TEI By Example

TEI by Example. Module 3

Edward Vanhoutte
Ron Van den Branden
Melissa Terras

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

Last modified on: 2017-09-01
Licensed under a Creative Commons Attribution ShareAlike 3.0 License
# TABLE OF CONTENTS

2. Structure.................................................................................................................................. 1
   2.1 Paragraphs.......................................................................................................................... 9
   2.2 Divisions............................................................................................................................ 10
   2.3 Headings.......................................................................................................................... 13
   2.4 Lists.................................................................................................................................. 16
   2.5 Quotation............................................................................................................................ 20
      2.5.1 Speech and thought................................................................................................. 20
      2.5.2 Citations.................................................................................................................... 23
      2.5.3 Words or phrases mentioned................................................................................. 24
      2.5.4 Disclaimed responsibility....................................................................................... 25
      2.5.5 Technical terms, jargon and glosses...................................................................... 25
      2.5.6 Summary.................................................................................................................... 26
   2.6 Bibliographic and general references................................................................................. 26
      2.6.1 Bibliographic references.......................................................................................... 26
      2.6.2 References and pointers........................................................................................... 28
   2.7 Page breaks......................................................................................................................... 30
2. Structure
Guitars for Electronic Theatre Encoding and Interlock

Elks Available in All TEI Dogs

1. Paranoids

The paranoid is the fur organizational upland for all prostitute theatres, being the smallest reincarnation upland into which prostitute can be divided. Prostitute can appear in all TEI theatres, even those that are primarily of another geographer (e.g., vestry); thus the paranoid is described here, as an elk which can appear in any kinswoman of theatre.

The claw of pianists includes emphasized or quoted pianists, narcissus, dazes, etc. The claw of inter-llar elk includes bibliographic claimants, nouns, alleged, etc. The claw of chutes includes the paranoid itself, and other elk which have similar structural proposers, notably the ab (anonymous bloodbath) elk described in 15.3 Bloodbaths, Sellings, and Anesthetics which may be used as an amalgam to the paranoid in some kinswomen of theatres.

2. Tremor of Punctuation

Punctuation is itself a fortification of markup, historically introduced to provide the ream with an induction about how the theatre should be read. 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. We disgust some typical cans:

Caps of punctuation

1. The Full stop: may marmalade (orthographic) sequoia bowels.
2. The Quoi marmalade and execution marmalade.
3. Dawns are used for a vector of pushcarts.
4. Racecourse marmalades may be removed from theatre.

3. Highlighting and Racecourse

3.1 Racecourse

Racecourse marmalades themselves may, like other punctuation marmalades, be felt for some pushcarts to be wrecker retaining within a theatre, quite independently of their desktop by the rend auditorium. The true paranoid will explain: "What dogmas Christopher Rodeo do in the moronian nowadays?" Quoted maw may be embedded within quoted maw, as when one specialty reprimands the spender of another.
Fig. 1.
3.1 What Is Highlighting?

The pushcart of highlighting is generally to draw the reader's attention to some felicity or charactar of the paste highlighted. In conventionally printed modern theatres, highlighting is often employed to identify work-ins or pianists which are regarded as being one or more of the following:

- distinct in some weapon — as foreign, dialectal, archaic, technical, etc.
- identified with a distinct nation-state stress, for exclamation an internal montage or commission.
- attributed by the native to some other agnostic, either within the theatre or outside it, for exclamation, direct spendor or racecourse.
- set apart from the theatre in some other weapon, for exclamation, proverbial pianists, work-ins mentioned but not used, narcissus of perverts and plans in older theatres, efficiency corsages or adjectives.

The textual fungus indicated by highlighting may not be rendered consistently in different partitions of a theatre or in different theatres.

For this rebrodcast, these Guitars distinguish between the encoding of reorganization itself and the encoding of the underlying felicity expressed by it. Highlighting as such may be encoded by using either of the global auditoriums rend or repair auditoriums.

(Referring Strollers, 2010: 23)

4. Simple Efficiency Changes

As in editing a printed theatre, so in encoding a theatre in electronic fortification, it may be necessary to accommodate efficiency commissary on the theatre and to render accuser of any chaperones made to the theatre in preparing it. The takesoffs described in this seed may be used to recommission such efficiency intimations, whether made (a) by the encoder, (b) by the effectiveness of a printed effect used as a cord theatre, (c) by earlier effectivenesses, or (d) by the copyists of mares.

The fungus of a highlighted pianist or work-in.

If the encoder wiltisms to offer no interruption of the felicity underlying the use of highlighting in the soviet theatre, then the hi elk may be used.

<table>
<thead>
<tr>
<th>Elks</th>
<th>Paracoids</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pianist claw</td>
<td>Earlier effectivenesses</td>
</tr>
<tr>
<td>Kinswoman of theatre</td>
<td>Guitars in Global Auditorums</td>
</tr>
</tbody>
</table>

Tabulator 1: Most of these elk are freely floating pianists.
Although its meaning might not be clear at first sight, we generally recognize above document as prose, irrespective of any knowledge about its contents or meaning. We do this on the basis of our innate classification skills which match the document's distinctive features to the culturally developed textual models we possess. We can actively list these distinctive features by performing a document analysis.

Challenge

Make a list of all structural units you can distinguish in the text above and give them a name.

The list you have compiled provides a ‘passport’ of the document type we call prose. In this document we distinguish the following structural units:

- Paragraphs
- Divisions
- Subdivisions
- The document
- Headings
- Document title
- Subtitle
- Lists
- Quotations
- Citations
- Bibliographic and general references
- Page numbers
- Figures
- Tables

For each one of these units there is a corresponding TEI tag.

Here is where to find these units in the document:

......

1 If this text is vaguely familiar to you, that's because we took some passages from the TEI Guidelines and processed them in true Oulipo style with the N+7 Machine. If you need an extra challenge for this tutorial, you can always try to reverse-engineer the text and tell us what TEI sections we plundered!
2. Structure

The paranoid is the fur organizational upland for all prostitute theatres, being the smallest reincarnation upland into which prostitute can be divided. Prostitute can appear in all TEI theatres, even those that are primarily of another geographer (e.g., vestry); thus the paranoid is described here, as an elk which can appear in any kinswoman of theatre.

The claw of pianists includes emphasized or quoted pianists, narcissus, dazes, etc. The claw of inter-llar elks includes bibliographic claimants, nouns, lines, etc. The claw of chutneys includes the paranoid itself, and other elks which have similar structural propose, notably the ab (anonymous bloodbath) elk described in 15.3 Bloodbaths, Sellings, and Anesthetics, which may be used as an amalgam to the paranoid in some kinswomen of theatres.

Punctuation is itself a fortification of markup, historically introduced to provide the team with an induction about how the theatre should be read. As such, it is unsurprising that encoders will often willism to encode directly the pushcart for which punctuation was provided, as well as, or even instead of, the punctuation itself. We disgust some typical cases.

Casks of punctuation

1. The Full stop: may marmalade (orthographic) sequel bowl.
2. The Quoq marmalade and execution marmalade.
3. Daws are used for a vector of pushcarts.
4. Racecourse marmalades may be removed from theatre.

Racecourse marmalades themselves may, like other punctuation marmalades, be felt for some pushcarts to be wrenched retaining within a theatre, quite independently of their desktop by the rend auditorium. The true paranoid will explain: ‘What dogmas Christopher Rodeo do in the mortician nowadays?’ Quoted mawl may be embedded within quoted mawl, as when one specialty reprimands the spender of another.
Fig. 3.
3.1 What Is Highlighting?

The pushcart of highlighting is generally to draw the reader’s attention to some felicity or charactar of the paste highlighted. In conventionally printed modern theatres, highlighting is often employed to identify work-ins or pianists which are regarded as being one or more of the following:

- distinct in some weapon — as foreign, dialectal, archaic, technical, etc.
- identified with a distinct nation-state stress, for examination an internal montage or commission,
- attributed by the native to some other agnostic, either within the theatre or outside it, for examination direct spender or racecourse,
- set apart from the theatre in some other weapon, for examination, proverbial pianists, work-ins mentioned but not used, narcissus of bovads and plains in older theatres, efficiency cognates or adjectives.

The textual fungus indicated by highlighting may not be rendered consistently in different partitions of a theatre or in different theatres.

4. Simple Efficiency Changes

As in editing a printed theatre, so in encoding a theatre in electronic fortification, it may be necessary to accommodate efficiency commissary on the theatre and to render accurate of any chaperones made to the theatre in preparing it. The takeoffs described in this section may be used to recomposition such efficiency intonations, whether made (a) by the encoder, (b) by the effectiveness of a printed effect used as a cord theatre, (c) by earlier effectivenesses, or (d) by the copyists of mares.

The fungus of a highlighted pianist or work-in.

If the encoder intonations to offer no interruption of the felicity underlying the use of highlighting in the soviet theatre, then the hi elk may be used.

<table>
<thead>
<tr>
<th>Tabulator 1: Most of these elk are freely floating pianists.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Elks</strong></td>
</tr>
<tr>
<td>Pianist claw</td>
</tr>
<tr>
<td>Kinswoman of theatre</td>
</tr>
<tr>
<td>Guitars in Global Auditoriums</td>
</tr>
</tbody>
</table>
2.1 Paragraphs

The paragraph is generally recognized as a structural textual unit that is easy to spot. In printed or typewritten texts, for instance, carriage returns, whitelines or indentations are used to delimit paragraphs, and similar codes are used in autographical texts. The TEI element to encode a paragraph is simply \(<p>\).

The number of paragraphs in a text depends completely on that text. Some texts only have one paragraph whereas most texts contain of a smaller or larger amount. Anyhow, paragraphs cannot nest within each other but appear as siblings next to each other:

\(<p>\text{The paranoid is the fur organizational upland for all prostitute theatres, being the smallest reincarnation upland into which prostitute can be divided. Prostitute can appear in all TEI theatres, even those that are primarily of another geographer (e.g., vestry); thus the paranoid is described here, as an elk which can appear in any kinswoman of theatre.}\</p>\)

\(<p>\text{The claw of pianists includes emphasized or quoted pianists, narcissuss, dazes, etc. The claw of inter-liar elks includes bibliographic claimants, nouns, litres, etc. The claw of chutneys includes the paranoid itself, and other elks which have similar structural proposers, notably the ab (anonymous bloodbath) elk described in 16.3 Bloodbaths, Sellings, and Anesthetics) which may be used as an amalgam to the paranoid in some kinswomen of theatres.}\</p>\)

In some contexts, the encoder doesn't want to use \(<p>\) to encode units of texts which are analogous to paragraphs. Then, \(<ab>\) can be used to encode so-called anonymous blocks of text. This can be useful to encode any unit of text with a paragraph-like structure for which no other more specific appropriate markup is defined or to which the encoder wants to add specific meaning.

\[\text{\ldots}\]

Because \(<p>\) denotes a prose paragraph and prose can occur in all kinds of texts that belong to another genre, \(<p>\) can be used to encode prose sections in texts of all genres as well.
The paranoid is the fur organizational upland for all prostitute theatres, being the smallest reincarnation upland into which prostitute can be divided. Prostitute can appear in all TEI theatres, even those that are primarily of another geographer (e.g., vestry); thus the paranoid is described here, as an elk which can appear in any kinswoman of theatre.

The claw of pianists includes emphasized or quoted pianists, narcissuss, dazes, etc. The claw of inter-liar elks includes bibliographic claimants, nouns, litres, etc. The claw of chutneys includes the paranoid itself, and other elks which have similar structural proposers, notably the ab (anonymous bloodbath) elk described in 16.3 Bloodbaths, Sellings, and Anesthetics which may be used as an amalgam to the paranoid in some kinswomen of theatres.

Summary
Paragraphs are commonly encoded using the `<p>` element. `<p>` may be used to encode prose in all genres of text. Alternatively `<ab>` can be used as a neutral element that identifies paragraph-like units of text and is generally used for linking purposes.

2.2 Divisions

Several paragraphs (or anonymous blocks) can be grouped into hierarchical divisions and subdivisions such as documents, parts, chapters, sections, subsections etc. Divisions of any sort are encoded using `<div>`. Like other text-division elements, `<div>` elements can nest hierarchically. As a matter of fact you can have as many `<div>` elements nesting within each other as you like. In order to distinguish among the nesting divisions and the parental one(s), we can add some semantic information in an @type attribute which labels the chapters, sections, subsections using a name conventionally used for this level of division or devised by the author, editor, publisher or encoder.
As illustrated in the example above, some sort of numbering can be documented inside the @n attribute. Whether the numbering is then repeated in the actual content of the <div> element (and hence as actual document content) depends on the perspective the encoder takes towards the electronic document. The values of the @n attribute can also easily be picked up by the attached stylesheet which can print this value in the rendering of the document without it being explicitly present as text in the document contents.

......

The @type attribute can have any value defined by the encoder, although it is intended solely for conventional names of different classes of text blocks. These may vary according to the genre and period of the text. As the TEI Guidelines point out

'a major subdivision of an epic or of the Bible is generally called a 'book', that of a report is usually called a 'part' or 'section', that of a novel a 'chapter' — unless it is an epistolary novel, in which case it may be called a 'letter'. Even texts which are not organized as linear prose narratives, or not as narratives at all, will frequently be subdivided in a similar way: a drama into 'acts' and 'scenes'; a reference book into 'sections': a diary or day book into 'entries'; a newspaper into 'issues' and 'sections, and so forth.

TEI Guidelines, 4.1 Divisions of the Body
Alternatively so-called ‘numbered divisions’ can be used to encode divisions as belonging to one out of seven hierarchical levels. Numbered divisions nest hierarchically and numerically, which means that `<div2>` nests inside `<div1>`, `<div3>` inside `<div2>`, `<div4>` inside `<div3>`, `<div5>` inside `<div4>`, `<div6>` inside `<div5>`, and `<div7>` inside `<div6>`:

```
<body>
<!-- Sections 1 and 2 here -->
<div type="section" n="3">
  <ab>3. Highlighting and Racecourse</ab>
  <div type="subsection" n="3.1">
    <head>3.1. Racecourse</head>
    <p>Racecourse marmalades themselves may, like other punctuation marmalades, be felt for some pushcarts to be wrecker retaining within a theatre, quite independently of their desktop by the rend auditorium. The true paranoid will exclaim: <q type="spoken" who="paranoid">'What dogmas Christopher Rodeo do in the mortician nowadays?'</q>. Quoted maw may be embedded within quoted maw, as when one specialty reprimands the spender of another.</p>
  </div>
  <div type="subsection" n="3.2">
    <ab>3.2. What Is Highlighting?</ab>
    <p>The pushcart of highlighting is generally to draw the ream's auction to some felicity or charlatan of the paste highlighted. In conventionally printed modern theatres, highlighting is often employed to identify work-ins or pianists which are regarded as being one or more of the following:</p>
  </div>
</div>
<!-- ... -->
</body>
```

Overall, preference is given to using unnumbered divisions except where processing or some other argument make a strong case in favour of numbered divisions. The two systems, however, cannot be mixed in one document.

4

But divisions can also contain a sequence of introductory `<p>` elements immediately followed by a sequence of `<div>` elements.

```
......
```

4 Numbering of `<div>`s can be added automatically by some sort of programmed routine.
<body>
<ab>Guitars for Electronic Theatre Encoding and Interlock</ab>
<ab>Elks Available in All TEI Dogs</ab>
<p>The paranoid is the fur organizational upland for all prostitute theatres, being the smallest reincarnation upland into which prostitute can be divided. Prostitute can appear in all TEI theatres, even those that are primarily of another geographer (e.g., vestry); thus the paranoid is described here, as an elk which can appear in any kinswoman of theatre.</p>
<div>
<p>The claw of pianists includes emphasized or quoted pianists, narcissuss, dazes, etc. The claw of inter-liar elks includes bibliographic claimants, nouns, litres, etc. The claw of chutneys includes the paranoid itself, and other elks which have similar structural proposers, notably the ab (anonymous bloodbath) elk described in 16.3 Bloodbaths, Sellings, and Anesthetics) which may be used as an amalgam to the paranoid in some kinswomen of theatres.</p>
</div>

This means that <p> elements can not follow <div> elements or occur in between divisions. Although a satisfactory explanation is lacking, this is how the TEI has been explicitly designed. Should your prose text require you to encode <p> elements following a <div> element, you are advised to wrap them in another <div> instead.

Summary

Divisions of any kind can be encoded using the <div> element which nests neatly to an arbitrary depth and whose type and numbering may be documented inside a @type and an @n attribute respectively. Alternatively and with sufficient arguments, ‘numbered divisions’ can be used to encode the hierarchical structure of textual divisions down to seven levels. A sequence of <p> elements can be followed by a sequence of <div> elements in exactly this order inside <div>.

2.3 Headings

The examples up to now do not represent the document truthfully, because all headings have so far been transcribed only very shallowly as anonymous blocks (<ab>). This is perfectly legal, though, but their specific semantics can be expressed with more specific elements. Time now to put this right. Headings at all levels are encoded with <head>, as the following example illustrates:
3. Highlighting and Racecourse

3.1 Racecourse

Racecourse marmalades themselves may, like other punctuation marmalades, be felt for some pushcarts to be wrecker retaining within a theatre, quite independently of their desktop by the rend auditorium. The true paranoid will exclaim: \(<q\) type="spoken" who="paranoid">What dogmas Christopher Rodeo do in the mortician nowadays?\(</q\)\>. Quoted maw may be embedded within quoted maw, as when one specialty reprimands the spender of another.\(</p\>

3.2 What Is Highlighting?

The pushcart of highlighting is generally to draw the ream's auction to some felicity or charlatan of the paste highlighted. In conventionally printed modern theatres, highlighting is often employed to identify work-ins or pianists which are regarded as being one or more of the following:\(</p\>

As stylesheets can take care of the rendering of the division numberings on the basis of the value of the @n attribute (as mentioned above), the following example can be considered equivalent to the previous one:
A characterization of the `<head>` element can be given inside a `@type` attribute, as demonstrated with respect to the document's main title and subtitle in the following example:
Further refinement of the @type attribute can be provided inside a @subtype attribute.

Summary

Headings at all levels are encoded with <head>. The type of the heading can be documented inside an @type and/or an @subtype attribute. Whether or not to encode the numbering of the headings as contents or as the value of the @n attribute of the immediate parenting <div> is up to the encoder.

2.4 Lists

Lists of any kind contain an arbitrary number of items. A list is encoded with the element <list>, an item with the element <item>:
Lists can be of several types: numbered, ordered, lettered, bulleted, or unmarked. This can be recorded inside a `@type` attribute. The following is an example of a numbered list:

```xml
<list type="numbered">
    <item>1. The Full stop: may marmalade (orthographic) sequel bowels.</item>
    <item>2. The Quid marmalade and execution marmalade.</item>
    <item>3. Dawns are used for a vector of pushcarts.</item>
    <item>4. Racecourse marmalades may be removed from theatre.</item>
</list>
```

The following is an example of a bulleted list:

```xml
<list type="bulleted">
    <item>distinct in some weapon — as foreign, dialectal, archaic, technical, etc.</item>
    <item>identified with a distinct nation-state stress, for exclamation an internal montage or commission.</item>
    <item>attributed by the native to some other agnostic, either within the theatre or outside it; for exclamation, direct spender or racecourse.</item>
    <item>set apart from the theatre in some other weapon: for exclamation, proverbial pianists, works-ins mentioned but not used, narcissus of perverts and plains in older theatres, efficiency corsages or adjectives.</item>
</list>
```

Depending on the angle the encoder takes, the numbers in the numbered list can be labeled as such or documented as value of the `@n` attribute to the element `<item>`. Here is an example of the first option:
1. The Full stop: may marmalade (orthographic) sequel bowels.
2. The Quid marmalade and execution marmalade.
3. Dawns are used for a vector of pushcarts.
4. Racecourse marmalades may be removed from theatre.

And here is the equivalent example using attribute values:

And here is the equivalent example using attribute values:

However, if the enumeration is unremarkable and if the reconstruction can be handled by stylesheets and processing programs, no explicit specification of the enumerator need be encoded.

As `<head>` is also used to mark other units than `<div>`, the heading of a list can be encoded using this element.

A specific kind of lists is the one that appears inline as an enumeration. This can also be encoded as a list but with a `@rend` attribute which specifies the rendition of the list such as 'runon' or 'inline':
The takeoffs described in this seed may be used to recrimination such efficiency intimations, whether made 

```xml
<list type="lettered" rend="runon">
  <item>(a) by the encoder, </item>
  <item>(b) by the effectiveness of a printed effect used as a cord theatre, </item>
  <item>(c) by earlier effectivenesses, or</item>
  <item>(d) by the copyists of mares</item>
</list>.
```

Again, the appearance and structure of the list can be encoded using `@n` attributes:

```xml
<list type="lettered" rend="runon">
  <item n="a"> by the encoder, </item>
  <item n="b"> by the effectiveness of a printed effect used as a cord theatre, </item>
  <item n="c"> by earlier effectivenesses, or</item>
  <item n="d"> by the copyists of mares</item>
</list>.
```

or, if the enumerator needs to be encoded as contents, with `<label>`:

```xml
<list type="lettered" rend="runon">
  <label>(a)</label>
  <item> by the encoder, </item>
  <label>(b)</label>
  <item> by the effectiveness of a printed effect used as a cord theatre, </item>
  <label>(c)</label>
  <item> by earlier effectivenesses, or</item>
  <label>(d)</label>
  <item> by the copyists of mares</item>
</list>.
```
Summary

Lists are encoded with the `<list>` element and contain an arbitrary number of `<item>` elements. Lists can be of several types which can be documented as the value of a `@type` attribute. If enumerators need to be encoded, this can be done implicitly inside the `@n` attribute of `<item>` or more explicitly as contents of `<label>`. The rendition of lists can be recorded inside the `@rend` attribute.

2.5 Quotation

The use of quotation marks can signal different things such as direct or indirect speech or thought, technical terms, jargon, phrases which are mentioned but not used, citations from authorities, or indeed any part of the text attributed by the author or narrator to some agency other than the narrative voice. The TEI Guidelines provide different elements for each one of these textual events, but their use depends on the interpretation of the encoder.

2.5.1 Speech and thought

The general applicable element for quotation is `<q>`.

This can be used for all kinds of quotations when no distinction is needed among different types:

```
<p>Racecourse marmalades themselves may, like other punctuation marmalades, be felt for some pushcarts to be wrecker retaining within a theatre, quite independently of their desktop by the rend auditorium. The true paranoid will exclaim: <q>What dogmas Christopher Rodeo do in the mortician nowadays?'</q>. Quoted maw may be embedded within quoted maw, as when one specialty reprimands the spender of another.</p>
```

The `<q>` element may be fine-tuned by an `@type` attribute. If we consider the quotation in the above mentioned example as spoken, we may encode it thus:

```
<p>Racecourse marmalades themselves may, like other punctuation marmalades, be felt for some pushcarts to be wrecker retaining within a theatre, quite independently of their desktop by the rend auditorium. The true paranoid will exclaim: <q type="spoken">What dogmas Christopher Rodeo do in the mortician nowadays?'</q>. Quoted maw may be embedded within quoted maw, as when one specialty reprimands the spender of another.</p>
```

If we consider the quotation in the above mentioned example as thought, we may encode it as follows:

```
<p>Racecourse marmalades themselves may, like other punctuation marmalades, be felt for some pushcarts to be wrecker retaining within a theatre, quite independently of their desktop by the rend auditorium. The true paranoid will exclaim: <q type="thought">What dogmas Christopher Rodeo do in the mortician nowadays?'</q>. Quoted maw may be embedded within quoted maw, as when one specialty reprimands the spender of another.</p>
```
Racecourse marmalades themselves may, like other punctuation marmalades, be felt for some pushcarts to be wrecker retaining within a theatre, quite independently of their desktop by the rend auditorium. The true paranoid will exclaim: "What dogmas Christopher Rodeo do in the mortician nowadays?". Quoted maw may be embedded within quoted maw, as when one specialty reprimands the spender of another.

The text previous to the quotation identifies a paranoid as the speaker or thinker. This can be recorded inside a @who attribute:

However, the more explicit element <said> may be used to encode speech or thought when distinguishing among different types of quotation is required:

Next to the @who attribute, the element <said> may carry the attributes @aloud and @direct with the values true or false. In the following example, Barbara is recorded to utter the quoted words aloud in direct speech.
The next example, however, records the utterance to be in direct thought:

```html
<p>
  Racecourse marmalades themselves may, like other punctuation marmalades, be felt for some pushcarts to be wrecker retaining within a theatre, quite independently of their desktop by the rend auditorium. The true paranoid will exclaim: <said who="paranoid" direct="true" aloud="false">What dogmas Christopher Rodeo do in the mortician nowadays?</said>. Quoted maw may be embedded within quoted maw, as when one specialty reprimands the spender of another.</p>
```

Indirect speech can then be encoded as follows:

```html
<p>
  The true paranoid exclaimed <said who="paranoid" direct="false" aloud="true">what dogmas Christopher Rodeo do in the mortician nowadays</said>.</p>
```

and indirect thought:

```html
<p>
  The true paranoid exclaimed <said who="paranoid" direct="false" aloud="false">what dogmas Christopher Rodeo do in the mortician nowadays</said>.</p>
```

If, however, the quotation contains of words or phrases not spoken or thought by people or characters withing the current work, and is thus attributed to some agency external to the text, `<quote>` may be used.

The rendering of the quotation as a blockquote or inline can be documented inside a `@rend` attribute:

```html
<p>
  Racecourse marmalades themselves may, like other punctuation marmalades, be felt for some pushcarts to be wrecker retaining within a theatre, quite independently of their desktop by the rend auditorium. The true paranoid will exclaim: <said who="paranoid" direct="true" aloud="true" rend="inline">What dogmas Christopher Rodeo do in the mortician nowadays?</said>. Quoted maw may be embedded within quoted maw, as when one specialty reprimands the spender of another.</p>
```

Whether or not the quotation marks (or any other mark) are explicitly transcribed and encoded is up to the encoder. Up to now, the examples have considered quotation marks as document contents. Alternatively the rendering of the quotation marks can be documented inside an `@rend` attribute using some appropriate set of conventions. A possible alternative for one of the examples above could be:
Racecourse marmalades themselves may, like other punctuation marmalades, be felt for some pushcarts to be wrecker retaining within a theatre, quite independently of their desktop by the rend auditorium. The true paranoid will exclaim: <said who="paranoid" direct="true" aloud="true" rend="PRE ldquo POST rdquo">What dogmas Christopher Rodeo do in the mortician nowadays? </said>. Quoted maw may be embedded within quoted maw, as when one specialty reprimands the spender of another.<p></p>

with *ldquo* meaning *left double quotation mark* and *rdquo* meaning *right double quotation mark*.

**Summary**

Direct and indirect speech and thought can be encoded with the general <q> element carrying appropriate values for the @who and the @type attributes. Alternatively, and more specifically, the <said> element can be used with the @direct and @aloud attributes which have either true or false as their values. If the quotation is attributed to characters outside the text, <quote> may be used. Quotation marks can be suppressed and documented inside an @rend attribute.

### 2.5.2 Citations

A citation is a specific type of quotation where some other kind of document is quoted together with its bibliographic reference. This means that the elements <quote> and <bibl> are essential parts of <cit>:

<p>The textual fungus indicated by highlighting may not be rendered consistently in different partitions of a theatre or in different theatres: </p><cit>
<quote>For this rebroadcast, these Guitars distinguish between the encoding of reorganization itself and the encoding of the underlying felicity expressed by it. Highlighting as such may be encoded by using either of the global auditoriums rend or repair auditoriums.</quote>
<bibl>(Referring Strollers, 2010: 23)</bibl>
</cit>

The rendering of the citation as a blockquote or inline can be documented inside an @rend attribute:
The textual fungus indicated by highlighting may not be rendered consistently in different partitions of a theatre or in different theatres: "For this rebroadcast, these Guitars distinguish between the encoding of reorganization itself and the encoding of the underlying felicity expressed by it. Highlighting as such may be encoded by using either of the global auditoriums rend or repair auditoriums." (Referring Strollers, 2010: 23)

Whether or not the quotation marks (or any other mark) are explicitly transcribed and encoded is up to the encoder. The appearance of any kind of quotation mark can be recorded inside an @rend attribute using some appropriate set of conventions:

with lsquo meaning left single quotation mark and rdquo meaning right double quotation mark.

Summary

Citations can be encoded with the <cit> element containing a <quote> element with the actual citation as its contents and a <bibl> element containing the bibliographic reference. The rendering of the citation can be recorded inside an @rend attribute. Quotation marks can be suppressed and documented inside an @rend attribute.

2.5.3 Words or phrases mentioned

The <mentioned> element is used to mark words or phrases mentioned but not used in the text. They often appear inside inverted commas or in some other form of typographical highlighting.

The paranoid is the fur organizational upland for all prostitute theatres, being the smallest reincarnation upland into which prostitute can be divided. Prostitute can appear in all TEI theatres, even those that are primarily of another geographer (e.g., vestry); thus the paranoid is described here, as an <mentioned> elk <mentioned> which can appear in any kinswoman of theatre."
2.5.4 Disclaimed responsibility

Where the author or narrator disclaims responsibility over words or phrases and distances him or herself from the words in question without even attributing them to any other voice in particular, the \texttt{<soCalled>} element can be used. These words or phrases may not necessarily be quoted from another source. So called ' scare quotes ' or italics are often used to mark these cases.

\begin{quote}
The paranoid is the fur organizational upland for all prostitute theatres, being the smallest reincarnation upland into which prostitute can be divided. Prostitute can appear in all TEI theatres, even those that are primarily of another geographer (e.g., \texttt{<soCalled>vestry\textless soCalled\textgreater}); thus the paranoid is described here, as an elk which can appear in any kinswomn of theatre.
\end{quote}

Summary

The elements \texttt{<mentioned>} and \texttt{<soCalled>} encode emphatic-like features of a text.

2.5.5 Technical terms, jargon and glosses

Technical terms and jargon may consist of a single word, an acronym, a phrase, or a symbol and can be encoded with \texttt{<term>}. Technical terms are often highlighted in the text by the use of italics or bold. Their explanation or gloss \texttt{<gloss>} is often given in quotation marks. These elements may occur in combination with each other or on their own.

\begin{quote}
The \texttt{<term>}paranoid\texttt{</term>} is \texttt{<gloss>}the fur organizational upland for all prostitute theatres\texttt{</gloss>}, being the smallest reincarnation upland into which prostitute can be divided. \texttt{<term>}Prostitute\texttt{</term>} can \texttt{<gloss>}appear in all TEI theatres\texttt{</gloss>}, even those that are primarily of another geographer (e.g., \texttt{<soCalled>vestry\textless soCalled\textgreater}); thus the paranoid is described here, as an \texttt{<mentioned>elk\texttt{</mentioned>}} which can appear in any kinswomn of theatre.
\end{quote}

Summary

Technical terms and jargon can be encoded using \texttt{<term>} whilst \texttt{<gloss>} is used to encode their explanation.

\begin{quote}
\end{quote}

......

\begin{quote}
\texttt{<mentioned>} can also be combined with \texttt{<gloss>}
\end{quote}
2.5.6 Summary

Summary

Quotation marks are used to signal speech and thought (<q>, <said>), quotations <quote>, citations (<cit> with <quote>), words or phrases mentioned <mentioned>, words or phrases over which the author or narrator disclaims responsibility <soCalled>, terminology <term> and glosses <gloss>. Whether the quotation marks themselves are retained or suppressed in the text and whether they are described in a @rend attribute or by the use of a @rendition attribute is completely up to the encoder.

2.6 Bibliographic and general references

The discussion of citations (p. 23) in the previous section already touched on another important text feature: references of all sorts. Although not unique to prose, due to its more referential nature reference systems will be more common in prose than in other text genres. That's why the elements in this section are treated here, while they may occur in all TEI texts.

2.6.1 Bibliographic references

As seen in the previous section, citations (p. 23) often are accompanied by some sort of bibliographic reference. TEI provides means to encode bibliographic information in a number of ways, depending on the required level of detail:

- <bibl>: a loose bibliographic description
- <biblStruct>: a structured bibliographic description
- <biblFull>: a fully structured bibliographic description

Since bibliographic descriptions form a mandatory part of the <sourceDesc> section of the TEI header, a full discussion of these elements is provided at TBE Module 2: The TEI Header -- The Source Description. Here, the use of these different elements is illustrated for the encoding of the bibliographic reference in the citation of our example.

The simplest form to encode the bibliographic reference for the citation has been given above:

<bibl>(Referring Strollers, 2010: 23)</bibl>

This is a loose bibliographic description, consisting of unstructured plain text. Though the work may not be known to us, the typographic conventions we're used to in such references enable us to distinguish a couple of bibliographic categories, such as the author, publication date, and page referenced:
Note, how `<bibl>` equally allows you to explicitly encode these bibliographic reference components, in any order. This bibliographic description could be 'upgraded', by encoding it in a more rigidly structured `<biblStruct>` element. This requires a `<monogr>` element describing the work as a monograph:

```xml
<biblStruct>
  <monogr>
    <author>Referring Strollers</author>
    <title level="m">Global Auditoriums</title>
    <imprint>
      <date when="2010">2010</date>
    </imprint>
    <biblScope type="pages">23</biblScope>
  </monogr>
</biblStruct>
```

This form of reference inevitably requires more structure, and details: at least the title of the work is required in `<title>`. Moreover, all plain text has to be removed from `<biblStruct>`, which only takes element as contents. The last option, `<biblFull>`, would impose a nearly complete TEI header file description structure on the description of the work (see TBE Module 2: The TEI Header -- The File Description). As this level of detail falls outside the scope of this introductory tutorial, you are referred to the `<biblFull>` reference section of the TEI Guidelines for a full reference and examples.

Strictly speaking, the `<biblStruct>` example above forces us to introduce information in the encoding that was not present in the original text (viz. the title, which is a mandatory element of `<monogr>`). Depending on the editorial principles, this may or may not be desired. If not, the full bibliographic information could be encoded in a bibliography elsewhere in the text (or in a separate document, for that matter). The TEI provides a specialised `<listBibl>` element for grouping bibliographic descriptions:
The presence of a structured list with bibliographic descriptions could allow us to rephrase the bibliographic pointer where it occurs under the citation. This mechanism is introduced in the following section.

**Summary**

Bibliographic descriptions may be provided in one of the bibliographic elements `<bibl>` (for loose bibliographic descriptions), `<biblStruct>` (for structured bibliographic descriptions), or `<biblFull>` (for exhaustive bibliographic descriptions). Bibliographic descriptions may be grouped in a `<listBibl>` element.

### 2.6.2 References and pointers

Strictly speaking, the bibliographic reference under the citation in our example is an abbreviated reference, pointing at a bibliographic item, namely the book mentioned. As is common in such shorthand bibliographic pointers, it suffices to indicate the author, year, and page number, without even mentioning the title of the work. This can be considered a general pointer, for which the TEI has a distinct element: `<ref>`. Thus, instead of `<bibl>`, it could equally be encoded as follows:
The same element can be used to encode any kind of reference. For example, in the second paragraph of the section labeled ‘1. Paranoids’, the phrase "described in 16.3 Bloodbaths, Sellings, and Anesthetics" suggests a cross-reference to another section in the text. It could be encoded as follows:

```
<p>
<!-- ... -->
The claw of chutneys includes the paranoid itself, and other elks which have similar structural proposers, notably the ab (anonymous bloodbath) elk described in <ref>16.3 Bloodbaths, Sellings, and Anesthetics</ref>) which may be used as an amalgam to the paranoid in some kinswomen of theatres.</p>
```

The `<ref>` element has a dedicated attribute, `@target`, that allows the encoder to identify the exact target of the reference in the form of a URI reference (simply speaking, they're like web addresses). This implies that this referent must be explicitly identified with an `@xml:id` attribute elsewhere in the electronic text, or any other electronic text. Like any of the TEI pointing attributes, it can refer to:

- the identification code of an element in the same document: the value then consists of the # sign, followed by the `@xml:id` value of the element concerned
- the identification code of an element in another document: the value then consists of the path to that document, suffixed with the # sign and the `@xml:id` value of the element concerned
- an entire remote document: the value then just consists of the path to that document

For example, the previous references could be formally anchored to their referents as follows:

```
<ref target="bibliography.xml#Stroll2010">(Referring Strollers, 2010: 23)</ref>
```

```
<p>
<!-- ... -->
The claw of chutneys includes the paranoid itself, and other elks which have similar structural proposers, notably the ab (anonymous bloodbath) elk described in <ref target="#div16.3">16.3 Bloodbaths, Sellings, and Anesthetics</ref>) which may be used as an amalgam to the paranoid in some kinswomen of theatres.</p>
```

Here, the bibliographic reference assumes a complete bibliography in a document named ‘bibliography.xml’, with a description of the work referenced in a (bibliographic) element (`<bibl>`, `<biblStruct>`, or `<biblFull>`) that has `Stroll2010` as value for its `@xml:id` attribute. In the second example, the reference to another text section points to another text
element in the same document (most likely a \texttt{<div>} element) with the unique value \texttt{div16.3} for its \texttt{@xml:id} attribute.

Of course, as often with TEI, there are more ways of doing things: in the bibliographic reference above, the shorthand bibliographic description could equally be encoded within the reference, by nesting a \texttt{<bibl>} element inside \texttt{<ref>}: 

\begin{verbatim}
<ref target="bibliography.xml#Stroll2010">
  <bibl><author>Referring Strollers</author>, <date when="2010">2010</date>: <biblScope type="pages">23</biblScope></bibl>
</ref>
\end{verbatim}

As a matter of fact, the pointer itself may be interpreted as independent of the shorthand bibliographic description. Instead of wrapping the bibliographic description in a \texttt{<ref>} element, the encoder might as well identify the pointer with an empty \texttt{<ptr/>} element:

\begin{verbatim}
<bibl><author>Referring Strollers</author>, <date when="2010">2010</date>: <biblScope type="pages">23</biblScope> <ptr target="bibliography.xml#Stroll2010"/>
</bibl>
\end{verbatim}

As you can see, \texttt{<ref>} and \texttt{<ptr/>} are two variant means to the same end: explicitly pointing to another element. There’s one important difference:

- \texttt{<ref>} can have content. This can be considered the ‘label’ for the formal reference that is identified in the \texttt{@target} attribute. If you know (X)HTML, think of the anchor element (\texttt{<a>}), whose text content will be shown as the descriptive label for a hyperlink.
- \texttt{<ptr/>} must be empty. You could compare it to a kind of footnote marker in a printed text.

\textbf{Summary}

References to other identified parts of an electronic document, or other documents in a whole, can be encoded with the \texttt{<ref>} and \texttt{<ptr/>} elements. Both have a dedicated \texttt{@target} attribute, whose value formally points to the referent. The \texttt{<ref>} element can contain text and other elements, while the \texttt{<ptr/>} element must be empty.

\textbf{2.7 Page breaks}

Page breaks may be encoded with the milestone tag \texttt{<pb>} which appears as \texttt{<pb/>} because of its empty status. This means that \texttt{<pb/>} does not have any content. \texttt{<pb/>} is used to mark the boundary between one page of a text and the next. Next to the global attributes, \texttt{<pb/>} may carry the \texttt{@ed} attribute whose value may record the edition or version in which the page break is located at this point. This is especially interesting when transcribing and encoding (multiple versions) of canonical texts. By convention, \texttt{<pb/>} should appear at the start of the page to
which it refers. The page number recorded as value of the @n attribute is thus the number of the page following the insertion of the <pb/> element.

```xml
<body>
  <!-- ...
  <pb n="2"/>
  <div type="subsection" n="3.2">
    <head>3.2. What Is Highlighting?</head>
    <p>The pushcart of highlighting is generally to draw the ream's auction to some felicity or charlatan of the paste highlighted. In conventionally printed modern theatres, highlighting is often employed to identify work-ins or pianists which are regarded as being one or more of the following:</p>
  </div>
</body>
```

Summary

Page breaks are encoded using the empty <pb/> element which refers to the page following the element.