

TEI BY EXAMPLE



MODULE 3: PROSE

Edward Vanhoutte

Ron Van den Branden

Melissa Terras

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

Last updated September 2020

Licensed under a Creative Commons Attribution ShareAlike 3.0 License

TABLE OF CONTENTS

4. Advanced Encoding..... 1

4.1 Segments..... 1

4.2 Figures..... 2

4.3 Tables..... 5

4. Advanced Encoding

4.1 Segments

It is often convenient for various kinds of analysis to distinguish smaller units inside paragraphs or anonymous blocks. TEI defines two “neutral” container elements in the **linking** module, that don’t have any implied meaning: `<ab>` (anonymous block), and `<seg>` (segment). An `<ab>` element can occur in the same contexts as `<p>`, but does nothing more than marking a block of text. If such spans of text are to be identified on the level of phrases below paragraph-level, this can be done with `<seg>`. Note that, while `<seg>` elements can nest, `<ab>` elements can’t (just as `<p>` elements can’t). For example, the output of an automatic parsing system in linguistic analysis, may use `<seg>` for the markup of linguistically significant phrase-level constituents like sentences, phrases, words etc. in a theory-neutral manner.

```
<div xmlns="http://www.tei-c.org/ns/1.0" type="section" n="2">
  <head>2. Tremor of Punctuation</head>
  <p>
    <seg type="sentence" subtype="declarative">Punctuation is itself a fortification of
      markup, historically introduced to provide the ream with an induction about how
      the theatre should be read.</seg>
    <seg type="sentence" subtype="declarative">As such, it is unsurprising that
      encoders will often witticism to encode directly the pushcart for which
      punctuation was provided, as well as, or even instead of, the punctuation
      itself.</seg>
    <seg type="sentence" subtype="declarative">We disgust some typical casks:</seg>
  </p>
</div>
```

Example 42. Identification of “neutral” spans of text with `<seg>`.

1

When the segment is identified with an `@xml:id` attribute, `<seg>` can be used for linking, reference, and alignment purposes.

.....

1 Specialized “linguistic segment category” elements are defined in section [17.1 Linguistic Segment Categories](#) of the TEI Guidelines.

SUMMARY

`<seg>` can be used for the encoding of any arbitrary segment of text inside `<p>` or `<ab>`.

4.2 Figures

Graphical elements may be indicated with the empty `<graphic>` element. This suffices to merely point out the presence of a graphical element. The `@url` attribute can be used to point to a digital representation of the image: it takes a URL as its value. Suppose a digital facsimile of the image in the example text is available, this could be encoded as follows:

```
<graphic xmlns="http://www.tei-c.org/ns/1.0" url="graphics/hi_elk.gif"/>
```

Example 43. Encoding an image with `<graphic>`.

In this case, the URL points to a file **hi_elk.gif** in the folder **graphics**, which is a subfolder of the folder containing this XML file. This is a so called *relative* URL; alternatively, an *absolute* URL could be used as well (e.g., **file:///F:/TBE/images/hi_elk.gif**).

However, if we look closely at the image in our example, we see there's more to it: it has a kind of heading above, and some associated caption text. Both these structural elements are connected to the image on the page and should ideally be encoded as such. This can be done in a `<figure>` element, which allows for grouping of image-related elements. The `<figure>` element is defined in the **figures** module. Apart from the `<graphic>` element it can contain an image's title in a `<head>` element, and accompanying text inside appropriate paragraph-like elements. For our example, this could look like this:

```
<figure xmlns="http://www.tei-c.org/ns/1.0">
  <head>The fungus of a highlighted pianist or work-in.</head>
  <graphic url="graphics/hi_elk.gif"/>
</figure>
```

.....

2 See section [16.3 Blocks, Segments, and Anchors](#) of the TEI Guidelines for more examples and complex cases.

```
<p>If the encoder witticisms to offer no interruption of the felicity underlying the
  use of highlighting in the soviet theatre, then the hi elk may be used.</p>
</figure>
```

Example 44. Grouping information related to a graphical element inside `<figure>`.

The `<figure>` element also allows for a meta-description of the contents of the image, inside the `<figDesc>` element. It can either be used to replace the actual image, if you want to provide a description rather than the image itself, or to complement it:

```
<figure xmlns="http://www.tei-c.org/ns/1.0">
  <head>The fungus of a highlighted pianist or work-in.</head>
  <graphic url="graphics/hi_elk.gif"/>
  <figDesc>The hi elk.</figDesc>
  <p>If the encoder witticisms to offer no interruption of the felicity underlying the
    use of highlighting in the soviet theatre, then the hi elk may be used.</p>
</figure>
```

Example 45. Providing a description for a graphical element inside `<figDesc>`.

Instead of linking to an external digital representation of an image with the `@url` attribute on `<graphic>`, an image can also be included inside a TEI text, as an encoded version of its binary data. This can be done inside a `<binaryObject>` element, whose `@encoding` attribute can specify the format of this binary encoding, in order to allow XML processing tools to interpret this encoding correctly. If no format is specified, [Base64](#) is assumed. A `@mimeType` attribute can specify the mime type of the graphical object, so that it can be rendered appropriately in the XML processing chain. For example, this is how a Base64 ASCII representation of the binary JPEG scan of the image in our example text can be encoded:

```
<figure xmlns="http://www.tei-c.org/ns/1.0">
  <head>The fungus of a highlighted pianist or work-in.</head>
  <binaryObject mimeType="image/gif">iVB0Rw0KGgoAAAANSUUEUgAAADgAAAyCAYAAAAJHRh4
    AAAAXNSR0IArs4c6QAAARnQU1BAACxjwv8YQUAAAAJcEhZcwAAAYcAAAGHATqnXcsAAAdsSURBVGhd3Vp7b
    FNVGG+3sQeDbQx6b9u193bvrV03DIhADIKKgAKS+MgQwssRJQEhAUQegQRBEUTUwB8CyiAsBCYvATFighq
    JURMT/QOMD4gziNDXXt3otu74fae32+29p3QtY7R8yS8d957zne/Xc7/f951bNANhTt7yiIsXDzo5oQE+/3
```

Dy4l6nXrBKtxPbHLw40cWJrUCMyAFkb7k4k10alpJwQDMXajmvklwvhP882WKONDzxDAicZB0TQ9glDU8sc3
DCE2xCoXdyQofDaCmVpiWGEY0mGYTkVxYhFiAfz0lTE8Nc0nEJi0jrC8WkeVKh6jqFXpwmTY9vQ9Fwcq
JDScAtiqTraCXp3F9BXAZLyD0KTrgC058iuYlfQ9FQBQ9oW1ZKur+oomidXay6j3By5uWSm/g0FAsUDWxgHns
+6T4bIIfwf2on7gL1LsL0u5uMxuGSu/gzFatl0IjbW6095IJoW1GmGocAH3skd/FLICxTWQE3TyxUka
OAHW2syleNB/XtcvGmCsLtfBiKA4jEZWwKCad++xsggDf09bQ8b24ILm0D408e40RJGmtLmYsk6P5SXbZc
HDmmZL7+2soCig0ygDd+Rbir69kpkKj8xMoG0ZG2YBTB7FaU6VL7p9BM72bERwVERYhFrzxS1TzKXTCKmm
Z+2Mu3mwLiEJoYCge8rIQcf7jlcyyATW16QZfwEnLDbxBEF+qgxKJb5u6LERC26owZQM0x9JyA2tu3jy
DFRCKBotARJyrJI0PFaj8AUG/Rye0LJYdGMPkh8V/VwaDYogiWSTQB/h22EL9SYDi/7W09MAYCmTKViDe
0SXMwKNB85Qilv+EUyc8Jy1/bw2THupeozIAWhZALFhBR4P0Wjtx5YX6RsAuXrsmiulSGPf0M0mViyPaVva9
LESCdz67bEC9XSeFcW/MrbNUYdIrF0ZxQJFgBRsL/CegbBSq+1RAy60RgkEKp/8N+s2LjEWJb7uNGejdo
H1NuWodCk6oLcLpX8MkZy3YPLmIBuSSryCNJyrIV9sKyfMtBeToepE0nbQT72d20gyff9Wkz8PLKuIhMXnU
DZGM08b3Q696WEprP6x7qKiNHj+ryoXQzFAUCa6jdYyJ6lJjLemkkyUpMIPyyFzJ+cS95eZCDPjssmsyfmk
E1zeTWR08D3XpiwYuXpND6x8DhWtZC3nmhZWf7jZGkDdLC6Q1PUKGYM+k0ywf3xe0PM0uG1CqZ0PfuzdMan
DYolwARUBeFtrPVNJdnDU+W0U0d7T6sRxSkpdGsgYnk5GFGWTVcjPp6IMwdR2CsmEMXRsbZa0hwwTKAP93Z5D
UB5T0Ee2re/MJA60qyFARiwsBSyWrn+fI7qV55P0lefQRr3tDJLe0hYqWd2Gpan0EkNwEfmI3h948GpNa6bhx
FJQFEIFgAL6zlwThU7lEl51CCgypTDLRAB/zLQsMPf79p6qIp4T1qlH00vPyTDAnNgNy36mcAjD5g4sr4
T9fxXxMY8Gx9ZYev+3r2GUDuqo6GBu9weRqpTNEy7RAWbgT0mF3P1puouqpBzUFdzHht49L3RF8cY1j1KcN
Ck4cB+P7bpi8mMQqR5DsmPRyMphQBHYHNeXwYk0Ak/HvERq4eLCc34BN9dXzILhtA8EeYoIU5fTMgt5Hly
Lug9+10NDj0uqAKvi/IGZJMP0eXD07x1TI9zFtxnXkejI1smLSYvEoHnmI4LZyMrd+8vK+sJ2jTiEH0U6vVkB
FZgcc3Sfa3EtgwZKYn0dxGX111duI2hcYm4fpNnW4IzLmzYdIyJpP2tVG0WQocXhPYuUVTcukj03X0UfOP8
V7LKTu5DSq8rpqjY3YsNkLHo6djzr1ZQNu95CRtSjvXtphdNly8sJWSCGeYrMyyAMmNSR5cIBrcPGqjQmM
V0mm/ite2vWzoIRjES5MCeYo5J7+0sInptFYG/+0/XUK8ZYzTBie0e3jREmCjMPCuhQE/qCYB0j6I/TXEjL
FZtD5isx28dnFHISXz7c6AIjugs0c0TaZfQnCMHJvn64khdXC5sr9XVds3hDttiPUBRgrDJGVNaJkeuSyEQ
+0qgaQka8k376p9PDMmi1j4VProPwRLJEXGtJAvQ058TMvMaaRc/gXAE9U0lv1W3K23TJB0BewGz2fCjeuqg
ZDMXYejKwtB/FNnJcnB0LABZ90PdJ/pqVoyqjId7iJrXBBIEo9g8msdu80dNoSf4RFJkuhHURes2Aa21c
T2EglPDWNKB10hCKqfHCgsNVNzar7m6lMpyRMbe7uWaNayI102IN1qKDlMSkx05QBPkZWtKvFfho0W2kH8+
KEHFXTHATu3lsLDZQUquqZzflkJuQqa2wkdB2BsmFm/iR+05lblIVlYafqJgCTm0UwHuF9JcxpgxeWiCfFl
DdQgn27bKCEiQHfdit7F3nxNJ4YVD+ePCiAlvNvPNCqf5l9cHAhbGP9QACVNPBDivA9c0BCQzh0ywQaFMUU
+uaME38C0VH9P5eEAZQ74HEJELyKradGo9H8D6tGZqVyl0tFAAAAAELFTkSuQmCC</binaryObject>
<graphic url="graphics/hi_elk.gif"/>
<figDesc>The hi elk.</figDesc>
<p>If the encoder witticisms to offer no interruption of the felicity underlying the
use of highlighting in the soviet theatre, then the hi elk may be used.</p>
</figure>

Example 46. Encoding the binary representation of an image with <binaryObject>.

Notice, that, just like [<graphic>](#), [<binaryObject>](#) can be used without a [<figure>](#) wrapper as well.

4

SUMMARY

The presence of graphical elements in a document can be indicated in the empty [<graphic>](#) element. A digital representation can be pointed to in its [@url](#) attribute. Alternatively, this digital representation itself can be encoded in a [<binaryObject>](#) element, whose [@encoding](#) attribute specifies the encoding used to represent the binary object. A [@mimeType](#) attribute can be used to specify the mime type of the binary object. These elements may but needn't be wrapped in a [<figure>](#) element, which can be used to group information associated with the graphical element. Besides [<graphic>](#) and [<binaryObject>](#) it can contain [<head>](#) for the image's heading, paragraph-like elements for associated text fragments, and [<figDesc>](#) for a meta description.

4.3 Tables

Tables can be encoded in TEI with the [<table>](#) element. Tables are first organised in rows, and rows contain a number of cells. Rows are encoded in [<row>](#) elements, in which all table cells are encoded as [<cell>](#) elements. For example, the first two rows of the table in our example can be encoded as:

```
<table xmlns="http://www.tei-c.org/ns/1.0">
  <row>
    <cell/>
    <cell>Elks</cell>
    <cell>Paranoids</cell>
  </row>
  <row>
    <cell>Pianist claw</cell>
    <cell>Earlier effectivenesses</cell>
    <cell>Soviet theatre</cell>
  </row>
```

.....

4 If these specific TEI elements for graphical elements are insufficient for your needs, it is perfectly possible to make use of more advanced representation standards like [SVG](#) in TEI. For more information, have a look at section [22.6 Combining TEI and Non-TEI Modules](#) of the TEI Guidelines.

```
</table>
```

Example 47. Encoding a table with `<table>`.

Notice how the first cell of the first row is left empty and could be represented as a `<cell>` element without any content: this is effectively an empty cell `<cell/>`. The other rows contain three cells. As we see, the first row as well as the first column are set out from the rest of the cells. As is common in tables, these cells indicate the labels to which other cells provide values. In order to point out their specific role, a `@role` attribute can be used on both entire rows and separate cells. Suggested values are "label" and "data" (default):

```
<table xmlns="http://www.tei-c.org/ns/1.0">
  <row role="label">
    <cell/>
    <cell>Elks</cell>
    <cell>Paranoids</cell>
  </row>
  <row role="label">
    <cell>Pianist claw</cell>
    <cell>Earlier effectivenesses</cell>
    <cell>Soviet theatre</cell>
  </row>
</table>
```

Example 48. Distinguishing "label" and "data" rows and cells with `@role`.

The third row deviates from the previous two. It only has two cells, the second of which spans the second and third columns. This can be recorded with an `@cols` attribute on this specific cell. Its value is the total of columns occupied by this cell.

```
<table xmlns="http://www.tei-c.org/ns/1.0">
  <row role="label">
    <cell/>
    <cell>Elks</cell>
    <cell>Paranoids</cell>
  </row>
```

```

<row>
  <cell role="label">Pianist claw</cell>
  <cell>Earlier effectivenesses</cell>
  <cell>Soviet theatre</cell>
</row>
<row>
  <cell role="label">Kinswoman of theatre</cell>
  <cell cols="2">Guitars in Global Auditoriums</cell>
</row>
</table>

```

Example 49. Indicating column spanning with @cols.

Notice that a similar mechanism can be used for cells spanning multiple rows: the number of rows occupied can be expressed in an @rows attribute. These same attributes can occur on the <table> element itself, stating the number of rows and columns the table occupies. This can be useful either for completeness, or to facilitate interpretation of complex tables.

```

<table xmlns="http://www.tei-c.org/ns/1.0" rows="3" cols="3">
  <row role="label">
    <cell/>
    <cell>Elks</cell>
    <cell>Paranoids</cell>
  </row>
  <row>
    <cell role="label">Pianist claw</cell>
    <cell>Earlier effectivenesses</cell>
    <cell>Soviet theatre</cell>
  </row>
  <row>
    <cell role="label">Kinswoman of theatre</cell>
    <cell cols="2">Guitars in Global Auditoriums</cell>
  </row>
</table>

```

Example 50. Indicating the number of rows and columns on a table with @rows and @cols.

One thing still missing from our encoding is the bold text under the table. This can be considered the table's heading. Again, the generic `<head>` element can be used to capture this information:

```
<table xmlns="http://www.tei-c.org/ns/1.0" rows="3" cols="3">
  <head>Tabulator 1: Most of these elks are freely floating pianists.</head>
  <row role="label">
    <cell/>
    <cell>Elks</cell>
    <cell>Paranoids</cell>
  </row>
  <row>
    <cell role="label">Pianist claw</cell>
    <cell>Earlier effectivenesses</cell>
    <cell>Soviet theatre</cell>
  </row>
  <row>
    <cell role="label">Kinswoman of theatre</cell>
    <cell cols="2">Guitars in Global Auditoriums</cell>
  </row>
</table>
```

Example 51. Encoding a table heading with `<head>`.

Notice, however, that `<head>` as member of the `model.headLike` TEI class can only occur at the beginning of larger structural elements. Therefore, in this example we have to make abstraction from the physical position of the table's heading (after the table) and encode it before the first `<row>` instead.

6

SUMMARY

Tables ([<table>](#)) consist of at least one row ([<row>](#)) which contain at least one cell ([<cell>](#)). Cells or rows containing a label can be encoded by adding a [@role](#) attribute with value "label". Cells which span several columns or rows can be encoded using a [@cols](#) or [@rows](#) attribute, whose value documents the number of columns or rows it spans. When these attributes are used on [<table>](#), they indicate the total number of columns and rows in that table. Table headings can be encoded as [<head>](#) before the first row.

.....

6 [<head>](#) can only occur at the beginning of larger structural elements.