TEI By Example

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

4. Encoding Textual Variants........................................................................................................ 1
   4.1 Basic Organisation of an Apparatus Entry.............................................................. 1
   4.2 Grouping Readings.................................................................................................... 4
   4.3 Classification.......................................................................................................... 6
   4.4 Reading Details....................................................................................................... 10
4. Encoding Textual Variants

4.1 Basic Organisation of an Apparatus Entry

Traditionally, printed critical editions have developed efficient mechanisms to represent textual variants on as little physical space as possible in what is commonly called a critical apparatus. Many types of apparatus exist, depending on the editorial theory, but all tend to put the different readings found in the different text witnesses on a par with one version of the text, which is commonly called the base text. The TEI Guidelines offer an analogous mechanism for representing textual variants in a concise way. A piece of text with corresponding variants in the different text witnesses, is encoded in an <app> (apparatus entry) element, which holds all different readings. Each reading must be encoded in a <rdg> (reading) element, which can be associated to its respective text witness by means of the @wit attribute. Its value should point to the definition of the text witness in a <listWit> element elsewhere in the edition (see 3. Describing Text Witnesses (p.______)). For example, let's have a closer look at the chapter title in our sample:

[witness p2] Chapter 2 A GENTLE INTRODUCTION TO SGML
[witness p3] Chapter 2 A Gentle Introduction to SGML
[witness p4] 2 A Gentle Introduction to XML
[witness p5] v A Gentle Introduction to XML

In above example, all text that differs from the corresponding fragment in any other witness is highlighted in yellow. Only the word “A” is shared between all text witnesses. In an electronic edition of our sample, these stretches of variant text could be encoded in two apparatus entries:

```xml
<app>
  <rdg wit="#p2">Chapter 2 A GENTLE INTRODUCTION TO SGML</rdg>
  <rdg wit="#p3">Chapter 2 A Gentle Introduction to SGML</rdg>
  <rdg wit="#p4">2 A Gentle Introduction to XML</rdg>
  <rdg wit="#p5">v A Gentle Introduction to XML</rdg>
</app>

<app>
  <rdg wit="#p2">GENTLE INTRODUCTION TO SGML</rdg>
  <rdg wit="#p3">Gentle Introduction to SGML</rdg>
  <rdg wit="#p4">Gentle Introduction to XML</rdg>
  <rdg wit="#p5">Gentle Introduction to XML</rdg>
</app>
```
In this example, both textual variants are encoded as two apparatus entries, with four readings each. Each `<rdg>` element points to the definition of its corresponding text witness by means of the *sigla* in its `@wit` attribute. Note how each sigil starts with a `#` sign, because it addresses the `@xml:id` value of a `<witness>` element in the edition.

In printed critical editions, the assumption of a base text against which all other versions are compared is quite common. Therefore, besides readings, a TEI apparatus entry can also contain a `<lem>` (lemma) element, identifying the reading it contains as a ‘preferred’ reading, according to the editor’s theory of the text. Note that if a `<lem>` element is used, it must occur as the first element inside `<app>`. If version #p2 were considered the base text to the edition of this sample, the previous example could be encoded as follows:

```xml
<app>
  <lem wit="#p2">Chapter 2</lem>
  <rdg wit="#p3">Chapter 2</rdg>
  <rdg wit="#p4">2</rdg>
  <rdg wit="#p5">v</rdg>
</app>

<app>
  <lem wit="#p2">GENTLE INTRODUCTION TO SGML</lem>
  <rdg wit="#p3">Gentle Introduction to SGML</rdg>
  <rdg wit="#p4">Gentle Introduction to XML</rdg>
  <rdg wit="#p5">Gentle Introduction to XML</rdg>
</app>
```

In order to make this representation more efficient, equal readings can be collapsed into one single `<rdg>` element, by combining the sigla into a list separated by white spaces in the `@wit` attribute:

```
<app>
  <rdg wit="#p2 #p3 #p4 #p5">Chapter 2 v 2</rdg>
</app>
```

Note, how the TEI Guidelines offer the means to encode textual variation, without imposing any theoretical assumptions on how to encode an apparatus for the variants in different texts. The treatment of variation in different text versions is an explicit theoretical act of interpretation, and it is up to the encoder to determine corresponding text fragments, and where to delimit stretches of variation. Likewise, the examples in this TBE tutorial module are fairly theory-neutral, in that they tend to use the maximal length of differing text fragments as guiding principle for the demarcation of textual variants.

Because in the context of electronic critical editing a ‘preferred’ reading in a `<lem>` element is fairly theory-dependent, the examples in this TBE tutorial module will mostly just list all variants as equal `<rdg>` elements. You have to know, however, that each `<app>` element may always specify one of its readings as lemma (`<lem>`) as well.
Remember how we distinguished different witness groups in the previous section of this tutorial? This allows us to rewrite the sigla of readings shared by the versions of the TEI Guidelines dealing with either SGML or XML, using the group identification code for the corresponding group of witnesses:

You should consider an `<app>` element as a cross section of a text fragment over all of the different text witnesses. This means that all `<lem>` and `<rdg>` contents should be interpreted as mutually exclusive alternatives. Therefore, each text witness listed in the `@wit` attributes inside an `<app>` element should occur only once. Ideally, this should be the minimal requirement as well, so that each apparatus entry contains one corresponding text fragment across all different text witnesses included in the edition (although this is not strictly necessary when the edition uses one base text: see 5. Encoding Variation in Texts (p. _______)).
<lem> or <rdg> element should indicate the text witness(es) it corresponds to in a @wit attribute. The value of this attribute consists of a whitespace separated list of pointers to the @xml:id code(s) of the <witness> element(s) describing the corresponding text witness(es).

### 4.2 Grouping Readings

In both variants considered so far, arguments could be made for (re)grouping the readings. In the first apparatus entry, reading #p5 is set apart from all others because of the different chapter number. In the second apparatus entry, one possible case for explicit grouping could be the ‘genetic’ similarity of the variants in those versions of the TEI Guidelines dealing with SGML or XML.

One way of grouping readings is provided by a specialised <rdgGrp> element. It can be wrapped around <rdg> elements in an apparatus entry, in order to indicate their relatedness in some way. This <rdgGrp> really is nothing more than a wrapper, that can list the sigla of the text witnesses it groups in an own @wit attribute. For example, the readings in the previous example could be grouped as follows:

```xml
<app>
  <rdg wit="#teiSGML">Chapter 2 <lb/></rdg>
  <rdgGrp wit="#teiXML">
    <rdg wit="#p4">2</rdg>
    <rdg wit="#p5">v <lb/></rdg>
  </rdgGrp>
</app>
<app>
  <rdgGrp wit="#teiSGML">
    <rdg wit="#p2">GENTLE INTRODUCTION TO SGML</rdg>
    <rdg wit="#p3">Gentle Introduction to SGML</rdg>
  </rdgGrp>
  <rdg wit="#teiXML">Gentle Introduction to XML</rdg>
</app>
```

When you have a closer look at these variants, you'll see that some of these readings contain common text as well. In the first variant, the number "2" is shared between both #teiSGML readings, and the #p4 reading. In the last variant, the #p2 and #p3 readings are set apart by the common phrase "SGML", as opposed to "XML" in the #teiXML readings. Yet, both #p2 and #p3 text witnesses vary internally in their use of capitals. Such refinements can't be expressed using the <rdgGrp> grouping mechanism, as a <rdgGrp> element can only contain <rdg> or <lem> elements. If this grouping is maintained, you could express them in a more fine grained manner using another grouping mechanism: introducing nesting <app> elements in the <rdg> elements that share common text as well as differing readings:
In the first variant, the apparatus distinguishes between those readings whose heading refers to the second chapter (#teiSGML and #p4), and reading #p5, which refers to chapter five. However, as the first group of readings shows internal variation, this can be expressed in further nesting <app> elements (see the nesting <app> elements for the "Chapter" sub-variant, and the line break). The common text can be encoded as plain text contents of the grouping <rdg> element (see the "2", which occurs in all readings of the group: #teiSGML, and #p4). In the second variant, the readings corresponding to the text witnesses dealing with SGML are set apart from those dealing with XML. Since the first group of readings contains internal variation, the variant text ("Gentle Introduction to") is wrapped in a nesting <app> element, while the common text ("SGML") appears as plain text inside the grouping <rdg> element.

**Summary**

When desired, related readings can be grouped using one of two mechanisms. The first one wraps a dedicated <rdgGrp> element around related readings. This element can only contain <lem> and <rdg> elements. A more sophisticated way of grouping readings is provided by using nesting <app> structures inside a <rdg> element.
4.3 Classification

So far, the most wrought out encoding of the chapter's title in the different text witnesses looks as follows:

```xml
<app>
  <rdg wit="#teiSGML #p4">Chapter</rdg>
</app>

2

<app>
  <rdg wit="#teiSGML">
    <lb/>
  </rdg>
</app>

<rdg wit="#p5">v</rdg>

<app>
  <rdg wit="#teiSGML">GENTLE INTRODUCTION TO</rdg>
  <rdg wit="#p2">Gentle Introduction to</rdg>
  <rdg wit="#p3">Gentle Introduction to XML</rdg>
</app>

Admittedly, this organisation is not the most intuitive one, mostly because it mixes different perspectives:

- a content-oriented one in the first apparatus entry, grouping those variants with a common reading (i.e. the chapter number referred to)
- a genetic-oriented one in the second apparatus entry, grouping the readings according to the groups of witnesses (i.e. those occurring in the versions of the TEI Guidelines dealing with SGML or XML)

However, this is not necessarily the most interesting perspective, for it obscures some obvious correspondences. For example, there is no way of deducting the correspondence between the `<lb/>` reading occurring in three of the four witnesses, as it is 'buried' in two different reading groups. There is no reason, however, not to reorganise these apparatus entries in more atomic units:
One could argue that on closer examination, not all of these variants have the same ‘status’: some are more substantive than others. This may be pointed out at the level of the individual readings, by means of a @type attribute. In this way, we could for example distinguish between orthographic readings (differing only in their spelling or presentation) and substantive readings (differing in meaning):
With this distinction in place, the type of reading could be adopted as guiding principle to derive larger stretches of variation: only when two subsequent variants only have orthographically different readings, they can be merged to one apparatus entry. Note also, how in this case all readings for the different apparatus entries share the same type. This can be encoded at the higher level of the apparatus entry as well, simply by providing a `@type` attribute for the `<app>` element:
The `<rdgGrp>` too can have a `@type` attribute for specifying the nature of the group of readings it holds. For example, we could revisit the earlier grouping example using `<rdgGrp>`:

```xml
<app type="substantive">
  <rdg wit="#teiSGML" type="substantive">Chapter 2</rdg>
  <rdgGrp wit="#teiXML" type="substantive">
    <rdg wit="#p4">2</rdg>
    <rdg wit="#p5">v</rdg>
  </rdgGrp>
</app>

<app>
  <rdgGrp wit="#teiSGML" type="orthographic">
    <rdg wit="#p2">GENTLE INTRODUCTION TO SGML</rdg>
    <rdg wit="#p3">Gentle Introduction to SGML</rdg>
  </rdgGrp>
  <rdgGrp>
    <rdgGrp wit="#teiXML" type="substantive">Gentle Introduction to XML</rdg>
  </rdgGrp>
</app>

4. Encoding Textual Variants
Summary

The readings inside `<rdg>` and `<lem>` can be categorised with a `@type` attribute, in order to indicate what type of variant they contain. When readings are grouped using `<rdgGrp>`, the `@type` attribute equally can indicate what type of variants the reading group consists of. When an apparatus entry only contains variants of the same type, this may be expressed by the `@type` attribute at the `<app>` level.

4.4 Reading Details

Besides witness (`@wit`) and type information (`@type`), readings and lemmas can provide more information about the readings they hold, in dedicated attributes. One type of information that is particularly useful for critical editions of manuscript source materials, is the identification of a document hand that is responsible for a certain reading, especially when its text witness has been written by different hands. This can be expressed in a `@hand` attribute, which points to the definition of that hand in the TEI header (see TBE Module 2: The TEI Header -- Document Hands). This could be applied to our example texts: although the TEI Guidelines are not manuscripts, they are written collaboratively by a team of editors who could be considered document hands. Suppose that we could determine who was responsible for what change in the different versions included in our example critical edition, this could be encoded as follows:
Of course this attribution is subject to a greater or lesser deal of interpretation (especially in this contrived example). Therefore, it makes sense to indicate who is responsible for it. This can be expressed in a @resp attribute, which can point to an individual responsible for some aspects of the electronic edition, as identified in the TEI header (see TBE Module 2: The TEI Header -- The Title Statement). As always, the @resp attribute applies to all aspects of the element it is attached to, and can equally be used to indicate the responsibility for an unsure transcription of a reading. As the hand attribution in the previous example can be considered quite putative, it makes sense to provide responsibility information as well:
Note, how the @hand and @resp attributes can only be provided for individual readings, corresponding to individual witnesses. It is thus illegal to use them when the witness list inside @wit contains more than one sigil, or a group sigil, as in following incorrect example:
This example is incorrect because the first reading of the first apparatus entry overgeneralises the hand information for the #p3 witness, and the last reading of the last entry incorrectly attributes the hand information for the #p5 witness. It can be done, however, using a dedicated <witDetail> element. This element provides additional information about a reading in a specific text witness. It must have two mandatory attributes: @wit, identifying the specific text witness about which more information is provided; and @target, pointing at the @xml:id of the concerned <rdg> element. This implies that the reading concerned must be formally identified with an @xml:id attribute. For example, the previous example could be corrected as:

```
<app>
  <rdg wit="#p2 #p3" xml:id="rdg1.1">Chapter 2 <lb/></rdg>
  <rdg wit="#p4" hand="#SB" resp="#TBEcrew">2</rdg>
  <rdg wit="#p5" resp="#TBEcrew">v <lb/></rdg>
</app>

<witDetail target="#rdg1.1" wit="#p2" resp="#TBEcrew">attributed to <ref target="#MSMQ">Michael Sperber-McQueen</ref></witDetail>

<app>
  <rdg wit="#p2" hand="#LB" resp="#TBEcrew">GENTLE INTRODUCTION TO SGML</rdg>
  <rdg wit="#p3">Gentle Introduction to SGML</rdg>
  
  <rdg wit="#teiXML" xml:id="rdg2.3">Gentle Introduction to XML</rdg>
</app>

<witDetail target="#rdg2.3" wit="#p4" resp="#TBEcrew">attributed to <ref target="#SB">Syd Bauman</ref></witDetail>
```

The <witDetail> element is a specialised type of <note>, which means it can occur nearly anywhere in the document: either inline at the place of the reading needing further specification, or grouped together elsewhere in the document.

When digitising an existing critical edition, the sigla associated with the different readings can (and should) be formally encoded in the @wit attribute, as discussed earlier (see 4.1 Basic
Organisation of an Apparatus Entry (p. 1)). However, they can be transcribed literally as well, using a specific TEI element: <wit>. This element can then contain the literal transcription of the sigla used in the source edition, which may be of interest when they differ from their formal equivalent in the @wit attribute. The <wit> element should appear after the <rdg> element containing the concerned reading. For example, if our critical edition of the TEI Guidelines were based on a previous edition, the original sigla could be encoded as follows:

```xml
<app>
  <rgd wit="#teiSGML">Chapter 2</rgd>
  <wit>teiP2, teiP3</wit>
  <rgd wit="#p4">2</rgd>
  <wit>teiP4</wit>
  <rgd wit="#p5">v</rgd>
  <wit>teiP5</wit>
</app>
<app>
  <rgd wit="#p2">GENTLE INTRODUCTION TO SGML</rgd>
  <wit>teiP2</wit>
  <rgd wit="#p3">Gentle Introduction to SGML</rgd>
  <wit>teiP3</wit>
  <rgd wit="#teiXML">Gentle Introduction to XML</rgd>
  <wit>teiP4, teiP5</wit>
</app>
```

Summary

Lemma (<lem>) and readings (<rdg>) can be further qualified by means of attributes. The @resp attribute can be used to identify the person responsible for the encoding of the reading, while the document hand responsible for that particular reading can be referred to in a @hand attribute. When more detailed information is to be given for a particular reading in a particular text witness, this can be done in a <witDetail> element, whose @wit attribute must point to the concerned text witness, and whose @target attribute must point to the identification code of the affected reading(s). Finally, when an existing critical edition is digitised, the original sigla information can be transcribed literally in a <wit> element, following the <rdg>.