TEI BY EXAMPLE



MODULE 7: CRITICAL EDITING

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. Textual Variation....2
3. Describing Text Witnesses....9
4. Encoding Textual Variants....13
4.1 Basic Organisation of an Apparatus Entry....13
4.2 Grouping Readings....17
4.3 Classification....19
4.4 Reading Details....24
5. Encoding Variation in Texts....28
5.1 The Location-Referenced Method....29
5.2 The Double End-Point Attachment Method....35
5.3 The Parallel Segmentation Method....42
6. Caveats....46
7. Summary....49
8. What's Next?....55

1. Introduction

When texts are considered random combinations of strings, this can entertain interesting predictions about the completion of the complete works of William Shakespeare by a(n army of) hypothetical monkey(s) randomly hitting the keys on a typewriter (see the infinite monkey theorem). Indeed, to a computer, the textual universe *is* a library of Babel, in which every possible text is as likely (or rather *un*likely) to exist as any other text. From this perspective, perfect duplicates may perfectly co-exist with gazillions of close approximations and texts that have nothing in common at all. In an intellectual, human context, where characters are ordered along (arbitrary) rules, two sensible texts will most likely have nothing in common at all, rather than being related in any way. In fact, the chances of identical texts can almost be ruled out to either perfect mechanical photocopies, or blatant cases of plagiarism, and as such be considered uninteresting from an intellectual point of view. However, especially in the context of greatly valued literary, cultural, and/or historical works, the odd chance that such a text has a closely resembling counterpart becomes quite interesting. It may at least indicate some kind of relationship between both texts, even provide insight in its transmission through time, shed light on its history and conception, perhaps tell us something about the creative process of its author, or by extension provide insights in The Creative Process in the working of the human mind. These domains of knowledge inform different kinds of theories of textual criticism, each with their own research interests, principles and practises. What they all have in common, however, is an attempt to represent related texts found in different physical *witnesses* as different *versions* of the same abstract *work*.

As we have seen already, in order to make this world of meaning accessible to/via computers, text encoding with TEI provides a sensible approach. Moreover, besides the general provisions for text encoding, the TEI Guidelines define a range of specific elements and mechanisms to represent textual variation in a sensible way for further analysis. The TEI Guidelines devote a complete chapter, 12. Critical Apparatus, to the documentation of specific elements that are grouped into the **`textcrit`** module for the encoding of textual variation. In order to use the elements covered in this tutorial module, you are thus required to include the dedicated **`textcrit`** TEI module in your TEI schema.[1]

.....

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

2. Textual Variation

Similar to all other TEI modules, the elements and attributes defined in the TEI **`textcrit`** module can be used for the encoding of existing source materials (be they in print or digital form), or the encoding of electronic documents from scratch. However, the *use* of this module in the context of electronic critical editing, adds another perspective to this traditional authorial/editorial angle (see Vanhoutte and Van den Branden 2009). Electronic or digital critical editions can be created from scratch either by encoding different primary source materials straight *as* a critical edition, or by generating the edition from previously encoded electronic transcriptions of those materials as independent texts in their own right. Therefore, the tags defined in the TEI **`textcrit`** module can be used to:

digitise

an existing print edition

create

a digital edition, e.g., by recording some or all of the known variations among different witnesses to the text in a critical apparatus of variants

generate

a digital edition from encoded transcriptions of the documentary source material

In the examples in this TBE module, "critical editing with TEI" will be understood as the act of encoding material sources in a TEI representation that allows for the creation or generation of a digital edition in some form (using any output format in the digital medium, e.g., HTML pages, PDF, flash movies,...), rather than digitising an existing critical edition. In this sense, the authorial/editorial angle of this TBE module differs from that of the other modules (focusing on the digitisation of a material source text in a certain genre). However, the strategies discussed in this tutorial for representing textual variation can equally be applied to the digitisation of existing critical editions. Where there are differences, these will be pointed out explicitly.

For example, consider following texts:

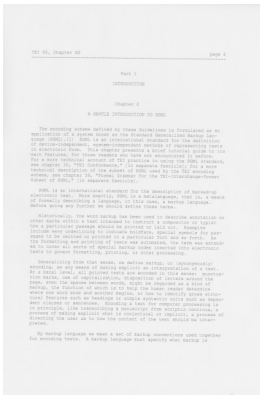


Figure 1. A page of version P2 of the TEI Guidelines.

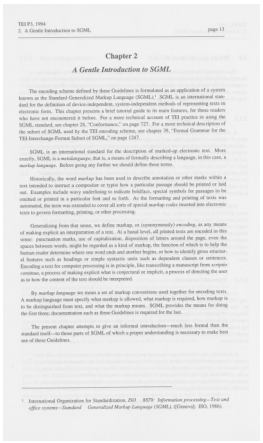


Figure 2. A page of version P3 of the TEI Guidelines.

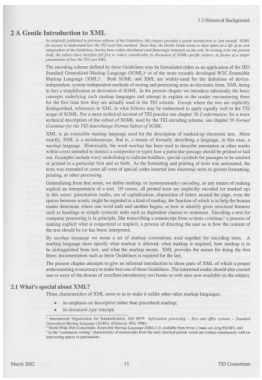


Figure 3. A page of version P4 of the TEI Guidelines.

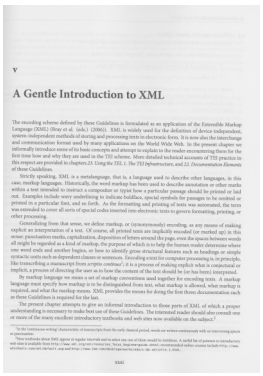


Figure 4. A page of version P5 of the TEI Guidelines.

Some of these images may look more or less familiar to you: they are facsimiles from the first page of chapter 2 of the printed TEI Guidelines throughout their different incarnations, from version P2 (1992) to the latest version, P5 (2009). As you can imagine, the technological evolutions of these 17 years have prompted considerable changes to this chapter that introduces the technological background of text encoding with TEI, ranging from rephrasing, addition or deletion of notes, changes in italicisation, restructuring of paragraphs, etc. One way of approaching this textual variation could consist of encoding these text versions as physically distinct TEI documents, in which corresponding text structures could be aligned by a common identification mechanism. For example, the first couple of paragraphs in these 4 text witnesses could be encoded in different TEI documents as follows:

P2 **P3**

```
<pb xmlns="http://www.tei-c.org/
ns/1.0" n="2"/>
<head xmlns="http://www.tei-c.org/
ns/1.0">Chapter 2 <lb/>A GENTLE I
NTRODUCTION TO SGML</head>
<p xmlns="http://www.tei-c.org/
ns/1.0" xml:id="p1" corresp="P3.xml
#p1 P4.xml#p1 P5.xml#p1">The encoding
 scheme defined by these Guidelines
 is formulated as an application
 of a system known as the Standard
 Generalized Markup Language (S
GML).<note place="foot" xml:id="n1" corresp="P3
.xml#n1 P4.xml
#n2"><bibl><editor>International
 Organization for
 Standardization</editor>, <title>I
SO 8879: Information processing
--Text and office systems--
Standard Generalized Mark-
up Language (SGML)</title>,
 ([<pubPlace>Geneva</pubPlace>]: <publisher>I
SO</publisher>, <date>1986</date>).</bibl>
 Although widely said to be
 short for the surnames of its
 progenitors, the official expansion
 of this abbreviation is "Standard
 Generalized Markup Language."</note>
 SGML is an international standard
 for the definition of device-
independent, system-independent
 methods of representing 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
```

```
<pb xmlns="http://www.tei-c.org/
ns/1.0" n="13"/>
<head xmlns="http://www.tei-c.org/
ns/1.0">Chapter 2 <lb/>A Gentle
 Introduction to SGML</head>
<p xmlns="http://www.tei-c.org/
ns/1.0" xml:id="p1" corresp="P2.xml
#p1 P4.xml#p1 P5.xml#p1">The encoding
 scheme defined by these Guidelines
 is formulated as an application
 of a system known as the Standard
 Generalized Markup Language (S
GML). <note place="foot" xml:id="n1" corresp="P2
.xml#n1 P4.xml#n2">
  <bibl><editor>International
   Organization for
   Standardization</editor>, <title>I
  SO 8879: Information processing
   - Text and office systems -
   Standard Generalized Markup
   Language (SGML)</title>,
   ([<pubPlace>Geneva</pubPlace>]: <publisher>I
  SO</publisher>, <date>1986</date>)</bibl>
</note> SGML is an international
 standard for the definition
 of device-independent, system-
independent methods of representing
 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 the SGML standard, see chapter
 28, "Conformance," on page 727.
 For a more technical description
 of the subset of SGML used by the
 TEI encoding scheme, see chapter
```

```
 the SGML standard, see chapter 30,
 "TEI Conformance," [in separate
 fascicle]; for a more technical
 description of the subset of SGML
 used by the TEI encoding scheme, see
 chapter 39, "Formal Grammar for the
 TEI-Interchange-Format Subset of S
GML," [in separate fascicle].</p>
<p xmlns="http://www.tei-c.org/
ns/1.0" xml:id="p2a" corresp="P3
.xml#p2a P4.xml#p2 P5.xml#p2">S
GML is an international standard
 for the description of marked-up
 electronic text. More exactly, SGML
 is a metalanguage, that is, a means
 of formally describing a language,
 in this case, a markup language.
 Before going any further we should
 define these terms.</p>
<p xmlns="http://www.tei-c.org/
ns/1.0" xml:id="p2b" corresp="P3
.xml#p2b P4.xml#p2 P5.xml
#p2">Historically, the word markup
 has been used to describe annotation
 or other marks within a text
 intended to instruct a compositor
 or typist how a particular passage
 should be printed or laid out.
 Examples include wavy underlining
 to indicate boldface, special
 symbols for passages to be omitted
 or printed in a particular font
 and so forth. As the formatting and
 printing of texts was automated,
 the term was extend-ed to cover
 all sorts of special markup codes
```

```
 39, "Formal Grammar for the TEI-
Interchange-Format Subset of SGML,"
 on page 1247.</p>
<p xmlns="http://www.tei-c.org/
ns/1.0" xml:id="p2a" corresp="P2
.xml#p2a P4.xml#p2 P5.xml#p2">S
GML is an international standard
 for the description of marked-up
 electronic text. More exactly, S
GML is a <hi>metalanguage</hi>, that
 is, a means of formally describing a
 language, in this case, a <hi>markup
 language</hi>. Before going any
 further we should define these
 terms.</p>
<p xmlns="http://www.tei-c.org/
ns/1.0" xml:id="p2b" corresp="P2
.xml#p2b P4.xml#p2 P5.xml
#p2">Historically, the
 word <hi>markup</hi> has been used
 to describe annotation or other
 marks within a text intended to
 instruct a compositor or typist
 how a particular passage should
 be printed or laid out. Examples
 include wavy underlining to indicate
 boldface, special symbols for
 passages to be omitted or printed
 in a particular font and so forth.
 As the formatting and printing
 of texts was automated, the term
 was extended to cover all sorts
 of special <hi>markup codes</hi>
 inserted into electronic texts to
 govern formatting, printing, or
 other processing.</p>
```

```
 inserted into electronic texts to
 govern formatting, printing, or
 other processing.</p>
```

Example 1. An encoded page of version P2 of the TEI Guidelines.

Example 2. An encoded page of version P3 of the TEI Guidelines.

P4

```
<pb xmlns="http://www.tei-c.org/
ns/1.0" n="13"/>
<head xmlns="http://www.tei-c.org/
ns/1.0">2 A Gentle Introduction to X
ML</head>
<note xmlns="http://www.tei-c.org/
ns/1.0" type="disclaimer" xml:id="n1">As
 originally published in previous
 editions of the Guidelines,
 this chapter provided a gentle
 introduction to 'just enough' SGML
 for anyone to understand how the T
EI used that standard. Since then,
 the Gentle Guide seems to have taken
 on a life of its own independent
 of the Guidelines, having been
 widely distributed (and flatteringly
 imitated) on the web. In revising it
 for the present draft, the editors
 have therefore felt free to reduce
 considerably its discussion of S
GML-specific matters, in favour of a
 simple presentation of how the TEI
 uses XML.</note>
<p xmlns="http://www.tei-c.org/
ns/1.0" xml:id="p1" corresp="P2.xml
#p1 P3.xml#p1 P5.xml#p1">The encoding
 scheme defined by these Guidelines
 may be formulated either as an
```

P5

```
<pb xmlns="http://www.tei-c.org/
ns/1.0" n="xxxi"/>
<head xmlns="http://www.tei-c.org/
ns/1.0">v <lb/>A Gentle Introduction
 to XML</head>
<p xmlns="http://www.tei-c.org/
ns/1.0" xml:id="p1" corresp="P2.xml
#p1 P3.xml#p1 P5.xml#p1">The encoding
 scheme defined by these Guidelines
 is formulated as an application of
 the Extensible Markup Language (X
ML) (Bray et al. (eds.) (2006)). X
ML 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 <hi>23.
 Using the TEI</hi>, <hi>1. The T
```

```
 application of the ISO Standard
 Generalized Markup Language (S
GML)<note place="foot" corresp="P2
.xml#n1 P3.xml#n1">
  <bibl><editor>International
   Organization for
   Standardization</editor>, <title>I
  SO 8879: Information processing
   - Text and office systems -
   Standard Generalized Markup
   Language (SGML)</title>,
   ([<pubPlace>Geneva</pubPlace>]: <publisher>I
  SO</publisher>, <date>1986</date>)</bibl>
</note> or of the more
 recently developed W3C
 Extensible Markup Language (X
ML)<note place="foot" xml:id="n3">
  <bibl><editor>World Wide Web
   Consortium</editor>: <title>Extensible
   Markup Language (XML) 1.0</title>,
   available from <ref target="http://
  www.w3.org/TR/REC-xml">http://www.w3
  .org/TR/REC-xml</ref></bibl>
</note>. Both SGML and XML are widely
-used for the definition of device
-independent, system-independent
 methods of storing and processing
 texts in electronic form; 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 they
 are actually used in the TEI scheme.
 Except where the two are explicitly
 distinguished, references to XML
```

```
EI Infrastructure</hi>, and <hi>22.
 Documentation Elements</hi> of these
 Guidelines.</p>
<p xmlns="http://www.tei-c.org/
ns/1.0" xml:id="p2" corresp="P2.xml
#p2a P2.xml#p2b P3.xml#p2a P3.xml#p2b
 P4.xml#p2">Strictly speaking, XML is
 a metalanguage, that is, a language
 used to describe other languages,
 in this case, markup languages.
 Historically, the word markup has
 been used to describe annotation or
 other marks within a text intended
 to instruct a compositor or typist
 how a particular passage should
 be printed or laid out. Examples
 include wavy underlining to indicate
 boldface, special symbols for
 passages to be omitted or printed in
 a particular font, and so forth. As
 the formatting and printing of texts
 was automated, the term was extended
 to cover all sorts of special codes
 inserted into electronic texts to
 govern formatting, printing, or
 other processing.</p>
```

Example 4. An encoded page of version P5 of the TEI Guidelines.

```
 in what follows may be understood
 to apply equally well to the TEI
 usage of SGML. For a more technical
 account of TEI practice see chapter
 28 <hi>Conformance</hi>; for a more
 technical description of the subset
 of SGML used by the TEI encoding
 scheme, see chapter 39 <hi>Formal
 Grammar for the TEI-Interchange-
Format Subset of SGML</hi>.</p>
<p xmlns="http://www.tei-c.org/
ns/1.0" xml:id="p2" corresp="P2
.xml#p2a P2.xml#p2b P3.xml#p2a
 P3.xml#p2b P5.xml#p2">XML is an
 extensible markup language used
 for the description of marked-up
 electronic text. More exactly, X
ML is a <hi>metalanguage</hi>, that
 is, a means of formally describing a
 language, in this case, a <hi>markup
 language</hi>. Historically, the
 word <hi>markup</hi> has been used
 to describe annotation or other
 marks within a text intended to
 instruct a compositor or typist
 how a particular passage should
 be printed or laid out. Examples
 include wavy underlining to indicate
 boldface, special symbols for
 passages to be omitted or printed in
 a particular font and so forth. As
 the formatting and printing of texts
 was automated, the term was extended
 to cover all sorts of special codes
 inserted into electronic texts to
 govern formatting, printing, or
 other processing.</p>
```

Example 3. An encoded page of version P4 of the TEI Guidelines.

This would allow for maximal representation of the distinct material sources, and leave the identification of the actual variation either to further processing or human inspection. A variant of this approach could integrate the transcriptions of the text in all material witnesses in a single TEI document, and make use of appropriate linking attributes to point out the alignment between the different text structures. In their naivety, such systems are both redundant and crude. While providing *all* text of all text witnesses, and aligning the corresponding text structures, they provide little *insight* in the places where the different witnesses actually differ.

In order to encode the actual textual variation between the different text versions in a meaningful way, the TEI Guidelines provide a specialised module of elements and attributes that allow you to encode textual variation at word level. This TBE tutorial will first discuss how to describe the different text witnesses represented in the critical edition; next deal with the encoding of textual variants between these witnesses in isolation; then treat different ways of integrating such records of variation within the encoding of the critical edition; and finally point out potential problems and pitfalls when creating a critical edition with TEI.

3. Describing Text Witnesses

When creating, generating or digitising a critical edition, it is of crucial importance to document the text witnesses whose transcriptions it contains. This can be done in a <listWit> (list of witnesses) element, which can be put either in the <sourceDesc> section of the TEI header (when creating or generating a critical edition), or somewhere in the <text>, usually in the <front> section (when digitising an existing critical edition). The <listWit> element should describe each text witness in its own <witness> element. This element can contain a prose description of the witness in plain text, possibly enriched with a specialised element for bibliographic description (<bibl>, <biblStruct>, or <biblFull>). The witness definitions should provide a unique identification code in the @xml:id attribute. This code is used as a *sigil* in the critical edition, in order to connect the textual variants with the respective witnesses in which they occur (see section 4). For example, the witness list for our critical edition of the TEI Guidelines could look as follows:

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

```
    <!-- ... -->
    <sourceDesc>
      <listWit>
        <witness xml:id="p2">
          <bibl><editor>Sperberg-McQueen, M.</editor>; <editor>Burnard, L.</editor>
           (eds.). <title>TEI P2 Guidelines for the Encoding and Interchange of
           Machine Readable Texts Draft P2</title> (published serially 1992-1993);
           Draft Version <date when="1993-04-02">2 of April 1993</date>: <extent>19
           chapters</extent>. Available from <ptr target="https://tei-c.org/Vault/Vault
          -GL.html"/> (accessed October 2008).</bibl>
        </witness>
        <witness xml:id="p3">
          <bibl><editor>Sperberg-McQueen, C.M.</editor>; <editor>Burnard,
           L.</editor> (eds.). <title>Guidelines for Electronic Text Encoding and
           Interchange. TEI P3. Revised reprint.</title> <publisher>Text Encoding
           Initiative</publisher>: <pubPlace>Oxford</pubPlace>, <pubPlace>Providence</pubPlace>, <pubP
        </witness>
        <witness xml:id="p4">
          <bibl><editor>Sperberg-McQueen, C.M.</editor>; <editor>Burnard,
           L.</editor> (eds.). <title>TEI P4: Guidelines for
           Electronic Text Encoding and Interchange. XML-compatible
           edition.</title> <publisher>Text Encoding Initiative
           Consortium</publisher>: <pubPlace>Oxford</pubPlace>, <pubPlace>Providence</pubPlace>, <pubP
        </witness>
        <witness xml:id="p5">
          <bibl><editor>Sperberg-McQueen, C.M.</editor>; <editor>Burnard,
           L.</editor> (eds.). <title>TEI P5: Guidelines for
           Electronic Text Encoding and Interchange. Revised and
           re-edited.</title> <publisher>Text Encoding Initiative
           Consortium</publisher>: <pubPlace>Oxford</pubPlace>, <pubPlace>Providence</pubPlace>, <pubP
        </witness>
      </listWit>
    </sourceDesc>
  </fileDesc>
  <!-- ... -->
 </teiHeader>
 <!-- ... -->
</TEI>
```

Example 5. A list of witness descriptions for four versions of the TEI Guidelines.

Such bibliographic descriptions of course are easier for printed works than for manuscripts; for the latter type of witnesses, some kind of description inside <listWit> is advised, preferably with a pointer (using <ptr> or <ref>) to a full description of the manuscript inside <msDescription>.

REFERENCE

For a full discussion of the <msDescription> element, see section 10.2 The Manuscript Description Element of the TEI Guidelines, and section 12.1.4.3 The Witness List for examples of describing manuscript witnesses in a digital edition.

In a critical edition, it may make sense to discern groups of witnesses that have many text variants in common in comparison to other witnesses and can often be conveniently summarised in one sigil. In the witness list, witnesses can be grouped by wrapping their <witness> descriptions in nesting <listWit> structures. The common sigil then can be provided as the value for an @xml:id attribute of the group's <listWit> element. The nested witness groups can be labelled with a <head> element. For example, in our sample text witnesses it may make sense to discern those versions of the TEI Guidelines dealing with SGML, and those dealing with XML. This could look as follows:

```
<TEI xmlns="http://www.tei-c.org/ns/1.0">
  <teiHeader>
    <fileDesc>
      <!-- ... -->
      <sourceDesc>
        <listWit>
          <listWit xml:id="teiSGML">
            <head>TEI Guidelines covering SGML</head>
            <witness xml:id="p2">
              <bibl><editor>Sperberg-McQueen, M.</editor>; <editor>Burnard, L.</editor>
                (eds.). <title>TEI P2 Guidelines for the Encoding and Interchange of
                Machine Readable Texts Draft P2</title> (published serially 1992-1993);
```

```
            Draft Version <date when="1993-04-02">2 of April 1993</date>: <extent>19
            chapters</extent>. Available from <ptr target="https://tei-c.org/Vault/
           Vault-GL.html"/> (accessed October 2008).</bibl>
         </witness>
         <witness xml:id="p3">
          <bibl><editor>Sperberg-McQueen, C.M.</editor>; <editor>Burnard,
           L.</editor> (eds.). <title>Guidelines for Electronic Text Encoding and
           Interchange. TEI P3. Revised reprint.</title> <publisher>Text Encoding
           Initiative</publisher>: <pubPlace>Oxford</pubPlace>, <pubPlace>Providence</pubPlace>, <pu
         </witness>
        </listWit>
        <listWit xml:id="teiXML">
         <head>TEI Guidelines covering XML</head>
         <witness xml:id="p4">
          <bibl><editor>Sperberg-McQueen, C.M.</editor>; <editor>Burnard,
           L.</editor> (eds.). <title>TEI P4: Guidelines for
           Electronic Text Encoding and Interchange. XML-compatible
           edition.</title> <publisher>Text Encoding Initiative
           Consortium</publisher>: <pubPlace>Oxford</pubPlace>, <pubPlace>Providence</pubPlace>, <pu
         </witness>
         <witness xml:id="p5">
          <bibl><editor>Sperberg-McQueen, C.M.</editor>; <editor>Burnard,
           L.</editor> (eds.). <title>TEI P5: Guidelines for
           Electronic Text Encoding and Interchange. Revised and
           re-edited.</title> <publisher>Text Encoding Initiative
           Consortium</publisher>: <pubPlace>Oxford</pubPlace>, <pubPlace>Providence</pubPlace>, <pu
         </witness>
        </listWit>
       </listWit>
      </sourceDesc>
    </fileDesc>
    <!-- ... -->
  </teiHeader>
  <!-- ... -->
</TEI>
```

Example 6. Grouping descriptions of related text witnesses in `<listWit>`.

SUMMARY

The different text witnesses included in a critical edition should be documented in a <listWit> element. Such a list may occur in the <sourceDesc> section of the TEI header (for digital editions created or generated from scratch), or in the <text> of the edition, usually in the <front> section (for digital editions digitised from an existing edition). Each text witness should be described in a <witness> element, containing either a prose description as plain text, possibly enriched with specific TEI elements for bibliographic description (<bibl>, <biblStruct>, <biblFull>). An @xml:id attribute must be provided for each witness, which is used as the sigil for this witness in the edition. Witness groups can be distinguished in separate nested <listWit> elements.

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 section 3). 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 a digital edition of our sample, these stretches of variant text could be encoded in two apparatus entries:

```
<app xmlns="http://www.tei-c.org/ns/1.0">
  <rdg wit="#p2">Chapter 2 <lb/></rdg>
  <rdg wit="#p3">Chapter 2 <lb/></rdg>
  <rdg wit="#p4">2</rdg>
  <rdg wit="#p5">v <lb/></rdg>
</app>
<app xmlns="http://www.tei-c.org/ns/1.0">
  <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>
```

Example 7. Apparatus entries without preferred readings, encoded in `<rdg>`.

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. Notice how each sigil starts with a # sign, because it addresses the @xml:id value of a <witness> element in the edition.[4]

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. Notice 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:[5]

```
<app xmlns="http://www.tei-c.org/ns/1.0">
  <lem wit="#p2">Chapter 2 <lb/></lem>
  <rdg wit="#p3">Chapter 2 <lb/></rdg>
  <rdg wit="#p4">2</rdg>
  <rdg wit="#p5">v <lb/></rdg>
</app>
<app xmlns="http://www.tei-c.org/ns/1.0">
  <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>
```

Example 8. Apparatus entries with a preferred reading in <lem>.

.....

4 Notice, 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.

5 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.

In order to make this representation more efficient, identical 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 xmlns="http://www.tei-c.org/ns/1.0">
  <rdg wit="#p2 #p3">Chapter 2 <lb/></rdg>
  <rdg wit="#p4">2</rdg>
  <rdg wit="#p5">v <lb/></rdg>
</app>
<app xmlns="http://www.tei-c.org/ns/1.0">
  <rdg wit="#p2">GENTLE INTRODUCTION TO SGML</rdg>
  <rdg wit="#p3">Gentle Introduction to SGML</rdg>
  <rdg wit="#p4 #p5">Gentle Introduction to XML</rdg>
</app>
```

Example 9. Grouping identical readings in a single <rdg> element.

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:

```
<app xmlns="http://www.tei-c.org/ns/1.0">
  <rdg wit="#teiSGML">Chapter 2 <lb/></rdg>
  <rdg wit="#p4">2</rdg>
  <rdg wit="#p5">v <lb/></rdg>
</app>
<app xmlns="http://www.tei-c.org/ns/1.0">
  <rdg wit="#p2">GENTLE INTRODUCTION TO SGML</rdg>
  <rdg wit="#p3">Gentle Introduction to SGML</rdg>
  <rdg wit="#teiXML">Gentle Introduction to XML</rdg>
</app>
```

Example 10. Referring to witness groups in <rdg>.

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 section 5).

SUMMARY

Each variant in a TEI encoded critical edition should be encoded as an apparatus entry, in an <app> element. An apparatus entry contains the different textual variants found in the text witnesses, encoded in different <rdg> (reading) elements. If the edition considers one of the text witnesses as the *base* text, the readings from that witness can be encoded as a lemma instead, in a <lem> element. Each <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 white space 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 diverging 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 <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 for grouping related readings. For example, the readings in the previous example could be grouped as follows:

```
<app xmlns="http://www.tei-c.org/ns/1.0">
  <rdg wit="#teiSGML">Chapter 2 <lb/></rdg>
  <rdgGrp>
```

```
    <rdg wit="#p4">2</rdg>
    <rdg wit="#p5">v <lb/></rdg>
  </rdgGrp>
</app>
<app xmlns="http://www.tei-c.org/ns/1.0">
  <rdgGrp>
    <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>
```

Example 11. Grouping related readings in `<rdgGrp>`.

When you take 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 to be 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 variant readings:

```
<app xmlns="http://www.tei-c.org/ns/1.0">
  <rdg wit="#teiSGML #p4"><app>
    <rdg wit="#teiSGML">Chapter</rdg>
    <rdg wit="#p4"/>
  </app> 2 <app>
    <rdg wit="#teiSGML">
      <lb/>
    </rdg>
    <rdg wit="#p4"/>
  </app></rdg>
  <rdg wit="#p5">v <lb/></rdg>
</app>
```

```
<app xmlns="http://www.tei-c.org/ns/1.0">
  <rdg wit="#teiSGML"><app>
    <rdg wit="#p2">GENTLE INTRODUCTION TO</rdg>
    <rdg wit="#p3">Gentle Introduction to</rdg>
  </app> SGML</rdg>
  <rdg wit="#teiXML">Gentle Introduction to XML</rdg>
</app>
```

Example 12. Grouping readings in nesting <app> elements.

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 so 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 elaborate encoding of the chapter's title in the different text witnesses looks as follows:

```
<app xmlns="http://www.tei-c.org/ns/1.0">
  <rdg wit="#teiSGML #p4"><app>
    <rdg wit="#teiSGML">Chapter</rdg>
    <rdg wit="#p4"/>
```

```
  </app> 2 <app>
    <rdg wit="#teiSGML">
      <lb/>
    </rdg>
    <rdg wit="#p4"/>
  </app></rdg>
  <rdg wit="#p5">v <lb/></rdg>
</app>
<app xmlns="http://www.tei-c.org/ns/1.0">
  <rdg wit="#teiSGML"><app>
    <rdg wit="#p2">GENTLE INTRODUCTION TO</rdg>
    <rdg wit="#p3">Gentle Introduction to</rdg>
  </app> SGML</rdg>
  <rdg wit="#teiXML">Gentle Introduction to XML</rdg>
</app>
```

Example 13. Heterogenous grouping of readings.

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:

```
<app xmlns="http://www.tei-c.org/ns/1.0">
  <rdg wit="#p2 #p3">Chapter</rdg>
  <rdg wit="#p4 #p5"/>
</app>
<app xmlns="http://www.tei-c.org/ns/1.0">
  <rdg wit="#p2 #p3 #p4">2</rdg>
  <rdg wit="#p5">v</rdg>
```

```
</app>
<app xmlns="http://www.tei-c.org/ns/1.0">
  <rdg wit="#p2 #p3 #p5">
    <lb/>
  </rdg>
  <rdg wit="#p4"/>
</app>
<app xmlns="http://www.tei-c.org/ns/1.0">
  <rdg wit="#p2">GENTLE INTRODUCTION TO</rdg>
  <rdg wit="#p3 #p4 #p5">Gentle Introduction to</rdg>
</app>
<app xmlns="http://www.tei-c.org/ns/1.0">
  <rdg wit="#p2 #p3">SGML</rdg>
  <rdg wit="#p4 #p5">XML</rdg>
</app>
```

Example 14. Atomic encoding of minimal stretches of variation.

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):

```
<app xmlns="http://www.tei-c.org/ns/1.0">
  <rdg wit="#p2 #p3" type="substantive">Chapter</rdg>
  <rdg wit="#p4 #p5" type="substantive"/>
</app>
<app xmlns="http://www.tei-c.org/ns/1.0">
  <rdg wit="#p2 #p3 #p4" type="substantive">2</rdg>
  <rdg wit="#p5" type="substantive">v</rdg>
</app>
<app xmlns="http://www.tei-c.org/ns/1.0">
  <rdg wit="#p2 #p3 #p5" type="orthographic">
    <lb/>
  </rdg>
  <rdg wit="#p4" type="orthographic"/>
```

```
</app>
<app xmlns="http://www.tei-c.org/ns/1.0">
  <rdg wit="#p2" type="orthographic">GENTLE INTRODUCTION TO</rdg>
  <rdg wit="#p3 #p4 #p5" type="substantive">Gentle Introduction to</rdg>
</app>
<app xmlns="http://www.tei-c.org/ns/1.0">
  <rdg wit="#p2 #p3" type="substantive">SGML</rdg>
  <rdg wit="#p4 #p5" type="substantive">XML</rdg>
</app>
```

Example 15. Categorising individual readings with `@type` on `<rdg>`.

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. Notice 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:

```
<app xmlns="http://www.tei-c.org/ns/1.0" type="substantive">
  <rdg wit="#p2 #p3 #p4"><app>
    <rdg wit="#p2 #p3">Chapter</rdg>
    <rdg wit="#p4"/>
  </app> 2 </rdg>
  <rdg wit="#p5">v</rdg>
</app>
<app xmlns="http://www.tei-c.org/ns/1.0" type="orthographic">
  <rdg wit="#p2 #p3 #p5">
    <lb/>
  </rdg>
  <rdg wit="#p4"/>
</app>
<app xmlns="http://www.tei-c.org/ns/1.0" type="orthographic">
  <rdg wit="#p2">GENTLE INTRODUCTION TO</rdg>
  <rdg wit="#p3 #p4 #p5">Gentle Introduction to</rdg>
</app>
<app xmlns="http://www.tei-c.org/ns/1.0" type="substantive">
```

```
  <rdg wit="#p2 #p3">SGML</rdg>
  <rdg wit="#p4 #p5">XML</rdg>
</app>
```

Example 16. Categorising variants with `@type` on `<app>`.

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>:

```
<app xmlns="http://www.tei-c.org/ns/1.0">
  <rdg type="substantive">Chapter 2 <lb/></rdg>
  <rdgGrp type="substantive">
    <rdg wit="#p4">2</rdg>
    <rdg wit="#p5">v <lb/></rdg>
  </rdgGrp>
</app>
<app xmlns="http://www.tei-c.org/ns/1.0">
  <rdgGrp type="orthographic">
    <rdg wit="#p2">GENTLE INTRODUCTION TO SGML</rdg>
    <rdg wit="#p3">Gentle Introduction to SGML</rdg>
  </rdgGrp>
  <rdg wit="#teiXML" type="substantive">Gentle Introduction to XML</rdg>
</app>
```

Example 17. Categorising variants with `@type` on `<rdgGrp>`.

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 Module 2: The TEI Header, section 3.3.4). 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:

```
<teiHeader xmlns="http://www.tei-c.org/ns/1.0">
  <!-- ... -->
  <profileDesc>
    <handNotes>
      <handNote xml:id="MSMQ">Michael Sperberg-McQueen</handNote>
      <handNote xml:id="LB">Lou Burnard</handNote>
      <handNote xml:id="SB">Syd Bauman</handNote>
      <handNote xml:id="SR">Sebastian Rahtz</handNote>
    </handNotes>
  </profileDesc>
  <!-- ... -->
</teiHeader>
<!-- ... -->
<app xmlns="http://www.tei-c.org/ns/1.0">
  <rdg wit="#p2" hand="#MSMQ">Chapter 2 <lb/></rdg>
  <rdg wit="#p3">Chapter 2 <lb/></rdg>
  <rdg wit="#p4" hand="#SB">2</rdg>
  <rdg wit="#p5" hand="#SR">v <lb/></rdg>
</app>
<app xmlns="http://www.tei-c.org/ns/1.0">
  <rdg wit="#p2" hand="#LB">GENTLE INTRODUCTION TO SGML</rdg>
  <rdg wit="#p3">Gentle Introduction to SGML</rdg>
  <rdg wit="#p4" hand="#SB">Gentle Introduction to XML</rdg>
  <rdg wit="#p5">Gentle Introduction to XML</rdg>
```

```
</app>
```

Example 18. Providing information on hands for individual readings.

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 this interpretation. This can be expressed in a @resp attribute, which can point to an individual responsible for some aspects of the digital edition, as identified in the TEI header (see Module 2: The TEI Header, section 3.1.1). 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:

```
<teiHeader xmlns="http://www.tei-c.org/ns/1.0">
  <fileDesc>
    <titleStmt>
      <title>The TEI Guidelines, an electronic critical edition</title>
      <editor xml:id="TBEcrew">The TBE crew</editor>
      <!-- ... -->
    </titleStmt>
    <!-- ... -->
  </fileDesc>
  <!-- ... -->
  <profileDesc>
    <handNotes>
      <handNote xml:id="MSMQ">Michael Sperberg-McQueen</handNote>
      <handNote xml:id="LB">Lou Burnard</handNote>
      <handNote xml:id="SB">Syd Bauman</handNote>
      <handNote xml:id="SR">Sebastian Rahtz</handNote>
    </handNotes>
  </profileDesc>
  <!-- ... -->
</teiHeader>
<!-- ... -->
<app xmlns="http://www.tei-c.org/ns/1.0">
  <rdg wit="#p2" hand="#MSMQ" resp="#TBEcrew">Chapter 2 <lb/></rdg>
  <rdg wit="#p3">Chapter 2 <lb/></rdg>
```

```
  <rdg wit="#p4" hand="#SB" resp="#TBEcrew">2</rdg>
  <rdg wit="#p5" hand="#SR" resp="#TBEcrew">v <lb/></rdg>
</app>
<app xmlns="http://www.tei-c.org/ns/1.0">
  <rdg wit="#p2" hand="#LB" resp="#TBEcrew">GENTLE INTRODUCTION TO SGML</rdg>
  <rdg wit="#p3">Gentle Introduction to SGML</rdg>
  <rdg wit="#p4" hand="#SB" resp="#TBEcrew">Gentle Introduction to XML</rdg>
  <rdg wit="#p5">Gentle Introduction to XML</rdg>
</app>
```

Example 19. Stating responsibility for interpretations.

Using attributes on <rdg> holds the danger of overgeneralisation, as in following example:

```
<app xmlns="http://www.tei-c.org/ns/1.0">
        <rdg wit="#p2 #p3" hand="#MSMQ" resp="#TBEcrew">Chapter 2 <lb/></rdg>
  <rdg wit="#p4" hand="#SB" resp="#TBEcrew">2</rdg>
  <rdg wit="#p5" hand="#SR" resp="#TBEcrew">v <lb/></rdg>
</app>
<app xmlns="http://www.tei-c.org/ns/1.0">
  <rdg wit="#p2" hand="#LB" resp="#TBEcrew">GENTLE INTRODUCTION TO SGML</rdg>
  <rdg wit="#p3">Gentle Introduction to SGML</rdg>
        <rdg wit="#teiXML" hand="#SB" resp="#TBEcrew">Gentle Introduction to XML</rdg>
</app>
```

Example 20. Imprecise attribution of reading details on "collapsed" <rdg> elements.

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, which is intended to provide more information about a specific reading in an apparatus entry. It must have a @wit attribute, identifying the specific text witness it provides more information for. In order to anchor it to a specific <rdg> element, a @target attribute can be used to point to 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 xmlns="http://www.tei-c.org/ns/1.0">
  <rdg wit="#p2 #p3" xml:id="rdg1.1">Chapter 2 <lb/></rdg>
  <witDetail target="#rdg1.1" wit="#p2" resp="#TBEcrew">attributed to <ref target="#M
  SMQ">Michael Sperberg-McQueen</ref></witDetail>
  <rdg wit="#p4" hand="#SB" resp="#TBEcrew">2</rdg>
  <rdg wit="#p5" resp="#TBEcrew">v <lb/></rdg>
</app>
<app xmlns="http://www.tei-c.org/ns/1.0">
  <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>
  <witDetail target="#rdg2.3" wit="#p4" resp="#TBEcrew">attributed
   to <ref target="#SB">Syd Bauman</ref></witDetail>
</app>
```

Example 21. Providing more information about specific readings with `<witDetail>`.

The <witDetail> element is a specialised type of <note>, which means it can occur at many places in the document: either inline at the place of the reading needing further specification, or grouped together elsewhere in the document. The TEI Guidelines recommend to place this element inside <app>, immediately after the <lem> or <rdg> element it provides more information for.

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 can be used to point to the identification code of the affected reading(s).

5. Encoding Variation in Texts

After this discussion of the encoding of textual variation itself, 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 attachment 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 attachment

[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 attachment, 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 Module 2: The TEI Header, section 3.2.7):

- @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 method chosen 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 where the variant occurs. 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:

```
<TEI xmlns="http://www.tei-c.org/ns/1.0">
  <teiHeader>
    <!-- ... -->
    <encodingDesc>
      <variantEncoding method="location-referenced" location="internal"/>
    </encodingDesc>
```

```
  <!-- ... -->
</teiHeader>
<text>
  <body>
    <!-- ... -->
    <p>The encoding scheme defined by these Guidelines is<app>
      <rdg wit="#p3">is</rdg>
      <rdg wit="#p4">may be</rdg>
    </app> formulated<app>
      <rdg wit="#p2 #p3"/>
      <rdg wit="#p4">either</rdg>
    </app> as an application of the Extensible<app>
      <rdg wit="#p2 #p3">a system known as the Standard Generalized</rdg>
      <rdg wit="#p4">the ISO Standard Generalized</rdg>
    </app> Markup Language (XML) (Bray et al. (eds.) (2006)). XML is widely used<app>
      <rdg wit="#p2">(SGML).<note place="foot"><bibl><editor>International
       Organization for Standardization</editor>, <title>ISO 8879: Information
       processing--Text and office systems--Standard Generalized Mark-up
       Language (SGML)</title>, ([<pubPlace>Geneva</pubPlace>]: <publisher>I
      SO</publisher>, <date>1986</date>).</bibl> Although widely said to be short for
       the surnames of its progenitors, the official expansion of this abbreviation
       is "Standard Generalized Markup Language."</note> SGML is an international
       standard</rdg>
      <rdg wit="#p3">(SGML). <note place="foot">
        <bibl><editor>International Organization for
         Standardization</editor>, <title>ISO 8879: Information processing
         - Text and office systems - Standard Generalized Markup Language
         (SGML)</title>, ([<pubPlace>Geneva</pubPlace>]: <publisher>I
        SO</publisher>, <date>1986</date>)</bibl>
      </note> SGML is an international standard</rdg>
      <rdg wit="#p4">(SGML)<note place="foot">
        <bibl><editor>International Organization for
         Standardization</editor>, <title>ISO 8879: Information processing
         - Text and office systems - Standard Generalized Markup Language
         (SGML)</title>, ([<pubPlace>Geneva</pubPlace>]: <publisher>I
        SO</publisher>, <date>1986</date>)</bibl>
      </note>or of the more recently developed W3C Extensible Markup Language (X
      ML)<note place="foot">
```

```
      <bibl><editor>World Wide Web Consortium</editor> : <title>Extensible Markup
       Language (XML) 1.0</title>, available from <ref target="http://www.w3.org/TR/
      REC-xml">http://www.w3.org/TR/REC-xml</ref></bibl>
    </note>. Both SGML and XML are widely-used</rdg>
  </app> for the definition of device-independent, system-independent methods of
   storing and processing<app>
    <rdg wit="#p4">storing and processing</rdg>
    <rdg wit="#p2 #p3">representing</rdg>
  </app> texts in electronic form. It is now also the interchange and communication
   format used by many applications on the World Wide Web. In<app>
    <rdg wit="#p2 #p3">. 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</rdg>
    <rdg wit="#p4">; 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</rdg>
  </app> 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<app>
    <rdg wit="#p2">SGML standard, see chapter 30, "TEI Conformance," [in separate
     fascicle]; for a more technical description of the subset of SGML</rdg>
    <rdg wit="#p3">SGML standard, see chapter 28, "Conformance," on page 727. For a
     more technical description of the subset of SGML</rdg>
    <rdg wit="#p4">reader encountering them for the first time how they are actually
     used in the TEI<app>
      <rdg wit="#p2 #p3 #p4">encoding</rdg>
    </app> scheme. Except where the two are explicitly distinguished, references
     to XML in what follows may be understood to apply equally well to the T
    EI usage of SGML. a more technical account of For TEI practice see chapter
     28 <hi>Conformance</hi> ; for a more technical description of the subset of S
    GML</rdg>
  </app> used in<app>
    <rdg wit="#p2 #p3 #p4">by</rdg>
  </app> the TEI scheme. More detailed technical accounts of TEI practice in
   this respect are provided in chapters <hi>23. Using the TEI</hi>, <hi>1. The
   TEI Infrastructure</hi>, and <hi>22. Documentation Elements</hi> of these
   Guidelines.<app>
```

```
        <rdg wit="#p2">, see chapter 39, "Formal Grammar for the TEI-Interchange-Format
         Subset of SGML," [in separate fascicle]</rdg>
        <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>
      <!-- ... -->
    </body>
  </text>
</TEI>
```

Example 22. Encoding variation with an internal "location-referenced" apparatus.

Notice how the apparatus entries can occur anywhere as long as it is inside the text structure (in this case, the <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: @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:[11]

```
<TEI xmlns="http://www.tei-c.org/ns/1.0">
  <teiHeader>
    <!-- ... -->
    <encodingDesc>
      <variantEncoding method="location-referenced" location="external"/>
    </encodingDesc>
    <!-- ... -->
  </teiHeader>
  <text>
    <body>
      <!-- ... -->
```

.....

11 Notice, 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 documentation of the <app> element, and section 2.3.5 The Reference System Declaration of the TEI Guidelines.

```
      <p n="par2">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 <hi>23. Using
       the TEI</hi>, <hi>1. The TEI Infrastructure</hi>, and <hi>22. Documentation
       Elements</hi> of these Guidelines.</p>
      <!-- ... -->
    </body>
    <back>
      <div type="apparatus">
        <p>
          <app loc="par2">
            <rdg wit="#p2 #p3">is</rdg>
            <rdg wit="#p4">may be</rdg>
          </app>
          <app loc="par2">
            <rdg wit="#p2 #p3"/>
            <rdg wit="#p4">either</rdg>
          </app>
          <app loc="par2">
            <rdg wit="#p2 #p3">a system known as the Standard Generalized</rdg>
            <rdg wit="#p4">the ISO Standard Generalized</rdg>
          </app>
          <!-- ... -->
        </p>
      </div>
    </back>
  </text>
</TEI>
```

Example 23. Encoding variation with an external “location-referenced” apparatus.

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 notice the difficulty with apparatus entries encoding *additions* to the base text, as in the second <app> element of following example):

```
<TEI xmlns="http://www.tei-c.org/ns/1.0">
  <teiHeader>
    <!-- ... -->
    <encodingDesc>
      <variantEncoding method="location-referenced" location="external"/>
    </encodingDesc>
    <!-- ... -->
  </teiHeader>
  <text>
    <body>
      <!-- ... -->
      <p n="par2">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 <hi>23. Using
       the TEI</hi>, <hi>1. The TEI Infrastructure</hi>, and <hi>22. Documentation
       Elements</hi> of these Guidelines.</p>
      <!-- ... -->
    </body>
    <back>
      <div type="apparatus">
        <p>
```

```
          <app loc="par2">
            <lem wit="#p3 #p5">is</lem>
            <rdg wit="#p4">may be</rdg>
          </app>
          <app loc="par2">
            <lem wit="#p2 #p3 #p5"/>
            <rdg wit="#p4">either</rdg>
          </app>
          <app loc="par2">
            <lem wit="#p5">the Extensible</lem>
            <rdg wit="#p2 #p3">a system known as the Standard Generalized</rdg>
            <rdg wit="#p4">the ISO Standard Generalized</rdg>
          </app>
          <!-- ... -->
        </p>
      </div>
    </back>
  </text>
</TEI>
```

Example 24. Including the base text in an external “location-referenced” apparatus.

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 Attachment Method

The double end-point attachment 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 attachment apparatus, the apparatus entries occur immediately after their lemma in the transcription of the base text. A specific @from attribute must be used to point exactly 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> element must be inserted in the base text, whose sole purpose is to provide a formal code in its @xml:id attribute. For example, an internal double end-point attachment apparatus for the example in the previous section could look as follows:

```
<TEI xmlns="http://www.tei-c.org/ns/1.0">
  <teiHeader>
    <!-- ... -->
    <encodingDesc>
      <variantEncoding method="double-end-point" location="internal"/>
    </encodingDesc>
    <!-- ... -->
  </teiHeader>
  <text>
    <body>
      <!-- ... -->
      <p>The encoding scheme defined by these Guidelines <anchor xml:id="lem1"/
      >is<app from="#lem1">
        <rdg wit="#p3">is</rdg>
        <rdg wit="#p4">may be</rdg>
      </app> formulated <anchor xml:id="lem2"/><app from="#lem2">
        <rdg wit="#p2 #p3"/>
        <rdg wit="#p4">either</rdg>
      </app> as an application of the <anchor xml:id="lem3"/
      >Extensible<app from="#lem3">
        <rdg wit="#p2 #p3">a system known as the Standard Generalized</rdg>
        <rdg wit="#p4">the ISO Standard Generalized</rdg>
      </app> Markup Language <anchor xml:id="lem4"/>(XML) (Bray et al. (eds.) (2006)). X
      ML is widely used<app from="#lem4">
        <rdg wit="#p2">(SGML).<note place="foot"><bibl><editor>International
          Organization for Standardization</editor>, <title>ISO 8879: Information
          processing--Text and office systems--Standard Generalized Mark-up
```

```
 Language (SGML)</title>, ([<pubPlace>Geneva</pubPlace>]: <publisher>I
SO</publisher>, <date>1986</date>).</bibl> Although widely said to be short for
 the surnames of its progenitors, the official expansion of this abbreviation
 is "Standard Generalized Markup Language."</note> SGML is an international
 standard</rdg>
<rdg wit="#p3">(SGML). <note place="foot">
 <bibl><editor>International Organization for
  Standardization</editor>, <title>ISO 8879: Information processing
  - Text and office systems - Standard Generalized Markup Language
  (SGML)</title>, ([<pubPlace>Geneva</pubPlace>]: <publisher>I
 SO</publisher>, <date>1986</date>)</bibl>
</note> SGML is an international standard</rdg>
<rdg wit="#p4">(SGML)<note place="foot">
 <bibl><editor>International Organization for
  Standardization</editor>, <title>ISO 8879: Information processing
  - Text and office systems - Standard Generalized Markup Language
  (SGML)</title>, ([<pubPlace>Geneva</pubPlace>]: <publisher>I
 SO</publisher>, <date>1986</date>)</bibl>
</note> or of the more recently developed W3C Extensible Markup Language (X
ML)<note place="foot">
 <bibl><editor>World Wide Web Consortium</editor>: <title>Extensible Markup
  Language (XML) 1.0</title>, available from <ref target="http://www.w3.org/TR/
  REC-xml">http://www.w3.org/TR/REC-xml</ref></bibl>
</note>. Both SGML and XML are widely-used</rdg>
</app> for the definition of device-independent, system-independent methods
 of <anchor xml:id="lem5"/>storing and processing<app from="#lem5">
 <rdg wit="#p2 #p3">representing</rdg>
 <rdg wit="#p4">storing and processing</rdg>
</app> texts in electronic form<anchor xml:id="lem6"/>. It is now also the
 interchange and communication format used by many applications on the World Wide
 Web. In<app from="#lem6">
 <rdg wit="#p2 #p3">. 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</rdg>
 <rdg wit="#p4">; 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</rdg>
```

```
      </app> the <anchor xml:id="lem7"/>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<app from="#lem7">
        <rdg wit="#p2">SGML standard, see chapter 30, "TEI Conformance," [in separate
         fascicle]; for a more technical description of the subset of SGML</rdg>
        <rdg wit="#p3">SGML standard, see chapter 28, "Conformance," on page 727. For a
         more technical description of the subset of SGML</rdg>
        <rdg wit="#p4">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 <hi>Conformance</hi> ; for a more technical description of the subset of S
        GML</rdg>
      </app> used <anchor xml:id="lem8"/>in<app from="#lem8">
        <rdg wit="#p2 #p3 #p4">by</rdg>
      </app> the TEI <anchor xml:id="lem9"/><app from="#lem9">
        <rdg wit="#p2 #p3 #p4">encoding</rdg>
      </app>scheme<anchor xml:id="lem10"/>. More detailed technical accounts of
       TEI practice in this respect are provided in chapters <hi>23. Using the
       TEI</hi>, <hi>1. The TEI Infrastructure</hi>, and <hi>22. Documentation
       Elements</hi> of these Guidelines<app from="#lem10">
        <rdg wit="#p2">, see chapter 39, "Formal Grammar for the TEI-Interchange-Format
         Subset of SGML," [in separate fascicle]</rdg>
        <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>
      <!-- ... -->
    </body>
  </text>
</TEI>
```

Example 25. Encoding variation with an internal “double end-point attachment” apparatus.

An external double end-point attachment 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:

```
<TEI xmlns="http://www.tei-c.org/ns/1.0">
  <teiHeader>
    <!-- ... -->
    <encodingDesc>
      <variantEncoding method="double-end-point" location="external"/>
    </encodingDesc>
    <!-- ... -->
  </teiHeader>
  <text>
    <body>
      <!-- ... -->
      <p>The encoding scheme defined by these Guidelines <anchor xml:id="lem1s"/
      >is<anchor xml:id="lem1e"/> formulated <anchor xml:id="lem2"/> as an
       application of the <anchor xml:id="lem3s"/>Extensible<anchor xml:id="lem3e"/
      > Markup Language <anchor xml:id="lem4s"/>(XML) (Bray et al. (eds.)
       (2006)). XML is widely used<anchor xml:id="lem4e"/> for the definition of
       device-independent, system-independent methods of <anchor xml:id="lem5"/
      >storing and processing<anchor xml:id="lem5e"/> texts in electronic
       form<anchor xml:id="lem6s"/>. It is now also the interchange and communication
       format used by many applications on the World Wide Web. In<anchor xml:id="lem6e"/
      > the <anchor xml:id="lem7s"/>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<anchor xml:id="lem7e"/
      > used <anchor xml:id="lem8s"/>in<anchor xml:id="lem8e"/> the T
      EI <anchor xml:id="lem9"/> scheme<anchor xml:id="lem10s"/>. More
       detailed technical accounts of TEI practice in this respect are
```

```
      provided in chapters <hi>23. Using the TEI</hi>, <hi>1. The TEI
      Infrastructure</hi>, and <hi>22. Documentation Elements</hi> of these
      Guidelines<anchor xml:id="lem10e"/>.</p>
    <!-- ... -->
  </body>
  <back>
    <div type="apparatus">
      <p>
        <app from="#lem1s" to="#lem1e">
          <rdg wit="#p3">is</rdg>
          <rdg wit="#p4">may be</rdg>
        </app>
        <app from="#lem2" to="#lem2">
          <rdg wit="#p2 #p3"/>
          <rdg wit="#p4">either</rdg>
        </app>
        <app from="#lem3s" to="#lem3e">
          <rdg wit="#p2 #p3">a system known as the Standard Generalized</rdg>
          <rdg wit="#p4">the ISO Standard Generalized</rdg>
        </app>
        <!-- ... -->
      </p>
    </div>
  </back>
 </text>
</TEI>
```

Example 26. Encoding variation with an external “double end-point attachment” apparatus.

Of course, here too, the lemma of the base text can be explicitly recorded in the apparatus entries as well:

```
<TEI xmlns="http://www.tei-c.org/ns/1.0">
 <teiHeader>
  <!-- ... -->
  <encodingDesc>
   <variantEncoding method="double-end-point" location="external"/>
  </encodingDesc>
```

```
  <!-- ... -->
</teiHeader>
<text>
  <body>
    <!-- ... -->
    <p>The encoding scheme defined by these Guidelines <anchor xml:id="lem1s"/
    >is<anchor xml:id="lem1e"/> formulated <anchor xml:id="lem2"/> as an
     application of the <anchor xml:id="lem3s"/>Extensible<anchor xml:id="lem3e"/
    > Markup Language <anchor xml:id="lem4s"/>(XML) (Bray et al. (eds.)
     (2006)). XML is widely used<anchor xml:id="lem4e"/> for the definition of
     device-independent, system-independent methods of <anchor xml:id="lem5"/
    >storing and processing<anchor xml:id="lem5e"/> texts in electronic
     form<anchor xml:id="lem6s"/>. It is now also the interchange and communication
     format used by many applications on the World Wide Web. In<anchor xml:id="lem6e"/
    > the <anchor xml:id="lem7s"/>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<anchor xml:id="lem7e"/
    > used <anchor xml:id="lem8s"/>in<anchor xml:id="lem8e"/> the T
    EI <anchor xml:id="lem9"/> scheme<anchor xml:id="lem10s"/>. More
     detailed technical accounts of TEI practice in this respect are
     provided in chapters <hi>23. Using the TEI</hi>, <hi>1. The TEI
     Infrastructure</hi>, and <hi>22. Documentation Elements</hi> of these
     Guidelines<anchor xml:id="lem10e"/>.</p>
    <!-- ... -->
  </body>
  <back>
    <div type="apparatus">
      <p>
        <app from="#lem1s" to="#lem1e">
          <lem wit="#p3 #p5">is</lem>
          <rdg wit="#p4">may be</rdg>
        </app>
        <app from="#lem2" to="#lem2">
          <lem wit="#p2 #p3 #p5"/>
          <rdg wit="#p4">either</rdg>
        </app>
        <app from="#lem3s" to="#lem3e">
          <lem wit="#p5">the Extensible</lem>
```

```
          <rdg wit="#p2 #p3">a system known as the Standard Generalized</rdg>
          <rdg wit="#p4">the ISO Standard Generalized</rdg>
        </app>
        <!-- ... -->
      </p>
    </div>
  </back>
 </text>
</TEI>
```

Example 27. Including the base text in an external “double end-point attachment” apparatus.

SUMMARY

The double end-point attachment 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 attachment 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:

```
<TEI xmlns="http://www.tei-c.org/ns/1.0">
  <teiHeader>
    <!-- ... -->
    <encodingDesc>
      <variantEncoding method="parallel-segmentation" location="internal"/>
    </encodingDesc>
    <!-- ... -->
  </teiHeader>
  <text>
    <body>
      <!-- ... -->
      <p>The encoding scheme defined by these Guidelines <app>
        <rdg wit="#p2 #p3 #p5">is</rdg>
        <rdg wit="#p4">may be</rdg>
      </app> formulated <app>
        <rdg wit="#p4">either </rdg>
        <rdg wit="#p2 #p3 #p5"/>
      </app>as an application of <app>
        <rdg wit="#p2 #p3">a system known as the Standard Generalized</rdg>
        <rdg wit="#p4">the ISO Standard Generalized</rdg>
        <rdg wit="#p5">the Extensible</rdg>
      </app> Markup Language <app>
        <rdg wit="#p2">(SGML).<note place="foot"><bibl><editor>International
         Organization for Standardization</editor>, <title>ISO 8879: Information
         processing--Text and office systems--Standard Generalized Mark-up
         Language (SGML)</title>, ([<pubPlace>Geneva</pubPlace>]: <publisher>I
        SO</publisher>, <date>1986</date>).</bibl> Although widely said to be short for
         the surnames of its progenitors, the official expansion of this abbreviation
         is "Standard Generalized Markup Language."</note> SGML is an international
         standard</rdg>
        <rdg wit="#p3">(SGML). <note place="foot">
```

```
      <bibl><editor>International Organization for
       Standardization</editor>, <title>ISO 8879: Information processing
       - Text and office systems - Standard Generalized Markup Language
       (SGML)</title>, ([<pubPlace>Geneva</pubPlace>]: <publisher>I
      SO</publisher>, <date>1986</date>)</bibl>
    </note> SGML is an international standard</rdg>
    <rdg wit="#p4">(SGML)<note place="foot">
      <bibl><editor>International Organization for
       Standardization</editor>, <title>ISO 8879: Information processing
       - Text and office systems - Standard Generalized Markup Language
       (SGML)</title>, ([<pubPlace>Geneva</pubPlace>]: <publisher>I
      SO</publisher>, <date>1986</date>)</bibl>
    </note>or of the more recently developed W3C Extensible Markup Language (X
    ML)<note place="foot">
      <bibl><editor>World Wide Web Consortium</editor> : <title>Extensible Markup
       Language (XML) 1.0</title>, available from <ref target="http://www.w3.org/TR/
      REC-xml">http://www.w3.org/TR/REC-xml</ref></bibl>
    </note>. Both SGML and XML are widely-used</rdg>
    <rdg wit="#p5">(XML) (Bray et al. (eds.) (2006)). XML is widely used</rdg>
  </app> for the definition of device-independent, system-independent methods
   of <app>
    <rdg wit="#p2 #p3">representing</rdg>
    <rdg wit="#p4 #p5">storing and processing</rdg>
  </app> texts in electronic form<app>
    <rdg wit="#p2 #p3">. 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</rdg>
    <rdg wit="#p4">; 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</rdg>
    <rdg wit="#p5">. It is now also the interchange and communication format used by
     many applications on the World Wide Web. In</rdg>
  </app> the <app>
    <rdg wit="#p2">SGML standard, see chapter 30, "TEI Conformance," [in separate
     fascicle]; for a more technical description of the subset of SGML</rdg>
    <rdg wit="#p3">SGML standard, see chapter 28, "Conformance," on page 727. For a
     more technical description of the subset of SGML</rdg>
```

```
      <rdg wit="#p4">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 <hi>Conformance</hi> ; for a more technical description of the subset of S
      GML</rdg>
      <rdg wit="#p5">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</rdg>
    </app> used <app>
      <rdg wit="#p2 #p3 #p4">by</rdg>
      <rdg wit="#p5">in</rdg>
    </app> the TEI <app>
      <rdg wit="#p2 #p3 #p4">encoding</rdg>
      <rdg wit="#p5"/>
    </app> scheme<app>
      <rdg wit="#p2">, see chapter 39, "Formal Grammar for the TEI-Interchange-Format
       Subset of SGML," [in separate fascicle]</rdg>
      <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>
      <rdg wit="#p5">. More detailed technical accounts of TEI practice in this
       respect are provided in chapters <hi>23. Using the TEI</hi>, <hi>1. The
       TEI Infrastructure</hi>, and <hi>22. Documentation Elements</hi> of these
       Guidelines</rdg>
    </app>.</p>
    <!-- ... -->
   </body>
  </text>
</TEI>
```

Example 28. Encoding variation with an (internal) “parallel segmentation” apparatus.

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.

6. Caveats

Until the release of version 2.9.1 of the TEI Guidelines in 2015, <lem> and <rdg> could only contain phrase-level elements. For a very long time, this had caused problems for variants that involve larger structural units. Yet, since version 2.9.1, <lem> and <rdg> can contain chunk-level elements such as , <p>, <ab>, <lg>, and <l>. This addition has greatly increased the use of <lem> and <rdg> for encoding real-life textual variation.

One tough problem remains, however, when textual variation occurs on a structural level. For example, if you look closely at the facsimiles of the TEI Guidelines above (see figures 1, 2, 3, and 4), you’ll notice that there is a paragraph shift at the sentence starting with “Historically, the word markup has been used ”:

- in the **p2** and **p3** versions, this sentence starts the third paragraph
- in the **p4** and **p5** versions, this sentence is part of the second paragraph

This poses a harder encoding problem, as it involves markup itself (i.e., the end and start tag of the third paragraph are the subject of variation). As XML requires proper nesting of elements, this is a problem in any XML representation of this kind of structural variation. Again, two strategies could be followed (none of which is ideal, however):

- Encode structural variants as variant structures. However, this may obscure their alignment.
- Encode structural variants using milestone elements instead of full-blown XML structures. However, depending on your view on the texts, this could be considered a less orthodox approach, as it implies some notion of a base text that determines the encoding of the others.

The first option would compare the individual transcriptions of these text witnesses, some of which spread more or less the same textual contents over 3 paragraphs, while others use only 2 paragraphs. In a parallel segmented apparatus, this might look as follows:

```
<app xmlns="http://www.tei-c.org/ns/1.0">
  <rdg wit="#p2_p #p3_p">
    <p>SGML is an international standard for the description of marked-up electronic
     text. More exactly, SGML is a <app>
      <rdg wit="#p2_p">metalanguage</rdg>
      <rdg wit="#p3_p">
        <hi>metalanguage</hi>
      </rdg>
    </app>, that is, a means of formally describing a language, in this case, a <app>
      <rdg wit="#p2_p">markup language</rdg>
      <rdg wit="#p3_p">
        <hi>markup language</hi>
      </rdg>
    </app>. Before going any further we should define these terms.</p>
  </rdg>
  <rdg wit="#p4 #p5"/>
</app>
<p xmlns="http://www.tei-c.org/ns/1.0"><app>
  <rdg wit="#p4">XML is an extensible markup language used for the description
   of marked-up electronic text. More exactly, XML is a <hi>metalanguage</hi>,
   that is, a means of formally describing a language, in this case, a <hi>markup
   language</hi>.</rdg>
  <rdg wit="#p5">Strictly speaking, XML is a metalanguage, that is, a language used to
   describe other languages, in this case, markup languages.</rdg>
  <rdg wit="#p2_p #p3_p"/>
</app>Historically, the word<!-- ... --></p>
```

Example 29. Encoding structural variants as variant structures.

This approach treats the shifting paragraph as a variant in its own right, that is present in some witnesses (**p2** and **p3**), while absent in the others (**p4** and **p5**). The second apparatus entry then omits the text of **p2** and **p3**, while including the (corresponding) text of **p4** and **p5**. However, as this example illustrates, the alignment of the

corresponding text fragments between both groups of witnesses (those starting a new paragraph and those that don't) is lost: there is no way of telling how the phrases "SGML is an international standard … . More exactly, SGML …" (in **p2** and **p3**) and "XML is an extensible markup language … . More exactly, XML …" correspond. This kind of encoding could be less problematic when *generating* an electronic critical edition (in which case the more complicated apparatus encoding could be generated by an automatic collation routine). When *creating* a digital edition, the construction of such a more complex apparatus entry could be less desirable.

The other solution would be to encode the paragraph break in the **p2** and **p3** versions using an empty "milestone" marker: an empty element that indicates some kind of structural boundary in the text where it occurs, as in this parallel segmented example:

```
<p xmlns="http://www.tei-c.org/ns/1.0"><app>
  <rdg wit="#p2 #p3">SGML is an international standard for the description of marked-up
   electronic text. More exactly</rdg>
  <rdg wit="#p4">XML is an extensible markup language used for the description of
   marked-up electronic text. More exactly</rdg>
  <rdg wit="#p5">Strictly speaking</rdg>
</app>, <app>
  <rdg wit="#p2 #p3">SGML</rdg>
  <rdg wit="#p4 #p5">XML</rdg>
</app> is a <app>
  <rdg wit="#p2 #p5">metalanguage</rdg>
  <rdg wit="#p3 #p4">
    <hi>metalanguage</hi>
  </rdg>
</app>, that is, a <app>
  <rdg wit="#p2 #p3 #p4">means of formally describing a language</rdg>
  <rdg wit="#p5">language used to describe other languages</rdg>
</app>, in this case, <app>
  <rdg wit="#p2">a markup language</rdg>
  <rdg wit="#p3 #p4">a <hi>markup language</hi></rdg>
  <rdg wit="#p5">markup languages</rdg>
</app>. <app>
  <rdg wit="#p2 #p3">Before going any further we should define these
   terms. <milestone unit="p"/></rdg>
  <rdg wit="#p4 #p5"/>
```

```
</app>Historically, the word <!-- ... --></p>
```

Example 30. Encoding structural variation with “milestone” markers.

Since the milestone paragraph boundary marker (`<milestone unit="p"/>`) removes the intrusive XML boundaries, this allows us to compare the text between all versions. However, this implies that the encoding of the third paragraph in the **p2** and **p3** versions is *suppressed*, in contrast to the other paragraphs in these text versions. This could be less a problem when *creating* an electronic critical edition, rather than when generating one. In the latter case, the milestone encoding would reflect a dependency on a base text (that does not have the paragraph break). Moreover, it presupposes some kind of structural alignment prior to the encoding of the individual texts.

SUMMARY

Problems can arise when the variation involves text structures as well, giving rise to problems of overlapping XML structures. This can be avoided by either ignoring the possible alignment of such structures in the apparatus, or paraphrasing some structural boundaries with empty milestone elements.

7. Summary

This tutorial module has focused on the encoding of textual variation in different text witnesses. Although the determination of textual variation itself can depend on the editorial theories for the critical edition, and the TEI Guidelines offer many possibilities to encode textual variation, we’ll conclude with a possible encoding as a critical edition of the text samples we used in this tutorial module. In this example, we chose for a parallel segmented internal apparatus, which could look as follows:

```
<TEI xmlns="http://www.tei-c.org/ns/1.0">
  <teiHeader>
    <fileDesc>
      <!-- ... -->
      <sourceDesc>
        <listWit>
          <witness xml:id="p2">
```

```
          <bibl><editor>Sperberg-McQueen, M.</editor>; <editor>Burnard, L.</editor>
           (eds.). <title>TEI P2 Guidelines for the Encoding and Interchange of
           Machine Readable Texts Draft P2</title> (published serially 1992-1993);
           Draft Version <date when="1993-04-02">2 of April 1993</date>: <extent>19
           chapters</extent>. Available from <ptr target="https://tei-c.org/Vault/Vault
          -GL.html"/> (accessed October 2008)</bibl>
        </witness>
        <witness xml:id="p3">
          <bibl><editor>Sperberg-McQueen, C.M.</editor>; <editor>Burnard,
           L.</editor> (eds.). <title>Guidelines for Electronic Text Encoding and
           Interchange. TEI P3. Revised reprint.</title> <publisher>Text Encoding
           Initiative</publisher>: <pubPlace>Oxford</pubPlace>, <pubPlace>Providence</pubPlace>, <pubP
        </witness>
        <witness xml:id="p4">
          <bibl><editor>Sperberg-McQueen, C.M.</editor>; <editor>Burnard,
           L.</editor> (eds.). <title>TEI P4: Guidelines for
           Electronic Text Encoding and Interchange. XML-compatible
           edition.</title> <publisher>Text Encoding Initiative
           Consortium</publisher>: <pubPlace>Oxford</pubPlace>, <pubPlace>Providence</pubPlace>, <pubP
        </witness>
        <witness xml:id="p5">
          <bibl><editor>Sperberg-McQueen, C.M.</editor>; <editor>Burnard,
           L.</editor> (eds.). <title>TEI P5: Guidelines for
           Electronic Text Encoding and Interchange. Revised and
           re-edited.</title> <publisher>Text Encoding Initiative
           Consortium</publisher>: <pubPlace>Oxford</pubPlace>, <pubPlace>Providence</pubPlace>, <pubP
        </witness>
      </listWit>
    </sourceDesc>
  </fileDesc>
  <!-- ... -->
  <encodingDesc>
    <variantEncoding method="parallel-segmentation" location="internal"/>
  </encodingDesc>
  <!-- ... -->
</teiHeader>
<text>
  <body>
```

```
<app>
  <rdg wit="#p2">
    <pb n="2"/>
  </rdg>
  <rdg wit="#p3 #p4">
    <pb n="13"/>
  </rdg>
  <rdg wit="#p5">
    <pb n="xxxi"/>
  </rdg>
</app>
<head><app>
  <rdg wit="#p2 #p3">Chapter 2 <lb/></rdg>
  <rdg wit="#p5">v <lb/></rdg>
  <rdg wit="#p4">2</rdg>
</app> A <app>
  <rdg wit="#p2">GENTLE INTRODUCTION TO SGML</rdg>
  <rdg wit="#p3">Gentle Introduction to SGML</rdg>
  <rdg wit="#p4 #p5">Gentle Introduction to XML</rdg>
</app></head>
<app>
  <rdg wit="#p4">
    <note type="disclaimer">As originally published in previous editions of the
     Guidelines, this chapter provided a gentle introduction to 'just enough' S
    GML for anyone to understand how the TEI used that standard. Since then, the
     Gentle Guide seems to have taken on a life of its own independent of the
     Guidelines, having been widely distributed (and flatteringly imitated) on
     the web. In revising it for the present draft, the editors have therefore
     felt free to reduce considerably its discussion of SGML-specific matters, in
     favour of a simple presentation of how the TEI uses XML.</note>
  </rdg>
  <rdg wit="#p2 #p3 #p5"/>
</app>
<p>The encoding scheme defined by these Guidelines <app>
  <rdg wit="#p2 #p3 #p5">is</rdg>
  <rdg wit="#p4">may be</rdg>
</app> formulated <app>
  <rdg wit="#p4">either </rdg>
```

```
  <rdg wit="#p2 #p3 #p5"/>
</app>as an application of <app>
  <rdg wit="#p2 #p3">a system known as the Standard Generalized</rdg>
  <rdg wit="#p4">the ISO Standard Generalized</rdg>
  <rdg wit="#p5">the Extensible</rdg>
</app> Markup Language <app>
  <rdg wit="#p2">(SGML).<note place="foot"><bibl><editor>International
   Organization for Standardization</editor>, <title>ISO 8879: Information
   processing--Text and office systems--Standard Generalized Mark-up
   Language (SGML)</title>, ([<pubPlace>Geneva</pubPlace>]: <publisher>I
  SO</publisher>, <date>1986</date>).</bibl> Although widely said to be short for
   the surnames of its progenitors, the official expansion of this abbreviation
   is "Standard Generalized Markup Language."</note> SGML is an international
   standard</rdg>
  <rdg wit="#p3">(SGML). <note place="foot">
    <bibl><editor>International Organization for
     Standardization</editor>, <title>ISO 8879: Information processing
     - Text and office systems - Standard Generalized Markup Language
     (SGML)</title>, ([<pubPlace>Geneva</pubPlace>]: <publisher>I
    SO</publisher>, <date>1986</date>)</bibl>
  </note> SGML is an international standard</rdg>
  <rdg wit="#p4">(SGML)<note place="foot">
    <bibl><editor>International Organization for
     Standardization</editor>, <title>ISO 8879: Information processing
     - Text and office systems - Standard Generalized Markup Language
     (SGML)</title>, ([<pubPlace>Geneva</pubPlace>]: <publisher>I
    SO</publisher>, <date>1986</date>)</bibl>
  </note>or of the more recently developed W3C Extensible Markup Language (X
  ML)<note place="foot">
    <bibl><editor>World Wide Web Consortium</editor> : <title>Extensible Markup
     Language (XML) 1.0</title>, available from <ref target="http://www.w3.org/TR/
    REC-xml">http://www.w3.org/TR/REC-xml</ref></bibl>
  </note>. Both SGML and XML are widely-used</rdg>
  <rdg wit="#p5">(XML) (Bray et al. (eds.) (2006)). XML is widely used</rdg>
</app> for the definition of device-independent, system-independent methods
 of <app>
  <rdg wit="#p2 #p3">representing</rdg>
  <rdg wit="#p4 #p5">storing and processing</rdg>
```

```
</app> texts in electronic form<app>
  <rdg wit="#p2 #p3">. 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</rdg>
  <rdg wit="#p4">; 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</rdg>
  <rdg wit="#p5">. It is now also the interchange and communication format used by
   many applications on the World Wide Web. In</rdg>
</app> the <app>
  <rdg wit="#p2">SGML standard, see chapter 30, "TEI Conformance," [in separate
   fascicle]; for a more technical description of the subset of SGML</rdg>
  <rdg wit="#p3">SGML standard, see chapter 28, "Conformance," on page 727. For a
   more technical description of the subset of SGML</rdg>
  <rdg wit="#p4">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 <hi>Conformance</hi> ; for a more technical description of the subset of S
  GML</rdg>
  <rdg wit="#p5">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</rdg>
</app> used <app>
  <rdg wit="#p2 #p3 #p4">by</rdg>
  <rdg wit="#p5">in</rdg>
</app> the TEI <app>
  <rdg wit="#p2 #p3 #p4">encoding</rdg>
  <rdg wit="#p5"/>
</app> scheme<app>
  <rdg wit="#p2">, see chapter 39, "Formal Grammar for the TEI-Interchange-Format
   Subset of SGML," [in separate fascicle]</rdg>
  <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>
```

```
  <rdg wit="#p5">. More detailed technical accounts of TEI practice in this
   respect are provided in chapters <hi>23. Using the TEI</hi>, <hi>1. The
   TEI Infrastructure</hi>, and <hi>22. Documentation Elements</hi> of these
   Guidelines</rdg>
</app>.</p>
<p><app>
  <rdg wit="#p2 #p3">SGML is an international standard for the description of
   marked-up electronic text. More exactly</rdg>
  <rdg wit="#p4">XML is an extensible markup language used for the description of
   marked-up electronic text. More exactly</rdg>
  <rdg wit="#p5">Strictly speaking</rdg>
</app>, <app>
  <rdg wit="#p2 #p3">SGML</rdg>
  <rdg wit="#p4 #p5">XML</rdg>
</app> is a <app>
  <rdg wit="#p2 #p5">metalanguage</rdg>
  <rdg wit="#p3 #p4">
    <hi>metalanguage</hi>
  </rdg>
</app>, that is, a <app>
  <rdg wit="#p2 #p3 #p4">means of formally describing a language</rdg>
  <rdg wit="#p5">language used to describe other languages</rdg>
</app>, in this case, <app>
  <rdg wit="#p2">a markup language</rdg>
  <rdg wit="#p3 #p4">a <hi>markup language</hi></rdg>
  <rdg wit="#p5">markup languages</rdg>
</app>. <app>
  <rdg wit="#p2 #p3">Before going any further we should define these
   terms. <milestone unit="p"/></rdg>
  <rdg wit="#p4 #p5"/>
</app>Historically, the word <app>
  <rdg wit="#p2 #p5">markup</rdg>
  <rdg wit="#p3 #p4">
    <hi>markup</hi>
  </rdg>
```

```
      </app> has been used to describe annotation or other marks within a text intended
       to instruct a compositor or typist how a particular passage should be printed or
       laid out. Examples include wavy underlining to indicate boldface, special symbols
       for passages to be omitted or printed in a particular <app>
        <rdg wit="#p2 #p3 #p4">font</rdg>
        <rdg wit="#p5">font,</rdg>
      </app> and so forth. As the formatting and printing of texts was automated, the
       term was <app>
        <rdg wit="#p2">extend-ed</rdg>
        <rdg wit="#p3 #p4 #p5">extended</rdg>
      </app> to cover all sorts of special <app>
        <rdg wit="#p2">markup codes</rdg>
        <rdg wit="#p3">
          <hi>markup codes</hi>
        </rdg>
        <rdg wit="#p4 #p5">codes</rdg>
      </app> inserted into electronic texts to govern formatting, printing, or other
       processing.</p>
    </body>
  </text>
</TEI>
```

Example 31. A parallel segmented encoding of the different text witnesses of the example.

8. What's Next?

You have reached the end of this tutorial module covering an introduction to critical editing with TEI. You can now either

- proceed with other TEI by Example modules
- have a look at the examples section for the critical editing 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!

BIBLIOGRAPHY

Vanhoutte, Edward, and Ron Van den Branden. 2009. “Describing, Transcribing, Encoding, and Editing Modern Correspondence Material: a Textbase Approach.” *Literary and Linguistic Computing* 24 (1): 77–98. 10.1093/llc/fqn035.