

# Talk 3: Core Elements of the TEI Guidelines

James Cummings

28 January 2014

## Core elements

The *core* module of the TEI 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

# Paragraphs

`<p>` (paragraph) marks paragraphs in prose

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

```
<p>Thanks for yours of this morning. I hope <lb/>you have had my card posted last Monday. <lb/>On Mond. next I lecture the <orgName ref="#Fieldclub">Field Club</orgName> - <lb/>a Nat. Hist. Association, in the lines of our <lb/>old Society - Geological, (you + me) + Botanical <lb/>(New) Do you remember: you<supplied>r</supplied> old <lb/>Black Molt?</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); `<distinct>` (linguistically distinct)

```
<p>Last week I wrote (to order) a strong <lb/>bit of Blank:
on <hi rend="ul">Antaeus v.
  Heracles</hi>. <lb/>These are the best lines, methinks:
<lb/>(N.B. Antaeus deriving
strength from his Mother Earth <lb/>nearly licked old
<distinct>Herk</distinct>.) </p>
```

- Other similar elements include: `<emph>`, `<mentioned>`, `<soCalled>`, `<term>` and `<gloss>`

# Emphasis, Foreign Words, and Unusual Language

- `<foreign>` identifies a word or phrase as belonging to some language other than that of the surrounding text.
- `<mentioned>` marks words or phrases mentioned, not used.
  - `<emph>` marks words or phrases which are stressed or emphasized for linguistic or rhetorical effect.
- `<distinct>` identifies any word or phrase which is regarded as linguistically distinct, for example as archaic, technical, dialectal, non-preferred, etc., or as forming part of a sublanguage.
  - `<term>` contains a single-word, multi-word, or symbolic designation which is regarded as a technical term.

## Foreign words

The *@xml:lang* attribute tells us what language is in use

John eats a `<foreign xml:lang="fr">croissant</foreign>` every morning.

`<mentioned xml:lang="fr">Croissant</mentioned>` is difficult to pronounce with your mouth full.

A `<term xml:lang="fr">croissant</term>` is a crescent-shaped piece of light, buttery, pastry that is usually eaten for breakfast, especially in France.

## Quotation

Quotation marks can be used to set off text for many reasons, so the TEI has the following elements:

- `<q>` (indicated by quotation marks)
- `<soCalled>` (disclaimed responsibility or scare quotes)
- `<said>` (speech or thought)
- `<quote>` (passage attributed to an external source)
- `<cit>` (groups a quotation and citation)
- `<bibl>` is used to give the source of a quote

```
<p>  
  <said who="#Celia">I know a lovely tin of potted grouse,</said> said Celia,  
  and she went offto cut some sandwiches.  
</p>
```



Using `<q>`

The `<q>` element should be used any time there are quotation marks where you are unable or don't want to give further semantic interpretation.

Mb 3 is endorsed:

`<q>`Aturnat R`<ex>`egis`</ex>` versus Batteley et `<gap/>`

`<lb/>` M: 12o

Ja`<ex>`cobi`</ex>`,`</q>` and in another place

`<q>`Martis octavo die Nouembris

`<lb/>` Anno duodecimo `<lb/>` Jacobis Regis `<lb/>` Harker.`</q>`

# 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)

## Choice Example

```
<p>...any might,  
<unclear reason="scribbled">majesty</unclear>,  
<choice>  
  <abbr>domin</abbr>  
  <expansion>domin<ex>ion</ex>  
</expansion>  
</choice> or power...</p>
```

# Choice Example

```

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>.

```

## 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>

```

<p>
  <add place="left">My </add>
  <del rend="stoked">It's </del>
  <add place="above">
    <del rend="stoked">The </del>
  </add> subject <del rend="stoked">of</del> is War, and the
  <unclear>pity </unclear>of <del rend="stoked">it</del> War.
  <lb/> The Poetry is in the pity.
</p>

```

## 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.

# Basic Names Example

```
<p>The scene opens at a party given by <name  
  nymRef="http://www.meanings-of-name.com/potiphar.html">Potiphar</name>  
in <name  
  ref="http://en.wikipedia.org/wiki/Venice"  
  type="place"> Venice</name>. </p>  
<p>It is when the natural end of the story is reached, and  
<name xml:id="SIMON">Simon</name> 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 <rs corresp="#SIMON">the  
young man</rs> apparently crippled for life.</p>
```



# 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

## Basic Address Example

The REED office address is:

```
<address>
  <addrLine>Jackman Humanities Building</addrLine>
  <addrLine>University of Toronto</addrLine>
  <street>170 St George St, Suite 810</street>
  <addrLine>Toronto, Ontario, Canada</addrLine>
  <postCode>M5R 2M8</postCode>
</address>
```

```
<address>
  <placeName role="building">Jackman Humanities Building</placeName>
  <orgName>University of Toronto</orgName>
  <street>170 St George St, Suite 810</street>
  <settlement>Toronto</settlement>
  <region>Ontario</region>
  <country>Canada</country>
  <postCode>M5R 2M8</postCode>
</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

## Number and Measure Examples

```
<l>With a <num value="1000">thousand</num> pains that  
vision's face was grained;</l>
```

```
... only  
<measure type="distance" unit="m" quantity="3218.69">two  
miles</measure> from the front....
```

## More Number and Measure Examples

```
<l>They went off at a pace I am bound to deplore,</l>
<l>For they did <num value="20">twenty</num> yards in a minute or more</l>
<l>And a yard or <num value="2">two</num> over, a capital score</l>
<l>For Farnaby Fullerton Rigby.</l>
```

```
<p>If neither of these values is available, a value of <num>20,35</num>
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.</p>
```

```
It is on these days that we travel to our Castle of Stopes; as the
crow flies, <measure quantity="24140" unit="m">fifteen miles</measure>
away. Indeed, that is the way we get to it, for it is a castle in the
air.
```

## Basic Dates and Times

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

```
<date when="1917-07">July 1917.<lb/> Wednesday</date>
```

# 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

## Simple Linking Example

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

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

The `<ref target="http://www.bbc.co.uk/">BBC web site</ref>`  
has a  
good sports section



# 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)

# Simple List Example

The previous slide contained only:

```
<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)</item>
    </list>
  </p>
</div>
```

## Simple List Example

```
<p>  
  <hi>To which is added,</hi> A Collection of LETTERS of  
  Friendship, and other Occasional LETTERS, written by  
  <list>  
    <item>Mr. <hi>Dryden,</hi>  
    </item>  
    <item>Mr. <hi>Wycherly,</hi>  
    </item>  
    <item>Mr.-</item>  
    <item>Mr. <hi>Congreve,</hi>  
    </item>  
    <item>Mr. <hi>Dennis,</hi> and other Hands.</item>  
  </list>  
</p>
```

## 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
- Notes should usually be encoded where the identifier/mark first appears; notes can also be kept separately and point back to their location with a *@target* attribute

```
<p>It is not only misfortune that makes strange bedfellows.  
<note place="foot">By-the-by, it is denied that Sir <name>Joseph  
Beecham</name> was in any way responsible for the Government's <title>Pills for  
Earthquakes</title>, by which it was hoped to avert the Irish crisis.</note>  
</p>
```

# Indexing

- If converting an existing index, use nested lists. For auto-generated indexes:
- `<index>` (marks an index entry) with optional *@indexName* attribute
- The `<term>` element is used to mark a term inside an `<index>` element
- The `<index>` element can self-nest for hierarchical index entries

# Indexing Example

```
<p>Last week I wrote (to order) a strong <lb/>bit of  
Blank<index>  
  <term>Verse</term>  
  <index>  
    <term>Blank Verse</term>  
  </index>  
</index>:</p>
```

# Graphics

- `<graphic>` (indicates the location of an inline graphic, illustration, or figure)
- `<binaryObject>` (encoded binary data embedding a graphic or other object)
- The *figures* module provides `<figure>` and `<figDesc>` for more complex graphics

```
<figure>  
  <graphic url="images/postcard-front.jpg"/>  
  <head>Postcard front</head>  
  <figDesc>A postcard image of two men relaxing at a table,  
  smoking pipes and drinking. A dog and potted fruit tree are  
  nearby with a house over the wall in the distance.</figDesc>  
</figure>
```

## Milestone Elements

- `<milestone>` marks a boundary point separating any kind of section of a text, typically but not necessarily indicating a point at which some part of a standard reference system changes, where the change is not represented by a structural element.
- `<gb>` (gathering begins) marks the point in a transcribed codex at which a new gathering or quire begins.
- `<pb>` (page break) marks the start of a new page in a paginated document.
- `<lb>` (line break) marks the start of a new (typographic) line in some edition or version of a text.
- `<cb>` (column break) marks the beginning of a new column of a text on a multi-column page.



## Simple Verse and Drama

- `<l>` (verse line) contains a single, possibly incomplete, line of verse.
- `<lg>` (line group) contains one or more verse lines functioning as a formal unit, e.g. a stanza, refrain, verse paragraph, etc.
- `<sp>` (speech) contains an individual speech in a performance text, or a passage presented as such in a prose or verse text.
- `<speaker>` contains a specialized form of heading or label, giving the name of one or more speakers in a dramatic text or fragment.
- `<stage>` (stage direction) contains any kind of stage direction within a dramatic text or fragment.

## Simple Verse Drama with a Milestone Example

```

<sp who="#Ann">
  <speaker>Anna</speaker>
  <pb n="38r"/>
  <l>And by thes letturs þ<ex>a</ex>
    <hi rend="sup">t</hi> be most reverre<ex>n</ex>t</l>
  <l>Take them in hand full agre þ<ex>er</ex>to</l>
  [...]
</sp>

```

## Simple Verse

```
<lg type="stanza">
  <l>It seemed that out of battle I escaped</l>
  <l>Down some profound dull tunnel, long since scooped</l>
  <l>Through granites which titanic wars had groined.</l>
</lg>
<lg type="stanza">
  <l>Yet also there encumbered sleepers groaned, </l>
  <l>Too fast in thought or death to be bestirred. </l>
  <l>Then, as I probed them, one sprang up, and stared </l>
  <l>With piteous recognition in fixed eyes, </l>
  <l>Lifting distressful hands, as if to bless. </l>
  <l>And by his smile, I knew that sullen hall, --- </l>
  <l>By his dead smile I knew we stood in Hell.</l>
</lg>
```

# Simple Drama

```
<sp>  
  <speaker>The reverend Doctor Opimiam</speaker>  
  <p>I do not think I have named a single unpresentable  
fish.</p>  
</sp>  
<sp>  
  <speaker>Mr Gryll</speaker>  
  <p>Bream, Doctor: there is not much to be said for  
bream.</p>  
</sp>
```

## Basic Bibliographic Records

- `<bibl>` (bibliographic citation) contains a loosely-structured bibliographic citation of which the sub-components may or may not be explicitly tagged.
- `<biblStruct>` (structured bibliographic citation) contains a structured bibliographic citation, in which only bibliographic sub-elements appear and in a specified order.
- `<listBibl>` (citation list) contains a list of bibliographic citations of any kind.
- `<cit>` (cited quotation) contains a quotation from some other document, together with a bibliographic reference to its source. In a dictionary it may contain an example text with at least one occurrence of the word form, used in the sense being described, or a translation of the headword, or an example.

## Other Bibliographic Elements

Elements such as `<analytic>`, `<author>`, `<biblScope>`, `<citedRange>`, `<editor>`, `<imprint>`, `<meeting>`, `<monogr>`, `<pubPlace>`, `<publisher>`, `<relatedItem>`, and `<title>` are also defined in the Core module.

# Basic Bibliographic Citation

```
<bibl>  
  <title level="a">Enigma</title>, <title level="j">Punch:  
  or the London Charivari</title>,  
  <date when="1914-07-01">July 1, 1914</date>, 147, p. 6  
</bibl>
```

## A <biblStruct> Example

```
<biblStruct xml:lang="fr">
  <monogr>
    <author>Eugène Sue</author>
    <title>Martin, l'enfant trouvé</title>
    <title type="sub">Mémoires d'un valet de chambre</title>
    <imprint>
      <pubPlace>Bruxelles et Leipzig</pubPlace>
      <publisher>C. Muquardt</publisher>
      <date when="1846">1846</date>
    </imprint>
  </monogr>
</biblStruct>
```



## A <cit> Example

In Punch there is an brief note which could be misconstrued as a slur upon Canadians.

```

<cit>
  <quote>
    <q>MEN FOR THE ANTARCTIC.<lb/>105 Canadian Dogs to go with
    Sir E. Shackleton.</q>
    <lb/>
    <title>Daily Express.</title>
    <p>A gay lot, these Canadians!</p>
  </quote>
  <bibl>
    <title level="a">Men for the Antarctic</title>
    <title level="j">Punch: or the London Charivari</title>,
    <date when="1914-07-01">1 July 1914</date>
    <biblScope>p. 6</biblScope>
  </bibl>
</cit>

```

It should not be understood as such.

## Another <biblStruct> Example

```
<biblStruct>
  <monogr>
    <title>Magnalia Christi Americana: or, The
      ecclesiastical history of New-England, ...</title>
    <author>Mather, Cotton (1663-1728)</author>
    <imprint>
      <publisher>Printed for Thomas Parkhurst, at the
        Bible and Three Crowns in Cheapside.</publisher>
      <pubPlace>London</pubPlace>
      <date when="1702">MDCCII</date>
    </imprint>
  </monogr>
</biblStruct>
```

# The Core Module

Elements defined: <abbr> <add> <addrLine> <address>  
<analytic> <author> <bibl> <biblScope> <biblStruct>  
<binaryObject> <cb> <choice> <cit> <citedRange> <corr>  
<date> <del> <desc> <distinct> <divGen> <editor> <email>  
<emph> <expan> <foreign> <gap> <gb> <gloss> <graphic>  
<head> <headItem> <headLabel> <hi> <imprint> <index>  
<item> <l> <label> <lb> <lg> <list> <listBibl> <measure>  
<measureGrp> <media> <meeting> <mentioned> <milestone>  
<monogr> <name> <note> <num> <orig> <p> <pb>  
<postBox> <postCode> <ptr> <pubPlace> <publisher> <q>  
<quote> <ref> <reg> <relatedItem> <resp> <respStmt>  
<rs> <said> <series> <sic> <soCalled> <sp> <speaker>  
<stage> <street> <teiCorpus> <term> <textLang> <time>  
<title> <unclear>

## Next

Any Questions? Next, the timetable says we're going to have another talk (maybe after a minute to catch my breath!).