TEI By Example

TEI by Example. Module 3

Edward Vanhoutte

Ron Van den Branden

Melissa Terras

TABLE OF CONTENTS

Module 3: Prose

1. Introduction

It's quite difficult to define prose. Prose is not poetry and not drama. Prose is the default way of communication in the Western World. It is the main genre for fictional and non-fictional writing in books, newspapers, flyers, reports, presentations etc. Novels, business reports, manuals, cookery books, glossy magazines, and transcriptions of conversations are all written in prose.

The OED defines prose as: 'Language in the form in which it is typically written (or spoken), usually characterized as having no deliberate metrical structure (in contrast with verse or poetry)' (*OED*, prose 1a) and further sums up a couple of features such as the avoidance of elaboration, metaphorical language and imaginative contents that distinguishes prose from poetry but do, however, characterize fictional prose.

As with all types of text, prose has structure and meaning. The TEI Guidelines do not devote a dedicated chapter on prose, since prose is the TEI default mode. However, the TEI does offer means to encode structural, semantic, and analytical features in prose. In this module, we bring together most of the encoding suggestions that are scattered throughout the TEI Guidelines and present a didactic approach towards encoding prose texts.

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, 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.

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 casks:

Casks of punctuation

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.

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: '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.

1

Fig. 1.

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:

- 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 spender 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 plains 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 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)

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 takeoffs described in this seed may be used to recrimination 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 witticisms to offer no interruption of the felicity underlying the use of highlighting in the soviet theatre, then the hi elk may be used.

	Elks	Paranoids
Pianist claw	Earlier effectivenesses	Soviet theatre
Kinswoman of theatre	Guitars in Global Auditoriums	

Tabulator 1: Most of these elks are freely floating pianists.

2

Fig. 2.

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.
[1]

> **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!

TITLE | **Guitars for Electronic Theatre Encoding and Interlock**

SUBTITLE | *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, 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.   **REFERENCE**

SECTION / **PARAGRAPH**

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 casks:

Casks of punctuation   **HEADING**

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.

LIST

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 '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.   **QUOTE**

SUBSECTION

PAGE NUMBER   1

Fig. 3.

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:

LIST

- 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 spender 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 plains 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:

CI-TA-TION 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)   **BIBLIOGRAPHIC REFERENCE**

SUBSECTION

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 takeoffs described in this seed may be used to recrimination 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.

LIST

The fungus of a highlighted pianist or work-in.   **HEADING**

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.

FIGURE

TABLE

	Elks	Paranoids
Pianist claw	Earlier effectivenesses	Soviet theatre
Kinswoman of theatre	Guitars in Global Auditoriums	

Tabulator 1: Most of these elks are freely floating pianists.   **HEADING**

2

Fig. 4.

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 lager amount. Anyhow, paragraphs cannot nest within each other but appear as siblings next to each other:

> <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>
>
> <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>

2

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.

......

2   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.

<ab>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.</ab>

<ab>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.</ab>

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.

```
<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>
```

3

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.

......

3   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 -->
  <div1 type="section" n="3">
    <ab>3. Highlighting and Racecourse</ab>
    <div2 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>
    </div2>
    <div2 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>
    </div2>
    <!-- ... -->
  </div1>
</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>
</body>
```

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:

```
<body>
  <head>Guitars for Electronic Theatre Encoding and Interlock</head>
  <head>Elks Available in All TEI Dogs</head>
  <!-- Sections 1 and 2 here -->
  <div type="section" n="3">
    <head>3. Highlighting and Racecourse</head>
    <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">
      <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>
    <!-- ... -->
  </div>
</body>
```

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:

```
<body>
  <head>Guitars for Electronic Theatre Encoding and Interlock</head>
  <head>Elks Available in All TEI Dogs</head>
  <!-- Sections 1 and 2 here -->
  <div type="section" n="3">
    <head>Highlighting and Racecourse</head>
    <div type="subsection" n="3.1">
      <head>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">
      <head>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>
    <!-- ... -->
  </div>
</body>
```

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:

```
<body>
   <head type="mainTitle">Guitars for Electronic Theatre Encoding and Interlock</head>
   <head type="subTitle">Elks Available in All TEI Dogs</head>
   <!-- Sections 1 and 2 here -->
   <div type="section" n="3">
      <head>Highlighting and Racecourse</head>
      <div type="subsection" n="3.1">
         <head>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">
         <head>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>
      <!-- ... -->
   </div>
</body>
```

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>**:

```
<list>
    <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>
```

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:

```
<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:

```
<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, work-
    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:

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

And here is the equivalent example using attribute values:

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

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.

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

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':

```
<p>
  <!-- ... -->
The takeoffs described in this seed may be used to recrimination such efficiency intimations, whether made <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>. </p>
```

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

```
<p>
  <!-- ... -->
The takeoffs described in this seed may be used to recrimination such efficiency intimations, whether made <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>. </p>
```

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

```
<p>
  <!-- ... -->
The takeoffs described in this seed may be used to recrimination such efficiency intimations, whether made <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>. </p>
```

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>

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

> \<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>

or

> \<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" 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>

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

> \<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>'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>

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.

> \<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">'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>

The next example, however, records the utterance to be in direct thought:

> \<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:

> \<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:

> \<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:

> \<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:

> <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="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>

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:

```
<p>The textual fungus indicated by highlighting may not be rendered consistently in different partitions of a
theatre or in different theatres: </p>
<cit rend="blockquote">
  <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>
```

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:

```
<p>The textual fungus indicated by highlighting may not be rendered consistently in different partitions of a
theatre or in different theatres: </p>
<cit rend="blockquote">
  <quote rend="PRE lsquo POST rsquo">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>
```

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.

```
<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 <mentioned>elk</mentioned> which
can appear in any kinswoman of theatre.</p>
```

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 **<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.

> <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., '<soCalled>vestry</soCalled>'); thus the paranoid is described here, as an elk which can appear in any kinswoman of theatre.</p>

Summary

The elements **<mentioned>** and **<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 **<term>**. Technical terms are often highlighted in the text by the use of italics or bold. Their explanation or gloss **<gloss>** is often given in quotation marks. These elements may occur in combination with each other or on their own.

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

Summary

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

5

......

5   **<mentioned>** can also be combined with **<gloss>**.

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 <u>citations (p. 23)</u> 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, <u>citations (p. 23)</u> 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 <u>TBE Module 2: The TEI Header -- The Source Description</u>. 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:

```
<bibl>(<author>Referring Strollers</author>, <date when="2010">2010</date>: <biblScope
type="pages">23</biblScope>)</bibl>
```

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:

```
<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:

```
<back>
  <div type="bibliography">
    <listBibl>
      <head>Bibliography</head>
      <biblStruct xml:id="Stroll2010">
        <monogr>
          <author>Referring Strollers</author>
          <title>Global Auditoriums</title>
          <imprint>
            <date when="2010">2010</date>
            <pubPlace>State of Grace</pubPlace>
            <publisher>Elks Inc.</publisher>
          </imprint>
          <biblScope type="pages">23</biblScope>
        </monogr>
      </biblStruct>
      <!-- ... -->
    </listBibl>
  </div>
</back>
```

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:

```
<ref>(Referring Strollers, 2010: 23)</ref>
```

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 'biblliography.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 **<div>** element) with the unique value `div16.3` for its **@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 **<bibl>** element inside **<ref>**:

```
<ref target="bibliography.xml#Stroll2010">
   <bibl>(<author>Referring Strollers</author>, <date when="2010">2010</date>: <biblScope
   type="pages">23</biblScope>)</bibl>
</ref>
```

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 **<ref>** element, the encoder might as well identify the pointer with an empty **<ptr/>** element:

```
<bibl>(<author>Referring Strollers</author>, <date when="2010">2010</date>: <biblScope
type="pages">23</biblScope>) <ptr target="bibliography.xml#Stroll2010"/> </bibl>
```

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

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

Summary

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

2.7 Page breaks

Page breaks may be encoded with the milestone tag **<pb>** which appears as **<pb/>** because of its empty status. This means that **<pb/>** does not have any content. **<pb/>** is used to mark the boundary between one page of a text and the next. Next to the global attributes, **<pb/>** may carry the **@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, **<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.

```
<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.

3. Semantics

Most encoders of prose want to encode more than just the document structure. Dates, names, places, events, organisations, titles of books, movies, and plays, for instance, frequently appear in prose and may need to be encoded. The TEI Guidelines propose appropriate elements for all these features. A tutorial on these elements and their use can be found in module 1: Introduction

4. Advanced Encoding

4.1 Segments

It is often convenient for various kinds of analysis to encode further subdivisions of paragraphs or anonymous blocks. This can be done using the **<seg>** element which contains any arbitrary phrase-level unit of text (including other **<seg>** elements). The output of an automatic parsing system in linguistic analysis, for instance, may use **<seg>** for the markup of linguistically significant constituents like sentences, phrases, words etc. in a theory neutral manner.

```
<div 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>
```

6

When the **@xml:id** attribute with **<seg>** identifies the segment, **<seg>** can be used for linking, reference, and alignment purposes.

7

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, but allows you to actually *point* towards a digital representation of the image as well. This can be done in an **@url** attribute, which takes a URL as its value. Suppose we have a scanned-in or otherwise gained digital facsimile of the image in the source text, this could be encoded as follows:

```
<graphic url="graphics/hi_elk.gif"/>
```

In this case, the URL points to a file 'hi_elk.gif' in the folder 'graphics', which is located 1 level below 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*).

......

6   Specialized 'linguistic segment category' elements are defined in TEI Guidelines, 17.1 Linguistic Segment Categories

7   See P5 16.3 Blocks, Segments, and Anchors for more examples and complex cases.

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 graphical element on the page and should ideally be encoded as such. For this, the TEI has a special **<figure>** element, allowing you to group image-related elements. 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>
    <head>The fungus of a highlighted pianist or work-in.</head>
    <graphic url="graphics/hi_elk.gif"/>
    <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>
```

<figure> 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>
    <head>The fungus of a highlighted pianist or work-in.</head>
    <graphic url="graphics/hi_elk.gif"/>
    <figDesc>the male 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>
```

A last option for dealing with images, is the literal inclusion of the image's binary representation in your XML document. This can be done inside the **<binaryObject>** element, whose **@mimeType** attribute can specify the mime type of the graphical object, so that processing applications can interpret it correctly. For example, this is how a base64 ASCII representation of the binary JPEG scan of the image may be encoded:

```
<figure>
  <head>The fungus of a highlighted pianist or work-in.</head>
  <binaryObject mimeType="image/
gif">/9j/4AAQSkZJRgABAQEADwAPAAD/2wBDAAUDBAQEAwUEBAQFBQUGBwwIBwcHBw8LCwkMEQ8SEhEP
ERETFhwXExEQaFRERGCEYGh0dHx8fExciJCIeJBweHx7/2wBDAQUFBQcGBw4ICA4eFBEUHh4eHh4e
Hh4eHh4eHh4eHh4eHh4eHh4eHh4eHh4eHh4eHh4eHh4eHh4eHh7/wAARCABAAC8DASIA
AhEBAxEB/8QAHAAAAQQDAQAAAAAAAAAAAAABwAEBQYBAgMI/8QAMBAAAgIBAwMDAgUDBQAAAAAA
AQIDBBEABRIGITETQVFhcRQVInKBMlKRBzNCYsH/
xAAYAQEBAQEBAAAAAAAAAAAAAAAAEAgUDAf/E
ACYRAAEDAwMEAgMAAAAAAAAAAEAAgMEESESFFETMTJhBUFCgZH/2gAMAwEAAhEDEQA/
APZelpaq /Wm92KsibXtzlLMiepLMAD6KZwAM/
wDJsHHwAT8al7g0XKpjC82Cmd03rZ9qYLuW60qbMvJVnnVG YfIBOTqD/
wBPes6nVi7isbVVkqW3hjEVgP6sQxxkA7EAgj2x/wCUsUE/USCzOcuzHkzH5JPcn6nW v5bGoQIpjMf
+20Z4Mh+VI7qfqNB3hv44T9k3T5ZRh0tVfone7Fpn2vcn52ok5xTYx60ecHP/AGUk Z+cg/
OLRpzHB4uEB7Sw2K5XJXhqqTTRxNM8cbMsa+XIGQB99C7Zb1neK0m7Xoo47VqVjKsYbivE8A
ByGfCj5Gc4JHfRW0NOsdsZr26bd6k0BsBrFZ4p3ib9Y7nkhB7Py7ft1xqqQS1d6YgOTLetph3SCJT
NJWmhlWWKeJI2kjIIJ481YDIBUnGcE4I1nbaJpCaa5cFieaU5lKCMcSx4LgdiQCFz5OBpdNyVZNk
qionpJGnptFjBjdezKw/uBBB+utOqmr/AJBBbhsRLN+IjMEcJ7+q7DCqPuf8AHn20H0ne113bdJNh
SDd4EieWCYAJI2A4YFSB8nB7fUZPYaKFOV56kM0kTQvJGrNNG3lCRkg/bQ66Q2atHulCjWrQIkGLF
kxxBQxQYUnHuX4nv5CnRK06mBDUGpILktDLrXqaPd5xX2iOLjUlPC+4LcmHZgigjKHwSTg4yB2Vt
ETdqSbjtdrb5ZZZokswvEzwvxdQwIyp9j30ERtu7pM+zTxwRYaSuLMD8THwyORjOCvbBXBYdwew0y
Mwi5lOFmVO5IApxnnhO/XovO1q1Ws0LjAB7FMsRIcYBIX+vA8c1IGsJPShn/ABNSravXcEC1cJAT
PnAbuoPwigHt/C3CtLRsRjMrwP8ApMkhTHM+AoHfHnzn2++tacMly8YVMqwoMO8bIGViMjIbORj4
H+fbwU9Jo61zp4UGt+REu20jVbvn+8Kd6R6kk2awy31jmr2HDWLAXi8fsD9UH9vkkDJyT2JRVgyhl
IKkZBB7HQKuVN0qGSnVMM8ihSk07EZ9RmCKFHdmyMAdge3caNGw7cu07PW25J5Z1gTiJJPJ9/wCB
8D2GB7aqR0DgDEqpW1TbioH7Qzvbxf3QV5ZLFiSzcQTJWWYpFBESvYoGXngN5OSSD4GAOKbW9Nq1
xbDyTQrxmLyOUcYwW45OPnAB8ds4wbbvHSlqvfkvbIY3ilPKanIeOD8xt4H7T2+oAA1GMltHEcu2
bkknuoqSOAf3KCp/gkay5GyA5ytuN0ZGMJX+it03aatZS5TrrGOXIoJQ/btxIwwBz5z/ABrNPofd
dvuS3W3CjMrJlgV9IKR8sQxIxj3AHfsc6s/TVbeamzwpIa2AWKV5EZWRCxKqWBI7DHbj28e2l1LV
3i3s08cbVsNx5wRxszOgYFlDEjORkf0/TTBiLp/XdALQZ+r+Xa/pDyXa5L9p9zaw9e0qGOpLXmkT 0xg/
ryOJOck4OO2ntLq3ddkSdrTzW/w45SVZZAxkRnKq6MSzjwCQxI8jGTnT9YrbnjHtu4s/sDUd cn7sAB/
JA1IbD0xbtXlv77BHDDGD6NLkHYkjHKQjtnHgAke+dEiEl8YCdKY9Qclf/9k=</binaryObject>
  <graphic url="graphics/hi_elk.gif"/>
  <figDesc>the male 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>
```

Of course, just like **<graphic/>**, **<binaryObject>** can be used without a **<figure>** wrapper as well.

8

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 **@mimeType** attribute specifies the binary format in which the file is encoded. 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. Inside tables, rows are considered the basic unit. 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>
  <row>
    <cell/>
    <cell>Elks</cell>
    <cell>Paranoids</cell>
  </row>
  <row>
    <cell>Pianist claw</cell>
    <cell>Earlier effectivenesses</cell>
    <cell>Soviet theatre</cell>
  </row>
</table>
```

Note 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, you can use an **@role** attribute on both entire rows and separate cells. Suggested values are `label` and `data` (default):

......

8   If these dedicated TEI mechanisms 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 P5 22.6 Combining TEI and Non-TEI Modules.

```
<table>
  <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>
```

The third row deviates from the previous two. It only has two cells, of which the second 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>
  <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>
```

Note 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 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>
```

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 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>
```

Note, however, that **<head>** 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.

9

Summary

Tables consist of an arbitrary number of rows which consist of an arbitrary number of cells. The respective markup for these is **<table>**, **<row>**, and **<cell>**. Empty cells can be encoded using **<cell/>**. Cells or rows containing a label can be encoded with the `label` value to the **@role** attribute. Cells which span over several columns can be encoded using a **@cols** attribute whose value documents the number of columns it spans. Table headings can be encoded as **<head>** before the first row.

5. Summary

We can now encode the document which served as an example throughout this module:

......

9   **<head>** can only occur at the beginning of larger structural elements.

```
<body>
  <head type="mainTitle">Guitars for Electronic Theatre Encoding and Interlock</head>
  <head type="subTitle">Elks Available in All TEI Dogs</head>
  <div type="section" n="1">
    <head>1. Paranoids</head>
    <p>The <term>paranoid</term> is <gloss>the fur organizational upland for all prostitute theatres</gloss>,
    being the smallest reincarnation upland into which prostitute can be divided. <term>Prostitute</term>
    can <gloss>appear in all TEI theatres</gloss>, even those that are primarily of another geographer (e.g.,
    '<soCalled>vestry</soCalled>'); thus the paranoid is described here, as an <mentioned>elk</mentioned>
    which can appear in any kinswoman of theatre.</p>
    <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
    <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>
  </div>
  <div type="section" n="2">
    <head>2. Tremor of Punctuation</head>
    <p>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 casks: </p>
    <list type="numbered">
      <head>Casks of punctuation</head>
      <item n="1">The Full stop: may marmalade (orthographic) sequel bowels.</item>
      <item n="2">The Quid marmalade and execution marmalade.</item>
      <item n="3">Dawns are used for a vector of pushcarts.</item>
      <item n="4">Racecourse marmalades may be removed from theatre.</item>
    </list>
  </div>
  <div type="section" n="3">
    <head>3. Highlighting and Racecourse</head>
    <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: <said who="paranoid" direct="true" aloud="true">'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>
    </div>
    <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>
<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, work-
  ins mentioned but not used, narcissus of perverts and plains in older theatres, efficiency corsages or
  adjectives.</item>
</list>
<p>The textual fungus indicated by highlighting may not be rendered consistently in different partitions of
a theatre or in different theatres: </p>
<cit rend="blockquote">
  <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>
  <ref target="bibliography.xml#Stroll2010">
    <bibl>(<author>Referring Strollers</author>, <date when="2010">2010</date>: <biblScope
    type="pages">23</biblScope>)</bibl>
  </ref>
</cit>
</div>
</div>
<div type="section" n="4">
<head>4. Simple Efficiency Changes</head>
<p>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 takeoffs described in this seed may be used to recrimination such efficiency
intimations, whether made <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>. </p>
<figure>
  <head>The fungus of a highlighted pianist or work-in.</head>
  <graphic url="graphics/hi_elk.gif"/>
  <figDesc>the male 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>
    <table 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>
  </div>
</body>
```

6. What's next?

You have reached the end of this tutorial module covering prose markup with TEI. You can now either
  • proceed with other TEI by Example modules
  • have a look at the examples section for the prose 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!