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

1. Introduction....1
2. Primary Source Material....2
3. *Representing* Primary Source Phenomena....7
3.1 Additions and Deletions....7
3.1.1 Simple Additions and Deletions....7
3.1.2 Complex Additions and Deletions....13
3.2 Facsimiles....16
3.3 Damage....17
4. Editorial Interventions....19
4.1 Unclear, Supplied, Omitted Text....19
4.2 Corrections....23
5. Summary....25
6. What's Next?....30

1. Introduction

Texts exist in many genres and forms, each with their particular structural and semantic features. Besides structural characteristics, texts can be roughly distinguished for their "editorial status." Typically, the majority of conserved documents have a "public" status: after scrupulous authoring and editing by an author and/or editorial instance, they have been published and multiplied, either as manuscript, in print or nowadays in electronic form. Still, the textual universe is wider than these published documents. Lots of texts were never intended to be published, because they have a private nature (letters, ego documents), were not considered "final" documents (but may have survived their published successors), were conceived as exclusive pieces of art,... Often, such texts are of great value, either because no other representations exist (anymore), because they reflect stages in the conception of a published literary work, because they are evidence of historical language use, or for many other reasons to many other types of research.

Because non-published texts typically are less editorially polished, they can contain many traces of the authoring or editing phase. Frequent phenomena in primary source materials are additions, deletions, restorations, errors, corrections,... Moreover, the condition of the material that carries the text may influence the transcription: damage may render a fragment illegible or incomplete. The TEI Guidelines offer specific elements to cover such phenomena in transcriptions. Notice, however, that these phenomena are not confined to handwritten texts: analogous phenomena may occur in typewritten documents, born-digital texts that have been printed for revision, or digital texts that include some form of electronic revision control information. Although many of the TEI elements discussed in this tutorial module are available in all TEI texts, some specific ones require inclusion of the dedicated **`transcr`** TEI module in your TEI schema, documented in chapter 11. Representation of Primary Sources of the TEI Guidelines.[1]

This tutorial focuses on transcription of *phenomena in primary sources* rather than *manuscripts*, because the latter term is often understood with a specific connotation: texts written by hand before the print age. Those have their own highly specialised textual phenomena and cataloguing needs, to which a specific chapter of the TEI Guidelines is devoted (10: Manuscript Description). Instead, this module will focus on the transcription of a

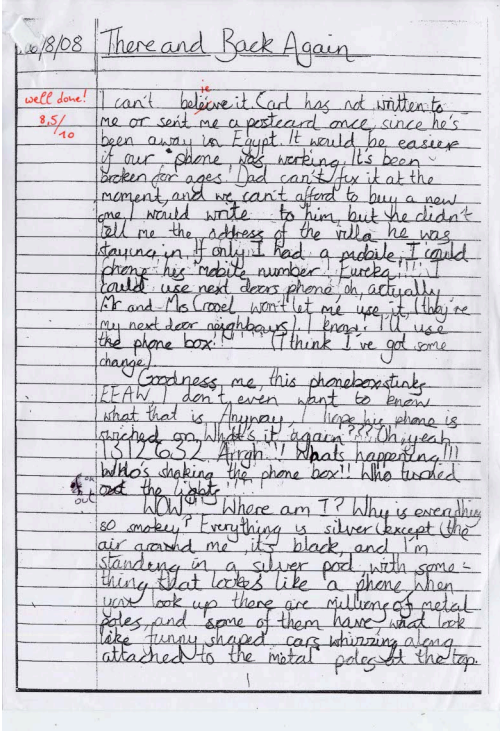
.....

1 For directions on composing a TEI schema by selecting TEI modules and elements, see Module 8: Customising TEI, ODD, Roma.

handwritten prose text. Notice, also, that this tutorial does not cover the advanced mechanisms for combining transcription with facsimiles, as discussed in section 11.2 Combining Transcription with Facsimile of the TEI Guidelines.

2. Primary Source Material

The text we use as an example throughout this module is a fragment of "There and Back Again," a story by the promising young talent Hannah Renton, in response to a writing assignment at school.



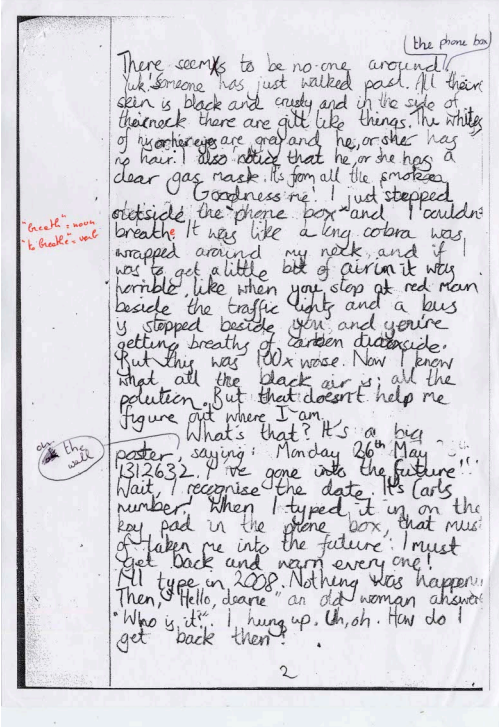
6/8/08 There and Back Again

well done! 8.5/10

I can't believe it. Carl has not written to me or sent me a postcard once, since he's been away in Egypt. It would be easier if our phone was working. It's been broken for ages. Dad can't fix it at the moment, and we can't afford to buy a new one. I would write to him, but he didn't tell me the address of the villa he was staying in. If only I had a mobile, I could phone his mobile number. Eureka!!! I could use next doors phone, oh, actually Mr and Mrs Crowel won't let me use it, (they're my next door neighbours). I know, I'll use the phone box!! (I think I've got some change).

Goodness me, this phonebox stinks EEAW. I don't even want to know what that is. Anyway, I hope his phone is switched on. What's it again?? Oh, yeah 1312632. Argh!! Whats happening!!! Who's shaking the phone box!! Who turned out the lights!!!

WOW!! Where am I? Why is everything so smokey? Everything is silver (except the air around me, it's black, and I'm standing in a silver pod, with something that looks like a phone. When you look up there are millions of metal poles, and some of them have, what look like funny shaped cars whizzing along attached to the metal poles at the top.



(the phone box)

There seems to be no-one around!
Yuk! Someone has just walked past. All their skin is black and crusty and in the side of their neck there are gill like things. The whites of his or her eyes are gray and he, or she has no hair. I also notice that he, or she has a clear gas mask. It's from all the smoke.

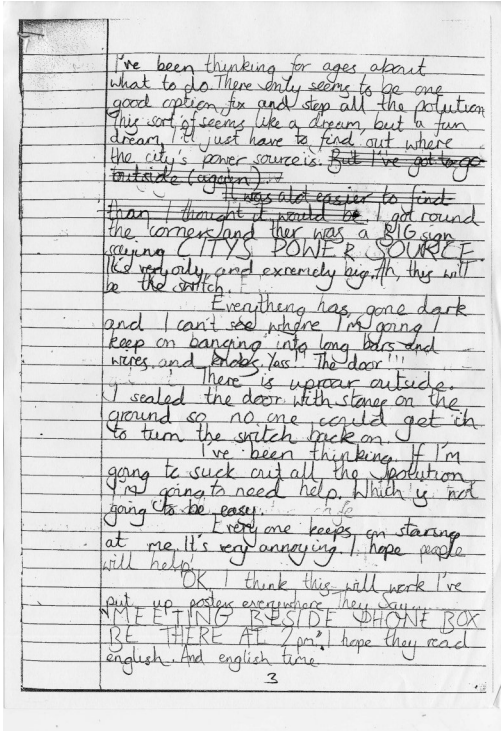
Goodness me! I just stepped outside the "phone box" and I couldn't breathe. It was like a long cobra was wrapped around my neck, and if I was to get a little bit of air in it was horrible, like when you stop at red man beside the traffic lights and a bus is stopped beside you and you're getting breaths of carbon dioxide. But this was 100x worse. Now I know what all the black air is; all the polution. But that doesn't help me figure out where I am.

"breath" = noun
"to breathe" = verb

What's that? It's a big poster, saying: Monday 26th May 1312632. I've gone into the future!! Wait, I recognise the date. It's Carls number! When I typed it in on the key pad in the phone box, that must of taken me into the future! I must get back and warn every one! I'll type in 2008. Nothing was happening. Then, "Hello, dearie" an old woman answers "Who is it?". I hung up. Uh, oh. How do I get back then?

oh the wall

2



I've been thinking for ages about what to do. There only seems to be one good option, fix and stop all the pollution. This sort of seems like a dream, but a fun dream, I'll just have to find out where the city's power source is. ~~But I've got to go outside (again).~~

~~It was alot easier to find than I thought it would be.~~ I got round the corners and ther was a BIG sign saying CITYS POWER SOURCE. It's very oily and exremely big. Ah, this will be the switch.

Everything has gone dark and I can't see where I'm going I keep on banging into long bars and wires and knobs. Yass!! The door!!!

There is uproar outside. I sealed the door with stones on the ground so no one could get in to turn the switch back on.

I've been thinking. If I'm going to suck out all the pollution I'm going to need help. Which is not going to be easy.

Every one keeps on staring at me. It's very annoying. I hope people will help!

OK, I think this will work I've put up posters everywhere they say... "MEETING BESIDE PHONE BOX BE THERE AT 7 pm". I hope they read english. And english time.

3

Figure 1. Some pages from a sample manuscript.

A first look at the document facsimiles above reveals that it is a prose document, with as most prominent structural features: a date, title, a graphic, and paragraphs. It could be transcribed as follows (for more details of prose transcription, see Module 3: Prose):

```
<body xmlns="http://www.tei-c.org/ns/1.0">
  <pb n="1"/>
  <dateline>
    <date when="2008-08-26">26/8/08</date>
  </dateline>
  <head>There and Back Again</head>
  <p>I can't believe it. Carl has not written to me or sent me a postcard once since he
  's been away in Egypt. It would be easier if our phone was working. It's been broken
   for ages! Dad can't fix it at the moment, and we can't afford to buy a new one. I
   would write to him, but he didn't tell me the address of the villa he was staying
   in. If only I had a mobile, I could phone his mobile number. Eureka!!! I could use
```

```
 next door's phone, oh, actually, Mr and Mrs Crooel won't let me use it (they're
 my next door neighbours). I know. I'll use the phone box!!! (I think I've got some
 change).</p>
<p>Goodness me, this phone box stinks. EEAW, I don't even want to know what that
 is. Anyway, I hope his phone is switched on. What's it again??? Oh, yeah, 1312632.
 Arrgh!! What's happening!!! Who's shaking the phone box!! Who turned out the
 lights!!!</p>
<p>WOW!!! Where am I? Why is everything so smokey? Everything is silver except the
 air around me, it's black, and I'm standing in a silver pod, with something that
 looks like a phone. When you look up there are millions of metal poles, and some
 of them have what look like funny shaped cars whizzing along attached to the metal
 poles at the top.</p>
<pb/>
<figure>
  <graphic url="phonebox_scan.jpg"/>
  <figDesc>the phone box travelling through time</figDesc>
</figure>
<pb n="2"/>
<p>There seems to be no-one around the phone box. Yuk! Someone has just walked past.
 All their skin is black and crusty and in the side of their neck are gitt like
 things. The whites of his or her eyes are grey and he, or she has no hair. I also
 notice that he, or she has a clear gas mask. It's from all the smoke.</p>
<p>Goodness me! I just stepped outside the phone box and I couldn't breathe. It was
 like a king cobra was wrapped around my neck and if I was to get a little bit of air
 in it was horrible, like when you stop at red man beside the traffic lights and a
 bus is stopped beside you and you're getting breaths of carbon dioxide. But this was
 100x worse. Now I know what all the black air is: all the pollution. But that doesn
't help me figure out where I am.</p>
<p>What's that? It's a big poster, saying: Monday 26th May 1312632. I've gone
 into the future!!! Wait, I recognise the date. It's Carl's number! When I typed
 it in on the key pad in the phone box, that must have taken me into the future:
 I must get back and warn everyone! I'll type in 2008. Nothing was happening.
 Then, <quote>Hello, dearie</quote>, an old woman answered. <quote>Who is it?</quote>
 I hung up. Uh, oh. How do I get back then?</p>
<pb n="3"/>
<p>I've been thinking for ages about what to do. There only seems to be one good
 option, fix and stop all the pollution. This sort of seems like a dream, but a fun
 dream. I'll just have to find out where the city's power source is.</p>
```

```
  <p>I got round the corner and there was a BIG sign saying CITY'S POWER SOURCE. It's
   very oily and extremely big. Ah, this will be the switch.</p>
  <p>Everything has gone dark and I can't see where I'm going. I keep on banging into
   long bars and wires and knobs. Yess!! The door!!!</p>
  <p>There is uproar outside. I sealed the door with stones on the ground so no one
   could get in to turn the switch back on.</p>
  <p>I've been thinking. If I'm going to suck out all the pollution I'm going to need
   help. Which is not going to be easy.</p>
  <p>Everyone keeps on staring at me. It's very annoying. I hope people will help!</p>
  <p>OK. I think this will work. I've put up posters everywhere. They say... <quote>M
  EETING BESIDE PHONE BOX BE THERE AT 2pm</quote>. I hope they read English. And
   English time.</p>
</body>
```

Example 1. A “normalised” encoding of the sample manuscript.

...or *could* it? When you compare the facsimiles and above transcription, you'll notice that a lot of phenomena of this particular hand written text have been filtered out or abstracted in the transcription. Mind you, it *could* be a plausible transcription that strives for a representation of the contents, rather than a faithful record of its actual realisation, provided these editorial choices are stated explicitly in the header's <editorialDecl> element.

CHALLENGE

Take a close look at the document facsimiles above, compare it to the transcription above, and make a list of things you observe that are specific to the non-published character of this document.

SOLUTION

- Additions
- Deletions
- Different hands
- Colour of ink
- Notes

- Errors
- Corrections
- Damage
- Unclear text
- Lacking text

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.

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>
```

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:[6]

```
<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 answere<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>
```

.....

6 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.

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.

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.[8] 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>
```

.....

8 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.

5. Summary

This tutorial module has focused on the encoding of specific phenomena of primary source texts in TEI. When all of the concepts discussed are applied to the example text, this is how its transcription could look:

```
<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>
    <!--...-->
    <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>
    <pb n="1" facs="scans/TBA001.jpg"/>
    <dateline>
      <date when="2008-08
      -26"><damage agent="stapling" hand="#teacher" unit="chars" quantity="3">
        <supplied resp="#TBE">26/</supplied>
      </damage>8/08</date>
    </dateline>
    <head>There and Back Again</head>
    <p><add place="margin" resp="#TBE" hand="#teacher">Well done! <lb/>8,5 / 10</add>I
     can't bel<subst hand="#teacher" resp="#TBE">
      <del rend="strikethrough">ei</del>
      <add>ie</add>
    </subst>ve it. Carl has not written to me or sent me a postcard once since he's
     been away in Egypt. It would be easier if our phone was working. It's been broken
     for ages! Dad can't fix it at the moment, and we can't afford to buy a new one. I
     would write to him, but he didn't tell me the address of the villa he was staying
     in. If only I had a mobile, I could phone his mobile number. Eureka!!! I could
     use next <choice>
      <sic>doors</sic>
      <corr>door's</corr>
    </choice> phone, oh, actually, Mr and Mrs Crooel won't let me use it (they're my
     next door neighbours). I know! I'll use the phone box!!! (I think I've got some
     change).</p>
    <p>Goodness me, this phone box stinks. EEAW, I don't even want to know what that
     is. Anyway, I hope his phone is switched on. What's it again??? Oh, yeah 1312632.
     Arrgh!! <choice>
      <sic>W<subst hand="#HR" resp="#TBE">
        <del rend="overwritten">a</del>
        <add>h</add>
      </subst>ats</sic>
      <corr>What's</corr>
```

```
</choice> happening!!! Who's shaking the phone box!! Who
 turned <restore hand="#HR" resp="#TBE">
  <subst>
    <del rend="crossout">out</del>
    <add place="margin">off</add>
  </subst>
</restore> the lights!!!!</p>
<p>WOW!!! Where am I? Why is everything so <choice>
  <sic>smokey</sic>
  <corr>smoky</corr>
</choice>? Everything is silver except the air around me, it's black, and I'm
 standing in a silver pod, with something that <choice>
  <sic>lookes</sic>
  <corr>looks</corr>
</choice> like a phone. When you look up there are millions of metal poles, and
 some of them have w<subst hand="#HR" resp="#TBE">
  <del rend="overwritten">a</del>
  <add>h</add>
</subst>at look like funny shaped cars whizzing along attached to the metal poles
 at the top.</p>
<pb facs="scans/TBA001b.jpg"/>
<pb n="2" facs="scans/TBA002.jpg"/>
<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>.
 Yuk! <subst hand="#HR" resp="#TBE">
  <del rend="writeover">s</del>
  <add>S</add>
</subst>omeone has just walked past. All their skin is black and crusty and in
 the side of their neck are <unclear reason="illegible" resp="#TBE">gitt</unclear>
 like things. The whites of his or her eyes are grey and he, or she has no hair. I
 also notice that he, or she has a clear gas mask. It's from all the smoke.</p>
<p>Goodness me! I just stepped outside the phone box and I
 couldn'<unclear reason="cutoff while xeroxing" resp="#TBE">t</unclear>
 breath<add hand="#teacher">e</add>. <add hand="#teacher" place="margin">"breath"
 = noun <lb/> "to breathe" = verb</add> It was like a <choice>
  <sic>kng</sic>
  <corr>king</corr>
```

```
</choice> cobra was wrapped around my neck and if I was to get a little bit of air
 in it was horrible, like when you stop at red man beside the traffic lights and a
 bus is stopped beside you and you're getting breaths of <choice>
  <sic>carben di<subst resp="#TBE">
    <del hand="#HR" rend="overwritten">ox</del>
    <add hand="#HR">acx</add>
  </subst>side</sic>
  <corr>carbon dioxide</corr>
</choice>. But this was 100x worse. Now I know what all the black air is: all
 the <choice>
  <sic>polution</sic>
  <corr>pollution</corr>
</choice>. But that doesn't help me figure out where I am.</p>
<p>What's that? It's a big poster <add hand="#HR" place="margin" resp="#T
BE"><subst>
  <del rend="crossout">at</del>
  <add place="above">on</add>
</subst> the wall</add>, saying: Monday 26th May 1312632. I've gone into the
 future!!! Wait, I recognise the date. It's <choice>
  <sic>Carls</sic>
  <corr>Carl's</corr>
</choice> number! When I typed it in on the key pad in the phone box, that
 must <choice>
  <sic>of</sic>
  <corr>have</corr>
</choice> taken me into the future: I must get back and warn everyone!
 I'll type in 2008. Nothing was happeni<supplied reason="cutoff while
 xeroxing" resp="#TBE">ng.</supplied> Then, <quote>Hello, dearie</quote>,
 an old woman answere<supplied reason="cutoff while xeroxing" resp="#T
BE">d.</supplied> <quote>Who is it?</quote> I hung up. Uh, oh. How do I get back
 then?</p>
<pb n="3" facs="scans/TBA003.jpg"/>
<p>I've been thinking for ages about what to do. There only seems to be one good
 option, fix and stop all the <choice>
  <sic>polution</sic>
  <corr>pollution</corr>
```

```
</choice>. This sort of seems like a dream, but a fun dream. I'll just have
 to find out where the city's power source is. <delSpan hand="#HR" resp="#T
BE" rend="crossout" spanTo="#delEnd"/> But I've got to go outside (again).</p>
<p>It was <choice>
  <sic>alot</sic>
  <corr>a lot</corr>
</choice> easier to find than I thought it would be. <anchor xml:id="delEnd"/> I
 got round the corner and <choice>
  <sic>ther</sic>
  <corr>there</corr>
</choice> was a BIG sign saying <choice>
  <sic>CITYS</sic>
  <corr>CITY'S</corr>
</choice> POWER SOURCE. It's very oily and extremely big. Ah, this will be the
 switch.<gap reason="erasure" unit="cm" quantity="2" resp="#TBE"/></p>
<p>Everything has gone dark and I can't see where I'm going. I
 keep on banging into long bars and wires and knobs. Yess!! The
 door!!! <unclear reason="erasure" resp="#TBE"><gap unit="cm" quantity="2.5"/>I
 got out.</unclear></p>
<p>There is uproar outside. I sealed the door with stones on the ground so no one
 could get <subst hand="#HR" resp="#TBE">
  <del rend="overwritten">it</del>
  <add>in</add>
</subst> to turn the switch back on.</p>
<p>I've been thinking. If I'm going to suck out all the <choice>
  <sic>polution</sic>
  <corr>pollution</corr>
</choice> I'm going to need help. Which is not going to be
 easy. <unclear reason="erasure" resp="#TBE"><gap unit="cm" quantity="2"/>
 cafe</unclear></p>
<p>Everyone keeps on staring at me. It's very annoying. I hope people will
 help!</p>
<p>OK. I think this will work. I've put up posters everywhere. They
 say... <quote>MEETING BESIDE PHONE BOX BE THERE AT 2pm</quote>. I hope they
 read <choice>
  <sic>english</sic>
  <corr>English</corr>
</choice>. And <choice>
```

```
        <sic>english</sic>
        <corr>English</corr>
      </choice> time.</p>
    </body>
  </text>
</TEI>
```

Example 29. A fully encoded transcription of the example text.

6. What's Next?

You have reached the end of this tutorial module covering the markup of primary source materials with TEI. You can now either

- proceed with other TEI by Example modules
- have a look at the examples section for the primary sources module.
- take an interactive test. This comes in the form of a set of multiple choice questions, each providing a number of possible answers. Throughout the quiz, your score is recorded and feedback is offered about right *and* wrong choices. Can you score 100%? Test it here!