TEI Lite: Encoding for Interchange: an introduction to the TEI

Final revised edition for TEI P5

Lou Burnard *C. M. Sperberg-McQueen* August 2012

Prefatory note

TEI Lite was the name adopted for what the TEI editors originally conceived of as a simple demonstration of how the TEI (Text Encoding Initiative) encoding scheme might be adopted to meet 90% of the needs of 90% of the TEI user community. In retrospect, it was predictable that many people should imagine TEI Lite to be all there is to TEI, or find TEI Lite to be far too heavy for their needs.

The original TEI Lite (1996) was based largely on observations of existing and previous practice in the encoding of texts, particularly as manifest in the collections of the Oxford Text Archive and other collections of the period. It is therefore unsurprising that it seems to have become, if not a *de facto* standard, at least a common point of departure for electronic text centres and encoding projects world wide. Maybe the fact that we actually produced this shortish, readable, manual for it also helped.

Early adopters of TEI Lite included a number of 'Electronic Text Centers' and digital library initiatives. It was also adopted as the basis for some early TEI-conformant authoring systems, and as the basis for introductory tutorials, many of them in languages other than English (see further the list of legacy versions at `http://www.tei-c.org/Vault/P4/Lite/`).

In 2002, following the publication of TEI P4, the XML version of the TEI Guidelines, which uses the generation of TEI Lite as an example of the TEI modification mechanism, the opportunity was taken to produce a lightly revised XML-conformant version. In 2006, a more substantially revised version based on TEI P5 was produced; this reflected the many changes between TEI P4 and TEI P5, but was not otherwise significantly different. In 2012, the TEI Technical Council, decided that a final revision should be undertaken to ensure that the documentation remained consistent with the latest (2.1) release of TEI P5. This version uses a recently added mechanism in the TEI customization architecture, which permits a customization to define only the TEI elements to be included in a schema, rather than the elements to be excluded from it. As such it is probably more resilient to change than earlier versions.

Lou Burnard, August 2012

TEI schema for text collections by the CLGS group (http://clgs.hypotheses.org).

1 Introduction

The Text Encoding Initiative (TEI) Guidelines are addressed to anyone who wants to interchange information stored in an electronic form. They emphasize the interchange of textual information, but other forms of information such as images and sound are also addressed. The Guidelines are equally applicable in the creation of new resources and in the interchange of existing ones.

The Guidelines provide a means of making explicit certain features of a text in such a way as to aid the processing of that text by computer software running on different machines. This process of making explicit we call *markup* or *encoding*. Any textual representation on a computer uses some form of markup; the TEI came into being partly because of the enormous variety of mutually incomprehensible encoding schemes currently besetting scholarship, and partly because of the expanding range of scholarly uses now being identified for texts in electronic form.

The TEI Guidelines describe an encoding scheme which can be expressed using a number of different formal languages. The first editions of the Guidelines used the *Standard Generalized Markup Language* (SGML); since 2002, this has been replaced by the use of the Extensible Markup Language (XML). These markup languages have in common the definition of text in terms of *elements* and *attributes*, and rules governing their appearance within a text. The TEI's use of XML is ambitious in its complexity and generality, but it is fundamentally no different from that of any other XML markup scheme, and so any general-purpose XML-aware software is able to process TEI-conformant texts.

Since 2001, the TEI has been a community initiative supported by an international membership consortium. It was originally an international research project sponsored by the Association for Computers and the Humanities, the Association for Computational Linguistics, and the Association for Literary and Linguistic Computing, with substantial funding over its first five years from the U.S. National Endowment for the Humanities, Directorate General XIII of the Commission of the European Communities, the Andrew W. Mellon Foundation, the Social Science and Humanities Research Council of Canada and others. The Guidelines were first published in May 1994, after six years of development involving many hundreds of scholars from different academic disciplines worldwide. During the years that followed, the Guidelines became increasingly influential in the development of the digital library, in the language industries, and even in the development of the World Wide Web itself. The TEI Consortium was set up in January 2001, and a year later produced an edition of the Guidelines entirely revised for XML compatibility. In 2004, it set about a major revision of the Guidelines to take full advantage of new schema languages, the first release of which appeared in 2005. This revision of the TEI Lite document conforms to version 2.1 of this most recent edition of the Guidelines, TEI P5, released in June 2012.

At the outset of its work, the overall goals of the TEI were defined by the closing statement of a planning conference held at Vassar College, N.Y., in November, 1987; these 'Poughkeepsie Principles' were further elaborated in a series of design documents. The Guidelines, say these design documents, should:

- suffice to represent the textual features needed for research;
- be simple, clear, and concrete;
- be easy for researchers to use without special-purpose software;
- allow the rigorous definition and efficient processing of texts;
- provide for user-defined extensions;

- conform to existing and emergent standards.

The world of scholarship is large and diverse. For the Guidelines to have wide acceptability, it was important to ensure that:

1. the common core of textual features be easily shared;
2. additional specialist features be easy to add to (or remove from) a text;
3. multiple parallel encodings of the same feature should be possible;
4. the richness of markup should be user-defined, with a very small minimal requirement;
5. adequate documentation of the text and its encoding should be provided.

The present document describes a manageable selection from the extensive set of elements and recommendations resulting from those design goals, which is called *TEI Lite*.

In selecting from the several hundred elements defined by the full TEI scheme, we have tried to identify a useful 'starter set', comprising the elements which almost every user should know about. Experience working with TEI Lite will be invaluable in understanding the full TEI scheme and in knowing how to integrate specialized parts of it into the general TEI framework.

Our goals in defining this subset may be summarized as follows:

- it should be able to handle adequately a reasonably wide variety of texts, at the level of detail found in existing practice (as demonstrated in, for example, the holdings of the Oxford Text Archive);
- it should be useful for the production of new documents (such as this one) as well as the encoding of existing texts;
- it should be usable with a wide range of existing XML software;
- it should be a pure subset of the full TEI scheme and defined using the customizaticon methods described in the TEI Guidelines;
- it should be as small and simple as is consistent with the other goals.

The reader may judge our success in meeting these goals for him or herself.

Although we have tried to make this document self-contained, as suits a tutorial text, the reader should be aware that it does not cover every detail of the TEI encoding scheme. All of the elements described here are fully documented in the TEI Guidelines themselves, which should be consulted for authoritative reference information on these, and on the many others which are not described here. Some basic knowledge of XML is assumed.

2 A Short Example

We begin with a short example, intended to show what happens when a passage of prose is typed into a computer by someone with little sense of the purpose of mark-up, or the potential of electronic texts. In an ideal world, such output might be generated by a very accurate optical scanner. It attempts to be faithful to the appearance of the printed text, by retaining the original line breaks, by introducing blanks to represent the layout of the original headings and page breaks, and so forth. Where characters not available on the keyboard are needed (such as the accented letter *a* in *faàl* or the long dash), it attempts to mimic their appearance.

CHAPTER 38

READER, I married him. A quiet wedding we had: he and I, the parson and clerk, were alone present. When we got back from church, I went into the kitchen of the manor-house, where Mary was cooking the dinner, and John cleaning the knives, and I said --
'Mary, I have been married to Mr Rochester this morning.' The housekeeper and her husband were of that decent, phlegmatic order of people, to whom one may at any time safely communicate a remarkable piece of news without incurring the danger of having one's ears pierced by some shrill ejaculation and subsequently stunned by a torrent of wordy wonderment. Mary did look up, and she did stare at me; the ladle with which she was basting a pair of chickens roasting at the fire, did for some three minutes hang suspended in air, and for the same space of time John's knives also had rest from the polishing process; but Mary, bending again over the roast, said only --
'Have you, miss? Well, for sure!'
A short time after she pursued, 'I seed you go out with the master, but I didn't know you were gone to church to be wed'; and she basted away. John, when I turned to him, was grinning from ear to ear.
'I telled Mary how it would be,' he said: 'I knew what Mr Edward' (John was an old servant, and had known his master when he was the cadet of the house, therefore he often gave him his Christian name) -- 'I knew what Mr Edward would do; and I was certain he would not wait long either: and he's done right, for aught I know. I wish you joy, miss!' and he politely pulled his forelock.
'Thank you, John. Mr Rochester told me to give you and Mary this.'
I put into his hand a five-pound note. Without waiting to hear more, I left the kitchen. In passing the door of that sanctum some time after, I caught the words --
'She'll happen do better for him nor ony o' t' grand ladies.' And again, 'If she ben't one o' th' handsomest, she's noan faa\l, and varry good-natured; and i' his een she's fair beautiful, onybody may see that.'
I wrote to Moor House and to Cambridge immediately, to say what I had done: fully explaining also why I had thus acted. Diana and

Mary approved the step unreservedly. Diana announced that she would just give me time to get over the honeymoon, and then she would come and see me.
'She had better not wait till then, Jane,' said Mr Rochester, when I read her letter to him; 'if she does, she will be too late, for our honeymoon will shine our life long: its beams will only fade over your grave or mine.'
How St John received the news I don't know: he never answered the letter in which I communicated it: yet six months after he wrote to me, without, however, mentioning Mr Rochester's name or alluding to my marriage. His letter was then calm, and though very serious, kind. He has maintained a regular, though not very frequent correspondence ever since: he hopes I am happy, and trusts I am not of those who live without God in the world, and only mind earthly things.

This transcription suffers from a number of shortcomings:

- the page numbers and running titles are intermingled with the text in a way which makes it difficult for software to disentangle them;
- no distinction is made between single quotation marks and apostrophe, so it is difficult to know exactly which passages are in direct speech;
- the preservation of the copy text's hyphenation means that simple-minded search programs will not find the broken words;
- the accented letter in *faàl* and the long dash have been rendered by ad hoc keying conventions which follow no standard pattern and will be processed correctly only if the transcriber remembers to mention them in the documentation;
- paragraph divisions are marked only by the use of white space, and hard carriage returns have been introduced at the end of each line. Consequently, if the size of type used to print the text changes, reformatting will be problematic.

We now present the same passage, as it might be encoded using the TEI Guidelines. As we shall see, there are many ways in which this encoding could be extended, but as a minimum, the TEI approach allows us to represent the following distinctions:

- Paragraph and chapter divisions are now marked explicitly.
- Apostrophes are distinguished from quotation marks; direct speech is explicitly marked.
- The accented letter and the long dash are correctly represented.
- Page divisions have been marked with an empty `<pb>` element alone.
- The lineation of the original has not been retained and words broken by typographic accident at the end of a line have been re-assembled without comment.
- For convenience of proof reading, a new line has been introduced at the start of each paragraph, but the indentation is removed.

```
<pb n="474"/>
<div n="38" type="chapter">
 <p>Reader, I married him. A quiet wedding we had: he and I, the parson and clerk,
were alone
   present. When we got back from church, I went into the kitchen of the manor-house,
where
   Mary was cooking the dinner, and John cleaning the knives, and I said —</p>
 <p>
  <q>Mary, I have been married to Mr Rochester this morning.</q> The housekeeper and
her
   husband were of that decent, phlegmatic order of people, to whom one may at any
time safely
    communicate a remarkable piece of news without incurring the danger of having
one's ears
   pierced by some shrill ejaculation and subsequently stunned by a torrent of wordy
   wonderment. Mary did look up, and she did stare at me; the ladle with which she
was basting
   a pair of chickens roasting at the fire, did for some three minutes hang suspended
in air,
    and for the same space of time John's knives also had rest from the polishing
process; but
   Mary, bending again over the roast, said only —</p>
 <p>
  <q>Have you, miss? Well, for sure!</q>
 </p>
```

```
 <p>A short time after she pursued, <q>I seed you go out with the master, but I didn't
know
     you were gone to church to be wed</q>; and she basted away. John, when I turned
to him, was
   grinning from ear to ear.  <q>I telled Mary how it would be,</q> he said:  <q>I
knew what Mr
     Edward</q> (John was an old servant, and had known his master when he was the
cadet of the
   house, therefore he often gave him his Christian name) — <q>I knew what Mr Edward
would do;
     and I was certain he would not wait long either: and he's done right, for aught
I know.  I
     wish you joy, miss!</q> and he politely pulled his forelock.</p>
 <p>
  <q>Thank you, John.  Mr Rochester told me to give you and Mary this.</q>
 </p>
 <p>I put into his hand a five-pound note.  Without waiting to hear more, I left the
kitchen.
   In passing the door of that sanctum some time after, I caught the words —</p>
 <p>
  <q>She'll happen do better for him nor ony o' t' grand ladies.</q> And again, <q>If
she
     ben't one o' th' handsomest, she's noan faàl, and varry good-natured; and i'
his een she's
     fair beautiful, onybody may see that.</q>
 </p>
 <p>I wrote to Moor House and to Cambridge immediately, to say what I had done:  fully
   explaining also why I had thus acted.  Diana and <pb n="475"/> Mary approved the
step
   unreservedly.  Diana announced that she would just give me time to get over the
honeymoon,
   and then she would come and see me.</p>
 <p>
  <q>She had better not wait till then, Jane,</q> said Mr Rochester, when I read her
letter
   to him; <q>if she does, she will be too late, for our honeymoon will shine our
life long:
     its beams will only fade over your grave or mine.</q>
 </p>
 <p>How St John received the news I don't know:  he never answered the letter in which
I
   communicated it:  yet six months after he wrote to me, without, however, mentioning
Mr
   Rochester's name or alluding to my marriage.  His letter was then calm, and though
very
   serious, kind.  He has maintained a regular, though not very frequent correspondence
ever
   since:  he hopes I am happy, and trusts I am not of those who live without God in
the world,
   and only mind earthly things.</p>
</div>
```

This particular encoding represents a set of choices or priorities. As a trivial example, note that in the second example, end-of-line hyphenation has been silently removed. Conceivably Brontë (or her printer) intended the word 'honeymoon' to appear as 'honey-moon' on its second appearance, though this seems unlikely: our decision to focus on Brontë's text, rather than on the printing of it in this particular edition, makes it impossible to be certain. This is an instance of the fundamental *selectivity* of any encoding. An encoding makes explicit only those textual features of importance to the encoder. It is not difficult to think of ways in which the encoding of even this short passage might readily be extended. For example:

- a regularized form of the passages in dialect could be provided;

- footnotes glossing or commenting on any passage could be added;
- pointers linking parts of this text to others could be added;
- proper names of various kinds could be distinguished from the surrounding text;
- detailed bibliographic information about the text's provenance and context could be prefixed to it;
- a linguistic analysis of the passage into sentences, clauses, words, etc., could be provided, each unit being associated with appropriate category codes;
- the text could be segmented into narrative or discourse units;
- systematic analysis or interpretation of the text could be included in the encoding, with potentially complex alignment or linkage between the text and the analysis, or between the text and one or more translations of it;
- passages in the text could be linked to images or sound held on other media.

TEI-recommended ways of carrying out most of these are described in the remainder of this document. The TEI scheme as a whole also provides for an enormous range of other possibilities, of which we cite only a few:

- detailed analysis of the components of names;
- detailed meta-information providing thesaurus-style information about the text's origins or topics;
- information about the printing history or manuscript variations exhibited by a particular series of versions of the text.

For recommendations on these and many other possibilities, the full Guidelines should be consulted.

3 The Structure of a TEI Text

All TEI-conformant texts contain (a) a *TEI header* (marked up as a <teiHeader> element) and (b) the transcription of the text proper (marked up as a <text> element). These two elements are combined together to form a single <TEI> element, which must be declared within the TEI namespace[1].

The TEI header provides information analogous to that provided by the title page of a printed text. It has up to four parts: a bibliographic description of the machine-readable text, a description of the way it has been encoded, a non-bibliographic description of the text (a *text profile*), and a revision history. The header is described in more detail in section U5-header.

A TEI text may be *unitary* (a single work) or *composite* (a collection of single works, such as an anthology). In either case, the text may have an optional *front* or *back*. In between is the *body* of the text, which, in the case of a composite text, may consist of *group*s, each containing more groups or texts.

A unitary text will be encoded using an overall structure like this:

[1] A *namespace* is an XML concept. Its function is to identify the vocabulary from which a group of element names are drawn, using a standard identifier resembling a web address. The namespace for all TEI elements is `http://www.tei-c.org/ns/1.0`

```
<TEI xmlns="http://www.tei-c.org/ns/1.0">
 <teiHeader>
<!-- [ TEI Header information ] -->
 </teiHeader>
 <text>
  <front>
<!-- [ front matter ... ] -->
  </front>
  <body>
<!-- [ body of text ... ] -->
  </body>
  <back>
<!-- [ back matter ... ] -->
  </back>
 </text>
</TEI>
```

A composite text also has an optional front and back. In between occur one or more groups of texts, each with its own optional front and back matter. A composite text will thus be encoded using an overall structure like this:

```
<TEI xmlns="http://www.tei-c.org/ns/1.0">
 <teiHeader>
<!--[ header information for the composite ]-->
 </teiHeader>
 <text>
  <front>
<!--[ front matter for the composite ]-->
  </front>
  <group>
   <text>
    <front>
<!--[ front matter of first text ]-->
    </front>
    <body>
<!--[ body of first text ]-->
    </body>
    <back>
<!--[ back matter of first text ]-->
    </back>
   </text>
   <text>
    <front>
<!--[ front matter of second text]-->
    </front>
    <body>
<!--[ body of second text ]-->
    </body>
    <back>
<!--[ back matter of second text ]-->
    </back>
   </text>
<!--[ more texts or groups of texts here ]-->
  </group>
  <back>
<!--[ back matter for the composite ]-->
  </back>
 </text>
</TEI>
```

It is also possible to define a composite of complete TEI texts, each with its own header. Such a collection is known as a *TEI corpus*, and may itself have a header:

```
<teiCorpus xmlns="http://www.tei-c.org/ns/1.0">
 <teiHeader>
<!--[header information for the corpus]-->
 </teiHeader>
 <TEI>
  <teiHeader>
<!--[header information for first text]-->
  </teiHeader>
  <text>
<!--[first text in corpus]-->
  </text>
 </TEI>
 <TEI>
  <teiHeader>
<!--[header information for second text]-->
  </teiHeader>
  <text>
<!--[second text in corpus]-->
  </text>
 </TEI>
</teiCorpus>
```

It is also possible to create a composite of corpora -- that is, one `<teiCorpus>` element may contain many nested `<teiCorpus>` elements rather than many nested <TEI> elements, to any depth considered necessary.

In the remainder of this document, we discuss chiefly simple text structures. The discussion in each case consists of a short list of relevant TEI *elements* with a brief definition of each, followed by definitions for any *attributes* specific to that element, and a reference to any *classes* of which the element is a member. These references are linked to full specifications for each object, as given in the TEI *Guidelines*. In most cases, short examples are also given.

For example, here are the elements discussed so far:

<TEI> (TEI document) contains a single TEI-conformant document, comprising a TEI header and a text, either in isolation or as part of a `<teiCorpus>` element.

<teiHeader> (TEI header) supplies the descriptive and declarative information making up an electronic title page for every TEI-conformant document.

<text> contains a single text of any kind, whether unitary or composite, for example a poem or drama, a collection of essays, a novel, a dictionary, or a corpus sample.

UNKNOWN ELEMENT teiCorpus

4 Encoding the Body

As indicated above, a simple TEI document at the textual level consists of the following elements:

<front> (front matter) contains any prefatory matter (headers, title page, prefaces, dedications, etc.) found at the start of a document, before the main body.

UNKNOWN ELEMENT group

<body> (text body) contains the whole body of a single unitary text, excluding any front or back matter.

<back> (back matter) contains any appendixes, etc. following the main part of a text.

Elements specific to front and back matter are described below in section U5-fronbac. In this section we discuss the elements making up the body of a text.

4.1 Text Division Elements

The body of a prose text may be just a series of paragraphs, or these paragraphs may be grouped together into chapters, sections, subsections, etc. Each paragraph is tagged using the <p> tag. The element is used to represent any such grouping of paragraphs.

<p> (paragraph) marks paragraphs in prose.

** ** (text division) contains a subdivision of the front, body, or back of a text.

The *type* attribute on the element may be used to supply a conventional name for this category of text division, or otherwise distinguish them. Typical values might be 'book', 'chapter', 'section', 'part', 'poem', 'song', etc. For a given project, it will usually be advisable to define and adhere to a specific list of such values.

A element may itself contain further, nested, s, thus mimicking the traditional structure of a book, which can be decomposed hierarchically into units such as parts, containing chapters, containing sections, and so on. TEI texts in general conform to this simple hierarchic model.

The *xml:id* attribute may be used to supply a unique identifier for the division, which may be used for cross references or other links to it, such as a commentary, as further discussed in section U5-ptrs. It is often useful to provide an *xml:id* attribute for every major structural unit in a text, and to derive its values in some systematic way, for example by appending a section number to a short code for the title of the work in question, as in the examples below. It is particularly useful to supply such identifiers if the resource concerned is to be made available over the web, since they make it much easier for other web-based applications to link directly to the corresponding parts of your text.

The *n* attribute may be used to supply (additionally or alternatively) a short mnemonic name or number for a division, or any other element. If a conventional form of reference or abbreviation for the parts of a work already exists (such as the book/chapter/verse pattern of Biblical citations), the *n* attribute is the place to record it; unlike the identifier supplied by *xml:id*, it does not need to be unique.

The *xml:lang* attribute may be used to specify the language of the division. Languages are identified by an internationally defined code, as further discussed in section z636 below.

The *rend* attribute may be used to supply information about the rendition (appearance) of a division, or any other element, as further discussed in section U5-hilites below. As with the *type* attribute, a project will often find it useful to predefine the possible values for this attribute, but TEI Lite does not constrain it in anyway.

These four attributes, *xml:id*, *n*, *xml:lang*, and *rend* are so widely useful that they are allowed on any element in any TEI schema: they are *global attributes*. Other global attributes defined in the TEI Lite scheme are discussed in section xatts.

The value of every *xml:id* attribute should be unique within a document. One simple way of ensuring that this is so is to make it reflect the hierarchic structure of the document. For example, Smith's *Wealth of Nations* as first published consists of five books, each of which is divided into chapters, while some chapters are further subdivided into parts. We might define *xml:id* values for this structure as follows:

```
<body>
 <div n="I" type="book" xml:id="WN1">
  <div n="I.1" type="chapter" xml:id="WN101">
<!-- ... -->
  </div>
  <div n="I.2" type="chapter" xml:id="WN102">
<!-- ... -->
  </div>
<!-- ... -->
  <div n="I.10" type="chapter"
   xml:id="WN110">
```

```
    <div n="I.10.1" type="part"
     xml:id="WN1101">
<!-- ... -->
    </div>
    <div n="I.10.2" type="part"
     xml:id="WN1102">
<!-- ... -->
    </div>
  </div>
<!-- ... -->
 </div>
 <div n="II" type="book" xml:id="WN2">
<!-- ... -->
 </div>
</body>
```

A different numbering scheme may be used for *xml:id* and *n* attributes: this is often useful where a canonical reference scheme is used which does not tally with the structure of the work. For example, in a novel divided into books each containing chapters, where the chapters are numbered sequentially through the whole work, rather than within each book, one might use a scheme such as the following:

```
<body>
 <div n="1" type="volume" xml:id="TS01">
  <div n="1" type="chapter" xml:id="TS011">
<!-- ... -->
  </div>
  <div n="2" type="chapter" xml:id="TS012">
<!-- ... -->
  </div>
 </div>
 <div n="2" type="volume" xml:id="TS02">
  <div n="3" type="chapter" xml:id="TS021">
<!-- ... -->
  </div>
  <div n="4" type="chapter" xml:id="TS022">
<!-- ... -->
  </div>
 </div>
</body>
```

Here the work has two volumes, each containing two chapters. The chapters are numbered conventionally 1 to 4, but the *xml:id* values specified allow them to be regarded additionally as if they were numbered 1.1, 1.2, 2.1, 2.2.

4.2 Headings and Closings

Every may have a title or heading at its start, and (less commonly) a trailer such as 'End of Chapter 1' at its end. The following elements may be used to transcribe them:

<head> (heading) contains any type of heading, for example the title of a section, or the heading of a list, glossary, manuscript description, etc.

UNKNOWN ELEMENT trailer

Some other elements which may be necessary at the beginning or ending of text divisions are discussed below in section h52.

Whether or not headings and trailers are included in a transcription is a matter for the individual transcriber to decide. Where a heading is completely regular (for example 'Chapter 1') or may be automatically constructed from attribute values (e.g. `<div type="chapter" n="1">`), it may be omitted; where it contains otherwise unrecoverable text it should always

be included. For example, the start of Hardy's *Under the Greenwood Tree* might be encoded as follows:

```
<div n="Winter" type="Part" xml:id="UGT1">
 <div n="1" type="Chapter" xml:id="UGT11">
  <head>Mellstock-Lane</head>
  <p>To dwellers in a wood almost every species of tree ... </p>
 </div>
</div>
```

4.3 Prose, Verse and Drama

As in the Bronte example above, the paragraphs making up a textual division are tagged with the <p> tag. In poetic or dramatic texts different tags are needed, to represent verse lines and stanzas in the first case, or individual speeches and stage directions in the second. :

<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.

Here, for example, is the start of a poetic text in which verse lines and stanzas are tagged:

```
<lg n="I">
 <l>I Sing the progresse of a
   deathlesse soule,</l>
 <l>Whom Fate, with God made, but doth not controule,</l>
 <l>Plac'd in
   most shapes; all times before the law</l>
 <l>Yoak'd us, and when, and since, in this I
   sing.</l>
 <l>And the great world to his aged evening;</l>
 <l>From infant morne, through manly
   noone I draw.</l>
 <l>What the gold Chaldee, of silver Persian saw,</l>
 <l>Greeke brass, or
   Roman iron, is in this one;</l>
 <l>A worke t'out weare Seths pillars, bricke and
   stone,</l>
 <l>And (holy writs excepted) made to yeeld to none,</l>
</lg>
```

Note that the <l> element marks verse lines, not typographic lines: the original lineation of the first few lines above has not therefore been made explicit by this encoding, and may be lost. The **<lb>** element described in section U5-pln might additionally be used to mark typographic lines if so desired.

Here is the end of a famous dramatic text, in which speeches and stage directions are marked:

```
<sp>
 <speaker>Vladimir</speaker>
 <p>Pull on your trousers.</p>
</sp>
<sp>
 <speaker>Estragon</speaker>
 <p>You want me to pull off my trousers?</p>
```

```
</sp>
<sp>
 <speaker>Vladimir</speaker>
 <p>Pull <emph>on</emph> your trousers.</p>
</sp>
<sp>
 <speaker>Vladimir</speaker>
 <p>
  <stage>(realizing his trousers are down)</stage>.
   True</p>
</sp>
<stage>He pulls up his trousers</stage>
<sp>
 <speaker>Vladimir</speaker>
 <p>Well? Shall we go?</p>
</sp>
<sp>
 <speaker>Estragon</speaker>
 <p>Yes, let's go.</p>
</sp>
<stage>They do not move.</stage>
```

Note that the <stage> (stage direction) element can appear either within a speech or between speeches. The <sp> ("speech") element contains, following an optional <speaker> element indicating who is speaking, either paragraphs (if the speech is in prose) or verse lines or stanzas as in the next example. In this case, it is quite common to find that verse lines are split between speakers. The easiest way of encoding this is to use the *part* attribute to indicate that the lines so fragmented are incomplete :

```
<div n="I" type="Act">
 <head>ACT I</head>
 <div n="1" type="Scene">
  <head>SCENE I</head>
  <stage rend="italic"> Enter Barnardo and Francisco, two Sentinels, at several
doors</stage>
  <sp>
   <speaker>Barn</speaker>
   <l part="Y">Who's there?</l>
  </sp>
  <sp>
   <speaker>Fran</speaker>
   <l>Nay, answer me. Stand and unfold yourself.</l>
  </sp>
  <sp>
   <speaker>Barn</speaker>
   <l part="I">Long live the King!</l>
  </sp>
  <sp>
   <speaker>Fran</speaker>
   <l part="M">Barnardo?</l>
  </sp>
  <sp>
   <speaker>Barn</speaker>
   <l part="F">He.</l>
  </sp>
  <sp>
   <speaker>Fran</speaker>
   <l>You come most carefully upon your hour.</l>
  </sp>
<!-- ... -->
```

```
 </div>
</div>
```

The same mechanism may be applied to stanzas which are divided between two speakers:

```
<div>
 <sp>
  <speaker>First voice</speaker>
  <lg part="I" type="stanza">
   <l>But why drives on that ship so fast</l>
   <l>Withouten wave or wind?</l>
  </lg>
 </sp>
 <sp>
  <speaker>Second Voice</speaker>
  <lg part="F">
   <l>The air is cut away before.</l>
   <l>And closes from behind.</l>
  </lg>
 </sp>
<!-- ... -->
</div>
```

The <sp> element can also be used for dialogue presented in a prose work as if it were drama, as in the next example, which also demonstrates the use of the *who* attribute to bear a code identifying the speaker of the piece of dialogue concerned:

```
<div>
 <sp who="#OPI">
  <speaker>The reverend Doctor Opimian</speaker>
  <p>I do not think I have named a single unpresentable fish.</p>
 </sp>
 <sp who="#GRM">
  <speaker>Mr Gryll</speaker>
  <p>Bream, Doctor: there is not much to be said for bream.</p>
 </sp>
 <sp who="#OPI">
  <speaker>The Reverend Doctor Opimian</speaker>
  <p>On the contrary, sir, I think there is much to be said for him. In the first
     place....</p>
  <p>Fish, Miss Gryll -- I could discourse to you on fish by the hour: but for the
present I
     will forbear.</p>
 </sp>
</div>
```

Here the *who* attribute values (`#OPI` etc.) are links, pointing to a list of the characters in the novel, each of which has an identifier:

```
<list>
 <head>Characters in the novel</head>
 <item xml:id="OPI">
  <name>Dr Opimian</name> : named for the famous Roman fine wine</item>
 <item xml:id="GRM">
  <name>Mr Gryll</name> : named for the mythical Gryllus, one of Ulysses'
   sailors transformed by Circe into a pig, who argues that he was happier in that
state than
   as a man</item>
</list>
```

5 Page and Line Numbers

Page and line breaks etc. may be marked with the following elements.

UNKNOWN ELEMENT pb

UNKNOWN ELEMENT lb

UNKNOWN ELEMENT milestone

These elements mark a single point in the text, not a span of text. The global *n* attribute should be used to supply the number of the page or line beginning at the tag.

When working from a paginated original, it is often useful to record its pagination, if only to simplify later proof-reading. It is also useful for synchronizing an encoded text with a set of page images. Recording the line breaks may be useful for similar reasons.

If features such as pagination or lineation are marked for more than one edition, specify the edition in question using the *ed* attribute, and supply as many tags are necessary. For example, in the following passage we indicate where the page breaks occur in two different editions (ED1 and ED2)

```
<p>I wrote to Moor House and to Cambridge immediately, to say what I had done: fully
explaining also why I had thus acted. Diana and <pb ed="ED1" n="475"/> Mary approved
the step
unreservedly. Diana announced that she would <pb ed="ED2" n="485"/>just give me time
to get
over the honeymoon, and then she would come and see me.</p>
```

A special attribute *break* may be used to indicate whether or not this empty element is considered as a word-breaking, irrespective of any adjacent whitespace. For example, in the following encoded sample:

The <pb> and <lb> elements are special cases of the general class of *milestone* elements which mark reference points within a text. The generic <milestone> element can mark any kind of reference point: for example, a column break, the start of a new kind of section not otherwise tagged, or in general any significant change in the text not marked by an XML element. The names used for types of unit and for editions referred to by the *ed* and *unit* attributes may be chosen freely, but should be documented in the header <refsDecl> element (see refsdecl). The <milestone> element may be used to replace the others, or the others may be used as a set; they should not be mixed arbitrarily.

6 Marking Highlighted Phrases

6.1 Changes of Typeface, etc.

Highlighted words or phrases are those made visibly different from the rest of the text, typically by a change of type font, handwriting style, ink colour etc., which is intended to draw the reader's attention to some associated change.

The global *rend* attribute can be attached to any element, and used wherever necessary to specify details of the highlighting used for it in the source. For example, a heading rendered in bold might be tagged <head rend="bold">, and one in italic <head rend="italic">.

The values to be used for the *rend* attribute are not specified by the TEI Guidelines, since they will depend entirely on the needs of the particular project. Some typical values might include italic, bold etc. for font variations; center, right etc. for alignment; large, small etc. for size; smallcaps, allcaps etc. for type variants and so on. Several such words may be used in combination as necessary, but no formal syntax is proposed. The full TEI Guidelines provide more rigorous mechanisms, using other W3C standards such as CSS, as an alternative to the use of *rend*.

It is not always possible or desirable to interpret the reasons for such changes of rendering in a text. In such cases, the element <hi> may be used to mark a sequence of highlighted text without making any claim as to its status.

<hi> (highlighted) marks a word or phrase as graphically distinct from the surrounding text, for reasons concerning which no claim is made.

In the following example, the use of a distinct typeface for the subheading and for the included name are recorded but not interpreted:

```
<p>
 <hi rend="gothic">And this Indenture further
   witnesseth</hi> that the said <hi rend="italic">Walter Shandy</hi>, merchant, in
consideration of the said intended marriage ...
</p>
```

Alternatively, where the cause for the highlighting can be identified with confidence, a number of other, more specific, elements are available.

UNKNOWN ELEMENT emph

UNKNOWN ELEMENT foreign

UNKNOWN ELEMENT gloss

UNKNOWN ELEMENT label

UNKNOWN ELEMENT mentioned

UNKNOWN ELEMENT term

<title> contains a title for any kind of work.

Some features (notably quotations and glosses) may be found in a text either marked by highlighting, or with quotation marks. In either case, the elements <q> and **<gloss>** (as discussed in the following section) should be used. If the highlighting is to be recorded, use the global *rend* attribute.

As an example of the elements defined here, consider the following sentence:

> On the one hand the *Nibelungenlied* is associated with the new rise of romance of twelfth-century France, the *romans d'antiquité*, the romances of Chrétien de Troyes, and the German adaptations of these works by Heinrich van Veldeke, Hartmann von Aue, and Wolfram von Eschenbach.

Interpreting the role of the highlighting, the sentence might look like this:

```
<p>On the one hand the <title>Nibelungenlied</title>
is associated with the new rise of romance of twelfth-century France, the
<foreign>romans
   d'antiquité</foreign>, the romances of Chrétien de Troyes, ...</p>
```

Describing only the appearance of the original, it might look like this:

```
<p>On the one hand the <hi rend="italic">Nibelungenlied</hi> is associated with the
new rise of romance of twelfth-century France,
the <hi rend="italic">romans d'antiquité</hi>, the romances of Chrétien de Troyes,
...</p>
```

6.2 Quotations and Related Features

Like changes of typeface, quotation marks are conventionally used to denote several different features within a text, of which the most frequent is quotation. When possible, we recommend that the underlying feature be tagged, rather than the simple fact that quotation marks appear in the text, using the following elements:

<q> (quoted) contains material which is distinguished from the surrounding text using quotation marks or a similar method, for any one of a variety of reasons including, but not limited to: direct speech or thought, technical terms or jargon, authorial distance, quotations from elsewhere, and passages that are mentioned but not used.

UNKNOWN ELEMENT mentioned

UNKNOWN ELEMENT soCalled

UNKNOWN ELEMENT gloss

Here is a simple example of a quotation:

```
<p>Few dictionary makers are likely to forget Dr. Johnson's description of the
lexicographer as <q>a harmless drudge.</q>
</p>
```

To record how a quotation was printed (for example, *in-line* or set off as a *display* or *block quotation*), the *rend* attribute should be used. This may also be used to indicate the kind of quotation marks used.

Direct speech interrupted by a narrator can be represented simply by ending the quotation and beginning it again after the interruption, as in the following example:

```
<p>
 <q>Who-e debel you?</q> — he at last said —
<q>you no speak-e, damme, I kill-e.</q> And so saying, the lighted tomahawk began
flourishing about me in the dark.
</p>
```

If it is important to convey the idea that the two <q> elements together make up a single speech, the linking attributes *next* and *prev* may be used, as described in section xatts.

Quotations may be accompanied by a reference to the source or speaker, using the *who* attribute, whether or not this is explicit in the text, as in the following example:

```
<q who="#Wilson">Spaulding, he came
down into the office just this day eight weeks with this very paper in his hand, and
he
says:—<q who="#Spaulding">I wish to the Lord, Mr. Wilson, that I was a red-headed
   man.</q>
</q>
```

This example also demonstrates how quotations may be embedded within other quotations: one speaker (Wilson) quotes another speaker (Spaulding).

The creator of the electronic text must decide whether quotation marks are replaced by the tags or whether the tags are added and the quotation marks kept. If the quotation marks are removed from the text, the *rend* attribute may be used to record the way in which they were rendered in the copy text.

The full TEI Guidelines provide additional elements to distinguish direct speech, quotation, and other typical uses of quotation mark although it is not always possible and may not be considered desirable to interpret the function of quotation marks in a text. For simplicity, only <q> (which may be used for any such case) has been included in TEI Lite.

6.3 Foreign Words or Expressions

Words or phrases which are not in the main language of the texts may be tagged as such in one of two ways. If the word or phrase is already tagged for some reason, the element indicated should bear a value for the global *xml:lang* attribute indicating the language used. Where there is no applicable element, the element **<foreign>** may be used, again using the *xml:lang* attribute. For example:

```
<p>John has real <foreign xml:lang="fr">savoir-faire</foreign>.</p>
<p>Have you read <title xml:lang="de">Die
   Dreigroschenoper</title>?</p>
<p>
```

```
 <mentioned xml:lang="fr">Savoir-faire</mentioned> is French
for know-how.
</p>
<p>The court issued a writ of <term xml:lang="la">mandamus</term>.</p>
```

As these examples show, the <foreign> element should not be used to tag foreign words if some other more specific element such as <title>, <mentioned>, or <term> applies. The global *xml:lang* attribute may be attached to any element to show that it uses some other language than that of the surrounding text.

The codes used to identify languages, supplied on the *xml:lang* attribute, must be constructed in a particular way, and must conform to common Internet standards[2], as further explained in the relevant section of the TEI Guidelines. Some simple example codes for a few languages are given here:

zh	Chinese	grc	Ancient Greek
en	English	el	Greek
enm	Middle English	ja	Japanese
fr	French	la	Latin
de	German	sa	Sanskrit

7 Notes

All notes, whether printed as footnotes, endnotes, marginalia, or elsewhere, should be marked using the same element:

<note> contains a note or annotation.

Where possible, the body of a note should be inserted in the text at the point at which its identifier or mark first appears. This may not be possible for example with marginalia, which may not be anchored to an exact location. For simplicity, it may be adequate to position marginal notes before the relevant paragraph or other element. Notes may also be placed in a separate division of the text (as end-notes are, in printed books) and linked to the relevant portion of the text using their *target* attribute.

The *n* attribute may be used to supply the number or identifier of a note if this is required. The *resp* attribute should be used consistently to distinguish between authorial and editorial notes, if the work has both kinds.

Examples:

```
<p>Collections are ensembles of
distinct entities or objects of any sort. <note n="1" place="foot"> We explain below
why we
   use the uncommon term <mentioned>collection</mentioned> instead of the expected
 <mentioned>set</mentioned>.            Our    usage    corresponds    to    the
<mentioned>aggregate</mentioned>
   of many mathematical writings and to the sense of <mentioned>class</mentioned>
found in
   older logical writings. </note> The elements ...</p>
```

```
<lg xml:id="RAM609">
 <note place="margin">The
```

[2]The relevant standard is *Best Current Practice 47* (http://tools.ietf.org/html/bcp47). The authoritative list of registered subtags is maintained by IANA and is available at http://www.iana.org/assignments/language-subtag-registry. For a general overview of the construction of language tags, see http://www.w3.org/International/articles/language-tags/, and for a practical step-by-step guide, see http://www.w3.org/International/questions/qa-choosing-language-tags.

```
   curse is finally expiated</note>
 <l>And now this spell was snapt: once more</l>
 <l>I viewed
   the ocean green,</l>
 <l>And looked far forth, yet little saw</l>
 <l>Of what had else been seen
   —</l>
</lg>
```

8 Cross References and Links

Explicit cross references or links from one point in a text to another in the same or another document may be encoded using the elements described in this section. Implicit links (such as the association between two parallel texts, or that between a text and its interpretation) may be encoded using the linking attributes discussed in section xatts.

8.1 Simple Cross References

A cross reference from one point within a single document to another can be encoded using either of the following elements:

UNKNOWN ELEMENT ref

UNKNOWN ELEMENT ptr

The difference between these two elements is that `<ptr>` is an empty element, simply marking a point from which a link is to be made, whereas `<ref>` may contain some text as well, typically identifying the target of the cross reference. The `<ptr>` element would be used for a cross reference which is to be indicated by some non-verbal means such as a symbol or icon, or in an electronic text by a button. It is also useful in document production systems, where the formatter can generate the correct verbal form of the cross reference.

The following two forms, for example, are logically equivalent :

```
See especially <ref target="#SEC12">section 12 on
page 34</ref>.
```

```
See especially <ptr target="#SEC12"/>.
```

The value of the *target* attribute on either element may be the identifier of some other element within the current document. The passage or phrase being pointed at must bear an identifier, and must therefore be tagged as an element of some kind. In the following example, the cross reference is to a element:

```
 ... see especially <ptr target="#SEC12"/>. ...
<div xml:id="SEC12">
 <head>Concerning Identifiers</head>
<!-- ... -->
</div>
```

Because the *xml:id* attribute is global, any element in a TEI document may be pointed to in this way. In the following example, a paragraph has been given an identifier so that it may be pointed at:

```
 ... this is
discussed in <ref target="#pspec">the paragraph on links</ref> ...
<p xml:id="pspec">Links
may be made to any kind of element ...</p>
```

Sometimes the target of a cross reference does not correspond with any particular feature of a text, and so may not be tagged as an element of some kind. If the desired target is simply a point in the current document, the easiest way to mark it is by introducing an <anchor> element at the appropriate spot. If the target is some sequence of words not otherwise tagged, the <seg> element may be introduced to mark them. These two elements are described as follows:

UNKNOWN ELEMENT anchor

UNKNOWN ELEMENT seg

In the following (imaginary) example, <ref> elements have been used to represent points in this text which are to be linked in some way to other parts of it; in the first case to a point, and in the second, to a sequence of words:

```
 Returning to <ref target="#ABCD">the point where I
dozed off</ref>, I noticed that <ref target="#EFGH">three words</ref> had been circled
in
red by a previous reader
```

This encoding requires that elements with the specified identifiers (ABCD and EFGH in this example) are to be found somewhere else in the current document. Assuming that no element already exists to carry these identifiers, the <anchor> and <seg> elements may be used:

```
 ....
<anchor type="bookmark" xml:id="ABCD"/> ....  ....<seg type="target" xml:id="EFGH">
...
</seg> ...
```

The *type* attribute should be used (as above) to distinguish amongst different purposes for which these general purpose elements might be used in a text. Some other uses are discussed in section xatts below.

8.2 Pointing to other documents

So far, we have shown how the elements <ptr> and <ref> may be used for cross-references or links whose targets occur within the same document as their source. However, the same elements may also be used to refer to elements in any other XML document or resource, such as a document on the web, or a database component. This is possible because the value of the *target* attribute may be any valid *universal resource indicator* (URI) *[Note: A full definition of this term, defined by the W3C (the consortium which manages the development and maintenance of the World Wide Web), is beyond the scope of this tutorial: however, the most frequently encountered version of a URI is the familiar 'URL' used to indicate a web page, such as* `http://www.tei-c.org/index.xml`*]* .

A URI may reference a web page or just a part of one, for example `http://www.tei-c.org/index.xml#SEC2`. The sharp sign indicates that what follows it is the identifier of an element to be located within the XML document identified by what precedes it: this example will therefore locate an element which has an *xml:id* attribute value of SEC2 within the document retrieved from `http://www.tei-c.org/index.xml`. In the examples we have discussed so far, the part to the left of the sharp sign has been omitted: this is understood to mean that the referenced element is to be located within the current document.

Parts of an XML document can be specified by means of other more sophisticated mechanisms using a special language called Xpath, also defined by the W3C. This is particularly useful where the elements to be linked to do not bear identifiers and must therefore be located by some other means.

8.3 Special kinds of Linking

The following special purpose *linking* attributes are defined for every element in the TEI Lite scheme:

ana links an element with its interpretation.

corresp links an element with one or more other corresponding elements.

next links an element to the next element in an aggregate.

prev links an element to the previous element in an aggregate.

The *ana* (analysis) attribute is intended for use where a set of abstract analyses or interpretations have been defined somewhere within a document, as further discussed in section U5-anal. For example, a linguistic analysis of the sentence 'John loves Nancy' might be encoded as follows:

```
<seg ana="SVO" type="sentence">
 <seg ana="#NP1" type="lex">John</seg>
 <seg ana="#VVI" type="lex">loves</seg>
 <seg ana="#NP1" type="lex">Nancy</seg>
</seg>
```

This encoding implies the existence elsewhere in the document of elements with identifiers SVO, NP1, and VV1 where the significance of these particular codes is explained. Note the use of the <seg> element to mark particular components of the analysis, distinguished by the *type* attribute.

The *corresp* (corresponding) attribute provides a simple way of representing some form of correspondence between two elements in a text. For example, in a multilingual text, it may be used to link translation equivalents, as in the following example

```
<seg corresp="#EN1" xml:id="FR1"
 xml:lang="fr">Jean
aime Nancy</seg>
<seg corresp="#FR1" xml:id="EN1"
 xml:lang="en">John loves
Nancy</seg>
```

The same mechanism may be used for a variety of purposes. In the following example, it has been used to represent the correspondences between 'the show' and 'Shirley', and between 'NBC' and 'the network':

```
<p>
 <title xml:id="shirley">Shirley</title>, which
made its Friday night debut only a month ago, was not
listed on <name xml:id="nbc">NBC</name>'s new schedule, although
<seg corresp="#nbc" xml:id="network">the network</seg>
says <seg corresp="#shirley" xml:id="show">the show</seg> still is being
considered.
</p>
```

The *next* and *prev* attributes provide a simple way of linking together the components of a discontinuous element, as in the following example:

```
<q next="#Q1b" xml:id="Q1a">Who-e debel you?</q> —
he at last said — <q prev="#Q1a" xml:id="Q1b">you no speak-e, damme, I kill-e.</q>
And so
saying, the lighted tomahawk began flourishing about me in the dark.
```

9 Editorial Interventions

The process of encoding an electronic text has much in common with the process of editing a manuscript or other text for printed publication. In either case a conscientious editor may wish to record both the original state of the source and any editorial correction or other change made in it. The elements discussed in this and the next section provide some facilities for meeting these needs.

9.1 Correction and Normalization

The following elements may be used to mark *correction*, that is editorial changes introduced where the editor believes the original to be erroneous:

UNKNOWN ELEMENT corr

UNKNOWN ELEMENT sic

The following elements may be used to mark *normalization*, that is editorial changes introduced for the sake of consistency or modernization of a text:

UNKNOWN ELEMENT orig

UNKNOWN ELEMENT reg

As an example, consider this extract from the quarto printing of Shakespeare's *Henry V*.

```
  ... for his nose was as sharp as a pen and a table of green feelds
```

A modern editor might wish to make a number of interventions here, specifically to modernize (or normalise) the Elizabethan spellings of *a'* and *feelds* for *he* and *fields* respectively. He or she might also want to emend *table* to *babbl'd*, following an editorial tradition that goes back to the 18th century Shakespearian scholar Lewis Theobald. The following encoding would then be appropriate:

```
... for his nose was as sharp as
a pen and <reg>he</reg>
<corr resp="#Theobald">babbl'd</corr> of green
<reg>fields</reg>
```

A more conservative or source-oriented editor, however, might want to retain the original, but at the same time signal that some of the readings it contains are in some sense anomalous:

```
... for his nose was as sharp as a pen and
<orig>a</orig>
<sic>table</sic> of green
<orig>feelds</orig>
```

Finally, a modern digital editor may decide to combine both possibilities in a single composite text, using the <choice> element.

UNKNOWN ELEMENT choice

This allows an editor to mark where alternative readings are possible:

```
... for his nose was as sharp as a pen and
<choice>
 <orig>a</orig>
 <reg>he</reg>
</choice>
<choice>
 <corr resp="#Theobald">babbl'd</corr>
 <sic>table</sic>
</choice> of green
```

```
<choice>
 <orig>feelds</orig>
 <reg>fields</reg>
</choice>
```

9.2 Omissions, Deletions, and Additions

In addition to correcting or normalizing words and phrases, editors and transcribers may also supply missing material, omit material, or transcribe material deleted or crossed out in the source. In addition, some material may be particularly hard to transcribe because it is hard to make out on the page. The following elements may be used to record such phenomena:

UNKNOWN ELEMENT add

<gap> indicates a point where material has been omitted in a transcription, whether for editorial reasons described in the TEI header, as part of sampling practice, or because the material is illegible, invisible, or inaudible.

UNKNOWN ELEMENT del

UNKNOWN ELEMENT unclear

These elements may be used to record changes made by an editor, by the transcriber, or (in manuscript material) by the author or scribe. For example, if the source for an electronic text read 'The following elements are provided for for simple editorial interventions.' then it might be felt desirable to correct the obvious error, but at the same time to record the deletion of the superfluous second *for*, thus:

```
The following elements are provided for <del resp="#LB">for</del> simple editorial
interventions.
```

The attribute value #LB on the *resp* attribute is used to point to a fuller definition (typically in a <respStmt> element) for the agency responsible for correcting the duplication of *for*.

If the source read 'The following elements provided for simple editorial interventions.' (i.e. if the verb had been inadvertently dropped) then the corrected text might read:

```
The following elements <add resp="#LB">are</add> provided for simple editorial
interventions.
```

These elements are also used to record authorial changes in manuscripts. A manuscript in which the author has first written 'How it galls me, what a galling shadow', then crossed out the word *galls* and inserted *dogs* might be encoded thus:

```
How it <del hand="#DHL" type="overstrike">galls</del>
<add hand="#DHL" place="supralinear">dogs</add> me, what a galling shadow
```

Again, the code #DHL points to another location where more information about the hand concerned is to be found[3].

Similarly, the <unclear> and <gap> elements may be used together to indicate the omission of illegible material; the following example also shows the use of <add> for a conjectural emendation:

```
One hundred
& twenty good regulars joined to me <unclear>
 <gap reason="indecipherable"/>
</unclear>
& instantly, would aid me signally <add hand="#ed">in?</add> an enterprise against
Wilmington.
```

[3]The full TEI provides a range of elements for encoding metadata about manuscript production and description, which are not however included in TEI Lite

The <del> element marks material which has been transcribed as part of the electronic text despite being marked as deleted, while <gap> marks the location of material which is omitted from the electronic text, whether it is legible or not. A language corpus, for example, might omit long quotations in foreign languages:

```
<p> ... An example of a list appearing in a fief
ledger of <name type="place">Koldinghus</name>
 <date>1611/12</date> is given below. It shows cash income from a sale of
honey.</p>
<gap>
 <desc>quotation from ledger (in Danish)</desc>
</gap>
<p>A description of the
overall structure of the account is once again ... </p>
```

Other corpora (particular those constructed before the widespread use of scanners) systematically omit figures and mathematics:

```
<p>At the bottom of your screen below the mode line is the <term>minibuffer</term>.
This is
the area where Emacs echoes the commands you enter and where you specify filenames
for Emacs
to find, values for search and replace, and so on. <gap reason="graphic">
  <desc>diagram of
     Emacs screen</desc>
 </gap>
</p>
```

The full TEI scheme provides more precise ways of capturing different aspects of a transcription, distinguishing for example between text added or supplied by the encoder and text indicated as supplied or deleted in the source. TEI Lite does not provide different tags for these purposes.

9.3 Abbreviations and their Expansion

Like names, dates, and numbers, abbreviations may be transcribed as they stand or expanded; they may be left unmarked, or encoded using the following elements:

UNKNOWN ELEMENT abbr

UNKNOWN ELEMENT expan

The <abbr> element is useful as a means of distinguishing semi-lexical items such as acronyms or jargon:

```
We can sum up the above
discussion as follows: the identity of a <abbr>CC</abbr> is defined by that calibration
of
values which motivates the elements of its <abbr>GSP</abbr>;
```

```
Every manufacturer of <abbr>3GL</abbr> or
<abbr>4GL</abbr> languages is currently nailing on <abbr>OOP</abbr> extensions
```

The *type* attribute may be used to distinguish types of abbreviation by their function.

The <expan> element is used to mark an expansion supplied by an encoder. This element is particularly useful in the transcription of manuscript materials. For example, the character p with a bar through its descender as a conventional representation for the word per is commonly encountered in Medieval European manuscripts. An encoder may choose to expand this as follows:

```
<expan>per</expan>
```

The expansion corresponding with an abbreviated form may not always contain the same letters as the abbreviation. Where it does, however, common editorial practice is to italicize or otherwise signal which letters have been supplied. The <expan> element should not be used for this purpose since its function is to indicate an expanded form, not a part of one. For example, consider the common abbreviation wt (for with) found in medieval texts. In a modern edition, an editor might wish to represent this as 'w*i*t*h*', italicising the letters not found in the source. One simple means of achieving that would be an encoding such as the follow

```
<expan>w<hi rend="it">i</hi>t<hi rend="it">h</hi>
</expan>
```

The full TEI also provides elements <ex> and <am> for use in this situation, but these are not included in the TEI Lite schema.

To record both an abbreviation and its expansion, the <choice> element mentioned above may be used to group the abbreviated form with its proposed expansion:

```
<choice>
 <abbr>wt</abbr>
 <expan>with</expan>
</choice>
```

10 Names, Dates, and Numbers

The TEI scheme defines elements for a large number of 'data-like' features which may appear almost anywhere within almost any kind of text. These features may be of particular interest in a range of disciplines; they all relate to objects external to the text itself, such as the names of persons and places, numbers and dates. They also pose particular problems for many natural language processing (NLP) applications because of the variety of ways in which they may be presented within a text. The elements described here, by making such features explicit, reduce the complexity of processing texts containing them.

10.1 Names and Referring Strings

A *referring string* is a phrase which refers to some person, place, object, etc. Two elements are provided to mark such strings:

UNKNOWN ELEMENT rs

<**name**> (name, proper noun) contains a proper noun or noun phrase.

The *type* attribute is used to distinguish amongst (for example) names of persons, places and organizations, where this is possible:

```
<q>My dear <rs type="person">Mr. Bennet</rs>, </q>
said his lady to him one day,
<q>have you heard that <rs type="place">Netherfield Park</rs>
is let at last?</q>
```

```
It being one of the principles of the <rs type="organization">Circumlocution
Office</rs> never, on any account whatsoever, to give a
straightforward answer, <rs type="person">Mr Barnacle</rs> said,
<q>Possibly.</q>
```

As the following example shows, the <rs> element may be used for any reference to a person, place, etc, not necessarily one in the form of a proper noun or noun phrase.

```
<q>My dear <rs type="person">Mr. Bennet</rs>,</q>
said <rs type="person">his lady</rs> to him one day...
```

The <name> element by contrast is provided for the special case of referencing strings which consist only of proper nouns; it may be used synonymously with the <rs> element, or nested within it if a referring string contains a mixture of common and proper nouns.

Simply tagging something as a name is rarely enough to enable automatic processing of personal names into the canonical forms usually required for reference purposes. The name as it appears in the text may be inconsistently spelled, partial, or vague. Moreover, name prefixes such as *van* or *de la*, may or may not be included as part of the reference form of a name, depending on the language and country of origin of the bearer.

The *key* attribute provides an alternative normalized identifier for the object being named, like a database record key. It may thus be useful as a means of gathering together all references to the same individual or location scattered throughout a document:

```
<q>My dear <rs key="BENM1" type="person">Mr.
   Bennet</rs>, </q> said <rs key="BENM2" type="person">his lady</rs> to him one day,

<q>have
you heard that <rs key="NETP1" type="place">Netherfield Park</rs> is let at
last?</q>
```

This use should be distinguished from the case of the <reg> (regularization) element, which provides a means of marking the standard form of a referencing string as demonstrated below:

```
<name key="WADLM1" type="person">
 <choice>
  <sic>Walter de la Mare</sic>
  <reg>de la Mare, Walter</reg>
 </choice>
</name> was
born          at          <name key="Ch1" type="place">Charlton</name>,            in
<name key="KT1" type="county">Kent</name>, in 1873.
```

The <index> element discussed in indexing may be more appropriate if the function of the regularization is to provide a consistent index:

```
<p>
 <name type="place">Montaillou</name> is not a
large    parish.          At    the    time    of    the    events    which    led    to
<name type="person">Fournier</name>'s
<index>
  <term>Benedict XII, Pope of Avignon (Jacques Fournier)</term>
 </index>
investigations, the local population consisted of between 200 and 250
inhabitants.
</p>
```

Although adequate for many simple applications, these methods have two inconveniences: if the name occurs many times, then its regularised form must be repeated many times; and the burden of additional XML markup in the body of the text may be inconvenient to maintain and complex to process. For applications such as onomastics, relating to persons or places named rather than the name itself, or wherever a detailed analysis of the component parts of a name is needed, the full TEI Guidelines provide a range of other solutions.

10.2 Dates and Times

Tags for the more detailed encoding of times and dates include the following:

<date> contains a date in any format.

UNKNOWN ELEMENT time

These elements have a number of attributes which can be used to provide normalised versions of their values.

att.datable provides attributes for normalization of elements that contain dates, times, or datable events.

@calendar indicates the system or calendar to which the date represented by the content of this element belongs.

@period supplies a pointer to some location defining a named period of time within which the datable item is understood to have occurred.

@when [att.datable.w3c] supplies the value of the date or time in a standard form, e.g. yyyy-mm-dd.

The *when* attribute specifies a normalized form for the date or time, using one of the standard formats defined by ISO 8601. Partial dates or times (e.g. '1990', 'September 1990', 'twelvish') can be expressed by omitting a part of the value supplied, as in the following examples:

```
<date when="1980-02-21">21
Feb 1980</date>
<date when="1990">1990</date>
<date when="1990-09">September 1990</date>
<date when="--09">September</date>
<date when="2001-09-11T12:48:00">Sept 11th, 12 minutes before 9
am</date>
```

Note in the last example the use of a normalized representation for the date string which includes a time: this example could thus equally well be tagged using the `<time>` element.

```
Given on the <date when="1977-06-12">Twelfth
Day of June in the Year of Our Lord One Thousand Nine Hundred and Seventy-seven of
the
Republic the Two Hundredth and first and of the University the Eighty-Sixth.</date>
```

```
<l>specially when it's nine below zero</l>
<l>and <time when="15:00:00">three o'clock in the afternoon</time>
</l>
```

10.3 Numbers

Numbers can be written with either letters or digits (`twenty-one`, `xxi`, and `21`) and their presentation is language-dependent (e.g. English *5th* becomes Greek *5.*; English *123,456.78* equals French *123.456,78*). In natural-language processing or machine-translation applications, it is often helpful to distinguish them from other, more 'lexical' parts of the text. In other applications, the ability to record a number's value in standard notation is important. The `<num>` element provides this possibility:

UNKNOWN ELEMENT num

For example:

```
<num value="33">xxxiii</num>
<num type="cardinal" value="21">twenty-one</num>
<num type="percentage" value="10">ten percent</num>
<num type="percentage" value="10">10%</num>
<num type="ordinal" value="5">5th</num>
```

11 Lists

The element <list> is used to mark any kind of *list*. A list is a sequence of text items, which may be numbered, bulleted, or arranged as a glossary list. Each item may be preceded by an item label (in a glossary list, this label is the term being defined):

UNKNOWN ELEMENT list

UNKNOWN ELEMENT item

UNKNOWN ELEMENT label

Individual list items are tagged with <item>. The first <item> may optionally be preceded by a <head>, which gives a heading for the list. The numbering of a list may be omitted, indicated using the *n* attribute on each item, or (rarely) tagged as content using the <label> element. The following are all thus equivalent:

```
<list>
 <head>A short list</head>
 <item>First item in list.</item>
 <item>Second item in list.</item>
 <item>Third item in list.</item>
</list>
<list>
 <head>A short list</head>
 <item n="1">First item in list.</item>
 <item n="2">Second item in list.</item>
 <item n="3">Third item in list.</item>
</list>
<list>
 <head>A short list</head>
 <label>1</label>
 <item>First item in list.</item>
 <label>2</label>
 <item>Second item in list.</item>
 <label>3</label>
 <item>Third item in list.</item>
</list>
```

The styles should not be mixed in the same list.

A simple two-column table may be treated as a *glossary list*, tagged <list type="gloss">. Here, each item comprises a *term* and a *gloss*, marked with <label> and <item> respectively. These correspond to the elements <term> and <gloss>, which can occur anywhere in prose text.

```
<list type="gloss">
 <head>Vocabulary</head>
 <label xml:lang="enm">nu</label>
 <item>now</item>
 <label xml:lang="enm">lhude</label>
 <item>loudly</item>
 <label xml:lang="enm">bloweth</label>
 <item>blooms</item>
 <label xml:lang="enm">med</label>
 <item>meadow</item>
 <label xml:lang="enm">wude</label>
 <item>wood</item>
 <label xml:lang="enm">awe</label>
 <item>ewe</item>
 <label xml:lang="enm">lhouth</label>
 <item>lows</item>
 <label xml:lang="enm">sterteth</label>
 <item>bounds, frisks</item>
 <label xml:lang="enm">verteth</label>
```

```
 <item xml:lang="la">pedit</item>
 <label xml:lang="enm">murie</label>
 <item>merrily</item>
 <label xml:lang="enm">swik</label>
 <item>cease</item>
 <label xml:lang="enm">naver</label>
 <item>never</item>
</list>
```

Where the internal structure of a list item is more complex, it may be preferable to regard the list as a *table*, for which special-purpose tagging is defined below (U5-tables).

Lists of whatever kind can, of course, nest within list items to any depth required. Here, for example, a glossary list contains two items, each of which is itself a simple list:

```
<list type="gloss">
 <label>EVIL</label>
 <item>
  <list type="simple">
   <item>I am cast upon a horrible desolate island, void of all hope of recov-
ery.</item>
   <item>I am singled out and separated as it were from all the world to be
miserable.</item>
   <item>I am divided from mankind — a solitaire; one banished from human soci-
ety.</item>
  </list>
 </item>
 <label>GOOD</label>
 <item>
  <list type="simple">
   <item>But I am alive; and not drowned, as all my ship's company were.</item>
   <item>But I am singled out, too, from all the ship's crew, to be spared from
       death...</item>
   <item>But I am not starved, and perishing on a barren place, affording no
       sustenances....</item>
  </list>
 </item>
</list>
```

A list need not necessarily be displayed in list format. For example,

```
<p>On those remote pages it is written that animals
are divided into <list rend="run-on">
  <item n="a">those that belong to the Emperor,</item>
  <item n="b"> embalmed ones, </item>
  <item n="c"> those that are trained, </item>
  <item n="d"> suckling pigs, </item>
  <item n="e"> mermaids, </item>
  <item n="f"> fabulous ones, </item>
  <item n="g"> stray dogs, </item>
  <item n="h"> those that are included in this classification, </item>
  <item n="i"> those that tremble as if they were mad, </item>
  <item n="j"> innumerable ones, </item>
  <item n="k"> those drawn with a very fine camel's-hair brush, </item>
  <item n="l"> others, </item>
  <item n="m"> those that have just broken a flower vase, </item>
  <item n="n"> those that resemble flies from a distance.</item>
 </list>
</p>
```

Lists of bibliographic items should be tagged using the **<listBibl>** element, described in the next section.

12 Bibliographic Citations

It is often useful to distinguish bibliographic citations where they occur within texts being transcribed for research, if only so that they will be properly formatted when the text is printed out. The element **<bibl>** is provided for this purpose. Where the components of a bibliographic reference are to be distinguished, the following elements may be used as appropriate. It is generally useful to mark at least those parts (such as the titles of articles, books, and journals) which will need special formatting. The other elements are provided for cases where particular interest attaches to such details.

UNKNOWN ELEMENT bibl

<author> in a bibliographic reference, contains the name(s) of an author, personal or corporate, of a work; for example in the same form as that provided by a recognized bibliographic name authority.

UNKNOWN ELEMENT biblScope

<date> contains a date in any format.

UNKNOWN ELEMENT editor

<publisher> provides the name of the organization responsible for the publication or distribution of a bibliographic item.

UNKNOWN ELEMENT pubPlace

<title> contains a title for any kind of work.

For example, the following editorial note might be transcribed as shown:

> He was a member of Parliament for Warwickshire in 1445, and died March 14, 1470 (according to Kittredge, *Harvard Studies* 5. 88ff).

```
He was a member of Parliament for Warwickshire
in 1445, and died March 14, 1470 (according to <bibl>
 <author>Kittredge</author>,
<title>Harvard Studies</title>
 <biblScope>5. 88ff</biblScope>
</bibl>).
```

For lists of bibliographic citations, the **<listBibl>** element should be used; it may contain a series of **<bibl>** elements.

13 Tables

Tables represent a challenge for any text processing system, but simple tables, at least, appear in so many texts that even in the simplified TEI tag set presented here, markup for tables is necessary. The following elements are provided for this purpose:

UNKNOWN ELEMENT table

UNKNOWN ELEMENT row

UNKNOWN ELEMENT cell

For example, Defoe uses mortality tables like the following in the *Journal of the Plague Year* to show the rise and ebb of the epidemic:

```
<p>It was indeed coming on amain, for the burials
that same week were in the next adjoining parishes thus:— <table cols="4" rows="5">
  <row role="data">
   <cell role="label">St. Leonard's, Shoreditch</cell>
   <cell>64</cell>
   <cell>84</cell>
   <cell>119</cell>
  </row>
  <row role="data">
```

```
    <cell role="label">St. Botolph's, Bishopsgate</cell>
    <cell>65</cell>
    <cell>105</cell>
    <cell>116</cell>
   </row>
   <row role="data">
    <cell role="label">St. Giles's, Cripplegate</cell>
    <cell>213</cell>
    <cell>421</cell>
    <cell>554</cell>
   </row>
  </table>
</p>
<p>This shutting up of houses was at first counted a very cruel and unchristian
method, and the poor people so confined made bitter lamentations. ... </p>
```

14 Figures and Graphics

Not all the components of a document are necessarily textual. The most straightforward text will often contain diagrams or illustrations, to say nothing of documents in which image and text are inextricably intertwined, or electronic resources in which the two are complementary.

The encoder may simply record the presence of a graphic within the text, possibly with a brief description of its content, and may also provide a link to a digitized version of the graphic, using the following elements:

UNKNOWN ELEMENT graphic

UNKNOWN ELEMENT figure

UNKNOWN ELEMENT figDesc

Any textual information accompanying the graphic, such as a heading and/or caption, may be included within the `<figure>` element itself, in a <head> and one or more <p> elements, as also may any text appearing within the graphic itself. It is strongly recommended that a prose description of the image be supplied, as the content of a `<figDesc>` element, for the use of applications which are not able to render the graphic, and to render the document accessible to vision-impaired readers. (Such text is not normally considered part of the document proper.)

The simplest use for these elements is to mark the position of a graphic and provide a link to it, as in this example;

```
<pb n="412"/>
<figure>
 <graphic url="p412fig.png"/>
</figure>
<pb n="413"/>
```

This indicates that the graphic contained by the file p412fig.png appears between pages 412 and 413.

The `<graphic>` element can appear anywhere that textual content is permitted, within but not between paragraphs or headings. In the following example, the encoder has decided to treat a specific printer's ornament as a heading:

```
<head>
 <graphic url="http://www.iath.virginia.edu/gants/Ornaments/Heads/hp-ral02.gif"/>
</head>
```

More usually, a graphic will have at the least an identifying title, which may be encoded using the <head> element, or a number of figures may be grouped together in a particular structure. It is also often convenient to include a brief description of the image. The `<figure>` element provides a means of wrapping one or more such elements together as a kind of graphic 'block':

```
<figure>
 <graphic url="fessipic.png"/>
 <head>Mr Fezziwig's Ball</head>
 <figDesc>A Cruikshank
   engraving showing Mr Fezziwig leading a group of revellers.</figDesc>
</figure>
```

These cases should be carefully distinguished from the case where an encoded text is complemented by a collection of digital images, maintained as a distinct resource. The *facs* attribute may be used to associate any element in an encoded text with a digital facsimile of it. In the simple case where only page images are available, the *facs* attribute on the <pb> element may be used to associate each image with an appropriate point in the text:

```
<text>
 <pb facs="page1.png" n="1"/>
<!-- text contained on page 1 is encoded here -->
 <pb facs="page2.png" n="2"/>
<!-- text contained on page 2 is encoded here -->
</text>
```

This method is only appropriate in the simple case where each digital image file page1.png etc. corresponds with a single transcribed and encoded page. If more detailed alignment of image and transcription is required, for example because the image files actually represent double page spreads, more sophisticated mechanisms are provided in the full TEI Guidelines.

15 Interpretation and Analysis

It is often said that *all* markup is a form of interpretation or analysis. While it is certainly difficult, and may be impossible, to distinguish firmly between 'objective' and 'subjective' information in any universal way, it remains true that judgments concerning the latter are typically regarded as more likely to provide controversy than those concerning the former. Many scholars therefore prefer to record such interpretations only if it is possible to alert the reader that they are considered more open to dispute, than the rest of the markup. This section describes some of the elements provided by the TEI scheme to meet this need.

15.1 Orthographic Sentences

Interpretation typically ranges across the whole of a text, with no particular respect to other structural units. A useful preliminary to intensive interpretation is therefore to segment the text into discrete and identifiable units, each of which can then bear a label for use as a sort of 'canonical reference'. To facilitate such uses, these units may not cross each other, nor nest within each other. They may conveniently be represented using the following element:
<s> (s-unit) contains a sentence-like division of a text.

As the name suggests, the <s> element is most commonly used (in linguistic applications at least) for marking *orthographic sentences*, that is, units defined by orthographic features such as punctuation. For example, the passage from *Jane Eyre* discussed earlier might be divided into s-units as follows:

```
<pb n="474"/>
<div n="38" type="chapter">
 <p>
  <s n="001">Reader, I married him.</s>
  <s n="002">A quiet wedding we had:</s>
  <s n="003">he
     and I, the parson and clerk, were alone present.</s>
  <s n="004">When we got back from
     church, I went into the kitchen of the manor-house, where Mary was cooking the
```

```
dinner, and
      John cleaning the knives, and I said —</s>
 </p>
 <p>
  <q>
   <s n="005">Mary, I have been married to Mr Rochester this morning.</s>
  </q> ... </p>
</div>
```

Note that <s> elements cannot nest: the beginning of one <s> element implies that the previous one has finished. When s-units are tagged as shown above, it is advisable to tag the entire text end-to-end, so that every word in the text being analysed will be contained by exactly one <s> element, whose identifier can then be used to specify a unique reference for it. If the identifiers used are unique within the document, then the *xml:id* attribute might be used in preference to the *n* used in the above example.

15.2 Words and punctuation

Tokenization, that is, the identification of lexical or non-lexical tokens within a text, is a very common requirement for all kinds of textual analysis, and not an entirely trivial one. The decision as to whether, for example, 'can't' in English or 'du' in French should be treated as one word or two is not simple. Consequently it is often useful to make explicit the preferred tokenization in a marked up text. The following elements are available for this purpose:
<w> (word) represents a grammatical (not necessarily orthographic) word.

UNKNOWN ELEMENT pc

For example, the output from a part of speech tagger might be recorded in TEI Lite as follows:

```
<s n="1">
 <w ana="#NP0">Marley</w>
 <w ana="#VBD">was</w>
 <w ana="#AJ0">dead</w>
 <pc>:</pc>
 <w ana="#TO0">to</w>
 <w ana="#VBB">begin</w>
 <w ana="#PRP">with</w>
 <pc>. </pc>
</s>
```

In this example, each word has been decorated with an automatically generated part of speech code, using the *ana* attribute discussed in section xatts above. The <w> also provides for each word to be associated with a root form or lemma, either explicitly using the *lemma* attribute, or by reference, using the *lemmaRef* attribute, as in this example:

```
...<w ana="#VBD" lemma="be"
 lemmaRef="http://www.myLexicon.com/be">was</w> ...
```

15.3 General-Purpose Interpretation Elements

The <w> element is a specialisation of the **<seg>** element which has already been introduced for use in identifying otherwise unmarked targets of cross references and hypertext links (see section U5-ptrs); it identifies some phrase-level portion of text to which the encoder may assign a user-specified *type*, as well as a unique identifier; it may thus be used to tag textual features for which there is no other provision in the published TEI Guidelines.

For example, the Guidelines provide no 'apostrophe' element to mark parts of a literary text in which the narrator addresses the reader (or hearer) directly. One approach might be to regard these as instances of the <q> element, distinguished from others by an appropriate value for

the *who* attribute. A possibly simpler, and certainly more general, solution would however be to use the <seg> element as follows:

```
<div n="38" type="chapter">
 <p>
  <seg type="apostrophe">Reader, I married him.</seg> A quiet wedding we had: ...</p>
</div>
```

The *type* attribute on the <seg> element can take any value, and so can be used to record phrase-level phenomena of any kind; it is good practice to record the values used and their significance in the header.

A <seg> element of one type (unlike the <s> element which it superficially resembles) can be nested within a <seg> element of the same or another type. This enables quite complex structures to be represented; some examples were given in section xatts above. However, because it must respect the requirement that elements be properly nested and may not cut across each other, it cannot cope with the common requirement to associate an interpretation with arbitrary segments of a text which may completely ignore the document hierarchy. It also requires that the interpretation itself be represented by a single coded value in the *type* attribute.

Neither restriction applies to the <interp> element, which provides powerful features for the encoding of quite complex interpretive information in a relatively straightforward manner.

UNKNOWN ELEMENT interp

UNKNOWN ELEMENT interpGrp

These elements allow the encoder to specify both the class of an interpretation, and the particular instance of that class which the interpretation involves. Thus, whereas with <seg> one can say simply that something is an apostrophe, with <interp> one can say that it is an instance (apostrophe) of a larger class (rhetorical figures).

Moreover, <interp> is a 'stand off' element: it does not surround the segments of text which it describes, but instead is linked to the passage in question either by means of the *ana* attribute discussed in section xatts above, or by means of its own *inst* attribute. This means that any kind of analysis can be represented, independently of the document hierarchy, as well as facilitating the grouping of analyses of a particular type together. A special purpose <interpGrp> element is provided for the latter purpose.

For example, suppose that you wish to mark such diverse aspects of a text as themes or subject matter, rhetorical figures, and the locations of individual scenes of the narrative. Different portions of our sample passage from *Jane Eyre* for example, might be associated with the rhetorical figures of apostrophe, hyperbole, and metaphor; with subject-matter references to churches, servants, cooking, postal service, and honeymoons; and with scenes located in the church, in the kitchen, and in an unspecified location (drawing room?).

These interpretations could be placed anywhere within the <text> element; it is however good practice to put them all in the same place (e.g. a separate section of the front or back matter), as in the following example:

```
<back>
 <div type="Interpretations">
  <p>
   <interp resp="#LB-MSM"
    type="figureOfSpeech" xml:id="fig-apos-1">apostrophe</interp>
   <interp resp="#LB-MSM"
    type="figureOfSpeech" xml:id="fig-hyp-1">hyperbole</interp>
   <interp resp="#LB-MSM" type="setting"
    xml:id="set-church-1">church</interp>
   <interp resp="#LB-MSM" type="reference"
    xml:id="ref-church-1">church</interp>
   <interp resp="#LB-MSM" type="reference"
    xml:id="ref-serv-1">servants</interp>
```

```
  </p>
 </div>
</back>
```

The evident redundancy of this encoding can be considerably reduced by using the <interpGrp> element to group together all those <interp> elements which share common attribute values, as follows:

```
<back>
 <div type="Interpretations">
  <p>
   <interpGrp resp="#LB-MSM"
    type="figureOfSpeech">
    <interp xml:id="fig-apos">apostrophe</interp>
    <interp xml:id="fig-hyp">hyperbole</interp>
    <interp xml:id="fig-meta">metaphor</interp>
   </interpGrp>
   <interpGrp resp="#LB-MSM"
    type="scene-setting">
    <interp xml:id="set-church">church</interp>
    <interp xml:id="set-kitch">kitchen</interp>
    <interp xml:id="set-unspec">unspecified</interp>
   </interpGrp>
   <interpGrp resp="#LB-MSM"
    type="reference">
    <interp xml:id="ref-church">church</interp>
    <interp xml:id="ref-serv">servants</interp>
    <interp xml:id="ref-cook">cooking</interp>
   </interpGrp>
  </p>
 </div>
</back>
```

Once these interpretation elements have been defined, they can be linked with the parts of the text to which they apply in either or both of two ways. The *ana* attribute can be used on whichever element is appropriate:

```
<div n="38" type="chapter">
 <p ana="#set-church #set-kitch"
  xml:id="P38.1">
  <s ana="#fig-apos" xml:id="P38.1.1">Reader, I
     married him.</s>
 </p>
</div>
```

Note in this example that since the paragraph has two settings (in the church and in the kitchen), the identifiers of both have been supplied.

Alternatively, the <interp> elements can point to all the parts of the text to which they apply, using their *inst* attribute:

```
<interp inst="#P38.1.1" resp="#LB-MSM"
 type="figureOfSpeech" xml:id="fig-apos-2">apostrophe</interp>
<interp inst="#P38.1" resp="#LB-MSM"
 type="scene-setting" xml:id="set-church-2">church</interp>
<interp inst="#P38.1" resp="#LB-MSM"
 type="scene-setting" xml:id="set-kitchen-2">kitchen</interp>
```

The <interp> element is not limited to any particular type of analysis. The literary analysis shown above is but one possibility; one could equally well use <interp> to capture a linguistic

part-of-speech analysis. For example, the example sentence given in section xatts assumes a linguistic analysis which might be represented as follows:

```
<interp type="pos" xml:id="NP1">noun
phrase, singular</interp>
<interp type="pos" xml:id="VV1">inflected verb, present-tense
singular</interp> ...
```

16 Technical Documentation

Although the focus of this document is on the use of the TEI scheme for the encoding of existing 'pre-electronic' documents, the same scheme may also be used for the encoding of new documents. In the preparation of new documents (such as this one), XML has much to recommend it: the document's structure can be clearly represented, and the same electronic text can be re-used for many purposes — to provide both online hypertext or browsable versions and well-formatted typeset versions from a common source for example.

To facilitate this, the TEI Lite schema includes some elements for marking features of technical documents in general, and of XML-related documents in particular.

16.1 Additional Elements for Technical Documents

The following elements may be used to mark particular features of technical documents:

UNKNOWN ELEMENT eg

UNKNOWN ELEMENT code

UNKNOWN ELEMENT ident

UNKNOWN ELEMENT gi

UNKNOWN ELEMENT att

UNKNOWN ELEMENT formula

UNKNOWN ELEMENT val

The following example shows how these elements might be used to encode a passage from a tutorial introducing the Fortran programming language:

```
<p>It is traditional to introduce a language with a
program like the following: <eg xml:space="preserve"> CHAR*12 GRTG
GRTG = 'HELLO WORLD'
PRINT *, GRTG
END
        </eg>
</p>
<p>This simple example first declares a variable <ident>GRTG</ident>, in the line
<code>CHAR*12 GRTG</code>, which identifies <ident>GRTG</ident> as consisting of 12
bytes
of type <ident>CHAR</ident>.  To this variable, the value <val>HELLO WORLD</val> is
then
assigned.</p>
```

A formatting application, given a text like that above, can be instructed to format examples appropriately (e.g. to preserve line breaks, or to use a distinctive font). Similarly, the use of tags such as `<ident>` greatly facilitates the construction of a useful index.

The `<formula>` element should be used to enclose a mathematical or chemical formula presented within the text as a distinct item. Since formulae generally include a large variety of special typographic features not otherwise present in ordinary text, it will usually be necessary to present the body of the formula in a specialized notation. The notation used should be specified by the *notation* attribute, as in the following example:

```
<formula notation="tex"> \begin{math}E =
mc^{2}\end{math} </formula>
```

A particular problem arises when XML encoding is the subject of discussion within a technical document, itself encoded in XML. In such a document, it is clearly essential to distinguish clearly the markup occurring within examples from that marking up the document itself, and end-tags are highly likely to occur. One simple solution is to use the predefined entity reference `<` to represent each < character which marks the start of an XML tag within the examples. A more general solution is to mark off the whole body of each example as containing data which is not to be scanned for XML mark-up by the parser. This is achieved by enclosing it within a special XML construct called a *CDATA marked section*, as in the following example:

```
  <p>A list should be encoded as
follows: <eg><![ CDATA [ <list> <item>First item in the
list</item> <item>Second item</item> </list> ]]> </eg> The
<gi>list</gi> element consists of a series of <gi>item</gi>
elements.
```

The `<list>` element used within the example above will not be regarded as forming part of the document proper, because it is embedded within a marked section (beginning with the special markup declaration <![CDATA[, and ending with]]>).

Note also the use of the `<gi>` element to tag references to element names (or *generic identifiers*) within the body of the text.

16.2 Generated Divisions

Most modern document production systems have the ability to generate automatically whole sections such as a table of contents or an index. The TEI Lite scheme provides an element to mark the location at which such a generated section should be placed.

UNKNOWN ELEMENT divGen

The `<divGen>` element can be placed anywhere that a division element would be legal, as in the following example:

```
<front>
 <titlePage>
<!-- ... -->
 </titlePage>
 <divGen type="toc"/>
 <div>
  <head>Preface</head>
<!-- ... -->
 </div>
</front>
<body>
<!-- ... -->
</body>
<back>
 <div>
  <head>Appendix</head>
<!-- ... -->
 </div>
 <divGen n="Index" type="index"/>
</back>
```

This example also demonstrates the use of the *type* attribute to distinguish the different kinds of division to be generated: in the first case a table of contents (a *toc*) and in the second an index.

When an existing index or table of contents is to be encoded (rather than one being generated) for some reason, the **<list>** element discussed in section U5-lists should be used.

16.3 Index Generation

While production of a table of contents from a properly tagged document is generally unproblematic for an automatic processor, the production of a good quality index will often require more careful tagging. It may not be enough simply to produce a list of all parts tagged in some particular way, although extracting (for example) all occurrences of elements such as **<term>** or <name> will often be a good departure point for an index.

The TEI schema provides a special purpose **<index>** tag which may be used to mark both the parts of the document which should be indexed, and how the indexing should be done.

UNKNOWN ELEMENT index

For example, the second paragraph of this section might include the following:

```
... TEI lite also provides a special purpose
<gi>index</gi> tag
<index>
 <term>indexing</term>
</index>
<index>
 <term>index (tag)</term>
 <index>
  <term>use in index generation</term>
 </index>
</index>
which may be used ...
```

The **<index>** element can also be used to provide a form of interpretive or analytic information. For example, in a study of Ovid, it might be desired to record all the poet's references to different figures, for comparative stylistic study. In the following lines of the *Metamorphoses*, such a study would record the poet's references to Jupiter (as *deus*, *se*, and as the subject of *confiteor* [in inflectional form number 227]), to Jupiter-in-the-guise-of-a-bull (as *imago tauri fallacis* and the subject of *teneo*), and so on.[4]

```
<l n="3.001">iamque deus posita fallacis
imagine tauri</l>
<l n="3.002">se confessus erat Dictaeaque rura tenebat</l>
```

This need might be met using the <note> element discussed in section in U5-notes, or with the **<interp>** element discussed in section U5-anal. Here we demonstrate how it might also be satisfied by using the **<index>** element.

We assume that the object is to generate more than one index: one for names of deities (called *dn*), another for onomastic references (called *on*), a third for pronominal references (called *pr*) and so forth. One way of achieving this might be as follows:

```
<l n="3.001">iamque deus posita
fallacis imagine tauri <index indexName="dn">
  <term>Iuppiter</term>
  <index>
   <term>deus</term>
  </index>
 </index>
 <index indexName="on">
```

[4]The analysis is taken, with permission, from Willard McCarty and Burton Wright, *An Analytical Onomasticon to the Metamorphoses of Ovid* (Princeton: Princeton University Press, forthcoming). Some simplifications have been undertaken.

```
  <term>Iuppiter (taurus)</term>
  <index>
   <term>imago tauri
       fallacis</term>
  </index>
 </index>
</l>
<l n="3.002">se confessus erat Dictaeaque rura tenebat
<index indexName="pr">
  <term>Iuppiter</term>
  <index>
   <term>se</term>
  </index>
 </index>
 <index indexName="v">
  <term>Iuppiter</term>
  <index>
   <term>confiteor
       (v227)</term>
  </index>
 </index>
</l>
```

For each <index> element above, an entry will be generated in the appropriate index, using as headword the content of the <term> element it contains; the <term> elements nested within the secondary <index> element in each case provide a secondary keyword. The actual reference will be taken from the context in which the <index> element appears, i.e. in this case the identifier of the <l> element containing it.

16.4 Addresses

The <address> element is used to mark a postal address of any kind. It contains one or more <addrLine> elements, one for each line of the address.

UNKNOWN ELEMENT address

UNKNOWN ELEMENT addrLine

Here is a simple example:

```
<address>
 <addrLine>Computer Center (M/C 135)</addrLine>
 <addrLine>1940 W. Taylor, Room 124</addrLine>
 <addrLine>Chicago, IL 60612-7352</addrLine>
 <addrLine>U.S.A.</addrLine>
</address>
```

The individual parts of an address may be further distinguished by using the <name> element discussed above (section nomen).

```
<address>
 <addrLine>Computer Center (M/C 135)</addrLine>
 <addrLine>1940 W. Taylor, Room 124</addrLine>
 <addrLine>
  <name type="city">Chicago</name>, IL 60612-7352</addrLine>
 <addrLine>
  <name type="country">USA</name>
 </addrLine>
</address>
```

17 Character Sets, Diacritics, etc.

With the advent of XML and its adoption of Unicode as the required character set for all documents, most problems previously associated with the representation of the divers languages

and writing systems of the world are greatly reduced. For those working with standard forms of the European languages in particular, almost no special action is needed: any XML editor should enable you to input accented letters or other 'non-ASCII' characters directly, and they should be stored in the resulting file in a way which is transferable directly between different systems.

There are two important exceptions: the characters & and < may not be entered directly in an XML document, since they have a special significance as initiating markup. They must always be represented as *entity references*, like this: `&` or `<`. Other characters may also be represented by means of entity reference where necessary, for example to retain compatibility with a pre-Unicode processing system.

18 Front and Back Matter

18.1 Front Matter

For many purposes, particularly in older texts, the preliminary material such as title pages, prefatory epistles, etc., may provide very useful additional linguistic or social information. P5 provides a set of recommendations for distinguishing the textual elements most commonly encountered in front matter, which are summarized here.

18.1.1 Title Page

The start of a title page should be marked with the element `<titlePage>`. All text contained on the page should be transcribed and tagged with the appropriate element from the following list:

UNKNOWN ELEMENT titlePage
UNKNOWN ELEMENT docTitle
UNKNOWN ELEMENT titlePart
UNKNOWN ELEMENT byline
UNKNOWN ELEMENT docAuthor
UNKNOWN ELEMENT docDate
UNKNOWN ELEMENT docEdition
UNKNOWN ELEMENT docImprint
UNKNOWN ELEMENT epigraph

Typeface distinctions should be marked with the *rend* attribute when necessary, as described above. Very detailed description of the letter spacing and sizing used in ornamental titles is not as yet provided for by the Guidelines. Changes of language should be marked by appropriate use of the *xml:lang* attribute or the `<foreign>` element, as necessary. Names of people, places, or organizations, may be tagged using the <name> element wherever they appear if no other more specific element is available.

Two example title pages follow:

```
<titlePage rend="Roman">
 <docTitle>
  <titlePart type="main">  PARADISE  REGAIN'D.  A  POEM  In  IV  <hi>BOOKS</hi>.
</titlePart>
  <titlePart> To which is added <title>SAMSON AGONISTES</title>. </titlePart>
 </docTitle>
 <byline>The Author <docAuthor>JOHN MILTON</docAuthor>
 </byline>
 <docImprint>
  <name>LONDON</name>, Printed by <name>J.M.</name> for <name>John Starkey</name>
   at the <name>Mitre</name> in <name>Fleetstreet</name>, near
 <name>Temple-Bar.</name>
 </docImprint>
```

```
 <docDate>MDCLXXI</docDate>
</titlePage>
```

```
<titlePage>
 <docTitle>
  <titlePart type="main"> Lives of the Queens of England, from the Norman
     Conquest;</titlePart>
  <titlePart type="sub">with anecdotes of their courts. </titlePart>
 </docTitle>
 <titlePart>Now first published from Official Records and other authentic documents
private
   as well as public.</titlePart>
 <docEdition>New edition, with corrections and additions</docEdition>
 <byline>By <docAuthor>Agnes Strickland</docAuthor>
 </byline>
 <epigraph>
  <q>The treasures of antiquity laid up in old historic rolls, I opened.</q>
  <bibl>BEAUMONT</bibl>
 </epigraph>
 <docImprint>Philadelphia: Blanchard and Lea</docImprint>
 <docDate>1860.</docDate>
</titlePage>
```

As elsewhere, the *ref* attribute may be used to link a name with a canonical definition of the entity being named. For example:

```
<byline>By <docAuthor>
  <name ref="http://en.wikipedia.org/wiki/Agnes_Strickland">Agnes
     Strickland</name>
 </docAuthor>
</byline>
```

18.1.2 Prefatory Matter

Major blocks of text within the front matter should be marked using elements; the following suggested values for the *type* attribute may be used to distinguish various common types of prefatory matter:

preface A foreword or preface addressed to the reader in which the author or publisher explains the content, purpose, or origin of the text

dedication A formal offering or dedication of a text to one or more persons or institutions by the author.

abstract A summary of the content of a text as continuous prose

ack A formal declaration of acknowledgment by the author in which persons and institutions are thanked for their part in the creation of a text

contents A table of contents, specifying the structure of a work and listing its constituents. The `<list>` element should be used to mark its structure.

frontispiece A pictorial frontispiece, possibly including some text.

Where other kinds of prefatory matter are encountered, the encoder is at liberty to invent other values for the *type* attribute.

Like any text division, those in front matter may contain low level structural or non-structural elements as described elsewhere. They will generally begin with a heading or title of some kind which should be tagged using the <head> element. Epistles will contain the following additional elements:

UNKNOWN ELEMENT salute
UNKNOWN ELEMENT signed
UNKNOWN ELEMENT byline
UNKNOWN ELEMENT dateline
UNKNOWN ELEMENT argument
UNKNOWN ELEMENT cit
UNKNOWN ELEMENT imprimatur
UNKNOWN ELEMENT opener
UNKNOWN ELEMENT closer

Epistles which appear elsewhere in a text will, of course, contain these same elements.

As an example, the dedication at the start of Milton's *Comus* should be marked up as follows:

```
<div type="dedication">
 <head>To the Right Honourable <name>JOHN Lord Viscount BRACLY</name>, Son and Heir
apparent
   to the Earl of Bridgewater, &c.</head>
 <salute>MY LORD,</salute>
 <p>THis <hi>Poem</hi>, which receiv'd its first occasion of Birth from your Self,
and
   others of your Noble Family .... and as in this representation your attendant
 <name>Thyrsis</name>, so now in all reall expression</p>
 <closer>
  <salute>Your faithfull, and most humble servant</salute>
  <signed>
   <name>H. LAWES.</name>
  </signed>
 </closer>
</div>
```

18.2 Back Matter

18.2.1 Structural Divisions of Back Matter

Because of variations in publishing practice, back matter can contain virtually any of the elements listed above for front matter, and the same elements should be used where this is so. Additionally, back matter may contain the following types of matter within the <back> element. Like the structural divisions of the body, these should be marked as elements, and distinguished by the following suggested values of the *type* attribute:

appendix An ancillary self-contained section of a work, often providing additional but in some sense extra-canonical text.

glossary A list of terms associated with definition texts ('glosses'): this should be encoded as a <<list type="gloss">> element

notes A section in which textual or other kinds of notes are gathered together.

bibliogr A list of bibliographic citations: this should be encoded as a `<listBibl>`

index Any form of pre-existing index to the work (An index may also be generated for a document by using the `<index>` element described above).

colophon A statement appearing at the end of a book describing the conditions of its physical production.

19 The Electronic Title Page

Every TEI text has a header which provides information analogous to that provided by the title page of printed text. The header is introduced by the element <teiHeader> and has four major parts:

<fileDesc> (file description) contains a full bibliographic description of an electronic file.

UNKNOWN ELEMENT encodingDesc

UNKNOWN ELEMENT profileDesc

<revisionDesc> (revision description) summarizes the revision history for a file.

A corpus or collection of texts with many shared characteristics may have one header for the corpus and individual headers for each component of the corpus. In this case the *type* attribute indicates the type of header. `<teiHeader type="corpus">` introduces the header for corpus-level information.

Some of the header elements contain running prose which consists of one or more <p>s. Others are grouped:

- Elements whose names end in *Stmt* (for statement) usually enclose a group of elements recording some structured information.
- Elements whose names end in *Decl* (for declaration) enclose information about specific encoding practices.
- Elements whose names end in *Desc* (for description) contain a prose description.

19.1 The File Description

The <fileDesc> element is mandatory. It contains a full bibliographic description of the file with the following elements:

<titleStmt> (title statement) groups information about the title of a work and those responsible for its content.

UNKNOWN ELEMENT editionStmt

UNKNOWN ELEMENT extent

<publicationStmt> (publication statement) groups information concerning the publication or distribution of an electronic or other text.

UNKNOWN ELEMENT seriesStmt

UNKNOWN ELEMENT notesStmt

<sourceDesc> (source description) describes the source from which an electronic text was derived or generated, typically a bibliographic description in the case of a digitized text, or a phrase such as "born digital" for a text which has no previous existence.

A minimal header has the following structure:

```
<teiHeader>
 <fileDesc>
  <titleStmt>
<!-- bibliographic description of the digital resource -->
  </titleStmt>
  <publicationStmt>
<!-- information about how the resource is distributed -->
  </publicationStmt>
  <sourceDesc>
<!-- information about the sources from which the digital resource is derived -->
  </sourceDesc>
 </fileDesc>
</teiHeader>
```

19.1.1 The Title Statement

The following elements can be used in the <titleStmt>:

<title> contains a title for any kind of work.

<author> in a bibliographic reference, contains the name(s) of an author, personal or corporate, of a work; for example in the same form as that provided by a recognized bibliographic name authority.

UNKNOWN ELEMENT sponsor

UNKNOWN ELEMENT funder

<principal> (principal researcher) supplies the name of the principal researcher responsible for the creation of an electronic text.

UNKNOWN ELEMENT respStmt

The title of a digital resource derived from a non-digital one will obviously be similar. However, it is important to distinguish the title of the computer file from that of the source text, for example:

```
  [title of source]: a machine readable transcription [title of source]: electronic
edition A machine readable version of: [title of source]
```

The <respStmt> element contains the following subcomponents:

UNKNOWN ELEMENT resp

<name> (name, proper noun) contains a proper noun or noun phrase.

Example:

```
<titleStmt>
 <title>Two stories by Edgar Allen Poe: a machine readable transcription</title>
 <author>Poe, Edgar Allen (1809-1849)</author>
 <respStmt>
  <resp>compiled by</resp>
  <name>James D. Benson</name>
 </respStmt>
</titleStmt>
```

19.1.2 The Edition Statement

The <editionStmt> groups information relating to one edition of the digital resource (where *edition* is used as elsewhere in bibliography), and may include the following elements:

UNKNOWN ELEMENT edition

UNKNOWN ELEMENT respStmt

Example:

```
<editionStmt>
 <edition n="U2">Third
   draft, substantially revised <date>1987</date>
 </edition>
</editionStmt>
```

Determining exactly what constitutes a new edition of an electronic text is left to the encoder.

19.1.3 The Extent Statement

The <extent> statement describes the approximate size of the digital resource.

Example:

```
<extent>4532
bytes</extent>
```

19.1.4 The Publication Statement

The <publicationStmt> is mandatory. It may contain a simple prose description or groups of the elements described below:

<publisher> provides the name of the organization responsible for the publication or distribution of a bibliographic item.

UNKNOWN ELEMENT distributor

UNKNOWN ELEMENT authority

At least one of these three elements must be present, unless the entire publication statement is in prose. The following elements may occur within them:

UNKNOWN ELEMENT pubPlace

UNKNOWN ELEMENT address

<idno> (identifier) supplies any form of identifier used to identify some object, such as a bibliographic item, a person, a title, an organization, etc. in a standardized way.

<availability> supplies information about the availability of a text, for example any restrictions on its use or distribution, its copyright status, any licence applying to it, etc.

<licence> contains information about a licence or other legal agreement applicable to the text.

<date> contains a date in any format.

Example:

```
<publicationStmt>
 <publisher>University of Victoria Humanities Computing and Media Centre</publisher>
 <pubPlace>Victoria, BC</pubPlace>
 <date>2011</date>
 <availability status="restricted">
  <licence target="http://creativecommons.org/licenses/by-sa/3.0/"> Distributed un-
der a
     Creative Commons Attribution-ShareAlike 3.0 Unported License </licence>
 </availability>
</publicationStmt>
```

19.1.5 Series and Notes Statements

The **<seriesStmt>** element groups information about the series, if any, to which a publication belongs. It may contain <title>, <idno>, or **<respStmt>** elements.

The **<notesStmt>**, if used, contains one or more <note> elements which contain a note or annotation. Some information found in the notes area in conventional bibliography has been assigned specific elements in the TEI scheme.

19.1.6 The Source Description

The <sourceDesc> is a mandatory element which records details of the source or sources from which the computer file is derived. It may contain simple prose or a bibliographic citation, using one or more of the following elements:

UNKNOWN ELEMENT bibl

UNKNOWN ELEMENT listBibl

Examples:

```
<sourceDesc>
 <bibl>The first folio of Shakespeare, prepared by Charlton Hinman (The Norton
Facsimile,
   1968)</bibl>
</sourceDesc>
```

```
<sourceDesc>
 <bibl>
  <author>CNN Network News</author>
  <title>News headlines</title>
  <date>12 Jun
     1989</date>
 </bibl>
</sourceDesc>
```

19.2 The Encoding Description

The <encodingDesc> element specifies the methods and editorial principles which governed the transcription of the text. Its use is highly recommended. It may be prose description or may contain elements from the following list:

UNKNOWN ELEMENT projectDesc

UNKNOWN ELEMENT samplingDecl

UNKNOWN ELEMENT editorialDecl

UNKNOWN ELEMENT refsDecl

UNKNOWN ELEMENT classDecl

19.2.1 Project and Sampling Descriptions

Examples of <projectDesc> and <samplingDesc>:

```
<encodingDesc>
 <projectDesc>
  <p>Texts collected for
     use in the Claremont Shakespeare Clinic, June 1990.
  </p>
 </projectDesc>
</encodingDesc>
```

```
<encodingDesc>
 <samplingDecl>
  <p>Samples of
     2000 words taken from the beginning of the text</p>
 </samplingDecl>
</encodingDesc>
```

19.2.2 Editorial Declarations

The <editorialDecl> contains a prose description of the practices used when encoding the text. Typically this description should cover such topics as the following, each of which may conveniently be given as a separate paragraph.

correction how and under what circumstances corrections have been made in the text.

normalization the extent to which the original source has been regularized or normalized.

quotation what has been done with quotation marks in the original -- have they been retained or replaced by entity references, are opening and closing quotes distinguished, etc.

hyphenation what has been done with hyphens (especially end-of-line hyphens) in the original -- have they been retained, replaced by entity references, etc.

segmentation how has the text has been segmented, for example into sentences, tone-units, graphemic strata, etc.

interpretation what analytic or interpretive information has been added to the text.

Example:

```
<editorialDecl>
 <p>The part of
   speech analysis applied throughout section 4 was added by hand and has not been
   checked.</p>
 <p>Errors in transcription controlled by using the WordPerfect spelling
   checker.</p>
 <p>All words converted to Modern American spelling using Webster's 9th
   Collegiate dictionary.</p>
</editorialDecl>
```

19.2.3 Reference and Classification Declarations

The <refsDecl> element is used to document the way in which any standard referencing scheme built into the encoding works. In its simplest form, it consists of prose description.

Example:

```
<refsDecl>
 <p>The <att>n</att>
   attribute on each <gi>div</gi> contains the canonical reference for each division
in the
   form XX.yyy where XX is the book number in roman numeral and yyy is the section
number in
   arabic.</p>
 <p>Milestone tags refer to the edition of 1830 as E30 and that of 1850 as E50.
 </p>
</refsDecl>
```

The <classDecl> element groups together definitions or sources for any descriptive classification schemes used by other parts of the header. At least one such scheme must be provided, encoded using the following elements:

UNKNOWN ELEMENT taxonomy

UNKNOWN ELEMENT bibl

UNKNOWN ELEMENT category

UNKNOWN ELEMENT catDesc

In the simplest case, the taxonomy may be defined by a bibliographic reference, as in the following example:

```
<classDecl>
 <taxonomy xml:id="LC-SH">
  <bibl>Library of Congress Subject Headings
  </bibl>
 </taxonomy>
</classDecl>
```

Alternatively, or in addition, the encoder may define a special purpose classification scheme, as in the following example:

```
<taxonomy xml:id="B">
 <bibl>Brown Corpus</bibl>
 <category xml:id="B.A">
  <catDesc>Press
     Reportage</catDesc>
  <category xml:id="B.A1">
   <catDesc>Daily</catDesc>
  </category>
  <category xml:id="B.A2">
   <catDesc>Sunday</catDesc>
```

```
  </category>
  <category xml:id="B.A3">
   <catDesc>National</catDesc>
  </category>
  <category xml:id="B.A4">
   <catDesc>Provincial</catDesc>
  </category>
  <category xml:id="B.A5">
   <catDesc>Political</catDesc>
  </category>
  <category xml:id="B.A6">
   <catDesc>Sports</catDesc>
  </category>
 </category>
 <category xml:id="B.D">
  <catDesc>Religion</catDesc>
  <category xml:id="B.D1">
   <catDesc>Books</catDesc>
  </category>
  <category xml:id="B.D2">
   <catDesc>Periodicals and
       tracts</catDesc>
  </category>
 </category>
</taxonomy>
```

Linkage between a particular text and a category within such a taxonomy is made by means of the <catRef> element within the <textClass> element, as described in the next section below.

19.3 The Profile Description

The <profileDesc> element enables information characterizing various descriptive aspects of a text to be recorded within a single framework. It has three optional components:

UNKNOWN ELEMENT creation

UNKNOWN ELEMENT langUsage

UNKNOWN ELEMENT textClass

The <creation> element is useful for documenting where a work was created, even though it may not have been published or recorded there.

Example:

```
<creation>
 <date when="1992-08">August 1992</date>
 <name type="place">Taos, New Mexico</name>
</creation>
```

The <langUsage> element is useful where a text contains many different languages. It may contain <language> elements to document each particular language used:

UNKNOWN ELEMENT language

For example, a text containing predominantly text in French as spoken in Quebec, but also smaller amounts of British and Canadian English might be documented as follows:

```
<langUsage>
 <language ident="fr-CA" usage="60">Québecois</language>
 <language ident="en-CA" usage="20">Canadian business English</language>
 <language ident="en-GB" usage="20">British English</language>
</langUsage>
```

The <textClass> element classifies a text. This may be done with reference to a classification system locally defined by means of the <classDecl> element, or by reference to

some externally defined established scheme such as the Universal Decimal Classification. Texts may also be classified using lists of keywords, which may themselves be drawn from locally or externally defined control lists. The following elements are used to supply such classifications:

UNKNOWN ELEMENT classCode

UNKNOWN ELEMENT catRef

UNKNOWN ELEMENT keywords

The simplest way of classifying a text is by means of the `<classCode>` element. For example, a text with classification 410 in the Universal Decimal Classification might be documented as follows:

```
<classCode scheme="http://www.udc.org">410</classCode>
```

When a classification scheme has been locally defined using the `<taxonomy>` element discussed in the preceding subsection, the `<catRef>` element should be used to reference it. To continue the earlier example, a work classified in the Brown Corpus as `Press reportage - Sunday` and also as `Religion` might be documented as follows:

```
<catRef target="#B.A3 #B.D"/>
```

The element `<keywords>` contains a list of keywords or phrases identifying the topic or nature of a text. As usual, the attribute *scheme* identifies the source from which these terms are taken. For example, if the LC Subject Headings are used, following declaration of that classification system in a `<taxonomy>` element as above :

```
<textClass>
 <keywords scheme="#LCSH">
  <list>
   <item>English literature -- History and criticism -- Data processing.</item>
   <item>English literature -- History and criticism -- Theory etc.</item>
   <item>English language -- Style -- Data processing.</item>
  </list>
 </keywords>
</textClass>
```

Multiple classifications may be supplied using any of the mechanisms described in this section.

19.4 The Revision Description

The <revisionDesc> element provides a change log in which each change made to a text may be recorded. The log may be recorded as a sequence of <change> elements each of which contains a brief description of the change. The attributes *when* and *who* may be used to identify when the change was carried out and the agency responsible for it.

Example:

```
<revisionDesc>
 <change when="1991-03-06" who="#EMB">File format updated</change>
 <change when="1990-05-25" who="#EMB">Stuart's corrections entered</change>
</revisionDesc>
```

In a production environment it will usually be found preferable to use some kind of automated system to track and record changes. Many such *version control systems*, as they are known, can also be configured to update the TEI Header of a file automatically.

A List of Elements Described

The TEI Lite schema is a pure subset of TEI P5. In the following list of elements and classes used, some information, notably the examples, derives from the canonical definition for the element in TEI P5 and may therefore refer to elements or attributes not provided by TEI Lite. Note however that only the elements listed here are available within the TEI Lite schema. These specifications also refer to many attributes which although available in TEI Lite are not discussed in this tutorial for lack of space.

A.1 Elements

<TEI> (TEI document) contains a single TEI-conformant document, comprising a TEI header and a text, either in isolation or as part of a `<teiCorpus>` element. [4. 15.1.]

Module textstructure

Attributes Attributes att.global (*@xml:id*, *@n*, *@xml:lang*, *@rend*, *@xml:space*) (att.global.analytic (*@ana*)) (att.global.facs (*@facs*))

Contained by —

May contain

header: teiHeader

textstructure: text

Declaration

```
element TEI
{
   att.global.attributes,
   ( teiHeader, ( ( model.resourceLike+, text? ) | text ) )
}
```

Schematron <s:ns prefix="tei" uri="http://www.tei-c.org/ns/1.0"/> <s:ns prefix="xs" uri="http://www.w3.org/2001/XMLSchema"/>

Schematron <s:ns prefix="rng" uri="http://relaxng.org/ns/structure/1.0"/>

Example

```
<TEI version="5.0" xmlns="http://www.tei-c.org/ns/1.0">
 <teiHeader>
  <fileDesc>
   <titleStmt>
    <title>The shortest TEI Document Imaginable</title>
   </titleStmt>
   <publicationStmt>
    <p>First published as part of TEI P2, this is the P5
         version using a name space.</p>
   </publicationStmt>
   <sourceDesc>
    <p>No source: this is an original work.</p>
   </sourceDesc>
  </fileDesc>
 </teiHeader>
 <text>
  <body>
   <p>This is about the shortest TEI document imaginable.</p>
  </body>
 </text>
</TEI>
```

Note This element is required.

<author>

in a bibliographic reference, contains the name(s) of an author, personal or corporate, of a work; for example in the same form as that provided by a recognized bibliographic name authority. [3.11.2.2. 2.2.1.]

Module core

Attributes Attributesatt.global (~~rend~~, ~~xml:space~~, @xml:id, @n, @xml:lang)

Member of model.respLike

Contained by

header: titleStmt

May contain

analysis: s w

core: date gap hi name note title

header: idno

namesdates: forename surname

Declaration

```
element author
{
   att.global.attribute.xmlid,
   att.global.attribute.n,
   att.global.attribute.xmllang,
   macro.phraseSeq}
```

Example

```
<author>British Broadcasting Corporation</author>
<author>La Fayette, Marie Madeleine Pioche de la Vergne, comtesse de
(1634–1693)</author>
<author>Anonymous</author>
<author>Bill and Melinda Gates Foundation</author>
<author>
 <persName>Beaumont, Francis</persName> and
<persName>John Fletcher</persName>
</author>
<author>
 <orgName key="BBC">British Broadcasting
   Corporation</orgName>: Radio 3 Network
</author>
```

Note Particularly where cataloguing is likely to be based on the content of the header, it is advisable to use a generally recognized name authority file to supply the content for this element. The attributes *key* or *ref* may also be used to reference canonical information about the author(s) intended from any appropriate authority, such as a library catalogue or online resource.In the case of a broadcast, use this element for the name of the company or network responsible for making the broadcast.
Where an author is unknown or unspecified, this element may contain text such as *Unknown* or *Anonymous*. When the appropriate TEI modules are in use, it may also contain detailed tagging of the names used for people, organizations or places, in particular where multiple names are given.

<availability>

supplies information about the availability of a text, for example any restrictions on its use or distribution, its copyright status, any licence applying to it, etc. [2.2.4.]

Module header

Attributes Attributesatt.global (~~rend~~, ~~xml:space~~, @xml:id, @n, @xml:lang)

`@status` "publicdomain" = Texts can be made available. "restricted" = Only for personal research use, cannot be made available.

Status Optional

Datatype `data.enumerated`

Legal values are: **publicdomain** [Default]

restricted

unknown

Member of model.publicationStmtPart.detail

Contained by

header: publicationStmt

May contain

core: p

header: licence

Declaration

```
element availability
{
   att.global.attribute.xmlid,
   att.global.attribute.n,
   att.global.attribute.xmllang,
   attribute status { "publicdomain" | "restricted" | "unknown" }?,
   ( model.availabilityPart | model.pLike )+
}
```

Example

```
<availability status="restricted">
 <p>Available for academic research purposes only.</p>
</availability>
<availability status="free">
 <p>In the public domain</p>
</availability>
<availability status="restricted">
 <p>Available under licence from the publishers.</p>
</availability>
```

Example

```
<availability>
 <licence target="http://opensource.org/licenses/MIT">
  <p>The MIT License
     applies to this document.</p>
  <p>Copyright (C) 2011 by The University of Victoria</p>
  <p>Permission is hereby granted, free of charge, to any person obtaining a copy
     of this software and associated documentation files (the "Software"), to deal
     in the Software without restriction, including without limitation the rights
     to use, copy, modify, merge, publish, distribute, sublicense, and/or sell
     copies of the Software, and to permit persons to whom the Software is
     furnished to do so, subject to the following conditions:</p>
  <p>The above copyright notice and this permission notice shall be included in
     all copies or substantial portions of the Software.</p>
  <p>THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR
     IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY,
     FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE
     AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER
     LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM,
```

```
     OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN
     THE SOFTWARE.</p>
 </licence>
</availability>
```

Note A consistent format should be adopted

<back>

<back> (back matter) contains any appendixes, etc. following the main part of a text. [4.7. 4.]

Module textstructure

Attributes Attributesatt.global (~~xml:space~~, @xml:id, @n, @xml:lang, @rend)
att.global.analytic (@ana)

Contained by

textstructure: text

May contain

core: gap head note p

drama: castList set

textstructure: div

Declaration

```
element back
{
   att.global.attribute.xmlid,
   att.global.attribute.n,
   att.global.attribute.xmllang,
   att.global.attribute.rend,
   att.global.analytic.attribute.ana,
   (
      (
         model.frontPart       | model.pLike.front       | model.pLike       | model.list
      (
         (
            (
               ( model.div1Like ),
               ( model.frontPart | model.div1Like | model.global )*
            )
          | (
               ( model.divLike ),
               ( model.frontPart | model.divLike | model.global )*
            )
         )?
      ),
      ( ( ( model.divBottomPart ), ( model.divBottomPart | model.global )* )? )
   )
}
```

Example

```
<back>
 <div1 type="appendix">
  <head>The Golden Dream or, the Ingenuous Confession</head>
  <p>To shew the Depravity of human Nature </p>
 </div1>
 <div1 type="epistle">
  <head>A letter from the Printer, which he desires may be inserted</head>
  <salute>Sir.</salute>
  <p>I have done with your Copy, so you may return it to the Vatican, if you please
```

```
</p>
 </div1>
 <div1 type="advert">
  <head>The Books usually read by the Scholars of Mrs Two-Shoes are these and are
sold at Mr
      Newbery's at the Bible and Sun in St Paul's Church-yard.</head>
  <list>
   <item n="1">The Christmas Box, Price 1d.</item>
   <item n="2">The History of Giles Gingerbread, 1d.</item>
   <item n="42">A Curious Collection of Travels, selected from the Writers of all
Nations,
       10 Vol, Pr. bound 1l.</item>
  </list>
 </div1>
 <div1 type="advert">
  <head>
   <hi rend="center">By the KING's Royal Patent,</hi> Are sold by J. NEWBERY, at the
     Bible and Sun in St. Paul's Church-Yard.</head>
  <list>
   <item n="1">Dr. James's Powders for Fevers, the Small-Pox, Measles, Colds, &c.
       2s. 6d</item>
   <item n="2">Dr. Hooper's Female Pills, 1s.</item>
  </list>
 </div1>
</back>
```

Note The content model of back matter is identical to that of front matter, reflecting the facts of cultural history.

<body>

<body> (text body) contains the whole body of a single unitary text, excluding any front or back matter. [4.]

Module textstructure

Attributes Attributesatt.global (~~xml:space~~, @xml:id, @n, @xml:lang, @rend) att.global.analytic (@ana)

Contained by

textstructure: text

May contain

core: gap head l lg note p q quote sp stage

drama: castList

textstructure: div

Declaration

```
element body
{
   att.global.attribute.xmlid,
   att.global.attribute.n,
   att.global.attribute.xmllang,
   att.global.attribute.rend,
   att.global.analytic.attribute.ana,
   (
      model.global*,
      ( ( model.divTop ), ( model.global | model.divTop )* )?,
      ( ( model.divGenLike ), ( model.global | model.divGenLike )* )?,
      (
         ( ( model.divLike ), ( model.global | model.divGenLike )* )+
       | ( ( model.div1Like ), ( model.global | model.divGenLike )* )+
       | (
```

```
            ( ( model.common ), model.global* )+,
            (
               ( ( model.divLike ), ( model.global | model.divGenLike )* )+
             | ( ( model.div1Like ), ( model.global | model.divGenLike )* )+
            )?
         )
      ),
      ( ( model.divBottom ), model.global* )*
   )
}
```

<castGroup> (cast list grouping) groups one or more individual castItem elements within a cast list. [7.1.4.]

Module drama

Attributes Attributes att.global (*@xml:id*, *@n*, *@xml:lang*, *@rend*, *@xml:space*) (att.global.analytic (*@ana*)) (att.global.facs (*@facs*))

Contained by

drama: castGroup castList

May contain

core: gap head note

drama: castGroup castItem roleDesc

Declaration

```
element castGroup
{
   att.global.attributes,
   (
      ( model.global | model.headLike )*,
      ( ( castItem | castGroup | roleDesc ), model.global* )+,
      ( trailer, model.global* )?
   )
}
```

Example

```
<castGroup rend="braced">
 <castItem>
  <role>Walter</role>
  <actor>Mr Frank Hall</actor>
 </castItem>
 <castItem>
  <role>Hans</role>
  <actor>Mr F.W. Irish</actor>
 </castItem>
 <roleDesc>friends of Mathias</roleDesc>
</castGroup>
```

Note The *rend* attribute may be used, as here, to indicate whether the grouping is indicated by a brace, whitespace, font change, etc.Note that in this example the role description 'friends of Mathias' is understood to apply to both roles equally.

<castItem> (cast list item) contains a single entry within a cast list, describing either a single role or a list of non-speaking roles. [7.1.4.]

Module drama

Attributes Attributes att.global (*@xml:id*, *@n*, *@xml:lang*, *@rend*, *@xml:space*) (att.global.analytic (*@ana*)) (att.global.facs (*@facs*))

`@type` characterizes the cast item.

Status Optional

Datatype `data.enumerated`

Legal values are: **role** the item describes a single role.[Default]

list the item describes a list of non-speaking roles.

Contained by drama: castGroup castList

May contain

analysis: s w

core: date gap hi name note title

drama: role roleDesc

header: idno

namesdates: forename surname

Declaration

```
element castItem
{
   att.global.attributes,
   attribute type { "role" | "list" }?,
   ( text | model.gLike | model.castItemPart | model.phrase | model.global )*
}
```

Example

```
<castItem>
 <role>Player</role>
 <actor>Mr Milward</actor>
</castItem>
```

Example

```
<castItem type="list">Constables, Drawer, Turnkey, etc.</castItem>
```

<castList>

(cast list) contains a single cast list or dramatis personae. [7.1.4. 7.1.]

Module drama

Attributes Attributes att.global (*@xml:id*, *@n*, *@xml:lang*, *@rend*, *@xml:space*) (att.global.analytic (*@ana*)) (att.global.facs (*@facs*))

Member of model.frontPart.drama model.inter

Contained by

core: head hi l note p q quote stage title

drama: castList set

header: change licence

textstructure: back body div front

May contain

core: gap head l lg note p q quote sp stage

drama: castGroup castItem castList

Declaration

```
element castList
{
```

```
   att.global.attributes,
   (
      ( model.divTop | model.global )*,
      ( ( model.common ), model.global* )*,
      ( ( castItem | castGroup ), model.global* )+,
      ( ( model.common ), model.global* )*
   )
}
```

Example

```
<castList>
 <castGroup>
  <head rend="braced">Mendicants</head>
  <castItem>
   <role>Aafaa</role>
   <actor>Femi Johnson</actor>
  </castItem>
  <castItem>
   <role>Blindman</role>
   <actor>Femi Osofisan</actor>
  </castItem>
  <castItem>
   <role>Goyi</role>
   <actor>Wale Ogunyemi</actor>
  </castItem>
  <castItem>
   <role>Cripple</role>
   <actor>Tunji Oyelana</actor>
  </castItem>
 </castGroup>
 <castItem>
  <role>Si Bero</role>
  <roleDesc>Sister to Dr Bero</roleDesc>
  <actor>Deolo Adedoyin</actor>
 </castItem>
 <castGroup>
  <head rend="braced">Two old women</head>
  <castItem>
   <role>Iya Agba</role>
   <actor>Nguba Agolia</actor>
  </castItem>
  <castItem>
   <role>Iya Mate</role>
   <actor>Bopo George</actor>
  </castItem>
 </castGroup>
 <castItem>
  <role>Dr Bero</role>
  <roleDesc>Specialist</roleDesc>
  <actor>Nat Okoro</actor>
 </castItem>
 <castItem>
  <role>Priest</role>
  <actor>Gbenga Sonuga</actor>
 </castItem>
 <castItem>
  <role>The old man</role>
  <roleDesc>Bero's father</roleDesc>
  <actor>Dapo Adelugba</actor>
 </castItem>
</castList>
```

```
<stage type="mix">The action takes place in and around the home surgery of
Dr Bero, lately returned from the wars.</stage>
```

<change> documents a change or set of changes made during the production of a source document, or during the revision of an electronic file. [2.5. 2.4.1. 11.7.]

Module header

Attributes Attributes att.ascribed (*@who*) att.datable.w3c (@when) att.global (~~rend~~, ~~xml:space~~, @xml:id, @n, @xml:lang)

Contained by

header: revisionDesc

May contain

analysis: s w

core: date gap hi l lg name note p q quote sp stage title

drama: castList

header: idno

namesdates: forename surname

Declaration

```
element change
{
   att.ascribed.attributes,
   att.datable.w3c.attribute.when,
   att.global.attribute.xmlid,
   att.global.attribute.n,
   att.global.attribute.xmllang,
   macro.specialPara}
```

Example

```
<titleStmt>
 <title> ... </title>
 <editor xml:id="LDB">Lou Burnard</editor>
 <respStmt xml:id="BZ">
  <resp>copy editing</resp>
  <name>Brett Zamir</name>
 </respStmt>
</titleStmt>
<!-- ... -->
<revisionDesc status="published">
 <change status="public" when="2008-02-02"
  who="#BZ">Finished chapter 23</change>
 <change status="draft" when="2008-01-02"
  who="#BZ">Finished chapter 2</change>
 <change n="P2.2" when="1991-12-21"
  who="#LDB">Added examples to section 3</change>
 <change when="1991-11-11" who="#MSM">Deleted chapter 10</change>
</revisionDesc>
```

Example

```
<profileDesc>
 <creation>
  <listChange>
   <change xml:id="DRAFT1">First draft in pencil</change>
   <change notBefore="1880-12-09"
    xml:id="DRAFT2">First revision, mostly
       using green ink</change>
```

```
   <change notBefore="1881-02-13"
    xml:id="DRAFT3">Final corrections as
       supplied to printer.</change>
  </listChange>
 </creation>
</profileDesc>
```

Note The *who* attribute may be used to point to any other element, but will typically specify a `<respStmt>` or `<person>` element elsewhere in the header, identifying the person responsible for the change and their role in making it.It is recommended that changes be recorded with the most recent first. The *status* attribute may be used to indicate the status of a document following the change documented.

<date>

<date> contains a date in any format. [3.5.4. 2.2.4. 2.5. 3.11.2.4. 15.2.3. 13.3.6.]

Module core

Attributes Attributesatt.global (~~xml:space~~, @xml:id, @n, @xml:lang, @rend) att.global.analytic (@ana) att.datable.w3c (@when) att.dimensions (~~unit~~, ~~quantity~~, ~~extent~~, ~~scope~~, @precision) att.typed (~~subtype~~, @type)

Member of model.dateLike model.publicationStmtPart.detail

Contained by

analysis: s

core: author date head hi l name note p publisher q quote speaker stage title

drama: castItem role roleDesc

header: change licence principal publicationStmt

namesdates: forename surname

May contain

analysis: s w

core: date gap hi name note title

header: idno

namesdates: forename surname

Declaration

```
element date
{
   att.global.attribute.xmlid,
   att.global.attribute.n,
   att.global.attribute.xmllang,
   att.global.attribute.rend,
   att.global.analytic.attribute.ana,
   att.datable.w3c.attribute.when,
   att.dimensions.attribute.precision,
   att.typed.attribute.type,
   ( text | model.gLike | model.phrase | model.global )*
}
```

Example

```
<date when="1980-02">early February 1980</date>
```

Example

```
Given on the <date when="1977-06-12">Twelfth Day
of June in the Year of Our Lord One Thousand Nine Hundred and Seventy-seven of the
```

```
Republic
the Two Hundredth and first and of the University the Eighty-Sixth.</date>
```

Example

```
<date when="1990-09">September 1990</date>
```

(text division) contains a subdivision of the front, body, or back of a text. [4.1.]

Module textstructure

Attributes Attributesatt.global (~~xml:space~~, @xml:id, @n, @xml:lang, @rend)
att.global.analytic (@ana) att.typed (~~subtype~~, @type)

Member of model.divLike

Contained by

textstructure: back body div front

May contain

core: gap head l lg note p q quote sp stage

drama: castList

textstructure: div

Declaration

```
element div
{
   att.global.attribute.xmlid,
   att.global.attribute.n,
   att.global.attribute.xmllang,
   att.global.attribute.rend,
   att.global.analytic.attribute.ana,
   att.typed.attribute.type,
   (
      ( model.divTop | model.global )*,
      (
         (
            ( ( ( model.divLike | model.divGenLike ), model.global* )+ )
          | (
               ( ( model.common ), model.global* )+,
               ( ( model.divLike | model.divGenLike ), model.global* )*
            )
         ),
         ( ( model.divBottom ), model.global* )*
      )?
   )
}
```

Example

```
<body>
 <div type="part">
  <head>Fallacies of Authority</head>
  <p>The subject of which is Authority in various shapes, and the object, to repress
all
     exercise of the reasoning faculty.</p>
  <div n="1" type="chapter">
   <head>The Nature of Authority</head>
   <p>With reference to any proposed measures having for their object the greatest
       happiness of the greatest number....</p>
   <div n="1.1" type="section">
    <head>Analysis of Authority</head>
    <p>What on any given occasion is the legitimate weight or influence to be
```

```
attached to
          authority ... </p>
    </div>
    <div n="1.2" type="section">
     <head>Appeal to Authority, in What Cases Fallacious.</head>
     <p>Reference to authority is open to the charge of fallacy when... </p>
    </div>
   </div>
 </div>
</body>
```

<fileDesc> (file description) contains a full bibliographic description of an electronic file. [2.2. 2.1.1.]

Module header

Attributes Attributes att.global (*@xml:id*, *@n*, *@xml:lang*, *@rend*, *@xml:space*) (att.global.analytic (*@ana*)) (att.global.facs (*@facs*))

Contained by

header: teiHeader

May contain

header: publicationStmt sourceDesc titleStmt

Declaration

```
element fileDesc
{
   att.global.attributes,
   (
      (
         titleStmt,
         editionStmt?,
         extent?,
         publicationStmt,
         seriesStmt?,
         notesStmt?
      ),
      sourceDesc+
   )
}
```

Example

```
<fileDesc>
 <titleStmt>
  <title>The shortest possible TEI document</title>
 </titleStmt>
 <publicationStmt>
  <p>Distributed as part of TEI P5</p>
 </publicationStmt>
 <sourceDesc>
  <p>No print source exists: this is an original digital text</p>
 </sourceDesc>
</fileDesc>
```

Note The major source of information for those seeking to create a catalogue entry or bibliographic citation for an electronic file. As such, it provides a title and statements of responsibility together with details of the publication or distribution of the file, of any series to which it belongs, and detailed bibliographic notes for

matters not addressed elsewhere in the header. It also contains a full bibliographic description for the source or sources from which the electronic text was derived.

<forename> contains a forename, given or baptismal name. [13.2.1.]

Module namesdates

Attributes Attributes att.global (*@xml:id*, *@n*, *@xml:lang*, *@rend*, *@xml:space*) (att.global.analytic (*@ana*)) (att.global.facs (*@facs*)) att.personal (*@full*, *@sort*) (att.naming (*@role*, *@nymRef*) (att.canonical (*@key*, *@ref*))) att.typed (*@type*, *@subtype*)

Member of model.persNamePart

Contained by

analysis: s

core: author date head hi l name note p publisher q quote speaker stage title

drama: castItem role roleDesc

header: change licence principal

namesdates: forename surname

May contain

analysis: s w

core: date gap hi name note title

header: idno

namesdates: forename surname

Declaration

```
element forename
{
   att.global.attributes,
   att.personal.attributes,
   att.typed.attributes,
   macro.phraseSeq}
```

Example

```
<persName>
 <roleName>Ex-President</roleName>
 <forename>George</forename>
 <surname>Bush</surname>
</persName>
```

<front> (front matter) contains any prefatory matter (headers, title page, prefaces, dedications, etc.) found at the start of a document, before the main body. [4.6. 4.]

Module textstructure

Attributes Attributesatt.global (~~xml:space~~, @xml:id, @n, @xml:lang, @rend) att.global.analytic (@ana)

Contained by

textstructure: text

May contain

core: gap head note p

drama: castList set

textstructure: div

Declaration

```
element front
{
   att.global.attribute.xmlid,
   att.global.attribute.n,
   att.global.attribute.xmllang,
   att.global.attribute.rend,
   att.global.analytic.attribute.ana,
   (
      ( ( model.frontPart | model.pLike | model.pLike.front | model.global )* ),
      (
         (
            (
               model.div1Like,
               ( model.div1Like | model.frontPart | model.global )*
            )
          | (
               model.divLike,
               ( model.divLike | model.frontPart | model.global )*
            )
         ),
         ( ( ( model.divBottom ), ( model.divBottom | model.global )* )? )
      )?
   )
}
```

Example

```
<front>
 <epigraph>
  <quote>Nam Sibyllam quidem Cumis ego ipse oculis meis
     vidi in ampulla pendere, et cum illi pueri dicerent:
  <q xml:lang="gr">Σίβυλλα τί θέλεις</q>; respondebat
     illa: <q xml:lang="gr">ἀποθανεῖν θέλω.</q>
  </quote>
 </epigraph>
 <div type="dedication">
  <p>For Ezra Pound <q xml:lang="it">il miglior fabbro.</q>
  </p>
 </div>
</front>
```

Example

```
<front>
 <div type="dedication">
  <p>To our three selves</p>
 </div>
 <div type="preface">
  <head>Author's Note</head>
  <p>All the characters in this book are purely imaginary, and if the
     author has used names that may suggest a reference to living persons
     she has done so inadvertently.
     ...</p>
 </div>
</front>
```

<gap> indicates a point where material has been omitted in a transcription, whether for editorial reasons described in the TEI header, as part of sampling practice, or because the material is illegible, invisible, or inaudible. [3.4.3.]

Module core

Attributes Attributesatt.global (~~xml:space~~, @xml:id, @n, @xml:lang, @rend)
att.global.analytic (@ana)

Member of model.global.edit

Contained by

analysis: s w

core: author date head hi l lg name note p publisher q quote sp speaker stage title

drama: castGroup castItem castList role roleDesc set

header: change licence principal

namesdates: forename surname

textstructure: back body div front text

May contain Empty element

Declaration

```
element gap
{
   att.global.attribute.xmlid,
   att.global.attribute.n,
   att.global.attribute.xmllang,
   att.global.attribute.rend,
   att.global.analytic.attribute.ana,
   ( model.descLike | model.certLike )*
}
```

Example

```
<gap quantity="4" reason="illegible"
 unit="chars"/>
```

Example

```
<gap quantity="1" reason="sampling"
 unit="essay"/>
```

Example

```
<del>
 <gap atLeast="4" atMost="8"
  reason="illegible" unit="chars"/>
</del>
```

Example

```
<gap extent="unknown" reason="lost"
 unit="lines"/>
```

Note The <gap>, **<unclear>**, and **<del>** core tag elements may be closely allied in use with the **<damage>** and **<supplied>** elements, available when using the additional tagset for transcription of primary sources. See section 11.3.3.2. for discussion of which element is appropriate for which circumstance.The <gap> tag simply signals the editors decision to omit or inability to transcribe a span of text. Other information, such as the interpretation that text was deliberately erased or covered, should be indicated using the relevant tags, such as **<del>** in the case of deliberate deletion.

<head>

(heading) contains any type of heading, for example the title of a section, or the heading of a list, glossary, manuscript description, etc. [4.2.1.]

Module core

Attributes Attributesatt.global (~~rend~~, ~~xml:space~~, @xml:id, @n, @xml:lang)
att.global.analytic (@ana)

Member of model.headLike model.pLike.front

Contained by

core: lg

drama: castGroup castList set

textstructure: back body div front

May contain

analysis: s w

core: date gap hi l lg name note q quote stage title

drama: castList

header: idno

namesdates: forename surname

Declaration

```
element head
{
   att.global.attribute.xmlid,
   att.global.attribute.n,
   att.global.attribute.xmllang,
   att.global.analytic.attribute.ana,
   (
      text
    | lg    | model.gLike    | model.phrase    | model.inter    | model.lLike    | model.
}
```

Example The most common use for the <head> element is to mark the headings of sections. In older writings, the headings or *incipits* may be rather longer than usual in modern works. If a section has an explicit ending as well as a heading, it should be marked as a `<trailer>`, as in this example:

```
<div1 n="I" type="book">
 <head>In the name of Christ here begins the first book of the ecclesiastical
history of
   Georgius Florentinus, known as Gregory, Bishop of Tours.</head>
 <div2 type="section">
  <head>In the name of Christ here begins Book I of the history.</head>
  <p>Proposing as I do ...</p>
  <p>From the Passion of our Lord until the death of Saint Martin four hundred and
twelve
     years passed.</p>
  <trailer>Here ends the first Book, which covers five thousand, five hundred and
ninety-six
     years from the beginning of the world down to the death of Saint
Martin.</trailer>
 </div2>
</div1>
```

Example The <head> element is also used to mark headings of other units, such as lists:

```
With a few exceptions, connectives are equally
useful in all kinds of discourse: description, narration, exposition, argument.
<list rend="bulleted">
 <head>Connectives</head>
 <item>above</item>
 <item>accordingly</item>
 <item>across from</item>
 <item>adjacent to</item>
 <item>again</item>
```

```
 <item>
<!-- ... -->
 </item>
</list>
```

Note The <head> element is used for headings at all levels; software which treats (e.g.) chapter headings, section headings, and list titles differently must determine the proper processing of a <head> element based on its structural position. A <head> occurring as the first element of a list is the title of that list; one occurring as the first element of a **<div1>** is the title of that chapter or section.

<hi>

(highlighted) marks a word or phrase as graphically distinct from the surrounding text, for reasons concerning which no claim is made. [3.3.2.2. 3.3.2.]

Module core

Attributes Attributesatt.global (~~xml:space~~, @xml:id, @n, @xml:lang, @rend)

Member of model.hiLike

Contained by

analysis: s w

core: author date head hi l name note p publisher q quote speaker stage title

drama: castItem role roleDesc

header: change licence principal

namesdates: forename surname

May contain

analysis: s w

core: date gap hi lg name note q quote stage title

drama: castList

header: idno

namesdates: forename surname

Declaration

```
element hi
{
   att.global.attribute.xmlid,
   att.global.attribute.n,
   att.global.attribute.xmllang,
   att.global.attribute.rend,
   macro.paraContent}
```

Example

```
<hi rend="gothic">And this Indenture further witnesseth</hi>
that the said <hi rend="italic">Walter Shandy</hi>, merchant,
in consideration of the said intended marriage ...
```

<idno>

(identifier) supplies any form of identifier used to identify some object, such as a bibliographic item, a person, a title, an organization, etc. in a standardized way. [2.2.4. 2.2.5. 3.11.2.4.]

Module header

Attributes Attributesatt.global (~~rend~~, ~~xml:space~~, @xml:id, @n, @xml:lang)

`@type` categorizes the identifier, for example as an ISBN, Social Security number, etc.
Status Optional
Datatype `data.enumerated`

Member of model.nameLike model.publicationStmtPart.detail
Contained by
analysis: s
core: author date head hi l name note p publisher q quote speaker stage title
drama: castItem role roleDesc
header: change idno licence principal publicationStmt
namesdates: forename surname
May contain
header: idno
Declaration

```
element idno
{
   att.global.attribute.xmlid,
   att.global.attribute.n,
   att.global.attribute.xmllang,
   attribute type { data.enumerated }?,
   ( text | model.gLike | idno )*
}
```

Example

```
<idno type="ISBN">978-1-906964-22-1</idno>
<idno type="ISSN">0143-3385</idno>
<idno type="DOI">http://dx.doi.org/10.1000/123</idno>
<idno type="URI">http://www.worldcat.org/oclc/185922478</idno>
<idno type="URI">http://authority.nzetc.org/463/</idno>
<idno type="LT">Thomason Tract E.537(17)</idno>
<idno type="Wing">C695</idno>
<idno type="oldCat">
 <g ref="#sym"/>345
</idno>
```

In the last case, the identifier includes a non-Unicode character which is defined elsewhere by means of a `<glyph>` or `<char>` element referenced here as `#sym`.

Note <idno> should be used for labels which identify an object or concept in a formal cataloguing system such as a database or an RDF store, or in a distributed system such as the World Wide Web. Some suggested values for *type* on <idno> are `ISBN`, `ISSN`, `DOI`, and `URI`.

<l>

(verse line) contains a single, possibly incomplete, line of verse. [3.12.1. 3.12. 7.2.5.]

Module core
Attributes Attributesatt.global (~~rend~~, ~~xml:space~~, @xml:id, @n, @xml:lang)
att.global.analytic (@ana)
Member of model.lLike
Contained by
core: head lg note q quote sp stage
drama: castList set

header: change licence
textstructure: body div
May contain
analysis: s w
core: date gap hi lg name note q quote stage title
drama: castList
header: idno
namesdates: forename surname
Declaration

```
element l
{
   att.global.attribute.xmlid,
   att.global.attribute.n,
   att.global.attribute.xmllang,
   att.global.analytic.attribute.ana,
   macro.paraContent}
```

Example

```
<l met="x/x/x/x/x/" real="/xx/x/x/x/">Shall I compare thee to a summer's day?</l>
```

<lg>

(line group) contains one or more verse lines functioning as a formal unit, e.g. a stanza, refrain, verse paragraph, etc. [3.12.1. 3.12. 7.2.5.]

Module core
Attributes Attributesatt.global (~~rend~~, ~~xml:space~~, @xml:id, @n, @xml:lang)
att.global.analytic (@ana)
Member of macro.paraContent model.divPart
Contained by
core: head hi l lg note p q quote sp stage title
drama: castList set
header: change licence
textstructure: body div
May contain
core: gap head l lg note stage
Declaration

```
element lg
{
   att.global.attribute.xmlid,
   att.global.attribute.n,
   att.global.attribute.xmllang,
   att.global.analytic.attribute.ana,
   (
      ( model.divTop | model.global )*,
      ( model.lLike | model.stageLike | model.labelLike | lg ),
      ( model.lLike | model.stageLike | model.labelLike | model.global | lg )*,
      ( ( model.divBottom ), model.global* )*
   )
}
```

Schematron

<sch:assert test="count(descendant::tei:lg|descendant::tei:l|descendant::tei:gap) > 0">An lg element must contain at least one child l, lg or gap element.</sch:assert>

Example

```
<lg type="free">
 <l>Let me be my own fool</l>
 <l>of my own making, the sum of it</l>
</lg>
<lg type="free">
 <l>is equivocal.</l>
 <l>One says of the drunken farmer:</l>
</lg>
<lg type="free">
 <l>leave him lay off it. And this is</l>
 <l>the explanation.</l>
</lg>
```

Note contains verse lines or nested line groups only, possibly prefixed by a heading.

<licence> contains information about a licence or other legal agreement applicable to the text. [2.2.4.]

Module header

Attributes Attributesatt.global (~~rend~~, ~~xml:space~~, @xml:id, @n, @xml:lang) att.pointing (~~targetLang~~, ~~evaluate~~, @target)

Member of model.availabilityPart

Contained by

header: availability

May contain

analysis: s w

core: date gap hi l lg name note p q quote sp stage title

drama: castList

header: idno

namesdates: forename surname

Declaration

```
element licence
{
   att.global.attribute.xmlid,
   att.global.attribute.n,
   att.global.attribute.xmllang,
   att.pointing.attribute.target,
   macro.specialPara}
```

Example

```
<licence target="http://www.nzetc.org/tm/scholarly/tei-NZETC-Help.html#licensing">
Licence: Creative Commons Attribution-Share Alike 3.0 New Zealand Licence
</licence>
```

Example

```
<availability>
 <licence notBefore="2013-01-01"
  target="http://creativecommons.org/licenses/by/3.0/">
  <p>The Creative Commons Attribution 3.0 Unported (CC BY 3.0) Licence
     applies to this document.</p>
```

```
  <p>The licence was added on January 1, 2013.</p>
 </licence>
</availability>
```

Note A <licence> element should be supplied for each licence agreement applicable to the text in question. The *target* attribute may be used to reference a full version of the licence. The *when*, *notBefore*, *notAfter*, *from* or *to* attributes may be used in combination to indicate the date or dates of applicability of the licence.

<name> (name, proper noun) contains a proper noun or noun phrase. [3.5.1.]

Module core

Attributes Attributesatt.global (~~xml:space~~, @xml:id, @n, @xml:lang, @rend) att.global.analytic (@ana) att.typed (~~subtype~~, @type)

Member of model.nameLike.agent

Contained by

analysis: s

core: author date head hi l name note p publisher q quote speaker stage title

drama: castItem role roleDesc

header: change licence principal

namesdates: forename surname

May contain

analysis: s w

core: date gap hi name note title

header: idno

namesdates: forename surname

Declaration

```
element name
{
   att.global.attribute.xmlid,
   att.global.attribute.n,
   att.global.attribute.xmllang,
   att.global.attribute.rend,
   att.global.analytic.attribute.ana,
   att.typed.attribute.type,
   macro.phraseSeq}
```

Example

```
<name type="person">Thomas Hoccleve</name>
<name type="place">Villingaholt</name>
<name type="org">Vetus Latina Institut</name>
<name ref="#HOC001" type="person">Occleve</name>
```

Note Proper nouns referring to people, places, and organizations may be tagged instead with <persName>, <placeName>, or <orgName>, when the TEI module for names and dates is included.

<note> contains a note or annotation. [3.8.1. 2.2.6. 3.11.2.8. 9.3.5.4.]

Module core

Attributes Attributesatt.global (~~rend~~, ~~xml:space~~, @xml:id, @n, @xml:lang) att.global.analytic (@ana) att.source (@source) att.typed (~~subtype~~, @type)

Member of model.noteLike

Contained by

analysis: s w

core: author date head hi l lg name note p publisher q quote sp speaker stage title

drama: castGroup castItem castList role roleDesc set

header: change licence principal

namesdates: forename surname

textstructure: back body div front text

May contain

analysis: s w

core: date gap hi l lg name note p q quote sp stage title

drama: castList

header: idno

namesdates: forename surname

Declaration

```
element note
{
   att.global.attribute.xmlid,
   att.global.attribute.n,
   att.global.attribute.xmllang,
   att.global.analytic.attribute.ana,
   att.source.attribute.source,
   att.typed.attribute.type,
   macro.specialPara}
```

Example In the following example, the translator has supplied a footnote containing an explanation of the term translated as "painterly":

```
And yet it is not only
in the great line of Italian renaissance art, but even in the
painterly <note place="bottom" resp="#MDMH"
 type="gloss">
 <term xml:lang="de">Malerisch</term>. This word has, in the German, two
distinct meanings, one objective, a quality residing in the object,
the other subjective, a mode of apprehension and creation. To avoid
confusion, they have been distinguished in English as
<mentioned>picturesque</mentioned> and
<mentioned>painterly</mentioned> respectively.
</note> style of the
Dutch genre painters of the seventeenth century that drapery has this
psychological significance.
```

For this example to be valid, the code MDMH must be defined elsewhere, for example by means of a responsibility statement in the associated TEI header:

```
<respStmt xml:id="MDMH">
 <resp>translation from German to English</resp>
 <name>Hottinger, Marie Donald Mackie</name>
</respStmt>
```

Example The global *n* attribute may be used to supply the symbol or number used to mark the note's point of attachment in the source text, as in the following example:

```
Mevorakh b. Saadya's mother, the matriarch of the
family during the second half of the eleventh century,
<note anchored="true" n="126"> The
```

```
alleged mention of Judah Nagid's mother in a letter from 1071 is, in fact, a
reference to
Judah's children; cf. above, nn. 111 and 54. </note> is well known from Geniza
documents
published by Jacob Mann.
```

However, if notes are numbered in sequence and their numbering can be reconstructed automatically by processing software, it may well be considered unnecessary to record the note numbers.

<p> (paragraph) marks paragraphs in prose. [3.1. 7.2.5.]

Module core

Attributes Attributesatt.global (~~xml:space~~, @xml:id, @n, @xml:lang, @rend)
att.global.analytic (@ana)

Member of model.pLike

Contained by

core: note q quote sp stage

drama: castList set

header: availability change licence publicationStmt sourceDesc

textstructure: back body div front

May contain

analysis: s w

core: date gap hi lg name note q quote stage title

drama: castList

header: idno

namesdates: forename surname

Declaration

```
element p
{
   att.global.attribute.xmlid,
   att.global.attribute.n,
   att.global.attribute.xmllang,
   att.global.attribute.rend,
   att.global.analytic.attribute.ana,
   macro.paraContent}
```

Example

```
<p>Hallgerd was outside. <q>There is blood on your axe,</q> she said. <q>What have
you
   done?</q>
</p>
<p>
 <q>I have now arranged that you can be married a second time,</q> replied Thjostolf.
</p>
<p>
 <q>Then you must mean that Thorvald is dead,</q> she said.
</p>
<p>
 <q>Yes,</q> said Thjostolf. <q>And now you must think up some plan for me.</q>
</p>
```

<principal> (principal researcher) supplies the name of the principal researcher responsible for the creation of an electronic text. [2.2.1.]

Module header

Attributes Attributes att.global (*@xml:id*, *@n*, *@xml:lang*, *@rend*, *@xml:space*) (att.global.analytic (*@ana*)) (att.global.facs (*@facs*)) att.canonical (*@key*, *@ref*)

Member of model.respLike

Contained by

header: titleStmt

May contain

core: date gap hi name note title

header: idno

namesdates: forename surname

Declaration

```
element principal
{
   att.global.attributes,
   att.canonical.attributes,
   macro.phraseSeq.limited}
```

Example

```
<principal ref="http://viaf.org/viaf/105517912">Gary Taylor</principal>
```

<publicationStmt> (publication statement) groups information concerning the publication or distribution of an electronic or other text. [2.2.4. 2.2.]

Module header

Attributes Attributes att.global (*@xml:id*, *@n*, *@xml:lang*, *@rend*, *@xml:space*) (att.global.analytic (*@ana*)) (att.global.facs (*@facs*))

Contained by

header: fileDesc

May contain

core: date p publisher

header: availability idno

Declaration

```
element publicationStmt
{
   att.global.attributes,
   (
      (
         ( model.publicationStmtPart.agency ),
         model.publicationStmtPart.detail*
      )+
    | model.pLike+
   )
}
```

Example

```
<publicationStmt>
 <publisher>C. Muquardt </publisher>
 <pubPlace>Bruxelles & Leipzig</pubPlace>
```

```
 <date when="1846"/>
</publicationStmt>
```

Example

```
<publicationStmt>
 <publisher>Chadwyck Healey</publisher>
 <pubPlace>Cambridge</pubPlace>
 <availability>
  <p>Available under licence only</p>
 </availability>
 <date when="1992">1992</date>
</publicationStmt>
```

Note Where a publication statement contains several members of the `model.publicationStmtPart` classes rather than one or more paragraphs or anonymous blocks, care should be taken to ensure that the repeated elements are presented in a meaningful order. It is a conformance requirement that elements supplying information about publication place, address, identifier, availability, and date be given following the name of the publisher, distributor, or authority concerned, and preferably in that order.

<publisher>

<publisher> provides the name of the organization responsible for the publication or distribution of a bibliographic item. [3.11.2.4. 2.2.4.]

Module core

Attributes Attributesatt.global (~~rend~~, ~~xml:space~~, @xml:id, @n, @xml:lang)

Member of model.publicationStmtPart.agency

Contained by

header: publicationStmt

May contain

analysis: s w

core: date gap hi name note title

header: idno

namesdates: forename surname

Declaration

```
element publisher
{
   att.global.attribute.xmlid,
   att.global.attribute.n,
   att.global.attribute.xmllang,
   macro.phraseSeq}
```

Example

```
<imprint>
 <pubPlace>Oxford</pubPlace>
 <publisher>Clarendon Press</publisher>
 <date>1987</date>
</imprint>
```

Note Use the full form of the name by which a company is usually referred to, rather than any abbreviation of it which may appear on a title page

---

<q> (quoted) contains material which is distinguished from the surrounding text using quotation marks or a similar method, for any one of a variety of reasons including, but not limited to: direct speech or thought, technical terms or jargon, authorial distance, quotations from elsewhere, and passages that are mentioned but not used. [3.3.3.]

Module core

Attributes Attributes att.ascribed (*@who*) att.global (~~rend~~, ~~xml:space~~, @xml:id, @n, @xml:lang) att.global.analytic (@ana)

`@type` may be used to indicate whether the offset passage is spoken or thought, or to characterize it more finely.

Status Optional

Datatype `data.enumerated`

Suggested values include: **spoken** representation of speech

thought representation of thought, e.g. internal monologue

written quotation from a written source

soCalled authorial distance

foreign

distinct linguistically distinct

term technical term

emph rhetorically emphasized

mentioned refering to itself, not its normal referent

Member of model.qLike

Contained by

core: head hi l note p q quote sp stage title

drama: castList set

header: change licence

textstructure: body div

May contain

analysis: s w

core: date gap hi l lg name note p q quote sp stage title

drama: castList

header: idno

namesdates: forename surname

Declaration

```
element q
{
   att.global.attribute.xmlid,
   att.global.attribute.n,
   att.global.attribute.xmllang,
   att.global.analytic.attribute.ana,
   att.ascribed.attributes,
   attribute type
   {
      "spoken"
    | "thought"
    | "written"
    | "soCalled"
    | "foreign"
    | "distinct"
    | "term"
```

```
    | "emph"
    | "mentioned"
    | xsd:Name
   }?,
   macro.specialPara}
```

Example

```
It is spelled <q>Tübingen</q> — to enter the
letter <q>u</q> with an umlaut hold down the <q>option</q> key and press
<q>0 0 f c</q>
```

Note May be used to indicate that a passage is distinguished from the surrounding text for reasons concerning which no claim is made. When used in this manner, <q> may be thought of as syntactic sugar for <hi> with a value of *rend* that indicates the use of such mechanisms as quotation marks.

<quote>

<quote> (quotation) contains a phrase or passage attributed by the narrator or author to some agency external to the text. [3.3.3. 4.3.1.]

Module core

Attributes Attributesatt.global (~~rend~~, ~~xml:space~~, @xml:id, @n, @xml:lang) att.global.analytic (@ana) att.typed (~~subtype~~, @type)

Member of model.quoteLike

Contained by

core: head hi l note p q quote sp stage title

drama: castList set

header: change licence

textstructure: body div

May contain

analysis: s w

core: date gap hi l lg name note p q quote sp stage title

drama: castList

header: idno

namesdates: forename surname

Declaration

```
element quote
{
   att.global.attribute.xmlid,
   att.global.attribute.n,
   att.global.attribute.xmllang,
   att.global.analytic.attribute.ana,
   att.typed.attribute.type,
   macro.specialPara}
```

Example

```
Lexicography has shown little sign of being affected by the
work of followers of J.R. Firth, probably best summarized in his
slogan, <quote>You shall know a word by the company it
keeps</quote>
<ref>(Firth, 1957)</ref>
```

Note If a bibliographic citation is supplied for the source of a quotation, the two may be grouped using the **<cit>** element.

<revisionDesc> (revision description) summarizes the revision history for a file. [2.5. 2.1.1.]

Module header

Attributes Attributes att.global (*@xml:id*, *@n*, *@xml:lang*, *@rend*, *@xml:space*) (att.global.analytic (*@ana*)) (att.global.facs (*@facs*)) att.docStatus (*@status*)

Contained by

header: teiHeader

May contain

header: change

Declaration

```
element revisionDesc
{
   att.global.attributes,
   att.docStatus.attributes,
   ( list | listChange | change+ )
}
```

Example

```
<revisionDesc status="embargoed">
 <change when="1991-11-11" who="#LB"> deleted chapter 10 </change>
</revisionDesc>
```

Note If present on this element, the *status* attribute should indicate the current status of the document. The same attribute may appear on any <change> to record the status at the time of that change. Conventionally change elements should be given in reverse date order, with the most recent change at the start of the list.

<role> contains the name of a dramatic role, as given in a cast list. [7.1.4.]

Module drama

Attributes Attributes att.global (*@xml:id*, *@n*, *@xml:lang*, *@rend*, *@xml:space*) (att.global.analytic (*@ana*)) (att.global.facs (*@facs*))

Member of model.castItemPart

Contained by

drama: castItem

May contain

analysis: s w

core: date gap hi name note title

header: idno

namesdates: forename surname

Declaration

```
element role { att.global.attributes, macro.phraseSeq }
```

Example

```
<role xml:id="jt">Joan Trash</role>
<roleDesc>A Ginger-bread-woman</roleDesc>
```

Note It is important to assign a meaningful ID attribute to the <role> element, since this ID is referred to by *who* attributes on many other elements.

<roleDesc> (role description) describes a character's role in a drama. [7.1.4.]

Module drama

Attributes Attributes att.global (*@xml:id*, *@n*, *@xml:lang*, *@rend*, *@xml:space*) (att.global.analytic (*@ana*)) (att.global.facs (*@facs*))

Member of model.castItemPart

Contained by

drama: castGroup castItem

May contain

analysis: s w

core: date gap hi name note title

header: idno

namesdates: forename surname

Declaration

```
element roleDesc { att.global.attributes, macro.phraseSeq }
```

Example

```
<roleDesc>gentlemen of leisure</roleDesc>
```

<s> (s-unit) contains a sentence-like division of a text. [17.1. 8.4.1.]

Module analysis

Attributes Attributesatt.global (~~xml:space~~, @xml:id, @n, @xml:lang, @rend) att.global.analytic (@ana) att.typed (~~subtype~~, @type)

Member of model.segLike

Contained by

core: author date head hi l name note p publisher q quote speaker stage title

drama: castItem role roleDesc

header: change licence

namesdates: forename surname

May contain

analysis: w

core: date gap hi name note title

header: idno

namesdates: forename surname

Declaration

```
element s
{
   att.global.attribute.xmlid,
   att.global.attribute.n,
   att.global.attribute.xmllang,
   att.global.attribute.rend,
   att.global.analytic.attribute.ana,
   att.typed.attribute.type,
   (
      text
    | model.gLike    | model.global    | binaryObject    | formula    | graphic    | medi
}
```

Schematron <s:report test="tei:s">You may not nest one s element within another: use seg instead</s:report>

Example

```
<head>
 <s>A short affair</s>
</head>
<s>When are you leaving?</s>
<s>Tomorrow.</s>
```

Note The <s> element may be used to mark orthographic sentences, or any other segmentation of a text, provided that the segmentation is end-to-end, complete, and non-nesting. For segmentation which is partial or recursive, the `<seg>` should be used instead.The *type* attribute may be used to indicate the type of segmentation intended, according to any convenient typology.

<set>

<set> (setting) contains a description of the setting, time, locale, appearance, etc., of the action of a play, typically found in the front matter of a printed performance text (not a stage direction). [7.1.]

Module drama

Attributes Attributes att.global (*@xml:id*, *@n*, *@xml:lang*, *@rend*, *@xml:space*) (att.global.analytic (*@ana*)) (att.global.facs (*@facs*))

Member of model.frontPart.drama

Contained by

textstructure: back front

May contain

core: gap head l lg note p q quote sp stage

drama: castList

Declaration

```
element set
{
   att.global.attributes,
   ( ( model.headLike | model.global )*, ( ( model.common ), model.global* )* )
}
```

Example

```
<set>
 <p>The action takes place on February 7th between the hours of noon and six in the
   afternoon, close to the Trenartha Tin Plate Works, on the borders of England and
Wales,
   where a strike has been in progress throughout the winter.</p>
</set>
```

Example

```
<set>
 <head>SCENE</head>
 <p>A Sub-Post Office on a late autumn evening</p>
</set>
```

Example

```
<front>
<!-- <titlePage>, <div type="Dedication">, etc. -->
 <set>
  <list type="gloss">
   <label>TIME</label>
```

```
    <item>1907</item>
    <label>PLACE</label>
    <item>East Coast village in England</item>
   </list>
  </set>
</front>
```

Note Contains paragraphs or phrase level tags.This element should not be used outside the front matter; for similar contextual descriptions within the body of the text, use the <stage> element.

<sourceDesc>

<sourceDesc> (source description) describes the source from which an electronic text was derived or generated, typically a bibliographic description in the case of a digitized text, or a phrase such as "born digital" for a text which has no previous existence. [2.2.7.]

Module header

Attributes Attributes att.global (*@xml:id*, *@n*, *@xml:lang*, *@rend*, *@xml:space*) (att.global.analytic (*@ana*)) (att.global.facs (*@facs*)) att.declarable (*@default*)

Contained by

header: fileDesc

May contain

core: p

Declaration

```
element sourceDesc
{
   att.global.attributes,
   att.declarable.attributes,
   (
      model.pLike+
    | ( model.biblLike | model.sourceDescPart | model.listLike )+
   )
}
```

Example

```
<sourceDesc>
 <bibl>
  <title level="a">The Interesting story of the Children in the Wood</title>. In
 <author>Victor E Neuberg</author>, <title>The Penny Histories</title>.
 <publisher>OUP</publisher>
  <date>1968</date>. </bibl>
</sourceDesc>
```

Example

```
<sourceDesc>
 <p>Born digital: no previous source exists.</p>
</sourceDesc>
```

<sp>

<sp> (speech) contains an individual speech in a performance text, or a passage presented as such in a prose or verse text. [3.12.2. 3.12. 7.2.2.]

Module core

Attributes Attributes att.ascribed (*@who*) att.global (~~rend~~, ~~xml:space~~, @xml:id, @n, @xml:lang) att.global.analytic (@ana)

Member of model.divPart
Contained by
core: note q quote stage
drama: castList set
header: change licence
textstructure: body div
May contain
core: gap l lg note p q quote speaker stage
Declaration

```
element sp
{
   att.global.attribute.xmlid,
   att.global.attribute.n,
   att.global.attribute.xmllang,
   att.global.analytic.attribute.ana,
   att.ascribed.attributes,
   (
      model.global*,
      ( speaker, model.global* )?,
      (
         (
            lg          | model.lLike          | model.pLike          | model.listLike
         model.global*
      )+
   )
}
```

Example

```
<sp>
 <speaker>The reverend Doctor Opimian</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>
<sp>
 <speaker>The Reverend Doctor Opimian</speaker>
 <p>On the contrary, sir, I think there is much to be said for him. In the first
place....</p>
 <p>Fish, Miss Gryll — I could discourse to you on fish by the hour: but for the
present I
   will forbear...</p>
</sp>
```

Note The *who* attribute on this element may be used either in addition to the <speaker> element or as an alternative.

Note Lines or paragraphs, stage directions, and phrase-level elements.The *who* attribute on this element may be used either in addition to the <speaker> element or as an alternative.

<speaker> contains a specialized form of heading or label, giving the name of one or more speakers in a dramatic text or fragment. [3.12.2.]

Module core

Attributes Attributesatt.global (~~rend~~, ~~xml:space~~, @xml:id, @n, @xml:lang) att.global.analytic (@ana)

Contained by

core: sp

May contain

analysis: s w

core: date gap hi name note title

header: idno

namesdates: forename surname

Declaration

```
element speaker
{
   att.global.attribute.xmlid,
   att.global.attribute.n,
   att.global.attribute.xmllang,
   att.global.analytic.attribute.ana,
   macro.phraseSeq}
```

Example

```
<sp who="#ni #rsa">
 <speaker>Nancy and Robert</speaker>
 <stage type="delivery">(speaking simultaneously)</stage>
 <p>The future? ...</p>
</sp>
<list type="speakers">
 <item xml:id="ni"/>
 <item xml:id="rsa"/>
</list>
```

<stage>

(stage direction) contains any kind of stage direction within a dramatic text or fragment. [3.12.2. 3.12. 7.2.4.]

Module core

Attributes Attributes att.ascribed (*@who*) att.placement (*@place*) att.global (~~rend~~, ~~xml:space~~, @xml:id, @n, @xml:lang) att.global.analytic (@ana)

`@type` indicates the kind of stage direction.

Status Recommended

Datatype `data.enumerated`

Suggested values include: **setting** describes a setting.

entrance describes an entrance.

exit describes an exit.

business describes stage business.

novelistic is a narrative, motivating stage direction.

delivery describes how a character speaks.

modifier gives some detail about a character.

location describes a location.

mixed more than one of the above

Member of model.stageLike

Contained by

core: head hi l lg note p q quote sp stage title

drama: castList set
header: change licence
textstructure: body div
May contain
analysis: s w
core: date gap hi l lg name note p q quote sp stage title
drama: castList
header: idno
namesdates: forename surname
Declaration

```
element stage
{
   att.ascribed.attributes,
   att.global.attribute.xmlid,
   att.global.attribute.n,
   att.global.attribute.xmllang,
   att.global.analytic.attribute.ana,
   att.placement.attributes,
   attribute type
   {
      "setting"
    | "entrance"
    | "exit"
    | "business"
    | "novelistic"
    | "delivery"
    | "modifier"
    | "location"
    | "mixed"
    | xsd:Name
   }?,
   macro.specialPara}
```

Example

```
<stage type="setting">A curtain being drawn.</stage>
<stage type="setting">Music</stage>
<stage type="entrance">Enter Husband as being thrown off his horse.</stage>
<stage type="exit">Exit pursued by a bear.</stage>
<stage type="business">He quickly takes the stone out.</stage>
<stage type="delivery">To Lussurioso.</stage>
<stage type="novelistic">Having had enough, and embarrassed for the family.</stage>
<stage type="modifier">Disguised as Ansaldo.</stage>
<stage type="location">At a window.</stage>
<stage rend="inline" type="delivery">Aside.</stage>
```

Example

```
<l>Behold. <stage n="*" place="margin">Here the vp<lb/>per part of the
<hi>Scene</hi> open'd; when
   straight appear'd a Heauen, and all the <hi>Pure Artes</hi> sitting on
   two semi<lb/>circular ben<lb/>ches, one a<lb/>boue another: who sate thus till
the rest of the
 <hi>Prologue</hi> was spoken, which being ended, they descended in
   order within the <hi>Scene,</hi> whiles the Musicke plaid</stage> Our
Poet knowing our free hearts</l>
```

Note The *who* attribute may be used to indicate more precisely the person or persons participating in the action described by the stage direction.

<surname>

<surname> contains a family (inherited) name, as opposed to a given, baptismal, or nick name. [13.2.1.]

Module namesdates

Attributes Attributes att.global (*@xml:id*, *@n*, *@xml:lang*, *@rend*, *@xml:space*) (att.global.analytic (*@ana*)) (att.global.facs (*@facs*)) att.personal (*@full*, *@sort*) (att.naming (*@role*, *@nymRef*) (att.canonical (*@key*, *@ref*))) att.typed (*@type*, *@subtype*)

Member of model.persNamePart

Contained by

analysis: s

core: author date head hi l name note p publisher q quote speaker stage title

drama: castItem role roleDesc

header: change licence principal

namesdates: forename surname

May contain

analysis: s w

core: date gap hi name note title

header: idno

namesdates: forename surname

Declaration

```
element surname
{
   att.global.attributes,
   att.personal.attributes,
   att.typed.attributes,
   macro.phraseSeq}
```

Example

```
<surname type="combine">St John Stevas</surname>
```

<teiHeader>

<teiHeader> (TEI header) supplies the descriptive and declarative information making up an electronic title page for every TEI-conformant document. [2.1.1. 15.1.]

Module header

Attributes Attributes att.global (*@xml:id*, *@n*, *@xml:lang*, *@rend*, *@xml:space*) (att.global.analytic (*@ana*)) (att.global.facs (*@facs*))

`@type` specifies the kind of document to which the header is attached, for example whether it is a corpus or individual text.

Status Optional

Datatype `data.enumerated`

Sample values include: **text** the header is attached to a single text.[Default]

corpus the header is attached to a corpus.

Contained by *tei* structure: TEI

May contain

header: fileDesc revisionDesc

Declaration

```
element teiHeader
{
   att.global.attributes,
   attribute type { data.enumerated }?,
   ( fileDesc, model.teiHeaderPart*, revisionDesc? )
}
```

Example

```
<teiHeader>
 <fileDesc>
  <titleStmt>
   <title>Shakespeare: the first folio (1623) in electronic form</title>
   <author>Shakespeare, William (1564–1616)</author>
   <respStmt>
    <resp>Originally prepared by</resp>
    <name>Trevor Howard-Hill</name>
   </respStmt>
   <respStmt>
    <resp>Revised and edited by</resp>
    <name>Christine Avern-Carr</name>
   </respStmt>
  </titleStmt>
  <publicationStmt>
   <distributor>Oxford Text Archive</distributor>
   <address>
    <addrLine>13 Banbury Road, Oxford OX2 6NN, UK</addrLine>
   </address>
   <idno type="OTA">119</idno>
   <availability>
    <p>Freely available on a non-commercial basis.</p>
   </availability>
   <date when="1968">1968</date>
  </publicationStmt>
  <sourceDesc>
   <bibl>The first folio of Shakespeare, prepared by Charlton Hinman (The Norton
Facsimile,
       1968)</bibl>
  </sourceDesc>
 </fileDesc>
 <encodingDesc>
  <projectDesc>
   <p>Originally prepared for use in the production of a series of old-spelling
       concordances in 1968, this text was extensively checked and revised for use
during the
       editing of the new Oxford Shakespeare (Wells and Taylor, 1989).</p>
  </projectDesc>
  <editorialDecl>
   <correction>
    <p>Turned letters are silently corrected.</p>
   </correction>
   <normalization>
    <p>Original spelling and typography is retained, except that long s and
ligatured
         forms are not encoded.</p>
   </normalization>
  </editorialDecl>
  <refsDecl xml:id="ASLREF">
   <cRefPattern matchPattern="(\S+) ([^.]+)\.(.*)"
    replacementPattern="#xpath(//div1[@n='$1']/div2/[@n='$2']//lb[@n='$3'])">
```

```
    <p>A reference is created by assembling the following, in the reverse order as
that
          listed here: <list>
      <item>the <att>n</att> value of the preceding <gi>lb</gi>
      </item>
      <item>a period</item>
      <item>the <att>n</att> value of the ancestor <gi>div2</gi>
      </item>
      <item>a space</item>
      <item>the <att>n</att> value of the parent <gi>div1</gi>
      </item>
     </list>
    </p>
   </cRefPattern>
  </refsDecl>
 </encodingDesc>
 <revisionDesc>
  <list>
   <item>
    <date when="1989-04-12">12 Apr 89</date> Last checked by CAC</item>
   <item>
    <date when="1989-03-01">1 Mar 89</date> LB made new file</item>
  </list>
 </revisionDesc>
</teiHeader>
```

Note One of the few elements unconditionally required in any TEI document.

<text>

<text> contains a single text of any kind, whether unitary or composite, for example a poem or drama, a collection of essays, a novel, a dictionary, or a corpus sample. [4. 15.1.]

Module textstructure

Attributes Attributesatt.global (~~rend~~, ~~xml:space~~, @xml:id, @n, @xml:lang)
att.global.analytic (@ana)

Contained by

textstructure: TEI

May contain

core: gap note

textstructure: back body front

Declaration

```
element text
{
   att.global.attribute.xmlid,
   att.global.attribute.n,
   att.global.attribute.xmllang,
   att.global.analytic.attribute.ana,
   (
      model.global*,
      ( front, model.global* )?,
      ( body | group ),
      model.global*,
      ( back, model.global* )?
   )
}
```

Example

```
<text>
 <front>
  <docTitle>
   <titlePart>Autumn Haze</titlePart>
  </docTitle>
 </front>
 <body>
  <l>Is it a dragonfly or a maple leaf</l>
  <l>That settles softly down upon the water?</l>
 </body>
</text>
```

Example The body of a text may be replaced by a group of nested texts, as in the following schematic:

```
<text>
 <front>
<!-- front matter for the whole group -->
 </front>
 <group>
  <text>
<!-- first text -->
  </text>
  <text>
<!-- second text -->
  </text>
 </group>
</text>
```

Note This element should not be used to represent a text which is inserted at an arbitrary point within the structure of another, for example as in an embedded or quoted narrative; the `<floatingText>` is provided for this purpose.

<title> contains a title for any kind of work. [3.11.2.2. 2.2.1. 2.2.5.]

Module core

Attributes Attributesatt.global (~~rend~~, ~~xml:space~~, @xml:id, @n, @xml:lang)
att.global.analytic (@ana)

`@type` classifies the title according to some convenient typology.

Derived from att.typed

Status Optional

Datatype `data.enumerated`

Sample values include: **main** main title

sub (subordinate) subtitle, title of part

alt (alternate) alternate title, often in another language, by which the work is also known

short abbreviated form of title

desc (descriptive) descriptive paraphrase of the work functioning as a title

Note This attribute is provided for convenience in analysing titles and processing them according to their type; where such specialized processing is not necessary, there is no need for such analysis, and the entire title, including subtitles and any parallel titles, may be enclosed within a single <title> element.

Member of model.emphLike

Contained by

analysis: s

core: author date head hi l name note p publisher q quote speaker stage title

drama: castItem role roleDesc

header: change licence principal titleStmt

namesdates: forename surname

May contain

analysis: s w

core: date gap hi lg name note q quote stage title

drama: castList

header: idno

namesdates: forename surname

Declaration

```
element title
{
   att.global.attribute.xmlid,
   att.global.attribute.n,
   att.global.attribute.xmllang,
   att.global.analytic.attribute.ana,
   attribute type { data.enumerated }?,
   macro.paraContent}
```

Example

```
<title>Information Technology and the Research Process: Proceedings of
a conference held at Cranfield Institute of Technology, UK,
18–21 July 1989</title>
```

Example

```
<title>Hardy's Tess of the D'Urbervilles: a machine readable
edition</title>
```

Example

```
<title type="full">
 <title type="main">Synthèse</title>
 <title type="sub">an international journal for
   epistemology, methodology and history of
   science</title>
</title>
```

Note The attributes *key* and *ref*, inherited from the class att.canonical may be used to indicate the canonical form for the title; the former, by supplying (for example) the identifier of a record in some external library system; the latter by pointing to an XML element somewhere containing the canonical form of the title.

<titleStmt>

(title statement) groups information about the title of a work and those responsible for its content. [2.2.1. 2.2.]

Module header

Attributes Attributes att.global (*@xml:id*, *@n*, *@xml:lang*, *@rend*, *@xml:space*) (att.global.analytic (*@ana*)) (att.global.facs (*@facs*))

Contained by

header: fileDesc

May contain

core: author title

header: principal

Declaration

```
element titleStmt { att.global.attributes, ( title+, model.respLike* ) }
```

Example

```
<titleStmt>
 <title>Capgrave's Life of St. John Norbert: a machine-readable
transcription</title>
 <respStmt>
  <resp>compiled by</resp>
  <name>P.J. Lucas</name>
 </respStmt>
</titleStmt>
```

<w> (word) represents a grammatical (not necessarily orthographic) word. [17.1.]

Module analysis

Attributes Attributes att.global (*@xml:id*, *@n*, *@xml:lang*, *@rend*, *@xml:space*) (att.global.analytic (*@ana*)) (att.global.facs (*@facs*)) att.segLike (*@function*) (att.datcat (*@datcat*, *@valueDatcat*)) (att.fragmentable (*@part*)) att.typed (*@type*, *@subtype*)

@lemma provides a lemma for the word, such as an uninflected dictionary entry form.

Status Optional

Datatype `data.text`

@lemmaRef provides a pointer to a definition of the lemma for the word, for example in an online lexicon.

Status Optional

Datatype `data.pointer`

Member of model.segLike

Contained by

analysis: s w

core: author date head hi l name note p publisher q quote speaker stage title

drama: castItem role roleDesc

header: change licence

namesdates: forename surname

May contain

analysis: w

core: gap hi note

Declaration

```
element w
{
   att.global.attributes,
   att.segLike.attributes,
   att.typed.attributes,
   attribute lemma { data.text }?,
   attribute lemmaRef { data.pointer }?,
   (
      text
```

```
    | model.gLike    | seg    | w    | m    | c    | pc    | model.global    | model.lPar
}
```

Example

```
<w lemma="hit"
 lemmaRef="http://www.example.com/lexicon/hitvb.xml" type="verb">hitt<m type="suffix">ing</m>
</w>
```

A.2 Model classes

---

model.availabilityPart groups elements such as licences and paragraphs of text which may appear as part of an availability statement [2.2.4.]

Module tei

Used by availability

Members licence

---

model.castItemPart groups component elements of an entry in a cast list, such as dramatic role or actor's name.

Module tei

Used by castItem

Members role roleDesc

---

model.common groups common chunk- and inter-level elements. [1.3.]

Module tei

Used by body castList div set

Members model.divPart[model.lLike[l] model.pLike[p] lg sp] model.inter[model.biblLike model.egLike model.labelLike model.listLike model.oddDecl model.qLike[model.quoteLike[quote] q] model.stageLike[stage] castList]

Note This class defines the set of chunk- and inter-level elements; it is used in many content models, including those for textual divisions.

---

model.dateLike groups elements containing temporal expressions. [3.5.4. 13.3.6.]

Module tei

Used by model.pPart.data

Members date

---

model.divBottom groups elements appearing at the end of a text division. [4.2.]

Module tei

Used by body div front lg

Members model.divBottomPart model.divWrapper

---

model.divLike groups elements used to represent un-numbered generic structural divisions.

Module tei

Used by back body div front

Members div

---

model.divPart groups paragraph-level elements appearing directly within divisions. [1.3.]

Module tei

Used by macro.specialPara model.common

Members model.lLike[l] model.pLike[p] lg sp

Note Note that this element class does not include members of the model.inter class, which can appear either within or between paragraph-level items.

---

model.divTop groups elements appearing at the beginning of a text division. [4.2.]

Module tei

Used by body castList div lg

Members model.divTopPart[model.headLike[head]] model.divWrapper

---

model.divTopPart groups elements which can occur only at the beginning of a text division. [4.6.]

Module tei

Used by model.divTop

Members model.headLike[head]

---

model.emphLike groups phrase-level elements which are typographically distinct and to which a specific function can be attributed. [3.3.]

Module tei

Used by model.highlighted model.limitedPhrase

Members title

---

model.frontPart groups elements which appear at the level of divisions within front or back matter. [7.1.]

Module tei

Used by back front

Members model.frontPart.drama[castList set]

---

model.frontPart.drama groups elements which appear at the level of divisions within front or back matter of performance texts only. [7.1.]

Module tei

Used by model.frontPart

Members castList set

---

model.global groups elements which may appear at any point within a TEI text. [1.3.]

Module tei

Used by back body castGroup castItem castList date div front head lg macro.paraContent macro.phraseSeq macro.phraseSeq.limited macro.specialPara s set sp text w

Members model.global.edit[gap] model.global.meta model.milestoneLike model.noteLike[note]

---

model.global.edit groups globally available elements which perform a specifically editorial function. [1.3.]

Module tei

Used by model.global

Members gap

---

model.headLike groups elements used to provide a title or heading at the start of a text division.

Module tei

Used by castGroup model.divTopPart set

Members head

---

model.hiLike groups phrase-level elements which are typographically distinct but to which no specific function can be attributed. [3.3.]

Module tei

Used by model.highlighted model.limitedPhrase w

Members hi

---

model.highlighted groups phrase-level elements which are typographically distinct. [3.3.]

Module tei

Used by model.phrase

Members model.emphLike[title] model.hiLike[hi]

---

model.inter groups elements which can appear either within or between paragraph-like elements. [1.3.]

Module tei

Used by head macro.paraContent macro.specialPara model.common

Members model.biblLike model.egLike model.labelLike model.listLike model.oddDecl model.qLike[model.quoteLike[quote] q] model.stageLike[stage] castList

---

model.lLike groups elements representing metrical components such as verse lines.

Module tei

Used by head lg model.divPart sp

Members l

---

model.limitedPhrase groups phrase-level elements excluding those elements primarily intended for transcription of existing sources. [1.3.]

Module tei

Used by macro.phraseSeq.limited

Members model.emphLike[title] model.hiLike[hi] model.pPart.data[model.addressLike model.dateLike[date] model.measureLike model.nameLike[model.nameLike.agent[name] model.offsetLike model.persNamePart[forename surname] model.placeStateLike[model.placeNamePart] idno]] model.pPart.editorial model.pPart.msdesc model.phrase.xml model.ptrLike

---

model.nameLike groups elements which name or refer to a person, place, or organization.

Module tei

Used by model.pPart.data

Members model.nameLike.agent[name] model.offsetLike model.persNamePart[forename surname] model.placeStateLike[model.placeNamePart] idno

Note A superset of the naming elements that may appear in datelines, addresses, statements of responsibility, etc.

---

model.nameLike.agent groups elements which contain names of individuals or corporate bodies. [3.5.]

Module tei

Used by model.nameLike

Members name

Note This class is used in the content model of elements which reference names of people or organizations.

---

model.noteLike groups globally-available note-like elements. [3.8.]

Module tei

Used by model.global

Members note

---

model.pLike groups paragraph-like elements.

Module tei

Used by availability back front model.divPart publicationStmt sourceDesc sp
Members p

---

model.pLike.front groups paragraph-like elements which can occur as direct constituents of front matter. [4.6.]
Module tei
Used by back front
Members head

---

model.pPart.data groups phrase-level elements containing names, dates, numbers, measures, and similar data. [3.5.]
Module tei
Used by model.limitedPhrase model.phrase
Members model.addressLike model.dateLike[date] model.measureLike model.nameLike[model.nameLike.agent[name] model.offsetLike model.persNamePart[forename surname] model.placeStateLike[model.placeNamePart] idno]

---

model.pPart.edit groups phrase-level elements for simple editorial correction and transcription. [3.4.]
Module tei
Used by model.phrase w
Members model.pPart.editorial model.pPart.transcriptional

---

model.persNamePart groups elements which form part of a personal name. [13.2.1.]
Module namesdates
Used by model.nameLike
Members forename surname

---

model.phrase groups elements which can occur at the level of individual words or phrases. [1.3.]
Module tei
Used by castItem date head macro.paraContent macro.phraseSeq macro.specialPara
Members model.graphicLike model.highlighted[model.emphLike[title] model.hiLike[hi]] model.lPart model.pPart.data[model.addressLike model.dateLike[date] model.measureLike model.nameLike[model.nameLike.agent[name] model.offsetLike model.persNamePart[forename surname] model.placeStateLike[model.placeNamePart] idno]] model.pPart.edit[model.pPart.editorial model.pPart.transcriptional] model.pPart.msdesc model.phrase.xml model.ptrLike model.segLike[s w] model.specDescLike

Note This class of elements can occur only within larger elements of the class *inter* or *chunk*. In prose, this means these elements can occur within paragraphs, list items, lines of verse, etc.

---

model.placeStateLike groups elements which describe changing states of a place.

Module tei

Used by model.nameLike

Members model.placeNamePart

---

model.publicationStmtPart.agency groups the child elements of a <publicationStmt> element of the TEI header that indicate an authorising agent. [2.2.4.]

Module tei

Used by publicationStmt

Members publisher

Note The 'agency' child elements, while not required, are required if one of the 'detail' child elements is to be used. It is not valid to have a 'detail' child element without a preceding 'agency' child element.See also model.publicationStmtPart.detail.

---

model.publicationStmtPart.detail groups the agency-specific child elements of the <publicationStmt> element of the TEI header. [2.2.4.]

Module tei

Used by publicationStmt

Members availability date idno

Note A 'detail' child element may not occur unless an 'agency' child element precedes it.See also model.publicationStmtPart.agency.

---

model.qLike groups elements related to highlighting which can appear either within or between chunk-level elements. [3.3.]

Module tei

Used by model.inter sp

Members model.quoteLike[quote] q

---

model.quoteLike groups elements used to directly contain quotations.

Module tei

Used by model.qLike

Members quote

---

model.respLike groups elements which are used to indicate intellectual or other significant responsibility, for example within a bibliographic element.

Module tei

Used by titleStmt

Members author principal

---

model.segLike groups elements used for arbitrary segmentation. [16.3. 17.1.]

Module tei

Used by model.phrase

Members s w

Note The principles on which segmentation is carried out, and any special codes or attribute values used, should be defined explicitly in the `<segmentation>` element of the `<encodingDesc>` within the associated TEI header.

---

model.stageLike groups elements containing stage directions or similar things defined by the module for performance texts. [7.3.]

Module tei

Used by lg model.inter sp

Members stage

Note Stage directions are members of class *inter*: that is, they can appear between or within component-level elements.

A.3 Attribute classes

---

att.ascribed provides attributes for elements representing speech or action that can be ascribed to a specific individual. [3.3.3. 8.3.]

Module tei

Members change q sp stage

Attributes Attributes

`@who` indicates the person, or group of people, to whom the element content is ascribed.

Status Optional

Datatype 1– occurrences of `data.pointer` separated by whitespace

In the following example from Hamlet, speeches (<sp>) in the body of the play are linked to <castItem> elements in the <castList> using the *who* attribute.

```
<castItem type="role">
 <role xml:id="Barnardo">Bernardo</role>
</castItem>
<castItem type="role">
 <role xml:id="Francisco">Francisco</role>
 <roleDesc>a soldier</roleDesc>
</castItem>
<!-- ... -->
<sp who="#Barnardo">
 <speaker>Bernardo</speaker>
 <l n="1">Who's there?</l>
</sp>
<sp who="#Francisco">
 <speaker>Francisco</speaker>
```

```
    <l n="2">Nay, answer me: stand, and unfold yourself.</l>
  </sp>
```

Note For transcribed speech, this will typically identify a participant or participant group; in other contexts, it will point to any identified <person> element.

---

att.canonical provides attributes which can be used to associate a representation such as a name or title with canonical information about the object being named or referenced.

Module tei

Members att.naming[att.personal[forename surname]] principal

Attributes Attributes

@key provides an externally-defined means of identifying the entity (or entities) being named, using a coded value of some kind.

Status Optional

Datatype data.text

```
<author>
 <name key="name 427308"
  type="organisation">[New Zealand Parliament, Legislative
Council]</name>
</author>
<author>
 <name key="Hugo, Victor (1802-1885)"
  ref="http://www.idref.fr/026927608">Victor Hugo</name>
</author>
```

Note The value may be a unique identifier from a database, or any other externally-defined string identifying the referent. No particular syntax is proposed for the values of the *key* attribute, since its form will depend entirely on practice within a given project. For the same reason, this attribute is not recommended in data interchange, since there is no way of ensuring that the values used by one project are distinct from those used by another. In such a situation, a preferable approach for magic tokens which follows standard practice on the Web is to use a *ref* attribute whose value is a tag URI as defined in RFC 4151.

@ref (reference) provides an explicit means of locating a full definition for the entity being named by means of one or more URIs.

Status Optional

Datatype 1– occurrences of data.pointer separated by whitespace

```
<name ref="http://viaf.org/viaf/109557338"
 type="person">Seamus Heaney</name>
```

Note The value must point directly to one or more XML elements or other resources by means of one or more URIs, separated by whitespace. If more than one is supplied the implication is that the name identifies several distinct entities.

---

att.datable.custom provides attributes for normalization of elements that contain datable events to a custom dating system (i.e. other than the Gregorian used by W3 and ISO). [13.3.6.]

Module namesdates

Members att.datable

Attributes Attributes

@when-custom supplies the value of a date or time in some standard form.

Status Optional

Datatype 1– occurrences of data.word separated by whitespace

The following are examples of custom date or time formats that are *not* valid ISO or W3C format normalizations, normalized to a different dating system

```
<p>Alhazen died in Cairo on the
<date when="1040-03-06"
  when-custom="431-06-12"> 12th day of Jumada t-Tania, 430
AH
 </date>.</p>
<p>The current world will end at the
<date when="2012-12-21"
  when-custom="13.0.0.0.0">end of B'ak'tun 13</date>.</p>
<p>The Battle of Meggidu
(<date when-custom="Thutmose_III:23">23rd year of reign of
Thutmose III</date>).</p>
<p>Esidorus bixit in pace annos LXX plus minus sub
<date when-custom="Ind:4-10-11">die XI mensis Octobris
indictione IIII</date>
</p>Not all custom date formulations will have Gregorian
equivalents.The when-custom attribute and other custom
dating are not contrained to a datatype by the TEI, but
individual projects are recommended to regularize and
document their dating formats.
```

@notBefore-custom specifies the earliest possible date for the event in some custom standard form.

Status Optional

Datatype 1– occurrences of data.word separated by whitespace

@notAfter-custom specifies the latest possible date for the event in some custom standard form.

Status Optional

Datatype 1– occurrences of data.word separated by whitespace

@from-custom indicates the starting point of the period in some standard form.

Status Optional

Datatype 1– occurrences of data.word separated by whitespace

```
<event datingMethod="#julian"
 from-custom="1666-09-02"
 to-custom="1666-09-05"
 xml:id="FIRE1">
 <head>The Great Fire of London</head>
 <p>The Great Fire of London burned through a large part
   of the city of London.</p>
</event>
```

@to-custom indicates the ending point of the period in some standard form.

Status Optional

Datatype 1– occurrences of data.word separated by whitespace

@datingPoint supplies a pointer to some location defining a named point in time with reference to which the datable item is understood to have occurred

Status Optional

Datatype data.pointer

@datingMethod supplies a pointer to a <calendar> element or other means of interpreting the values of the custom dating attributes.

Status Optional

Datatype data.pointer

```
Contayning the Originall, Antiquity, Increaſe, Moderne
eſtate, and deſcription of that Citie, written in the yeare
<date calendar="#julian"
 datingMethod="#julian"
 when-custom="1598">1598</date>. by Iohn Stow
Citizen of London.
```

Note Here the *calendar* attribute points to a <calendar> element for the Julian calendar, specifying that the text content of the <date> element is a Julian date, and the *datingMethod* attribute also points to the Julian calendar to indicate that the content of the *when-custom* attribute value is Julian too.

---

att.datable.iso provides attributes for normalization of elements that contain datable events using the ISO 8601 standard. [3.5.4. 13.3.6.]

Module namesdates

Members att.datable

Attributes Attributes

@when-iso supplies the value of a date or time in a standard form.

Status Optional

Datatype data.temporal.iso

The following are examples of ISO date, time, and date & time formats that are *not* valid W3C format normalizations.

```
<date when-iso="1996-09-24T07:25+00">Sept. 24th, 1996 at
3:25 in the morning</date>
<date when-iso="1996-09-24T03:25-04">Sept. 24th, 1996 at
3:25 in the morning</date>
<time when-iso="1999-01-04T20:42-05">4 Jan 1999 at 8:42
pm</time>
<time when-iso="1999-W01-1T20,70-05">4 Jan 1999 at 8:42
pm</time>
<date when-iso="2006-05-18T10:03">a few minutes after ten in
the morning on Thu 18 May</date>
<time when-iso="03:00">3 A.M.</time>
<time when-iso="14">around two</time>
<time when-iso="15,5">half past three</time>All of the
examples of the when attribute in the att.datable.w3c class are
also valid with respect to this attribute.
He likes to be punctual. I said <q>
 <time when-iso="12">around noon</time>
</q>, and he showed up at <time when-iso="12:00:00">12
O'clock</time> on the dot.The second occurence of <time>
could have been encoded with the when attribute, as 12:00:00
is a valid time with respect to the W3C XML Schema Part 2:
Datatypes Second Edition specification. The first occurence
could not.
```

@notBefore-iso specifies the earliest possible date for the event in standard form, e.g. yyyy-mm-dd.

Status Optional

Datatype data.temporal.iso

@notAfter-iso specifies the latest possible date for the event in standard form, e.g. yyyy-mm-dd.

Status Optional

Datatype data.temporal.iso

@from-iso indicates the starting point of the period in standard form.

Status Optional

Datatype data.temporal.iso

@to-iso indicates the ending point of the period in standard form.

Status Optional

Datatype data.temporal.iso

Note The value of these attributes should be a normalized representation of the date, time, or combined date & time intended, in any of the standard formats specified by ISO 8601, using the Gregorian calendar.

Note If both *when-iso* and *dur-iso* are specified, the values should be interpreted as indicating a span of time by its starting time (or date) and duration. That is,

```
<date dur-iso="P8D" when-iso="2007-06-01"/>
```

indicates the same time period as

```
<date when-iso="2007-06-01/P8D"/>
```

In providing a 'regularized' form, no claim is made that the form in the source text is incorrect; the regularized form is simply that chosen as the main form for purposes of unifying variant forms under a single heading.

---

att.datable.w3c provides attributes for normalization of elements that contain datable events conforming to the W3C *XML Schema Part 2: Datatypes Second Edition*.

Module tei

Members att.datable

Attributes Attributes

@when supplies the value of the date or time in a standard form, e.g. yyyy-mm-dd.

Status Optional

Datatype data.temporal.w3c

Examples of W3C date, time, and date & time formats.

```
<p>
 <date when="1945-10-24">24 Oct 45</date>
 <date when="1996-09-24T07:25:00Z">September 24th, 1996 at
3:25 in the morning</date>
 <time when="1999-01-04T20:42:00-05:00">Jan 4 1999 at 8
pm</time>
 <time when="14:12:38">fourteen twelve and 38 seconds</time>
 <date when="1962-10">October of 1962</date>
 <date when="--06-12">June 12th</date>
 <date when="---01">the first of the month</date>
 <date when="--08">August</date>
 <date when="2006">MMVI</date>
 <date when="0056">AD 56</date>
 <date when="-0056">56 BC</date>
</p>
This list begins in
the year 1632, more precisely on Trinity Sunday, i.e. the
Sunday after
Pentecost, in that year the <date calendar="#Julian"
 when="1632-06-06">27th of May (old style)</date>.
<opener>
 <dateline>
  <placeName>Dorchester, Village,</placeName>
  <date when="1828-03-02">March 2d. 1828.</date>
 </dateline>
```

```
 <salute>To
   Mrs. Cornell,</salute> Sunday
<time when="12:00:00">noon.</time>
</opener>
```

---

att.datcat provides the *dcr:datacat* and *dcr:ValueDatacat* attributes which are used to align XML elements or attributes with the appropriate Data Categories (DCs) defined by the ISO 12620:2009 standard and stored in the Web repository called ISOCat at http://www.isocat.org/. [9.5.2. 18.3.]

Module tei

Members att.segLike[w]

Attributes Attributes

`@datcat` contains a PID (persistent identifier) that aligns the given element with the appropriate Data Category (or categories) in ISOcat.
Status Optional
Datatype 1– occurrences of `data.pointer` separated by whitespace

`@valueDatcat` contains a PID (persistent identifier) that aligns the content of the given element or the value of the given attribute with the appropriate simple Data Category (or categories) in ISOcat.
Status Optional
Datatype 1– occurrences of `data.pointer` separated by whitespace

Example In this example *dcr:datcat* relates the feature name to the data category "partOfSpeech" and *dcr:valueDatcat* the feature value to the data category "commonNoun". Both these data categories reside in the ISOcat DCR at www.isocat.org, which is the DCR used by ISO TC37 and hosted by its registration authority, the MPI for Psycholinguistics in Nijmegen.

```
<fs
   xmlns:dcr="http://www.isocat.org/ns/dcr">
 <f dcr:datcat="http://www.isocat.org/datcat/DC-1345"
  dcr:valueDatcat="http://www.isocat.org/datcat/DC-
1256" fVal="#commonNoun" name="POS"/>
</fs>
```

Note ISO 12620:2009 is a standard describing the data model and procedures for a Data Category Registry (DCR). Data categories are defined as elementary descriptors in a linguistic structure. In the DCR data model each data category gets assigned a unique Peristent IDentifier (PID), i.e., an URI. Linguistic resources or preferably their schemas that make use of data categories from a DCR should refer to them using this PID. For XML-based resources, like TEI documents, ISO 12620:2009 normative Annex A gives a small Data Category Reference XML vocabulary (also available online at http://www.isocat.org/12620/), which provides two attributes, *dcr:datcat* and *dcr:valueDatcat*.

---

att.declarable provides attributes for those elements in the TEI header which may be independently selected by means of the special purpose *decls* attribute. [15.3.]

Module tei

Members sourceDesc

Attributes Attributes

@default indicates whether or not this element is selected by default when its parent is selected.
Status Optional
Datatype data.truthValue
Legal values are: **true** This element is selected if its parent is selected
false This element can only be selected explicitly, unless it is the only one of its kind, in which case it is selected if its parent is selected.[Default]

Note The rules governing the association of declarable elements with individual parts of a TEI text are fully defined in chapter 15.3.. Only one element of a particular type may have a *default* attribute with a value of true.

---

att.dimensions provides attributes for describing the size of physical objects.

Module tei
Members att.editLike
Attributes Attributes att.ranging (*@atLeast*, *@atMost*, *@min*, *@max*, *@confidence*)

@unit names the unit used for the measurement
Status Optional
Datatype data.enumerated
Suggested values include: **cm** (centimetres)
mm (millimetres)
in (inches)
lines lines of text
chars (characters) characters of text

@quantity specifies the length in the units specified
Status Optional
Datatype data.numeric

@extent indicates the size of the object concerned using a project-specific vocabulary combining quantity and units in a single string of words.
Status Optional
Datatype data.text

```
<gap extent="5 words"/>
<height extent="half the page"/>
```

@precision characterizes the precision of the values specified by the other attributes.
Status Optional
Datatype data.certainty

@scope where the measurement summarizes more than one observation, specifies the applicability of this measurement.
Status Optional
Datatype data.enumerated
Sample values include: **all** measurement applies to all instances.
most measurement applies to most of the instances inspected.
range measurement applies to only the specified range of instances.

---

att.docStatus provides attributes for use on metadata elements describing the status of a document.

Module tei

Members revisionDesc

Attributes Attributes

`@status` describes the status of a document either currently or, when associated with a dated element, at the time indicated.

Status Optional

Datatype `data.enumerated`

Sample values include: **approved**

candidate

cleared

deprecated

draft [Default]

embargoed

expired

frozen

galley

proposed

published

recommendation

submitted

unfinished

withdrawn

Example

```
<revisionDesc status="published">
 <change status="published"
  when="2010-10-21"/>
 <change status="cleared" when="2010-10-02"/>
 <change status="embargoed"
  when="2010-08-02"/>
 <change status="frozen" when="2010-05-01"
  who="#MSM"/>
 <change status="draft" when="2010-03-01"
  who="#LB"/>
</revisionDesc>
```

---

att.fragmentable groups structural elements which may be fragmented, usually as a consequence of some overlapping hierarchy.

Module tei

Members att.divLike att.segLike[w]

Attributes Attributes

`@part` specifies whether or not its parent element is fragmented in some way, typically by some other overlapping structure: for example a speech which is divided between two or more verse stanzas, a paragraph which is split across a page division, a verse line which is divided between two speakers.

Status Optional

Datatype `data.enumerated`

Legal values are: **Y** (yes) the element is fragmented in some (unspecified) respect

N (no) either the element is not fragmented, or no claim is made as to its completeness.[Default]

I (initial) this is the initial part of a fragmented element

M (medial) this is a medial part of a fragmented element

F (final) this is the final part of a fragmented element

Note The values I, M, or F should be used only where it is clear how the element may be be reconstituted.

---

att.global provides attributes common to all elements in the TEI encoding scheme.

Module tei

Members TEI castGroup castItem castList fileDesc forename principal publicationStmt revisionDesc role roleDesc set sourceDesc surname teiHeader titleStmt w

Attributes Attributes att.global.analytic (*@ana*) att.global.facs (*@facs*)

@xml:id (identifier) provides a unique identifier for the element bearing the attribute.

Status Optional

Datatype `xsd:ID`

Note The *xml:id* attribute may be used to specify a canonical reference for an element; see section 3.10..

@n (number) gives a number (or other label) for an element, which is not necessarily unique within the document.

Status Optional

Datatype `data.text`

Note The value of this attribute is always understood to be a single token, even if it contains space or other punctuation characters, and need not be composed of numbers only. It is typically used to specify the numbering of chapters, sections, list items, etc.; it may also be used in the specification of a standard reference system for the text.

@xml:lang (language) indicates the language of the element content using a 'tag' generated according to BCP 47.

Status Optional

Datatype `data.language`

```
<p> … The consequences of
this rapid depopulation were the loss of the last
<foreign xml:lang="rap">ariki</foreign> or chief
(Routledge 1920:205,210) and their connections to
ancestral territorial organization.</p>
```

Note The xml:lang value will be inherited from the immediately enclosing element, or from its parent, and so on up the document hierarchy. It is generally good practice to specify xml:lang at the highest appropriate level, noticing that a different default may be needed for the teiHeader from that needed for the associated resource element or elements, and that a single TEI document may contain texts in many languages.The authoritative list of registered language subtags is maintained by IANA and is available at http://www.iana.org/assignments/language-subtag-registry.

For a good general overview of the construction of language tags, see http://www.w3.org/International/articles/language-tags/, and for a practical step-by-step guide, see http://www.w3.org/International/questions/qa-choosing-language-tags. The value used must conform with BCP 47. If the value is a private use code (i.e., starts with x- or contains -x-), a <language> element with a matching value for its *ident* attribute should be supplied in the TEI header to document this value. Such documentation may also optionally be supplied for non-private-use codes, though these must remain consistent with their (IETF) Internet Engineering Task Force definitions.

@rend (rendition) indicates how the element in question was rendered or presented in the source text.

Status Optional

Datatype 1– occurrences of data.word separated by whitespace

```
<head rend="align(center) case(allcaps)">
 <lb/>To The <lb/>Duchesse <lb/>of <lb/>Newcastle,
<lb/>On Her <lb/>
 <hi rend="case(mixed)">New Blazing-World</hi>.
</head>
```

Note These Guidelines make no binding recommendations for the values of the *rend* attribute; the characteristics of visual presentation vary too much from text to text and the decision to record or ignore individual characteristics varies too much from project to project. Some potentially useful conventions are noted from time to time at appropriate points in the Guidelines. The values of the *rend* attribute are a set of sequence-indeterminate individual tokens separated by whitespace.

@xml:space signals an intention about how white space should be managed by applications.

Status Optional

Datatype data.enumerated

Legal values are: **default** signals that the application's default white-space processing modes are acceptable

preserve indicates the intent that applications preserve all white space

Note The XML specification provides further guidance on the use of this attribute. Note that many parsers may not handle xml:space correctly.

---

att.global.analytic provides additional global attributes for associating specific analyses or interpretations with appropriate portions of a text. [17.3.]

Module analysis

Members att.global[TEI castGroup castItem castList fileDesc forename principal publicationStmt revisionDesc role roleDesc set sourceDesc surname teiHeader titleStmt w]

Attributes Attributes

@ana (analysis) indicates one or more elements containing interpretations of the element on which the *ana* attribute appears.

Status Optional

Datatype 1– occurrences of `data.pointer` separated by whitespace

Note When multiple values are given, they may reflect either multiple divergent interpretations of an ambiguous text, or multiple mutually consistent interpretations of the same passage in different contexts.

---

att.global.facs groups elements corresponding with all or part of an image, because they contain an alternative representation of it, typically but not necessarily a transcription of it. [11.1.]

Module transcr

Members att.global[TEI castGroup castItem castList fileDesc forename principal publicationStmt revisionDesc role roleDesc set sourceDesc surname teiHeader titleStmt w]

Attributes Attributes

@facs (facsimile) points to all or part of an image which corresponds with the content of the element.

Status Optional

Datatype 1– occurrences of `data.pointer` separated by whitespace

---

att.naming provides attributes common to elements which refer to named persons, places, organizations etc. [3.5.1. 13.3.5.]

Module tei

Members att.personal[forename surname]

Attributes Attributes att.canonical (*@key*, *@ref*)

@role may be used to specify further information about the entity referenced by this name in the form of a set of whitespace-separated values, for example the occupation of a person, or the status of a place.

Status Optional

Datatype 1– occurrences of `data.enumerated` separated by whitespace

@nymRef (reference to the canonical name) provides a means of locating the canonical form (*nym*) of the names associated with the object named by the element bearing it.

Status Optional

Datatype 1– occurrences of `data.pointer` separated by whitespace

Note The value must point directly to one or more XML elements by means of one or more URIs, separated by whitespace. If more than one is supplied, the implication is that the name is associated with several distinct canonical names.

---

att.personal (attributes for components of names usually, but not necessarily, personal names) common attributes for those elements which form part of a name usually, but not necessarily, a personal name. [13.2.1.]

Module tei

Members forename surname

Attributes Attributes att.naming (*@role*, *@nymRef*) (att.canonical (*@key*, *@ref*))

@full indicates whether the name component is given in full, as an abbreviation or simply as an initial.

Status Optional

Datatype data.enumerated

Legal values are: **yes** the name component is spelled out in full.[Default]

abb (abbreviated) the name component is given in an abbreviated form.

init (initial letter) the name component is indicated only by one initial.

@sort specifies the sort order of the name component in relation to others within the name.

Status Optional

Datatype data.count

---

att.placement provides attributes for describing where on the source page or object a textual element appears. [3.4.3. 11.3.1.4.]

Module tei

Members stage

Attributes Attributes

@place specifies where this item is placed

Status Recommended

Datatype 1– occurrences of data.enumerated separated by whitespace

Suggested values include: **below** below the line

bottom at the foot of the page

margin in the margin (left, right, or both)

top at the top of the page

opposite on the opposite, i.e. facing, page

overleaf on the other side of the leaf

above above the line

end at the end of e.g. chapter or volume.

inline within the body of the text.

inspace in a predefined space, for example left by an earlier scribe.

```
<add place="margin">[An addition written in the
margin]</add>
<add place="bottom opposite">[An addition written at the
foot of the current page and also on the facing page]</add>
<note place="bottom">Ibid, p.7</note>
```

---

att.ranging provides attributes for describing numerical ranges.

Module tei

Members att.dimensions[att.editLike]

Attributes Attributes

@atLeast gives a minimum estimated value for the approximate measurement.
Status Optional
Datatype data.numeric

@atMost gives a maximum estimated value for the approximate measurement.
Status Optional
Datatype data.numeric

@min where the measurement summarizes more than one observation or a range, supplies the minimum value observed.
Status Optional
Datatype data.numeric

@max where the measurement summarizes more than one observation or a range, supplies the maximum value observed.
Status Optional
Datatype data.numeric

@confidence specifies the degree of statistical confidence (between zero and one) that a value falls within the range specified by *min* and *max*, or the proportion of observed values that fall within that range.
Status Optional
Datatype data.probability

Example

```
The MS. was lost in transmission by mail from <del rend="overstrike">
 <gap atLeast="1" atMost="2"
  extent="one or two letters" reason="illegible" unit="chars"/>
</del> Philadelphia to the Graphic office, New York.
```

---

att.responsibility provides attributes indicating who is responsible for something asserted by the markup and the degree of certainty associated with it. [3.4. 17.3. 13.1.1.]

Module tei

Members att.editLike

Attributes Attributes att.source (*@source*)

@cert (certainty) signifies the degree of certainty associated with the intervention or interpretation.
Status Optional
Datatype data.certainty

@resp (responsible party) indicates the agency responsible for the intervention or interpretation, for example an editor or transcriber.
Status Optional
Datatype 1– occurrences of data.pointer separated by whitespace

Example

```
Blessed are the
<choice>
 <sic>placemakers</sic>
 <corr cert="high" resp="#editor">peacemakers</corr>
</choice>: for they shall be called the children of God.
```

---

att.segLike provides attributes for elements used for arbitrary segmentation. [16.3. 17.1.]

Module tei

Members w

Attributes Attributes att.datcat (*@datcat*, *@valueDatcat*) att.fragmentable (*@part*)

`@function` characterizes the function of the segment.

Status Optional

Datatype `data.enumerated`

Note Attribute values will often vary depending on the type of element to which they are attached. For example, a `<cl>`, may take values such as coordinate, subject, adverbial etc. For a `<phr>`, such values as subject, predicate etc. may be more appropriate. Such constraints will typically be implemented by a project-defined customization.

---

att.source provides attributes for pointing to the source of a bibliographic reference. [3.3.3. 8.3.4.]

Module tei

Members att.responsibility[att.editLike]

Attributes Attributes

`@source` provides a pointer to the bibliographical source from which a quotation or citation is drawn.

Status Optional

Datatype 1– occurrences of `data.pointer` separated by whitespace

Example

```
<p>
<!-- ... -->
As Willard McCarty (<bibl xml:id="mcc_2012">2012, p.2</bibl>)
tells us, <quote source="#mcc_2012">‘Collaboration’ is a
   problematic and should be a contested term.</quote>
<!-- ... -->
</p>
```

Example

```
<p>
<!-- ... -->
 <quote source="#chicago_15_ed">Grammatical theories
   are in flux, and the more we learn, the less we
   seem to know.</quote>
<!-- ... -->
</p>
<!-- ... -->
<bibl xml:id="chicago_15_ed">
 <title level="m">The Chicago Manual of Style</title>,
<edition>15th edition</edition>.
<pubPlace>Chicago</pubPlace>:
<publisher>University of Chicago Press</publisher>
(<date>2003</date>),
<biblScope unit="page">p.147</biblScope>.

</bibl>
```

---

att.typed provides attributes which can be used to classify or subclassify elements in any way. [1.3.1. 17.1.1. 3.5.1. 3.6. 3.5.5. 3.12.1. 7.2.5. 4.1.1. 4.1.2. 4.2.1. 4.4. 13.3.2.3. 11.3.1.1. 16.1.1. 16.3. 12.2. 22.4.4.2. 8.3. 23.3.1.4.]

Module tei

Members forename surname w

Attributes Attributes

`@type` characterizes the element in some sense, using any convenient classification scheme or typology.

Status Optional

Datatype `data.enumerated`

```
<div type="verse">
 <head>Night in Tarras</head>
 <lg type="stanza">
  <l>At evening tramping on the hot white road</l>
  <l>…</l>
 </lg>
 <lg type="stanza">
  <l>A wind sprang up from nowhere as the sky</l>
  <l>…</l>
 </lg>
</div>
```

Note The *type* attribute is present on a number of elements, not all of which are members of att.typed, usually because these elements restrict the possible values for the attribute in a specific way.

`@subtype` provides a sub-categorization of the element, if needed

Status Optional

Datatype `data.enumerated`

Note The *subtype* attribute may be used to provide any sub-classification for the element additional to that provided by its *type* attribute.

Schematron <sch:rule context="*[@subtype]"> <sch:assert test="@type">The <sch:name/> element should not be categorized in detail with @subtype unless also categorized in general with @type</sch:assert></sch:rule>

Note When appropriate, values from an established typology should be used. Alternatively a typology may be defined in the associated TEI header. If values are to be taken from a project-specific list, this should be defined using the `<valList>` element in the project-specific schema description, as described in 23.3.1.4. .

A.4 Macros

---

data.certainty defines the range of attribute values expressing a degree of certainty.

Module tei

Used by

Declaration `data.certainty = "high" | "medium" | "low" | "unknown"`

Note Certainty may be expressed by one of the predefined symbolic values high, medium, or low. The value unknown should be used in cases where the encoder does not wish to assert an opinion about the matter. For more precise indication, data.probability may be used instead or in addition.

---

data.count defines the range of attribute values used for a non-negative integer value used as a count.

Module tei

Used by

Declaration
```
data.count = xsd:nonNegativeInteger
```

Note Only positive integer values (including zero) are permitted

---

data.duration.iso defines the range of attribute values available for representation of a duration in time using ISO 8601 standard formats

Module tei

Used by

Declaration
```
data.duration.iso = token { pattern = "[0-9.,DHMPRSTWYZ/:+\-]+" }
```

Example
```
<time dur-iso="PT0,75H">three-quarters of an hour</time>
```

Example
```
<date dur-iso="P1,5D">a day and a half</date>
```

Example
```
<date dur-iso="P14D">a fortnight</date>
```

Example
```
<time dur-iso="PT0.02S">20 ms</time>
```

Note A duration is expressed as a sequence of number-letter pairs, preceded by the letter P; the letter gives the unit and may be Y (year), M (month), D (day), H (hour), M (minute), or S (second), in that order. The numbers are all unsigned integers, except for the last, which may have a decimal component (using either `.` or `,` as the decimal point; the latter is preferred). If any number is *0*, then that number-letter pair may be omitted. If any of the H (hour), M (minute), or S (second) number-letter pairs are present, then the separator `T` must precede the first 'time' number-letter pair.For complete details, see ISO 8601 *Data elements and interchange formats — Information interchange — Representation of dates and times.*

---

data.duration.w3c defines the range of attribute values available for representation of a duration in time using W3C datatypes.

Module tei

Used by

Declaration
```
data.duration.w3c = xsd:duration
```

Example
```
<time dur="PT45M">forty-five minutes</time>
```

Example
```
<date dur="P1DT12H">a day and a half</date>
```

Example

```
<date dur="P7D">a week</date>
```

Example

```
<time dur="PT0.02S">20 ms</time>
```

Note A duration is expressed as a sequence of number-letter pairs, preceded by the letter P; the letter gives the unit and may be Y (year), M (month), D (day), H (hour), M (minute), or S (second), in that order. The numbers are all unsigned integers, except for the `S` number, which may have a decimal component (using `.` as the decimal point). If any number is *0*, then that number-letter pair may be omitted. If any of the H (hour), M (minute), or S (second) number-letter pairs are present, then the separator `T` must precede the first 'time' number-letter pair.For complete details, see the W3C specification.

---

data.enumerated defines the range of attribute values expressed as a single XML name taken from a list of documented possibilities.

Module tei

Used by Element:

- availability/@status
- castItem/@type
- idno/@type
- q/@type
- stage/@type
- teiHeader/@type
- title/@type

Declaration `data.enumerated = data.name`

Note Attributes using this datatype must contain a word which follows the rules defining a legal XML name (see `http://www.w3.org/TR/REC-xml/#dt-name`): for example they cannot include whitespace or begin with digits.Typically, the list of documented possibilities will be provided (or exemplified) by a value list in the associated attribute specification, expressed with a `<valList>` element.

---

data.language defines the range of attribute values used to identify a particular combination of human language and writing system. [6.1.]

Module tei

Used by

Declaration `data.language = xsd:language | ""`

Note The values for this attribute are language 'tags' as defined in BCP 47. Currently BCP 47 comprises RFC 4646 and RFC 4647; over time, other IETF documents may succeed these as the best current practice.A 'language tag', per BCP 47, is assembled from a sequence of components or *subtags* separated by the hyphen character (-, U+002D). The tag is made of the following subtags, in the following order. Every subtag except the first is optional. If present, each occurs only once, except the fourth and fifth components (variant and extension), which are repeatable.

language The IANA-registered code for the language. This is almost always the same as the ISO 639 2-letter language code if there is one. The list of available

registered language subtags can be found at `http://www.iana.org/assignments/language-subtag-registry`. It is recommended that this code be written in lower case.

script The ISO 15924 code for the script. These codes consist of 4 letters, and it is recommended they be written with an initial capital, the other three letters in lower case. The canonical list of codes is maintained by the Unicode Consortium, and is available at `http://unicode.org/iso15924/iso15924-codes.html`. The IETF recommends this code be omitted unless it is necessary to make a distinction you need.

region Either an ISO 3166 country code or a UN M.49 region code that is registered with IANA (not all such codes are registered, e.g. UN codes for economic groupings or codes for countries for which there is already an ISO 3166 2-letter code are not registered). The former consist of 2 letters, and it is recommended they be written in upper case. The list of codes can be found at `http://www.iso.org/iso/home/standards/country_codes/iso-3166-1_decoding_table.htm`. The latter consist of 3 digits; the list of codes can be found at `http://unstats.un.org/unsd/methods/m49/m49.htm`.

variant An IANA-registered variation. These codes 'are used to indicate additional, well-recognized variations that define a language or its dialects that are not covered by other available subtags'.

extension An extension has the format of a single letter followed by a hyphen followed by additional subtags. These exist to allow for future extension to BCP 47, but as of this writing no such extensions are in use.

private use An extension that uses the initial subtag of the single letter x (i.e., starts with `x-`) has no meaning except as negotiated among the parties involved. These should be used with great care, since they interfere with the interoperability that use of RFC 4646 is intended to promote. In order for a document that makes use of these subtags to be TEI-conformant, a corresponding `<language>` element must be present in the TEI header.

There are two exceptions to the above format. First, there are language tags in the IANA registry that do not match the above syntax, but are present because they have been 'grandfathered' from previous specifications.
Second, an entire language tag can consist of only a private use subtag. These tags start with `x-`, and do not need to follow any further rules established by the IETF and endorsed by these Guidelines. Like all language tags that make use of private use subtags, the language in question must be documented in a corresponding `<language>` element in the TEI header.
Examples include

sn Shona

zh-TW Taiwanese

zh-Hant-HK Chinese written in traditional script as used in Hong Kong

en-SL English as spoken in Sierra Leone

pl Polish

es-MX Spanish as spoken in Mexico

es-419 Spanish as spoken in Latin America

The W3C Internationalization Activity has published a useful introduction to BCP 47, Language tags in HTML and XML.

---

data.name defines the range of attribute values expressed as an XML Name.

Module tei

Used by data.enumerated

Declaration

```
data.name = xsd:Name
```

Note Attributes using this datatype must contain a single word which follows the rules defining a legal XML name (see `http://www.w3.org/TR/REC-xml/#dt-name`): for example they cannot include whitespace or begin with digits.

---

data.numeric defines the range of attribute values used for numeric values.

Module tei

Used by

Declaration

```
data.numeric =
   xsd:double | token { pattern = "(\-?[\d]+/\-?[\d]+)" } | xsd:decimal
```

Note Any numeric value, represented as a decimal number, in floating point format, or as a ratio.To represent a floating point number, expressed in scientific notation, 'E notation', a variant of 'exponential notation', may be used. In this format, the value is expressed as two numbers separated by the letter E. The first number, the significand (sometimes called the mantissa) is given in decimal format, while the second is an integer. The value is obtained by multiplying the mantissa by 10 the number of times indicated by the integer. Thus the value represented in decimal notation as 1000.0 might be represented in scientific notation as 10E3.
A value expressed as a ratio is represented by two integer values separated by a solidus (/) character. Thus, the value represented in decimal notation as 0.5 might be represented as a ratio by the string 1/2.

---

data.outputMeasurement defines a range of values for use in specifying the size of an object that is intended for display.

Module tei

Used by

Declaration

```
data.outputMeasurement =
   token
   {
      pattern = "[\-+]?\d+(\.\d+)?(%|cm|mm|in|pt|pc|px|em|ex|gd|rem|vw|vh|vm)"
   }
```

Example

```
<figure>
 <head>The TEI Logo</head>
 <figDesc>Stylized yellow angle brackets with the letters
<mentioned>TEI</mentioned> in
```

```
   between and <mentioned>text encoding initiative</mentioned> underneath, all on a
white
   background.</figDesc>
 <graphic height="600px"
  url="http://www.tei-c.org/logos/TEI-600.jpg" width="600px"/>
</figure>
```

Note These values map directly onto the values used by XSL-FO and CSS. For definitions of the units see those specifications; at the time of this writing the most complete list is in the CSS3 working draft.

---

data.pattern (regular expression pattern) defines attribute values which are expressed as a regular expression.

Module tei

Used by

Declaration `data.pattern = token`

Note A regular expression, often called a *pattern*, is an expression that describes a set of strings. They are usually used to give a concise description of a set, without having to list all elements. For example, the set containing the three strings *Handel*, *Händel*, and *Haendel* can be described by the pattern `H(ä|ae?)ndel` (or alternatively, it is said that the pattern `H(ä|ae?)ndel` *matches* each of the three strings)Wikipedia

---

data.pointer defines the range of attribute values used to provide a single URI, absolute or relative, pointing to some other resource, either within the current document or elsewhere.

Module tei

Used by Element:

- w/@lemmaRef

Declaration `data.pointer = xsd:anyURI`

Note The range of syntactically valid values is defined by RFC 3986 *Uniform Resource Identifier (URI): Generic Syntax*. Note that the values themselves are encoded using RFC 3987 *Internationalized Resource Identifiers* (IRIs) mapping to URIs. For example, `https://secure.wikimedia.org/wikipedia/en/wiki/%` is encoded as `https://secure.wikimedia.org/wikipedia/en/wiki/%25` while `http://موقع.وزارة-الاتصالات.مصر/` is encoded as `http://xn--4gbrim.xn----rmckbbajlc6dj7bxne2c.xn--wgbh1c/`

---

data.probability defines the range of attribute values expressing a probability.

Module tei

Used by

Declaration

```
data.probability = xsd:double { minInclusive = "0" maxInclusive = "1" }
```

Note Probability is expressed as a real number between 0 and 1; 0 representing *certainly false* and 1 representing *certainly true*.

---

data.replacement defines attribute values which contain a replacement template.

Module tei

Used by

Declaration

```
data.replacement = text
```

---

data.temporal.iso defines the range of attribute values expressing a temporal expression such as a date, a time, or a combination of them, that conform to the international standard *Data elements and interchange formats – Information interchange – Representation of dates and times.*

Module tei

Used by

Declaration

```
data.temporal.iso =
   xsd:date
 | xsd:gYear
 | xsd:gMonth
 | xsd:gDay
 | xsd:gYearMonth
 | xsd:gMonthDay
 | xsd:time
 | xsd:dateTime
 | token { pattern = "[0-9.,DHMPRSTWYZ/:+\-]+" }
```

Note If it is likely that the value used is to be compared with another, then a time zone indicator should always be included, and only the dateTime representation should be used.For all representations for which ISO 8601 describes both a *basic* and an *extended* format, these Guidelines recommend use of the extended format.
While ISO 8601 permits the use of both `00:00` and `24:00` to represent midnight, these Guidelines strongly recommend against the use of `24:00`.

---

data.temporal.w3c defines the range of attribute values expressing a temporal expression such as a date, a time, or a combination of them, that conform to the W3C *XML Schema Part 2: Datatypes Second Edition* specification.

Module tei

Used by

Declaration

```
data.temporal.w3c =
   xsd:date
 | xsd:gYear
 | xsd:gMonth
 | xsd:gDay
 | xsd:gYearMonth
 | xsd:gMonthDay
 | xsd:time
 | xsd:dateTime
```

Note If it is likely that the value used is to be compared with another, then a time zone indicator should always be included, and only the dateTime representation should be used.

---

data.text defines the range of attribute values used to express some kind of identifying string as a single sequence of unicode characters possibly including whitespace.

Module tei

Used by Element:

- w/@lemma

Declaration

```
data.text = string
```

Note Attributes using this datatype must contain a single 'token' in which whitespace and other punctuation characters are permitted.

---

data.truthValue defines the range of attribute values used to express a truth value.

Module tei

Used by

Declaration

```
data.truthValue = xsd:boolean
```

Note The possible values of this datatype are 1 or true, or 0 or false.This datatype applies only for cases where uncertainty is inappropriate; if the attribute concerned may have a value other than true or false, e.g. unknown, or inapplicable, it should have the extended version of this datatype: data.xTruthValue.

---

data.versionNumber defines the range of attribute values used for version numbers.

Module tei

Used by

Declaration

```
data.versionNumber =
   token { pattern = "[\d]+[a-z]*[\d]*(\.[\d]+[a-z]*[\d]*){0,3}" }
```

---

data.word defines the range of attribute values expressed as a single word or token.

Module tei

Used by

Declaration

```
data.word = token { pattern = "(\p{L}|\p{N}|\p{P}|\p{S})+" }
```

Note Attributes using this datatype must contain a single 'word' which contains only letters, digits, punctuation characters, or symbols: thus it cannot include whitespace.

---

data.xTruthValue (extended truth value) defines the range of attribute values used to express a truth value which may be unknown.

Module tei

Used by

Declaration

```
data.xTruthValue = xsd:boolean | "unknown" | "inapplicable"
```

Note In cases where where uncertainty is inappropriate, use the datatype data.TruthValue.

---

macro.paraContent (paragraph content) defines the content of paragraphs and similar elements. [1.3.]

Module tei

Used by hi l p title

Declaration

```
macro.paraContent =
   ( text | model.gLike | model.phrase | model.inter | model.global | lg )*
```

---

macro.phraseSeq (phrase sequence) defines a sequence of character data and phrase-level elements. [1.4.1.]

Module tei

Used by author forename name publisher role roleDesc speaker surname

Declaration

```
macro.phraseSeq = ( text | model.gLike | model.phrase | model.global )*
```

---

macro.phraseSeq.limited (limited phrase sequence) defines a sequence of character data and those phrase-level elements that are not typically used for transcribing extant documents. [1.4.1.]

Module tei

Used by principal

Declaration

```
macro.phraseSeq.limited = ( text | model.limitedPhrase | model.global )*
```

---

macro.specialPara ('special' paragraph content) defines the content model of elements such as notes or list items, which either contain a series of component-level elements or else have the same structure as a paragraph, containing a series of phrase-level and inter-level elements. [1.3.]

Module tei

Used by change licence note q quote stage

Declaration

```
macro.specialPara =
   (
```

```
      text
    | model.gLike    | model.phrase    | model.inter    | model.divPart    | model.global
```