URI pointer, of which the local part is preceded with a # sign. For this example, more details about the hand could be given as follows:

```
<TEI xmlns="http://www.tei-c.org/ns/1.0">
  <teiHeader>
    <!--...-->
    <profileDesc>
      <!--...-->
```

```
      <handNotes>
        <handNote xml:id="HR" scribe="Hanna
        Renton" script="handwritten" medium="pencil" scope="major">
          <p>the document's main hand, Hanna Renton</p>
        </handNote>
      </handNotes>
    </profileDesc>
  </teiHeader>
  <text>
    <body>
      <!--...-->
      <p>There seem<del rend="crossout" hand="#HR">e</del>s to be no-one
       around <add place="above" hand="#HR">the phone box</add>.</p>
      <!--...-->
    </body>
  </text>
</TEI>
```

Example 5. Providing information about document hands.

Notice how the @hand attribute is the means to distinguish between additions or deletions in a text made by different persons, if they can be distinguished. Our sample text contains other interventions in a different ink, made by a different hand. For example, on page two, text has been added both in the margin and inline, near the original word “breath.” A detailed study of the genesis of this work could identify the person responsible for these additions as the author’s teacher. With proper identification of this hand in the header, this attribution can be recorded in the transcription:

```
<TEI xmlns="http://www.tei-c.org/ns/1.0">
  <teiHeader>
    <!--...-->
    <profileDesc>
      <!--...-->
      <handNotes>
        <handNote xml:id="HR" scribe="Hanna
        Renton" script="handwritten" medium="pencil" scope="major">
          <p>the document's main hand, Hanna Renton</p>
```

```
        </handNote>
        <handNote xml:id="teacher" scribe="class
        Teacher" script="handwritten" medium="red_ballpen" scope="minor">
          <p>the document author's teacher</p>
        </handNote>
      </handNotes>
    </profileDesc>
  </teiHeader>
  <text>
    <body>
      <!--...-->
      <p>Goodness me! I just stepped outside the phone box and I couldn't
       breath<add hand="#teacher">e</add>. <add hand="#teacher" place="margin">"breath"
       = noun <lb/>"to breathe" = verb</add> <!--...--></p>
      <!--...-->
    </body>
  </text>
</TEI>
```

Example 6. Identifying multple hands.

Notice how, although the content of both <add> elements was most probably added in the same addition, they are split in order to capture the different positioning on the page. Another thing of notice, is the slight abstraction that has been made of the actual occurrence of the marginal addition on the page: instead of interrupting the words “couldn’t breath,” the encoder has opted to transcribe the annotation at the end of the sentence. The inline addition, on the other hand, is transcribed where it appears in the original, and therefore not specified with a @place attribute.

Additions and deletions may come in isolation like in the examples above, but often occur in combination, when existing text is deleted and new text is added. Such a case of juxtaposed deletion and addition can be found on the second page of the example, in the word that should read “dioxide.” Apparently, a first version was started correctly, with “diox,” which the author has revised to “diacxside,” by overwriting the original “ox” and adding the “acx,” resulting in the final reading “diacxside.” This can be represented with a simple sequence of <del> and <add>:

```
di<del xmlns="http://www.tei-c.org/ns/1.0" hand="#HR" rend="overwritten">ox</del><add
xmlns="http://www.tei-c.org/ns/1.0" hand="#HR">acx</add>side
```

Example 7. A simple encoding of subsequent deletions and additions.

Such combinations of deletions and deletions can be grouped in a dedicated <subst> (substitution) element, in order to identify them as a single editorial intervention.

```
di<subst xmlns="http://www.tei-c.org/ns/1.0" hand="#HR">
 <del rend="overwritten">ox</del>
 <add>acx</add>
</subst>side
```

Example 8. Grouping subsequent deletions and additions in `<subst>`.

Notice, how the identification of the responsible hand in the @hand attribute has been moved upward to the <subst> element. Because <subst> contains a deletion and addition by the same hand, the hand identification is inherited by the corresponding <del> and <add> elements in the transcription.

It must be acknowledged, that this analysis of the word “dioxide” in the text involves a fair amount of interpretation. Responsibility for these kinds of interpretation can be taken by means of a dedicated @resp (responsibility) attribute. It can occur on <add>, <del>, and <subst> elements, and points to an identified person in the TEI header of an electronic document. In this case, the TBE crew, who edited this electronic text, can be held responsible for this interpretation as follows:

```
<TEI xmlns="http://www.tei-c.org/ns/1.0">
  <teiHeader>
    <fileDesc>
      <titleStmt>
        <title>There and Back Again: digital edition</title>
        <author xml:id="HannaRenton">Hanna Renton</author>
        <editor xml:id="TBE">The TBE crew</editor>
        <!--...-->
      </titleStmt>
      <!--...-->
```

```
    </fileDesc>
    <!--...-->
  </teiHeader>
  <text>
    <body>
      <!--...-->
      <p><!--...-->di<subst resp="#TBE" hand="#HR">
        <del rend="overwritten">ox</del>
        <add>acx</add>
      </subst>side<!--...--></p>
      <!--...-->
    </body>
  </text>
</TEI>
```

Example 9. Stating responsibility for editorial interpretations in `@resp`.

Similarly, for the addition and correction in the first sentence on the first page, the authority can be indicated with @hand, and the responsibility for the identification of this addition with @resp:

```
<p xmlns="http://www.tei-c.org/ns/1.0"><add place="margin" resp="#T
BE" hand="#teacher">Well done! <lb/>8,5 / 10</add>I can't bel<subst resp="#T
BE" hand="#teacher">
  <del rend="strikethrough">ei</del>
  <add>ie</add>
</subst>ve it. <!--...--></p>
```

Example 10. Identifying hands and stating editorial responsibility for other additions.

SUMMARY

Additions and deletions can be encoded with the <add> and <del> elements, respectively. While the @rend attribute can be used to record general visual aspects of their realisation in the source, the <add> element has a specific @place attribute. This can be used to indicate where the addition is located (e.g., `"inline"`, `"above"`

or "below" the line; at the "bottom" or "top" of the page; in the "margin", "overleaf", on the "opposite" page). Specific characteristics of the hand can be encoded by referring to a hand definition in the header, using the @hand attribute, while responsibility for the encoding of additions and deletions can be stated in the @resp attribute, referring to an identified person in the TEI header. Sequences of deletions and additions originating from one single intervention can be wrapped in a <subst> element.

3.1.2 Complex Additions and Deletions

Deletions and additions are not limited to a single "layer" of a document, as in the previous examples. They may as well nest, when, for example, an added fragment itself contains further deletions and/or additions. Take, for example, the fragment on page 2 of the example text, that originally read: "It's a big poster, saying: ...." The author later had added the phrase "at the wall" in the margin, but has later corrected this to "on the wall." This can be encoded as a single addition ("at the wall"), containing a nesting substitution, consisting of a deletion ("at") and an addition ("on"):

```
<p xmlns="http://www.tei-c.org/ns/1.0">What's that? It's a big
 poster <add hand="#HR" place="margin" resp="#TBE"><subst>
  <del rend="crossout">at</del>
  <add place="above">on</add>
</subst> the wall</add>, saying <!-.....--></p>
```

Example 11. Encoding complex additions and deletions.

Besides substitutions, often in hand written texts, an original reading can be restored after being rejected first. The TEI Guidelines provide a specific element for marking such restorations: <restore>. This element can be wrapped around prior deletions. Take, for example, the phrase that should read "Who turned out the lights!!!" on the first page in the example. This had first been substituted for "Who turned off the lights!!!" but has afterwards been restored to the original reading, indicated both by an "OK" marker, and the addition of the original word "out":

```
Who turned <restore xmlns="http://www.tei-c.org/ns/1.0" hand="#HR" resp="#TBE">
 <subst>
  <del rend="crossout">out</del>
  <add place="margin">off</add>
```

```
  </subst>
</restore> the lights!!!
```

Example 12. Encoding a restoration of a previous intervention.

Apart from having internal structure themselves, additions and deletions may overlap with other logical structures of a text, for example when crossing a paragraph boundary, or a phrase that is transcribed as a name or title. Take a look, for example, at page 3 of the sample document, which features two entire sentences being crossed out. This deletion, however, runs over two paragraphs. This could impossibly be encoded with a simple deletion:

```
<p xmlns="http://www.tei-c.org/ns/1.0"><!--...--> <del>But I've got to go outside
 (again). </p>
<p xmlns="http://www.tei-c.org/ns/1.0"> It was a lot easier to find than I thought it
 would be.</del> <!--...--></p>
```

Example 13. Incorrect encoding of a deletion crossing paragraph boundaries.

That is, the markup itself (a paragraph boundary) is involved in the deletion, and can not just be enclosed in another container, which would produce overlapping hierarchies as in the previous incorrect example. Other cases that would result in invalid TEI occur when long deletions or additions that nest properly inside bigger structures encompass text structures that are illegal inside <add> or <del> (such as entire paragraphs or divisions: <add> and <del> can only contain phrase-level elements). In order to facilitate the encoding of such cases (that are to be expected, given the tension between the unedited nature of primary sources and the formalism of the TEI markup vocabulary), the TEI Guidelines provide two specific elements: <delSpan> and <addSpan>. These are empty elements marking the beginning of a longer deletion or addition, respectively. The scope of the addition or deletion is made explicit by means of a specific @spanTo attribute, which points to an identified end point, which comes later in the transcription. This end point can be represented with an empty <anchor> element, which is an all-purpose empty element for identifying a certain point in a text, via its @xml:id attribute. Although they can't contain any text, <addSpan> and <delSpan> can have all attributes of their <add> and <del> counterparts. The deletion in the example document can thus be encoded as follows:

```
<p xmlns="http://www.tei-c.org/ns/1.0"><!--...--> where the city's power source
 is. <delSpan hand="#HR" resp="#TBE" rend="crossout" spanTo="#delEnd"/>But I've got to
 go outside (again).</p>
<p xmlns="http://www.tei-c.org/ns/1.0">It was a lot easier to find than I thought it
 would be. <anchor xml:id="delEnd"/>I got round the corner and <!--...--></p>
```

Example 14. Encoding of boundary-crossing deletions with `<delSpan>`.

SUMMARY

Complex deletions and additions may be represented using nesting <del> and <add> elements. Deleted text that has been restored again can be encoded with a <restore> element. When deletions or additions contain logical structures that cannot be transcribed as valid content of the <del> or <add> elements, or cross structural boundaries that would lead to overlapping hierarchies, these can be represented with empty elements. The <delSpan> and <addSpan> elements can indicate the start of such deletions or additions, respectively, and point to their end point with a @spanTo attribute. The value of this attribute must point towards a following <anchor> element, which is an empty element identifying a point in the document with its @xml:id attribute.

3.2 Facsimiles

The first page of our example text is followed by an unnumbered page containing a drawing. As seen in Module 3: Prose, section 4.2 of this tutorial series, this can be represented with an empty <graphic> element, whose @url attribute points to a digital representation of the image. When the graphic is enclosed in a larger <figure> element, a description of the image can be provided in <figDesc>:

```
<pb xmlns="http://www.tei-c.org/ns/1.0"/>
<figure xmlns="http://www.tei-c.org/ns/1.0">
  <graphic url="phonebox_scan.jpg"/>
  <figDesc>the phone box travelling through time</figDesc>
</figure>
<pb xmlns="http://www.tei-c.org/ns/1.0" n="2"/>
```

Example 15. Encoding a page facsimile in `<graphic>`.

Alternatively, it could be interesting for primary source materials to provide access to digital facsimiles of the entire document, in order to complement the transcription with visual evidence. The TEI Guidelines provide a mechanism to link each element in a digital transcription with a (part of) a facsimile: the global @facs attribute (for "facsimile"). This attribute can be attached to any TEI element (if the **transcr** module is included in the TEI schema), and point to a digital scan by means of an URI. Typically, scans are made page by page (or folio by folio), which makes it most convenient to attach a @facs attribute to the corresponding <pb> (page break) elements in the electronic transcription. If we have digital facsimiles for each page of our example text available under the folder **scans**, whose file names consist of the letters TBA + 3 digits, these facsimiles could be referenced from the transcription as follows:

```
<pb xmlns="http://www.tei-c.org/ns/1.0" facs="scans/TBA001b.jpg"/>
<pb xmlns="http://www.tei-c.org/ns/1.0" n="2" facs="scans/TBA002.jpg"/>
```

Example 16. Linking the encoding to facsimiles with @facs.

Notice how this link makes the encoding of the figure inside <graphic> redundant, so that it can be left out of the transcription. Besides this simple mechanism of pointing to entire images, the TEI provides a more refined system to define specific zones inside facsimiles, that can be associated with specific elements in the transcription. As this is a more advanced topic, you are referred to the detailed discussion in section 11.1 Digital Facsimiles of the TEI Guidelines.

SUMMARY

Digital facsimiles can be referenced from any element with the global @facs attribute. It can hold a URI pointer to a digital scan of a document fragment.

3.3 Damage

Transcription of primary source texts is determined by the state of the source material. Depending on the quality of the source materials and handwriting, some passages may be unclear, which may hamper straightforward transcription. Transcription of such passages may involve higher degrees of interpretation, or even editorial intervention, which may be signalled with appropriate elements that will be discussed further in this module

(see section 4). However, if such interpretative elements are caused by material damage to the source text, this damage can be signalled in the transcription with the <damage> element. Many aspects of the damage can be expressed in attributes, most important of which are:

- @hand: the hand that caused the damage (when it is caused by identifiable human intervention)
- @agent: the cause of the damage
- @extent: the amount of damaged text, expressed in prose descriptions like `"2 words"`, `"3 letters"`
- @quantity: the length of the damage in a specific unit (specified with the @unit attribute)
- @unit: the unit in which the length of the damage is expressed (with the @quantity attribute), e.g., `"cm"` (centimetres), `"chars"` (text characters)
- @type: a characterisation of the type of the damage

One important condition for the use of the <damage> element, is that it should contain text that is more or less legible, or can be reconstructed. When no further (interpretative) claims about this text are made, this text can be enclosed as such inside the <damage> element. For example, in our example text, the stapling of the sheets of the writing assignment has caused damage to the top left part of the document, rendering the date line partly illegible. This can be recorded in the transcription:

```
<body xmlns="http://www.tei-c.org/ns/1.0">
  <pb n="1"/>
  <dateline>
    <date when="2008-08
    -26"><damage agent="stapling" hand="#teacher" unit="chars" quantity="3">26/</damage>8
    /08</date>
  </dateline>
  <head>There and Back Again</head>
  <!--...-->
</body>
```

Example 17. Signalling damage in the source with `<damage>`.

Similar to the <del> and <add> elements, <damage> has an empty counterpart: <damageSpan>, that should be used when damage runs over different structural boundaries, or contains document structures too large to be valid inside <damage>. It has the same attributes as <damage>, as well as an extra attribute @spanTo, whose value points to a following identified element in the transcription.

SUMMARY

As it may influence the transcription and (degree of) interpretation of the source text, damage can be transcribed as a primary source phenomenon in its own right with the <damage> element. It can specify several characteristics associated with the damage, such as the responsible hand (if any) in the @hand attribute, the cause of the damage in the @agent attribute, and a classification of the kind of damage in the @type attribute. The extent of the damage can either be given implicitly in an @extent attribute, or explicitly with a combination of the @quantity and @unit attributes, recording the number of measured units, respectively. Damage crossing logical structures or encompassing large text structures can be encoded using an empty <damageSpan> element, whose @spanTo attribute can point to the end point of the damage identified further in the transcription.