TEI by Example. Module 7

Edward Vanhoutte
Ron Van den Branden
Melissa Terras

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

Last modified on: 2018-04-28
Licensed under a Creative Commons Attribution ShareAlike 3.0 License
TABLE OF CONTENTS

5. Encoding Variation in Texts ........................................................................................................ 1
   5.1 The Location-Referenced Method .................................................................................. 2
   5.2 The Double-End-Point-Attached Method ...................................................................... 9
   5.3 The Parallel Segmentation Method .............................................................................. 16
5. Encoding Variation in Texts

After this discussion of encoding textual variants themselves, it is time to have a look at the bigger picture: how do you integrate these variants into an electronic critical edition? The TEI Guidelines provide 3 different mechanisms for integrating apparatus entries in the encoding of texts (don't let the names intimidate you):

- **location-referenced method**: apparatus entries are linked to the identified text blocks in a base text that contain the respective lemmas [I, E]
- **double-end-point-attached method**: apparatus entries are linked to explicitly identified start and end positions in a base text [I, E]
- **parallel segmentation method**: apparatus entries are encoded inside a transcription of the common (invariant) text of all text witnesses [I]

In this overview, the [I] and [E] labels indicate where an apparatus encoded with that method can be physically located with regards to the transcription of the (base) text it is linked to:

- [E]: external apparatus: the apparatus is located outside the transcription of a base text, either in some other part of the TEI document containing the transcription, or in a physically distinct document
  # location-referenced, double-end-point-attached
- [I]: internal apparatus: each apparatus entry is located inline in the transcription of a (base) text, at the place where the variant occurs
  # location-referenced, double-end-point-attached, parallel segmentation

The method chosen and the physical location of the apparatus must be encoded in the TEI Header, in the `<variantEncoding/>` element inside the `<encodingDesc>` section. This is an empty element with two mandatory attributes (see TBE Module 2: The TEI Header -- The Variant Encoding):

- **@method**: indicates the method of linking the critical apparatus to the text: either location-referenced, double-end-point, or parallel-segmentation.
- **@location**: indicates the location of the critical apparatus with regards to the text: either external or internal.

**Summary**

The TEI Guidelines offer 3 methods for linking the critical apparatus to the text. The chosen method must be documented in the `<encodingDesc>` section of the TEI header, in a special `<variantEncoding/>` element. This is an empty element with 2 mandatory attributes. The `@method` attribute specifies the method of linking the apparatus to the text (either location-referenced, double-end-point, or parallel-segmentation). The `@location` attribute specifies the location of the apparatus relative to the text (either external or internal).
5.1 The Location-Referenced Method

The location-referenced method links an apparatus entry to a base text, by anchoring it to the text structure in the base text that contains the variant. This can be done either internally (inside the running text), or externally (outside the running text).

In an internal location-referenced apparatus, the apparatus entries are encoded within the text structures in which the variants occur. The exact location, however, is unimportant. For example, the second paragraph could be encoded as follows:

5.1 The Location-Referenced Method
The encoding scheme defined by these Guidelines is formulated as an application of the Extensible Markup Language (XML) (Bray et al. (eds.) (2006)). XML is widely used for the definition of device-independent, system-independent methods of storing and processing texts in electronic form. It is now also the interchange and communication format used by many applications on the World Wide Web. In the present chapter we informally introduce some of its basic concepts and attempt to explain to the reader encountering them for the first time how and why they are used in the TEI scheme. More detailed technical accounts of TEI practice in this respect are provided in chapters 23. Using the TEI, 1. The TEI Infrastructure, and 22. Documentation Elements of these Guidelines.

5. Encoding Variation in Texts
SGML is an international standard of the more recently developed W3C Extensible Markup Language (XML). Both SGML and XML are widely-used.

This chapter presents a brief tutorial guide to its main features, for those readers who have not encountered it before. For a more technical account of TEI practice see chapter 28, "Conformance," on page 727. For a more technical description of the subset of SGML see chapter 28, "Conformance," on page 727. For a more technical description of the subset of SGML see chapter 30, "TEI Conformance," [in separate fascicle]; for a more technical description of the subset of SGML see chapter 30, "TEI Conformance," [in separate fascicle]; for a more technical description of the subset of SGML see chapter 30, "TEI Conformance," [in separate fascicle]; for a more technical description of the subset of SGML see chapter 30, "TEI Conformance," [in separate fascicle]; for a more technical description of the subset of SGML see chapter 30, "TEI Conformance," [in separate fascicle]; for a more technical description of the subset of SGML see chapter 30, "TEI Conformance," [in separate fascicle]; for a more technical description of the subset of SGML see chapter 30, "TEI Conformance," [in separate fascicle]; for a more technical description of the subset of SGML see chapter 30, "TEI Conformance," [in separate fascicle]; for a more technical description of the subset of SGML see chapter 30, "TEI Conformance," [in separate fascicle]; for a more technical description of the subset of SGML see chapter 30, "TEI Conformance," [in separate fascicle]; for a more technical description of the subset of SGML see chapter 30, "TEI Conformance," [in separate fascicle]; for a more technical description of the subset of SGML see chapter 30, "TEI Conformance," [in separate fascicle]; for a more technical description of the subset of SGML see chapter 30, "TEI Conformance," [in separate fascicle]; for a more technical description of the subset of SGML see chapter 30, "TEI Conformance," [in separate fascicle]; for a more technical description of the subset of SGML see chapter 30, "TEI Conformance," [in separate fascicle]; for a more technical description of the subset of SGML see chapter 30, "TEI Conformance," [in separate fascicle]; for a more technical description of the subset of SGML see chapter 30, "TEI Conformance," [in separate fascicle]; for a more technical description of the subset of SGML see chapter 30, "TEI Conformance," [in separate fascicle]; for a more technical description of the subset of SGML see chapter 30, "TEI Conformance," [in separate fascicle]; for a more technical description of the subset of SGML see chapter 30, "TEI Conformance," [in separate fascicle]; for a more technical description of the subset of SGML see chapter 30, "TEI Conformance," [in separate fascicle]; for a more technical description of the subset of SGML see chapter 30, "TEI Conformance," [in separate fascicle]; for a more technical description of the subset of SGML see chapter 30, "TEI Conformance," [in separate fascicle]
Note how the apparatus entries can occur anywhere as long as it is inside the text structure (in this case, the \texttt{<p>} element) that contains their variants. The same method can be used for an external apparatus, in which the textual variants are encoded either at a different place inside the base text, or in a physically distinct TEI document. In this external apparatus, each apparatus entry must have a specific attribute: \texttt{@loc}. Its value should refer to the canonical reference of the text structure that contains the variants concerned. In an external apparatus, the previous example could look as follows:
The encoding scheme defined by these Guidelines is formulated as an application of the Extensible Markup Language (XML) (Bray et al. (eds.) (2006)). XML is widely used for the definition of device-independent, system-independent methods of storing and processing texts in electronic form. It is now also the interchange and communication format used by many applications on the World Wide Web. In the present chapter we informally introduce some of its basic concepts and attempt to explain to the reader encountering them for the first time how and why they are used in the TEI scheme. More detailed technical accounts of TEI practice in this respect are provided in chapters 23. Using the TEI, 1. The TEI Infrastructure, and 22. Documentation Elements of these Guidelines.
In these examples, the #p5 version of the TEI Guidelines is adopted as the base text to which the apparatus entries are linked. This is the sole text witness for which a full transcription is provided in the electronic critical edition using this reference method. Because of this, the reading of this base text may be omitted from the `<app>` elements, as in the examples above. Due to the implicit nature of the location references of the apparatus entries, it may be hard to identify the exact places with textual variation. Therefore, the reading of the base text may equally be provided in the apparatus entries inside a `<lem>` element; combined with string matching, this can help the user of the edition to find out where the actual variation occurs (but note the difficulty with apparatus entries encoding additions to the base text, as in the second `<app>` element):

......

1 Note, how the `@loc` attribute does *not* refer to an `@xml:id` value of the text structure concerned, but to its ‘canonical reference’. For more information, see the `<app>` reference section, and 2.3.5 The Reference System Declaration of the TEI Guidelines.
The encoding scheme defined by these Guidelines is formulated as an application of the Extensible Markup Language (XML) (Bray et al. (eds.) (2006)). XML is widely used for the definition of device-independent, system-independent methods of storing and processing texts in electronic form. It is now also the interchange and communication format used by many applications on the World Wide Web. In the present chapter we informally introduce some of its basic concepts and attempt to explain to the reader encountering them for the first time how and why they are used in the TEI scheme. More detailed technical accounts of TEI practice in this respect are provided in chapters 23. Using the TEI, 1. The TEI Infrastructure, and 22. Documentation Elements of these Guidelines.
Summary
The location-referenced method uses an implicit anchoring technique to link the apparatus entries with the base text. In an internal apparatus, the apparatus entries can occur anywhere inside the text structure in which their variants occur. In an external apparatus, the link is established through the use of the @loc attribute on the <app> elements, which points to a canonical reference of the relevant text structures in the base text.

5.2 The Double-End-Point-Attached Method
The double-end-point-attached method links an apparatus entry to a base text, by anchoring it to the exact start and end positions of its lemma in the base text. This can be done either internally (inside the running text), or externally (outside the running text).

In an internal double-end-point-attached apparatus, the apparatus entries occur immediately after their lemma in the transcription of the base text. A specific @from attribute must be used to exactly point at the starting point of the preceding lemma in the text. Its value should be a pointer to the formal identification code of an element in the base text that corresponds to the start of the lemma. If this point coincides with the start of an existing text structure, the identification code of its element may be used; otherwise, an empty <anchor/> milestone element must be inserted in the base text, whose sole purpose is to provide a formal code for its @xml:id attribute. For example, an internal apparatus for the example in the previous section could look as follows:
The encoding scheme defined by these Guidelines is formulated as an application of the Extensible Markup Language (XML) (Bray et al. (eds.) (2006)). XML is widely used as an application of the Extensible Markup Language (SGML). Although widely said to be short for the surnames of its progenitors, the official expansion of this abbreviation is "Standard Generalized Markup Language."
or of the more recently developed W3C Extensible Markup Language (XML). Both SGML and XML are widely-used for the definition of device-independent, system-independent methods of storing and processing texts in electronic form. It is now also the interchange and communication format used by many applications on the World Wide Web. This chapter presents a brief tutorial guide to its main features, for those readers who have not encountered it before. For a more technical account of TEI practice in using XML being in fact a simplification or derivation of SGML. In the present chapter we introduce informally the basic concepts underlying such markup languages and attempt to explain to the reader encountering them for the first time how and why they are actually used in the TEI scheme. Except where the two are explicitly distinguished, references to XML in what follows may be understood to apply equally well to the TEI usage of SGML. More detailed technical accounts of TEI practice in this respect are provided in chapters 22. Documentation Elements. For TEI practice see chapter 28, "Conformance," ; for a more technical description of the subset of SGML used by the TEI scheme. More detailed technical accounts of TEI practice in this respect are provided in chapters 23. Using the TEI, 1. The TEI Infrastructure, and 22. Documentation Elements.
of these Guidelines<app from="#lem10">

<rdg wit="#p3">, see chapter 39, "Formal Grammar for the TEI-Interchange-Format Subset of SGML," on page 1247</rdg>
<rdg wit="#p4">, see chapter 39 <hi>Formal Grammar for the TEI-Interchange-Format Subset of SGML</hi></rdg>
</app>. </p>

An external double-end-point-attached apparatus is very similar to its internal equivalent, apart from the fact that the apparatus entries are located outside of the running text. Due to this physical separation, the need arises to explicitly point out the end point of the lemma in the base text as well (again, either using the @xml:id attribute of an existing text structure, or that of an explicit <anchor/> element). In order to refer to this end point of the textual variation, the <app> element must have another attribute: @to, pointing at the identification code of the relevant point in the base text. For example, an external apparatus for the previous example could look as follows:
The encoding scheme defined by these Guidelines is formulated as an application of the Extensible Markup Language (XML) (Bray et al. (eds.) (2006)). XML is widely used for the definition of device-independent, system-independent methods of storing and processing texts in electronic form. It is now also the interchange and communication format used by many applications on the World Wide Web. In the present chapter we informally introduce some of its basic concepts and attempt to explain to the reader encountering them for the first time how and why they are used in the TEI scheme. More detailed technical accounts of TEI practice in this respect are provided in chapters 23. Using the TEI, 1. The TEI Infrastructure, and 22. Documentation Elements of these Guidelines.
Of course, here too, the lemma of the base text can be explicitly recorded in the apparatus entries as well:
The encoding scheme defined by these Guidelines is formulated as an application of the Extensible Markup Language (XML) (Bray et al. (eds.) (2006)). XML is widely used for the definition of device-independent, system-independent methods of storing and processing texts in electronic form. It is now also the interchange and communication format used by many applications on the World Wide Web. In the present chapter we informally introduce some of its basic concepts and attempt to explain to the reader encountering them for the first time how and why they are used in the TEI scheme. More detailed technical accounts of TEI practice in this respect are provided in chapters 23. Using the TEI, 1. The TEI Infrastructure, and 22. Documentation Elements of these Guidelines. 

5. Encoding Variation in Texts
Summary

The double-end-point-attached method provides a means to explicitly anchor an apparatus entry to the exact position where its lemma in the base text differs from one of the other readings. In an internal apparatus, the apparatus entries should be placed immediately after the base text's lemma. Each `<app>` element must have a `@from` attribute pointing to the `@xml:id` identification code of an element indicating the start of the lemma in the base text. In an external apparatus, the apparatus entries must formally identify the end point of the lemma as well, using a `@to` attribute that points to the `@xml:id` identification code of an element indicating the end of the lemma in the base text. If no other elements are available, these `@xml:id` attributes may be encoded on empty `<anchor/>` elements inside the base text.

5.3 The Parallel Segmentation Method

Contrary to both other methods, the parallel segmentation method only allows for the encoding of an inline apparatus. Similarly to an internal double-end-point-attached apparatus entry, a parallel segmented apparatus entry is encoded inline, at the exact place where the variation occurs. However, a parallel segmented apparatus entry encodes all readings as equal variants, thus interweaving the common (invariant) text of all text witnesses with apparatus entries that contain all different alternative readings. In this sense, the notions of a base text and lemma become obsolete: all text that is common, is shared; all varying text is encoded as a separate reading in an apparatus entry. Because of this exact anchoring at the place of occurrence in the ‘palimpsest’ text, no specific attributes are necessary for the `<app>` element. For example, the preceding example can be expressed as a parallel segmented apparatus as follows:
The encoding scheme defined by these Guidelines is formulated as an application of a system known as the Standard Generalized Markup Language (SGML). Although widely said to be short for the surnames of its progenitors, the official expansion of this abbreviation is "Standard Generalized Markup Language." SGML is an international standard.
Both SGML and XML are widely-used. XML is widely used for the definition of device-independent, system-independent methods of representing, storing and processing texts in electronic form. This chapter presents a brief tutorial guide to its main features, for those readers who have not encountered it before. For a more technical account of TEI practice in using XML being in fact a simplification or derivation of SGML, in the present chapter we introduce informally the basic concepts underlying such markup languages and attempt to explain to reader encountering them for the first time how they are actually used in the TEI scheme. Except where the two are explicitly distinguished, references to XML in what follows may be understood to apply equally well to the TEI usage of SGML. a more technical account of For TEI practice see chapter 28, "Conformance," on page 727. For a more technical description of the subset of SGML present chapter we informally introduce some of its concepts and attempt to explain to the reader encountering them basic for the first time how and why they are used by in the TEI scheme, see chapter 39, "Formal Grammar for the TEI-Interchange-Format Subset of SGML," [in separate fascicle].
Summary

The parallel segmentation method encodes all variants as equal readings inside apparatus entries that are located at their precise place of occurrence in all texts. This results in a single text that contains an integral view on both the common text and the textual variants. Because of this, the notions of base text and lemma become irrelevant.