Talk 4: Core TEI Elements

Dot Porter

July 2009
Elements Available in All TEI Documents

The so-called 'Core' module groups together elements which may appear in any kind of text and the tags used to mark them in all TEI documents. This includes:

- paragraphs
- highlighting, emphasis and quotation
- simple editorial changes
- basic names numbers, dates, addresses
- simple links and cross-references
- lists, notes, annotation, indexing
- graphics
- reference systems, bibliographic citations
- simple verse and drama
<p>(paragraph) marks paragraphs in prose</p>

- Fundamental unit for prose texts
- <p>can contain all the phrase-level elements in the core</p>
- <p>can appear directly inside <body> or inside <div> (divisions)</p>

It was a cottage, the cottage of a dream. And by a cottage I mean, not four plain rooms and a kitchen, but one surprising room opening into another; rooms all on different levels and of different shapes, with delightful places to bump your head on; open fireplaces; a large square hall, oak-beamed, where your guests can hang about after breakfast, while deciding whether to play golf or sit in the garden. Yet all so cunningly disposed that from outside it looks only a cottage or, at most, two cottages persuaded into one. </p>
Highlighting

By *highlighting* we mean the use of any combination of typographic features (font, size, hue, etc.) in a printed or written text in order to distinguish some passage of a text from its surroundings. For words and phrases which are:

- distinct in some way (e.g. foreign, archaic, technical)
- emphatic or stressed when spoken
- not really part of the text (e.g. cross references, titles, headings)
- a distinct narrative stream (e.g. an internal monologue, commentary)
- attributed to some other agency inside or outside the text (e.g. direct speech, quotation)
- set apart in another way (e.g. proverbial phrases, words mentioned but not used)
Highlighting Examples

- `<hi>` (general purpose highlighting)

```html
<p>The rest of this communication is omitted owing to considerations of space. --><hi rend="sc">Ed</hi>.]</p>
```

- `<distinct>` (linguistically distinct)

```
But then I remind myself that the Russian ballet is nothing if not <distinct>bizarre</distinct>.
```

- Other similar elements include: `<emph>`, `<mentioned>`, `<soCalled>`, `<term>` and `<gloss>`
Quotation marks can be used to set off text for many reasons, so the TEI has the following elements:

- `<q>` (separated from the surrounding text with quotation marks)
- `<said>` (speech or thought)
- `<quote>` (passage attributed to an external source)
- `<cit>` (groups a quotation and citation)

```xml
<p>
  <said who="#Celia">I know a lovely tin of potted grouse,</said> said Celia, and she went off to cut some sandwiches. By twelve o'clock we were getting out of the train.
</p>
```
Simple Editorial Changes: `<choice>` and Friends

- `<choice>` (groups alternative editorial encodings)
- Errors:
  - `<sic>` (apparent error)
  - `<corr>` (corrected error)
- Regularization:
  - `<orig>` (original form)
  - `<reg>` (regularized form)
- Abbreviation:
  - `<abbr>` (abbreviated form)
  - `<expan>` (expanded form)
I profess not to know how women's
<choice>
  <orig>heartes</orig>
  <reg>hearts</reg>
</choice> are wooed and won. To me they have always been <choice>
  <sic>maters</sic>
  <corr>matters</corr>
</choice> of riddle and <choice>
  <abbr>admirat'n</abbr>
  <expan>admiration</expan>
</choice>.
Elements Available in All TEI Documents

Additions, Deletions, and Omissions

• `<add>` (addition to the text, e.g. marginal gloss)
• `<del>` (phrase marked as deleted in the text)
• `<gap>` (indicates point where material is omitted)
• `<unclear>` (contains text unable to be transcribed clearly)
Example of `<add>`, `<del>`, `<gap>`, and `<unclear>`

`<add place="left">`The Cause`</add>` The immediate cause, however, of the prevalence of supernatural tales`</del>`
`<add place="supra">stories</add>` in these parts, was doubtless owing to the `<unclear reason="blood splatter">vicinity</unclear>` of Sleepy Hollow. `<gap reason="illegible">`The rest of this paragraph is covered in dried blood. `</gap>`
Basic Names

- `<name>` (a name in the text, contains a proper noun or noun phrase)
- `<rs>` (a general-purpose name or referencing string)

The `@type` attribute is useful for categorizing these, and they both also have `@key`, `@ref`, and `@nymRef` attributes.
The scene opens at a party given by Potiphar in Venice. It is when the natural end of the story is reached, and Simon has come into his own and has just been wedded to his proper affinity, that the structure seems to me to fall with a crash. I might perhaps, though not without reluctance, have pardoned an impertinent railway accident which leaves the young man apparently crippled for life.
Addresses

- `<email>` (an electronic mail address)
- `<address>` (a postal address)
- `<addrLine>` (a non-specific address line)
- `<street>` (a full street address)
- `<postCode>` (a postal (or zip) code)
- `<postBox>` (a postal box number)
- `<name>` can also be used
- and the 'namesdates' module extends this with more geographic names
<email>gbs@heaven.com</email>

<address>
  <name>George Bernard Shaw</name>
  <addrLine>Shaw's Corner</addrLine>
  <settlement>Ayot St Lawrence</settlement>
  <district>Hertfordshire</district>
  <postCode>HE 1 XXX</postCode>
  <country>England</country>
</address>
Basic Numbers and Measures

- `<num>` (marks a number of any sort)
- `<measure>` (marks a quantity or commodity)
- `<measureGrp>` (groups specifications relating to a single object)
- While `<num>` has simple `@type` and `@value` attributes, `<measure>` has `@type`, `@quantity`, `@unit` and `@commodity` attributes
They went off at a pace I am bound to deplore,

For they did twenty yards in a minute or more

And a yard or two over, a capital score

For Farnaby Fullerton Rigby.

If neither of these values is available, a value of 20,35 for ash content can be assumed initially and checked, after the sampling has been carried out, using one of the methods described in ISO 13909-7.

It is on these days that we travel to our Castle of Stopes; as the crow flies, fifteen miles away. Indeed, that is the way we get to it, for it is a castle in the air.
Elements Available in All TEI Documents

Dates

- `<date>` (contains a date in any format and includes a `@when` attribute for a regularised form and a `@calendar` attribute to specify what calendar system)
- `<time>` (contains a time in any format and includes a `@when` attribute for a regularised form)

```xml
<p>At <time when="09:30:00">9.30 o'clock</time>, as the fog lifted somewhat, the rescuing steamer Lyonnesse had sighted the Gothland, fast on the rocks, with a bad list to starboard, and apparently partly filled with water.</p>
<p>House of Commons, <date when="1914-06-22">Monday, June 22, 1914</date>.</p>
```
Simple Linking

- `<ptr>` (defines a pointer to another location)
- `<ref>` (defines a reference to another location, with optional linking text)
- Both elements have:
  - `@target` attribute taking a URI reference
  - `@cRef` attribute for canonical referencing schemes
- If the linking text is able to be generated, `<ptr>` and `<ref>` might be used in the same place.
Simple Linking Example

See `<ref target="#Section12">section 12 on page 34</ref>.

See `<ptr target="#Section12"/>.`
Lists

- `<list>` (a sequence of items forming a list)
- `<item>` (one component of a list)
- `<label>` (label associated with an item)
- `<headLabel>` (heading for column of labels)
- `<headItem>` (heading for column of items)
The previous slide contained only:

```xml
<div>
  <head>Lists</head>
  <p>
    <list>
      <item>
        <gi>list</gi> (a sequence of items forming a list)</item>
      <item>
        <gi>item</gi> (one component of a list)</item>
      <item>
        <gi>label</gi> (label associated with an item)</item>
      <item>
        <gi>headLabel</gi> (heading for column of labels)</item>
      <item>
        <gi>headItem</gi> (heading for column of items)
      </list>
    </p>
  </div>
```
Notes

- `<note>` (contains a note or annotation)
- Notes can be those existing in the text, or provided by the editor of the electronic text
- A `@place` attribute can be used to indicate the physical location of the note
- Although notes should usually be encoded where its identifier/mark first appears, notes can also be kept separately and point back to their location with a `@target` attribute
It is not only misfortune that makes strange bedfellows. 

By-the-by, it is denied that Sir Joseph Beecham was in any way responsible for the Government's Pills for Earthquakes, by which it was hoped to avert the Irish crisis.
• `<graphic>` (indicates the location of an inline graphic, illustration, or figure)

• `<binaryObject>` (encoded binary data embedding a graphic or other object)

• The `figure` module provides `<figure>` and `<figDesc>` for more complex graphics

```xml
<figure>
  <graphic url="images/014.png"/>
  <head>Garden City Washing-day.</head>
  <p>Our sensitive artist insists on a harmonious colour-scheme.</p>
  <figDesc>A bearded man sits in a deckchair and wags his finger at a woman hanging up washing</figDesc>
</figure>
```