

# 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

|                                          |   |
|------------------------------------------|---|
| 4. Editorial Interventions.....          | 1 |
| 4.1 Unclear, Supplied, Omitted Text..... | 1 |
| 4.2 Corrections.....                     | 5 |



## 4. Editorial Interventions

### 4.1 Unclear, Supplied, Omitted Text

Depending on the quality of the source material or the handwriting, transcription of primary source texts may be more or less straightforward. As any further interpretation of an electronic transcription depends on this first interpretative act, it may be desirable for an encoder to indicate places of uncertainty, either for further inspection or to take intellectual responsibility. Text for which the reading is uncertain can be encoded in an `<unclear>` element. The reason for the unclear reading can be stated in a `@reason` attribute, which takes either a single keyword, or a white space separated list of keywords. If the legibility is affected by damage, the cause of the damage can be described in the `@agent` attribute. For example, as our previous transcription of the word “dioxide” as “diacxside” was quite uncertain, this could be indicated with the `<unclear>` element:

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

```
</body>
</text>
</TEI>
```

**Example 18. Signalling unclear text with `<unclear>`.**

Similarly, if we decided that the damaged dateline at the start of the document could still be deciphered, the uncertain status of this part of the text could be indicated with an `<unclear>` element:

```
<dateline xmlns="http://www.tei-c.org/ns/1.0">
  <date when="2008-08
-26"><damage agent="stapling" hand="#teacher" unit="chars" quantity="3">
    <unclear>26/</unclear>
  </damage>8/08</date>
</dateline>
```

**Example 19. Combining `<damage>` with `<unclear>`.**

...or without the `<damage>` element, if this is deemed less important to the transcription:

```
<dateline xmlns="http://www.tei-c.org/ns/1.0">
  <date when="2008-08
-26"><unclear reason="damage" agent="stapling" unit="chars" quantity="3">26/</unclear>8
/08</date>
</dateline>
```

**Example 20. Indicating damage in a `@reason` attribute on `<unclear>`.**

Notice, however, that the use of the `<unclear>` element implies that the text it encloses must still be present in the document source, and still be legible to some degree. If the encoder considers text too unclear to be transcribed in any way, he or she may opt to omit this part of the text, and indicate this editorial intervention with a `<gap>` element. This is an empty element, whose sole purpose is to indicate the omission, possibly with

characterisation of the reason (@[reason](#)), or the cause of the damage causing this omission, if any (@[agent](#)). As with the [<damage>](#) element, the extent of the omission can be specified implicitly with the @[extent](#) attribute, or more explicitly by combining the @[unit](#) and @[quantity](#) attributes.

For example, on page 3 of our sample text, the phrase “Yess!! The door!!!” is followed by some words that can’t all be deciphered confidently, as they appear to have been erased by the author. When transcribing this passage, we could opt to encode an informed guess and mark it with the [<unclear>](#) element, while leaving out the truly illegible words. This omission can be marked with [<gap>](#):

```
Yess!! The door!!! <unclear xmlns="http://www.tei-c.org/
ns/1.0" reason="erasure" resp="#TBE"><gap unit="cm" quantity="2.5"/> I got
out.</unclear>
```

**Example 21. Omitting illegible text with [<gap>](#).**

Similarly, the damage in the dateline could be deemed too destructive for a confident reading of the day, which may motivate the encoder to leave it out. In this case, too, a [<gap>](#) element can be used, either within or without a surrounding [<damage>](#) element:

```
<dateline xmlns="http://www.tei-c.org/ns/1.0">
  <date when="2008-08-26"><damage agent="stapling" unit="chars" quantity="3">
    <gap unit="chars" quantity="3"/>
  </damage>8/08</date>
</dateline>
```

**Example 22. Omitting illegible text in a damaged region.**

In contrast, the editor may wish to make a stronger intervention, by supplying text that is lacking from or illegible in the document source. This can be done by wrapping the added text in a [<supplied>](#) element. In a @[reason](#) attribute, the reason for this editorial addition can be given. For the dateline example, if the text is considered illegible, but the encoder feels able to reconstruct the date, this can result in following encoding:

```
<dateline xmlns="http://www.tei-c.org/ns/1.0">
  <date when="2008-08-26"><damage agent="stapling" unit="chars" quantity="3">
```

```
<supplied resp="#TBE">26/</supplied>
</damage>8/08</date>
</dateline>
```

**Example 23. Encoding editorial additions with <supplied>.**

This allows us to encode other lacking text as well: at the end of page 2, a couple of final words on some lines are incomplete due to xeroxing. These can be reconstructed fairly straightforwardly for the transcription. However, these reconstructions are best signalled with the <supplied> element:<sup>1</sup>

```
<p xmlns="http://www.tei-c.org/ns/1.0"><!--...--> Nothing was
happeni<supplied reason="cutoff-while-xeroxing" resp="#TBE">ng.</supplied>
Then, <quote>Hello, dearie</quote>, an old woman answe<supplied reason="cutoff-while
-xeroxing" resp="#TBE">d.</supplied> <quote>Who is it?</quote> I hung up. Uh, oh. How
do I get back then?</p>
```

**Example 24. Providing a reason for an editorial addition, with @reason on <supplied>.**

## SUMMARY

When text in the source document is still partly legible, but needs interpretation in order to be transcribed, this uncertainty can be expressed by enclosing the text in an <unclear> element. If text is deemed totally illegible, it can be omitted from the transcription, but signalled with a <gap> element (without any content). Both elements can indicate the reason for the editorial intervention (@reason), and the nature of the damage (@agent). An editor wishing to supply text in the electronic transcription for illegible or lacking text in the source text, can encode this supplied text with the <supplied> element. In a @reason attribute, the reason for this intervention can be stated.

.....

- 1 Notice, the crucial difference between the encoding of text added or deleted by the author or editor of the *source document* on the one hand, and by the encoder of the *electronic* transcription on the other hand. Additions or deletions present in the source may only be encoded respectively as <add> and <del>, while text that has been added or deleted by editorial emendation must be encoded as <supplied> or <gap>, respectively.

## 4.2 Corrections

If we look back at the comparison between the facsimiles (see [figure 1](#)) and the initial transcription (see [example 1](#)), we notice that a lot of words have been silently corrected by the transcriber. Although some errors had been corrected by the teacher (who can be considered an editor or corrector of the source document), many have slipped through. Depending on the aim of the transcription, such apparent errors may be transcribed unmediatedly, corrected silently, marked explicitly, or corrected explicitly. All of these practices are perfectly legitimate as long as they are applied consistently and motivated in the `<editorialDecl>` element of the electronic document's header. An encoder adhering to a more explicit practice would like to at least signal apparent errors, editorial corrections, or both. The TEI provides specific elements for this purpose: `<sic>`, for indicating apparent errors, and `<corr>`, for indicating editorial corrections. For example, the sentence “Now I know what all the black air is: all the polution.” on page 2 could be transcribed as follows:

```
Now I know what all the black air is: all the <sic xmlns="http://www.tei-c.org/ns/1.0">polution</sic>.
```

**Example 25. Encoding an apparent error with `<sic>`.**

...if the encoder would be interested in transcribing the source text as accurately as possible, or:

```
Now I know what all the black air is: all the <corr xmlns="http://www.tei-c.org/ns/1.0">pollution</corr>.
```

**Example 26. Encoding the correction of an apparent error with `<corr>`.**

...if the content matters most (perhaps to ease searching operations in a digital edition of the text). However, both “views” on the text can be combined in a `<choice>` element.<sup>3</sup> This enables an encoder to express alternative encodings of the same text. Both views could thus be combined as:

```
Now I know what all the black air is: all the <choice xmlns="http://www.tei-c.org/ns/1.0">
  <sic>polution</sic>
  <corr>pollution</corr>
</choice>.
```

**Example 27. Combining errors and corrections in `<choice>`.**

The `<sic>` element may contain all elements that are necessary to represent the original source text, like deletions, damage, and so on. The “diacxside” fragment on page two can thus be corrected as follows:

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

.....

- 3 Again, a crucial distinction must be pointed out between encoding of corrections present in the *source document*, and editorial corrections in the *electronic transcription*. The latter must always be encoded using `<sic>` and `<corr>`, possibly wrapped in a `<choice>` element. Corrections present in the source document must be encoded using combinations of `<del>` and `<add>`, possibly grouped in a `<subst>` element, and preferably specified with attributes identifying the responsible document hand (`@hand`), and the editor responsible for this identification (`@resp`).

**Example 28. Encoding authorial phenomena inside <sic>.**

## SUMMARY

Apparent errors in the source text may be indicated explicitly in a [<sic>](#) element, or corrected with a [<corr>](#) element. Both the original and the correction can be included in the transcription, if they are wrapped in a [<choice>](#) element.