Elements for basic TEI documents

This is more of a brief reference sheet than an exhaustive list of TEI elements: it is intended to provide you with a way to look up the most commonly used elements, grouped together for the exercises in which we'll be encountering them. For detailed information about the contents and semantics of these elements (and for other more arcane elements), have a look at the TEI Guidelines `(http://www.tei-c.org/release/doc/tei-p5-doc/html/)`.

Simple prose

`<div>`
A division of a text: for instance, an act, a chapter, a section, a poem, a letter… Use the `type=` attribute to indicate what kind of division.

`<head>`
The heading of a division: contains words and phrase-level encoding. `<head>` may appear at the start of `<div>`, but also at the start of `<body>`, `<front>`, `<back>`, `<list>`, and `<lg>`.

`<p>`
A prose paragraph: contains words and phrase-level encoding.

`<list>`
A list: contains a series of `<item>` elements.

`<item>`
An item in a list: contains an optional `<label>` followed by words and phrase-level encoding, or a series of paragraphs.

`<label>`
The label of an item (e.g. a letter, number, or word indicating its order or other facts about it): contains words and phrase-level encoding. Note that `<label>` can also be the first element inside a paragraph.

`<said>`
Passages spoken aloud or thought, e.g. by a character on a novel

`<quote>`
Used to encode quotations from other sources; contains words and phrase-level encoding.

Phrase-level encoding

<name>
Used to encode all kinds of names. If you want to distinguish between different kinds of names, you can use the `type=` attribute (e.g. `<name type="person">`). TEI also includes specific elements for different kinds of names (e.g. `<persName>`) for projects that need more detailed encoding.

<date>
Used to encode dates. The `when=` attribute can be used to encode a regularized form of the date (e.g. `<date when="2001">The first year of the new century</date>` or `<date when="2005-05-29">Sun, 29 May 05</date>`).

<foreign>
Used for foreign-language words when no other element (e.g. `<quote>`) is already present.

<distinct>
Used for linguistically distinct words (e.g. dialect words, regionally accented words).

<mentioned>
Used for words which are mentioned but not used (for instance, for spelling or definition purposes).

<term>
Used to encode specialized terminology; often associated with a `<gloss>`.

<emph>
Used to encode emphasized words or phrases.

<soCalled>
Used to encode (or express) authorial distance; e.g., phrases that were or should be in scare quotes.

<hi>
Used to encode words or phrases which are highlighted for reasons which the encoder either does not know or chooses not to analyse.

<q>
Used to encode passages surrounded by quotation marks, when you don't want to bother with a more precise element like `<said>`. Roughly the same as `<hi rend="surrounded-with-quotation-marks">`.

Poetry

`<lg>`
> A group of verse lines: contains one or more `<l>` elements.

`rhyme=`
> May be optionally used to specify the rhyme scheme of the line group.

`<l>`
> A single verse line: contains words and phrase-level elements.

`met=`
> May be optionally used to specify the metrical pattern of the line.

`<rhyme>`
> May be optionally used to indicate the portion of the metrical line that rhymes, and with its `label=` attribute which part of the rhyme scheme is in play.

Simple drama

`<sp>`
> A dramatic speech

`<speaker>`
> A speaker identification printed in the text

`<stage>`
> A stage direction. The `type=` attribute may be used to identify the kind of stage direction; suggested values include:
>> ```
>> business
>> costume
>> delivery
>> entrance
>> exit
>> location
>> narrative
>> novelistic
>> ```

`<castList>`
> A cast list in a dramatic text, listing the roles in the drama. It consists of one or more `<castItem>` or `<castGroup>` elements.

`<castGroup>`
> A grouping of related items in a cast list, containing one or more `<castItem>` elements and an optional `<head>` and `<trailer>`.

`<castItem>`

An item in a cast list, containing a `<role>` and an optional `<roleDesc>`.

`<role>`

The name of a role in a cast list

`<roleDesc>`

The description of a role in a cast list

Text structure

`<TEI>`

The outermost (or *'root'*) element for any TEI P5 conformant document. It groups together the TEI header and the document text. It must have the TEI namespace specifed, and should have an `xml:lang=` attribute, i.e. `<TEI xmlns="http://www.tei-c.org/ns/1.0" xml:lang="en">`.

`<teiHeader>`

The wrapper for all of the document's metadata. The elements that go inside the TEI header are too numerous to list usefully here; see the templates for details.

`<text>`

The wrapper element which contains all of the document's content. The `<text>` element is most often used for a single work (i.e. a single published document, or a single aesthetic unit such as a play or a work of fiction). Terms like "single work" and "aesthetic unit" need to be defined by the individual project. A `<text>` element contains an optional `<front>`, a mandatory `<body>`, and an optional `<back>`.

`<front>`

Contains the front matter of the document, if any: title pages, tables of contents, introductory essays, and so forth. The `<front>` element contains an optional `<titlePage>` and may be subdivided into `<div>` elements.

`<body>`

Contains the main body of the document, not including front matter and back matter. The `<body>` element typically includes one or more `<div>` elements. It may start with a `<head>`. (Think about where the `<head>` belongs—is it the heading for the body, or the heading for the first division?)

`<back>`

Contains the back matter of the document, if any: indices, appendices, epilogues, colophons, errata lists, etc. May be subdivided into `<div>` elements if necessary.

<group>
An element which groups together multiple <text> elements, with an optional <front> and <back>.

Complex prose

<note>
A note (a footnote, endnote, marginal note, or inline note). Link the note to the point where it's anchored using xml:id= and target=. <note> contains words and phrase-level encoding.

<anchor>
An anchor point, usually used as a place for some other element (such as a note) to point to, using the anchor's xml:id= attribute.

<opener>
This element may appear at the start of a <div>, <text>, <front>, or <back>, and it groups together the elements that appear at the start of a letter or similar document: the date and place of writing (using <dateLine>, and the salutation to the person being addressed (using <salute>).

<closer>
Very similar to <opener>, but located at the end of the <div> instead of at the beginning.

<trailer>
This element is used for things that come at the very end of the document or section, such as "The End".

<dateline>
Used within <opener> and <closer> to encode the date and place of writing. Contains words and phrase-level encoding.

<salute>
Used within <opener> and <closer> to encode the salutation to the person being addressed (e.g. "Dear Sir", or "I remain faithfully yours…"). Contains words and phrase-level encoding.

<signed>
Used within <closer> to encode the signature or name of the person writing. Contains words and phrase-level encoding.

<postscript>
Used to encode a postscript, e.g. of a letter.

<bibl>
Used to encode bibliographical references, either in a list (using <listBibl>) or in running prose.

Alternative Encodings

`<choice>`
Groups together two or more alternate encodings of a phrase-level passage, using the elements listed below.

`<abbr>`
An abbreviation; may be used alone or, when inside `<choice>`, in combination with `<expan>` which holds an expanded reading.

`<expan>`
The expanded reading of an abbreviation; typically used inside `<choice>`, in combination with `<abbr>` which holds the corresponding abbreviated reading. Rarely used alone.

`<sic>`
A typographical error or oddity in the original; may be used alone or, when inside `<choice>`, in combination with `<corr>`, which holds a corrected reading.

`<corr>`
A corrected reading of a typographical error or oddity in the original; may be used alone or, when inside `<choice>`, in combination with `<sic>`, which holds the original reading.

`<orig>`
An unmodernized reading in the original; may be used alone or, when inside `<choice>`, in combination with `<reg>`, which holds a regularized reading.

`<reg>`
A modernization of a reading in the original; may be used alone or, when inside `<choice>`, in combination with `<orig>`, which holds the corresponding unmodernized reading.

Manuscripts and Encoding Physical Documents

`<pb>`
An empty element which marks the break between one page and another. By convention, information stored in the attributes of `<pb>` refer to the page that *follows* the break. Equivalent to `<milestone unit="page">`.

`<lb>`
An empty element which marks a typographical line break. Equivalent to `<milestone unit="line">`.

<cb>

An empty element which marks the break between one column and the next. Equivalent to `<milestone unit="column">`.

<milestone>

An empty element which marks a boundary point in the text according to some standard reference system, such as signatures, scrolls, leaves. Use the `unit=` attribute to indicate the reference system whose units are being marked at this point.

<add>

A handwritten addition. The `hand=` attribute indicates the handwriting in which the addition is made. This attribute contains an identifier which points to a `<hand>` element in the `<profileDesc>` of the TEI header; this `<hand>` element contains an extended description of the handwriting, ink, and other details.

<addSpan>

An empty element which marks the starting point for a handwritten addition that either is too long to be encoded with `<add>`, or overlaps an element boundary. Its `spanTo=` attribute points to an `<anchor>` element which marks the endpoint of the added material. The `hand=` attribute indicates the handwriting in which the addition is made (see above for details).

<del>

A deletion. The `hand=` attribute indicates the handwriting in which the addition is made (see above for details).

<delSpan>

An empty element which marks the starting point for a deletion that is either too long to be encoded with `<del>` or that overlaps an element boundary. Its `spanTo=` attribute points to an `<anchor>` element which marks the endpoint of the deleted material. The `hand=` attribute indicates the handwriting in which the deletion is made (see above for details).

<handShift>

An empty element which marks the boundary point at which a change of handwriting takes place. Its `new=` attribute indicates the handwriting that begins at the point being marked. The `new=` attribute functions just like the `hand=` attribute, in pointing to a `<hand>` element in the TEI header, which provides detailed information on the handwriting in question.

Transcriptional complexities

`<supplied>`
>Indicates that a given word or passage cannot be read in the original and is being supplied (either through editorial judgment or from some other textual source).

`<unclear>`
>Indicates that a given word or passage is unclear, but not entirely illegible (expresses uncertainty rather than absolute lack of information); multiple alternative readings may be grouped in a `<choice>` element.

`<damage>`
>A damaged portion of the original text; the `type=` attribute allows you to classify the damage, and the `extent=` attribute allows you to indicate the extent of the damage.

`<gap>`
>A gap in the original text (either from damage, deletion, excerption, or some other cause). The `<desc>` child element provides a description of what is missing, and the `reason=` attribute provides the reason for the omission.

`<subst>`
>Combines an addition and a deletion so that the `<add>` is understood as being a substitution for the `<del>`.

`<restore>`
>Indicates restoration of text to an earlier state by cancellation of a marking or instruction; in particular, useful to indicate that a deletion was restored, e.g. by the notation "stet".

`<app>`
>Contains one entry in a critical apparatus, with an optional lemma and at least one reading.

`<rdg>`
>A single reading, e.g. from a particular witness.

`<lem>`
>The reading from the base text.

Attributes

`xml:id=`
>Provides a unique identifier for this particular element, thus allowing other elements to point to it (using their `target=`, `next=`, `prev=`, etc.).

n=
> Provides a label or identifier for this particular element, not necessarily unique.

target=
> Provides a URI (e.g.
> `http://bauman.zapto.org/gallery/Niagara_Falls_2008-01/2008_01_07T16_35_39`
> or `#sect08`) that points to either another document or an element within an XML document (including the current one).

next= and prev=
> Allow what is logically a single text object (e.g. a quotation) to be encoded as a series of two or more discrete XML elements, as a work-around for overlap problems. These attributes represent the connections between these fragmentary elements, by pointing to a prior or subsequent element in the chain of fragments. They do so by referring to that element's `xml:id=` value. That is, if `next=` is specified on a `<said>` element, then its value should be a hash mark (#) followed by the value of the `xml:id=` of another `<said>` element, the one that is the next part of the spoken passage. For example, `<said xml:id="s01" next="#s02">Hey</said>, he said, <said xml:id="s02" prev="#s01">What's up?</said>`

xml:lang=
> Used to indicate the language of an element's content. Its value conforms to BCP 47 (a standard system for defining language codes). For information on how BCP 47 codes are constructed, see the note in the data.language
> (http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-data.language.html)
> documentation. Some sample values for the `xml:lang=` attribute are:

English	en
French	fr
German	de
Italian	it
Latin	la
Arabic as spoken in Iraq	ar-IQ
Chinese	zh
simplified Chinese	zh-Hans
Taiwanese	zh-TW

> If further explanation is required, a `<language>` element with an `ident=` attribute of the same BCP 47 code can be specified in the TEI header.