TEI Lite: Encoding for Interchange: an introduction to the TEI — Revised for TEI P5 release

Lou Burnard and C. M. Sperberg-McQueen

February 2006

1 Prefatory note

TEI Lite was the name adopted for what the TEI editors originally conceived of as a simple demonstration of how the TEI 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 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 in our own experience. 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", many of whom produced their own documentation and tutorial materials (some examples are listed in the TEI Tutorials pages). It was also widely adopted as the basis for TEI-conformant authoring systems. Documentation introducing TEI Lite has been widely used for tutorial purposes and has been widely translated (see further the list of versions at `http://www.tei-c.org/Lite/`).

With the publication of TEI P4, the XML version of the TEI Guidelines, which uses the generation of TEI Lite as an example of the modification mechanism built into the TEI Guidelines, the opportunity was taken to produce a lightly revised XML-conformant version, but the present revision is the first substantively changed version since its first appearance in 1997. This revision takes advantage of the many new features introduced into the TEI Guidelines at release P5. A brief list of those changes likely to affect users of previous versions of this document is given below (*Substantive changes from the P4 version*).

Lou Burnard, February 2006

This document provides an introduction to the recommendations of the Text Encoding Initiative (TEI), by describing a specific subset of the full TEI encoding scheme. The scheme documented here can be used to encode a wide variety of commonly encountered textual features, in such a way as to maximize the usability of electronic transcriptions and to facilitate their interchange among scholars using different computer systems. It is fully compatible with the full TEI scheme, as defined by TEI document P5, *Guidelines for Electronic Text Encoding and Interchange*, as of February 2006, and available from the TEI Consortium website at `http://www.tei-c.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 programs 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.

The TEI was sponsored by the Association for Computers and the Humanities, the Association for Computational Linguistics, and the Association for Literary and Linguistic Computing, and is now maintained and developed by an independent membership consortium, hosted by four major Universities. Funding has been provided in part from the U.S. National Endowment for the Humanities, Directorate General XIII of the Commission of the European Communities, the Andrew W. Mellon Foundation, and the Social Science and Humanities Research Council of Canada. 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 were 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 manual conforms to version 0.3 of this most recent edition of the Guidelines, TEI P5.

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 derivable from the full TEI scheme using the extension mechanisms 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. At the time of first writing (1995), our confidence that we have at least partially done so is borne out by its use in practice for the encoding of real texts. The Oxford Text Archive uses TEI Lite when it translates texts from its holdings from their original markup schemes into SGML; the Electronic Text Centers at the University of Virginia and the University of Michigan have used TEI Lite to encode their holdings. And the Text Encoding Initiative itself uses TEI Lite, in its current technical documentation — including this document.

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 par-
son 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 Ed-
ward' (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
474
JANE EYRE                      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.
'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 honey-
moon 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 allud-
ing to my marriage. His letter was then calm, and though very serious,
kind. He has maintained a regular, though not very frequent correspond-
```

```
ence 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  type="chapter"  n="38">
  <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. The decision to focus on Brontë's text, rather than on the printing of it in this particular edition, 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.

A TEI-recommended way of carrying out most of these is 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.

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 *19. The Electronic Title Page*.

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:

```
<TEI>
  <teiHeader/>
  <text>
    <front/>
    <body/>
    <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>
  <teiHeader/>
  <text>
    <front/>
    <group>
      <text>
        <front/>
        <body/>
        <back/>
      </text>
      <text>
        <front/>
        <body/>
        <back/>
      </text>
    </group>
    <back/>
  </text>
</TEI>
```

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

```
<teiCorpus>
  <teiHeader/>
  <TEI>
    <teiHeader/>
    <text/>
  </TEI>
  <TEI>
    <teiHeader/>
    <text/>
  </TEI>
</teiCorpus>
```

It is not however possible to create a composite of corpora – that is, a number of <teiCorpus> elements combined together and treated as a single object. This is a restriction of the current version of the TEI Guidelines.

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 prefixed to every TEI-conformant 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 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.

<group> contains the body of a composite text, grouping together a sequence of distinct texts (or groups of such texts) which are regarded as a unit for some purpose, for example the collected works of an author, a sequence of prose essays, etc.

<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 *18. Front and Back Matter*. 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 *8. Cross References and Links*. 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.

The n attribute may be used to supply (additionally or alternatively) a short mnemonic name or number for the division. 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.

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 *6.3. Foreign Words or Expressions* 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 *6. Marking Highlighted Phrases* 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 *8.3. Special kinds of Linking*.

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 xml:id="WN1" n="I" type="book">
    <div xml:id="WN101" n="I.1" type="chapter"/>
    <div xml:id="WN102" n="I.2" type="chapter"/>
    <div xml:id="WN110" n="I.10" type="chapter">
      <div xml:id="WN1101" n="I.10.1" type="part"/>
      <div xml:id="WN1102" n="I.10.2" type="part"/>
    </div>
  </div>
  <div xml:id="WN2" n="II" type="book"/>
</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 xml:id="TS01" n="1" type="Volume">
    <div xml:id="TS011" n="1" type="Chapter"/>
    <div xml:id="TS012" n="2"/>
  </div>
  <div xml:id="TS02" n="2" type="Volume">
    <div xml:id="TS021" n="3" type="Chapter"/>
    <div xml:id="TS022" n="4"/>
  </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 closing such as "End of Chapter 1". 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.

<trailer> contains a closing title or footer appearing at the end of a division of a text.

Some other elements which may be necessary at the beginning or ending of text divisions are discussed below in section *18.1.2. Prefatory Matter*.

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.), 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 xml:id="UGT1" n="Winter" type="Part">
  <div xml:id="UGT11" n="1" type="Chapter">
    <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 noted above, the paragraphs making up a textual division should be tagged with the <p> tag. For example:

```
<p>I fully appreciate Gen. Pope's splendid achievements
with their invaluable results; but you must know that
Major Generalships in the Regular Army, are not as
plenty as blackberries.
</p>
```

A number of different tags are provided for the encoding of the structural components of verse and performance texts (drama, film, etc.):

<l> (verse line) contains a single, possibly incomplete, line of verse.

<lg> (line group) contains a group of verse lines functioning as a formal unit, e.g. a stanza, refrain, verse paragraph, etc.

<sp> (speech) An individual speech in a performance text, or a passage presented as such in a prose or verse text.

<speaker> 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 *5. Page and Line Numbers* may be used to mark typographic lines if so desired.

Sometimes, particularly in dramatic texts, 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, as in this example:

```
<div type="Act" n="I">
  <head>ACT I</head>
  <div type="Scene" n="1">
    <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 type="stanza" part="I">
      <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>
```

This example shows how dialogue presented in a prose work as if it were drama should be encoded. It 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 Opimiam</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 Opimiam</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>
```

5 Page and Line Numbers

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

<pb/> (page break) marks the boundary between one page of a text and the next in a standard reference system.

<lb/> (line break) marks the start of a new (typographic) line in some edition or version of a text.

<milestone/> marks a boundary point separating any kind of section of a text, typically but not necessarily indicating a point at which some part of a standard reference system changes, where the change is not represented by a structural element.

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. Recording the line breaks may be useful for the same reason; treatment of end-of-line hyphenation in printed source texts will require some consideration.

If pagination, etc., 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>
```

The <pb> and <lb> elements are special cases of the general class of *milestone* elements which mark reference points within a text. TEI Lite also includes a generic <milestone> element, which is not restricted to special cases but 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. 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. For example, a heading rendered in bold might be tagged <head rend="bold">, and one in italic <head rend="italic">.

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.

<emph> (emphasized) marks words or phrases which are stressed or emphasized for linguistic or rhetorical effect.

<foreign> (foreign) identifies a word or phrase as belonging to some language other than that of the surrounding text.

<gloss> identifies a phrase or word used to provide a gloss or definition for some other word or phrase.

<label> contains the label associated with an item in a list; in glossaries, marks the term being defined.

<mentioned> marks words or phrases mentioned, not used.

<term> contains a single-word, multi-word, or symbolic designation which is regarded as a technical 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 rendition 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> (separated from the surrounding text with quotation marks) contains material which is marked as (ostensibly) being somehow different than the surrounding text, 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.

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

<mentioned> marks words or phrases mentioned, not used.

<soCalled> contains a word or phrase for which the author or narrator indicates a disclaiming of responsibility, for example by the use of scare quotes or italics.

<gloss> identifies a phrase or word used to provide a gloss or definition for some other word or phrase.

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 *8.3. Special kinds of Linking*.

Quotations may be accompanied by a reference to the source or speaker, using the who attribute, whether or not the source is given 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.

As with highlighting, it is not always possible and may not be considered desirable to interpret the function of quotation marks in a text in this way. In such cases, the tag <hi rend=”quoted”> might be used to mark quoted text without making any claim as to its status.

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="fra">savoir-faire</foreign>.</p>
<p>Have you read <title  xml:lang="deu">Die Dreigroschenoper</title>?</p>
<p>
  <mentioned  xml:lang="fra">Savoir-faire</mentioned> is French for
know-how.
</p>
<p>The court issued a writ of <term  xml:lang="lat">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[1], as further explained in the relevant section of the TEI Guidelines. Some simple example codes for a few languages are given here:

zh or zho	Chinese	grc	Ancient Greek
en	English	ell or el	Greek
enm	Middle English	ja or jpn	Japanese
fr or fra	French	la or lat	Latin
de or deu	German	sa or san	Sanskrit

[1]The relevant standards are RFC 3066, and the lists of two and three language identifiers maintained as part of ISO 639 (see http://www.w3.org/WAI/ER/IG/ert/iso639.htm)

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; otherwise, the TEI header should state which kind they are.

Examples:

```
<p>Collections are ensembles of distinct
entities or objects of any sort.
<note  place="foot"  n="1">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 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 *8.3. Special kinds of Linking*.

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:

<ref> (reference) defines a reference to another location, possibly modified by additional text or comment.

<ptr/> (pointer) defines a pointer to another location.

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 the text of the cross-reference itself. 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 (assuming we have documented somewhere the exact verbal form of cross references represented by <ptr> elements):

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

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

The value of the target attribute must have been used as the identifier of some other element within the current document. This implies that 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 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:

<anchor/> (anchor point) attaches an identifier to a point within a text, whether or not it corresponds with a textual element.

<seg> (arbitrary segment) represents any segmentation of text below the "chunk" level.

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 *8.3. Special kinds of Linking* 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). 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 URL 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.

Within a URL, 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. A full specification of the language is well beyond the scope of this document; here we provide only a flavour of its power.

In the XPath language, locations are defined as a series of *steps*, each one identifying some part of the document, often in terms of the locations identified by the previous step. For example, you would point to the third sentence of the second paragraph of chapter two by selecting chapter two in the first step, the second paragraph in the second step, and the third sentence in the last step. A step can be defined in terms of the document tree itself, using such concepts as parent, descendent, preceding, etc. or, more loosely, in terms of text patterns, word or character positions.

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 *15. Interpretation and Analysis*. For example, a linguistic analysis of the sentence "John loves Nancy" might be encoded as follows:

```
<seg type="sentence" ana="SVO">
  <seg type="lex" ana="#NP1">John</seg>
  <seg type="lex" ana="#VVI">loves</seg>
  <seg type="lex" ana="#NP1">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 xml:lang="fra" xml:id="FR1" corresp="#EN1">Jean aime Nancy</seg>
<seg xml:lang="en" xml:id="EN1" corresp="#FR1">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 anaphoric 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 xml:id="network" corresp="#nbc">the network</seg>
says <seg xml:id="show" corresp="#shirley">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 xml:id="Q1a" next="#Q1b">Who-e debel you?</q>
— he at last said — <q xml:id="Q1b" prev="#Q1a">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:

<corr> (correction) contains the correct form of a passage apparently erroneous in the copy text.

<sic> (latin for thus or so) contains text reproduced although apparently incorrect or inaccurate.

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

<orig> (original form) contains a reading which is marked as following the original, rather than being normalized or corrected.

<reg> (regularization) contains a reading which has been regularized or normalized in some sense.

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 Shakesperean scholar 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.

<choice> groups a number of alternative encodings for the same point in a text.

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:

<add> (addition) contains letters, words, or phrases inserted in the text by an author, scribe, annotator, or corrector.

<gap> (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.

<del> (deletion) contains a letter, word, or passage deleted, marked as deleted, or otherwise indicated as superfluous or spurious in the copy text by an author, scribe, annotator, or corrector.

<unclear> contains a word, phrase, or passage which cannot be transcribed with certainty because it is illegible or inaudible in the source.

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 indicates that "LB" corrected 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 not limited to changes made by an editor; they can also be 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
```

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.
```

The <del> element marks material which is 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>
```

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:

<abbr> (abbreviation) contains an abbreviation of any sort.

<expan> (expansion) contains the expansion of an abbreviation.

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**ith**", italicising the letters not found in the source. An appropriate encoding for this purpose would be

```
<expan>w<hi>i</hi>t<hi>h</hi>
</expan>
```

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:

<rs> (referencing string) contains a general purpose name or referring string.

<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  type="person"  key="BENM1">Mr. Bennet</rs>,
</q> said <rs  type="person"  key="BENM2">his lady</rs>
to him one day,
<q>have you heard that
<rs  type="place"  key="NETP1">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  type="person"  key="WADLM1">
  <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.

<time> contains a phrase defining a time of day in any format.

The value 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:

<num> (number) contains a number, written in any form.

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 ordered, unordered, or a glossary list. Each item may be preceded by an item label (in a glossary list, this label is the term being defined):

<list> (list) contains any sequence of items organized as a list.

<item> contains one component of a list.

<label> contains the label associated with an item in a list; in glossaries, marks the term being defined.

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="lat">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 (*13. 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 recovery.</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 society.</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.

<bibl> (bibliographic citation) contains a loosely-structured bibliographic citation of which the sub-components may or may not be explicitly tagged.

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

<biblScope> (scope of citation) defines the scope of a bibliographic reference, for example as a list of page numbers, or a named subdivision of a larger work.

<date> contains a date in any format.

<editor> secondary statement of responsibility for a bibliographic item, for example the name of an individual, institution or organization, (or of several such) acting as editor, compiler, translator, etc.

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

<pubPlace> (publication place) contains the name of the place where a bibliographic item was published.

<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:

<table> contains text displayed in tabular form, in rows and columns.

<row> contains one row of a table.

<cell> contains one cell of a table.

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  rows="5"  cols="4">
   <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, by using the elements described in this section. The same elements may also be used to embed digitized versions of the graphic within an electronic document.

<graphic> indicates the location of an inline graphic, illustration, or figure.

<figure> groups elements representing or containing graphic information such as an illustration or figure.

<figDesc> (description of figure) contains a brief prose description of the appearance or content of a graphic figure, for use when documenting an image without displaying it.

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 may also 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"/>
<graphic  url="p412fig.png"/>
<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>
```

When a digitized version of the graphic concerned is available, it may be embedded at the appropriate point within the document in this way.

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  type="chapter"  n="38">
  <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 General-Purpose Interpretation Elements

A more general purpose segmentation element, the <seg> has already been introduced for use in identifying otherwise unmarked targets of cross references and hypertext links (see section *8. Cross References and Links*); 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 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 type="chapter" n="38">
  <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 *8.3. Special kinds of Linking* 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.

<interp> (interpretation) summarizes a specific interpretative annotation which can be linked to a span of text.

<interpGrp> (interpretation group) collects together a set of related interpretations which share responsibility or type.

These elements allows 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 an empty element, which must be linked to the passage to which it applies either by means of the ana attribute discussed in section *8.3. Special kinds of Linking* above, or by means of its own inst attribute. This means that any kind of analysis can be represented, with no need to respect the document hierarchy, and also facilitates 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 xml:id="fig-apos-1" resp="#LB-MSM" type="figureOfSpeech">apostrophe</interp>
      <interp xml:id="fig-hyp-1" resp="#LB-MSM" type="figureOfSpeech">hyperbole</interp>
      <interp xml:id="set-church-1" resp="#LB-MSM" type="setting">church</interp>
      <interp xml:id="ref-church-1" resp="#LB-MSM" type="reference">church</interp>
      <interp xml:id="ref-serv-1" resp="#LB-MSM" type="reference">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 type="figureOfSpeech" resp="#LB-MSM">
        <interp xml:id="fig-apos">apostrophe</interp>
        <interp xml:id="fig-hyp">hyperbole</interp>
        <interp xml:id="fig-meta">metaphor</interp>
      </interpGrp>
      <interpGrp type="scene-setting" resp="#LB-MSM">
        <interp xml:id="set-church">church</interp>
        <interp xml:id="set-kitch">kitchen</interp>
        <interp xml:id="set-unspec">unspecified</interp>
      </interpGrp>
      <interpGrp type="reference" resp="#LB-MSM">
        <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 type="chapter" n="38">
  <p xml:id="P38.1" ana="#set-church #set-kitch">
    <s xml:id="P38.1.1" ana="#fig-apos">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
   xml:id="fig-apos-2"
   type="figureOfSpeech"
   resp="#LB-MSM"
   inst="#P38.1.1">apostrophe</interp>
<interp
   xml:id="set-church-2"
   type="scene-setting"
   inst="#P38.1"
   resp="#LB-MSM">church</interp>
<interp
   xml:id="set-kitchen-2"
   type="scene-setting"
   inst="#P38.1"
   resp="#LB-MSM">kitchen</interp>
```

The <interp> 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 *8.3. Special kinds of Linking* assumes a linguistic analysis which might be represented as follows:

```
<interp  xml:id="NP1"  type="pos">noun phrase, singular</interp>
<interp  xml:id="VV1"  type="pos">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:

<eg> (example) contains any kind of illustrative example.

<code> contains literal code from some formal language such as a programming language.

<ident> (identifier) contains an identifier or name for an object of some kind in a formal language.

<gi> (element name) contains the name (generic identifier) of an element.

<att> (attribute) contains the name of an attribute appearing within running text.

<formula> contains a mathematical or other formula.

<val> (value) contains a single attribute value.

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> 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.

<divGen> (automatically generated text division) indicates the location at which a textual division generated automatically by a text-processing application is to appear.

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

```
<front>
  <titlePage/>
  <divGen  type="toc"/>
  <div>
    <head>Preface</head>
  </div>
</front>
<body/>
<back>
  <div>
    <head>Appendix</head>
  </div>
  <divGen  type="index"  n="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 *11. 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.

<index> (index entry) marks a location to be indexed for whatever purpose.

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.[2]

```
<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 *7. Notes*, or with the <interp> element discussed in section *15. Interpretation and Analysis*. 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">
    <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.

[2]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.

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.

<address> contains a postal address, for example of a publisher, an organization, or an individual.

<addrLine> (address line) contains one line of a postal address.

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 *10.1. Names and Referring Strings*).

```
<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:

<titlePage> (title page) contains the title page of a text, appearing within the front or back matter.

<docTitle> (document title) contains the title of a document, including all its constituents, as given on a title page.

<titlePart> contains a subsection or division of the title of a work, as indicated on a title page.

<byline> contains the primary statement of responsibility given for a work on its title page or at the head or end of the work.

<docAuthor> (document author) contains the name of the author of the document, as given on the title page (often but not always contained in a byline).

<docDate> (document date) contains the date of a document, as given (usually) on a title page.

<docEdition> (document edition) contains an edition statement as presented on a title page of a document.

<docImprint> (document imprint) contains the imprint statement (place and date of publication, publisher name), as given (usually) at the foot of a title page.

<epigraph> contains a quotation, anonymous or attributed, appearing at the start of a section or chapter, or on a title page.

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 lang attribute or the <foreign> element, as necessary. Names, wherever they appear, should be tagged using the <name>, as elsewhere.

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>
```

18.1.2 Prefatory Matter

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

foreword a text addressed to the reader, by the author, editor or publisher, possibly in the form of a letter.

preface a text addressed to the reader, by the author, editor or publisher, possibly in the form of a letter.

dedication a text (often a letter) addressed to someone other than the reader in which the author typically commends the work in hand to the attention of the person concerned.

abstract a prose argument summarizing the content of the work.

ack Acknowledgements.

contents a table of contents (typically this should be tagged as a <list>).

frontispiece a pictorial frontispiece, possibly including some text.

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:

<salute> (salutation) contains a salutation or greeting prefixed to a foreword, dedicatory epistle, or other division of a text, or the salutation in the closing of a letter, preface, etc.

<signed> (signature) contains the closing salutation, etc., appended to a foreword, dedicatory epistle, or other division of a text.

<byline> contains the primary statement of responsibility given for a work on its title page or at the head or end of the work.

<dateline> contains a brief description of the place, date, time, etc. of production of a letter, newspaper story, or other work, prefixed or suffixed to it as a kind of heading or trailer.

<argument> A formal list or prose description of the topics addressed by a subdivision of a text.

<cit> (cited quotation) contains a quotation from some other document, together with a bibliographic reference to its source. In a dictionary it may contain an example text with at least one occurrence of the word form, used in the sense being described, or a translation of the headword, or an example.

<opener> groups together dateline, byline, salutation, and similar phrases appearing as a preliminary group at the start of a division, especially of a letter.

<closer> groups together salutations, datelines, and similar phrases appearing as a final group at the end of a division, especially of a letter.

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 appendix.

glossary a list of words and definitions, typically marked up as a <list type="gloss"> element .

notes a series of <note> elements.

bibliography a series of bibliographic references, typically in the form of a special bibliographic-list element <listBibl>, whose items are individual <bibl> elements.

index a set of index entries, possibly represented as a structured list or glossary list, with optional leading <head> and perhaps some paragraphs of introductory or closing text (An index may also be generated for a document by using the <index> element, described above in section index: (index entry) marks a location to be indexed for whatever purpose.).

colophon a description at the back of the book describing where, when, and by whom it was printed; in modern books it also often gives production details and identifies the type faces used.

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.

<encodingDesc> (encoding description) documents the relationship between an electronic text and the source or sources from which it was derived.

<profileDesc> (text-profile description) provides a detailed description of non-bibliographic aspects of a text, specifically the languages and sublanguages used, the situation in which it was produced, the participants and their setting.

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

A corpus or collection of texts, which share many 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 intellectual content.

<editionStmt> (edition statement) groups information relating to one edition of a text.

<extent> describes the approximate size of a text as stored on some carrier medium, whether digital or non-digital, specified in any convenient units.

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

<seriesStmt> (series statement) groups information about the series, if any, to which a publication belongs.

<notesStmt> (notes statement) collects together any notes providing information about a text additional to that recorded in other parts of the bibliographic description.

<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/>
    <publicationStmt/>
```

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

<sponsor> specifies the name of a sponsoring organization or institution.

<funder> (funding body) specifies the name of an individual, institution, or organization responsible for the funding of a project or text.

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

<respStmt> (statement of responsibility) supplies a statement of responsibility for the intellectual content of a text, edition, recording, or series, where the specialized elements for authors, editors, etc. do not suffice or do not apply.

It is recommended that the title should distinguish the computer file from 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:

<resp> (responsibility) contains a phrase describing the nature of a person's intellectual responsibility.

<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 a text (where *edition* is used as elsewhere in bibliography), and may include the following elements:

<edition> (edition) describes the particularities of one edition of a text.

<respStmt> (statement of responsibility) supplies a statement of responsibility for the intellectual content of a text, edition, recording, or series, where the specialized elements for authors, editors, etc. do not suffice or do not apply.

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 describe the approximate size of a file.

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.

<distributor> supplies the name of a person or other agency responsible for the distribution of a text.

<authority> (release authority) supplies the name of a person or other agency responsible for making an electronic file available, other than a publisher or distributor.

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:

<pubPlace> (publication place) contains the name of the place where a bibliographic item was published.

<address> contains a postal address, for example of a publisher, an organization, or an individual.

<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, etc.

<date> contains a date in any format.

Example:

```
<publicationStmt>
  <publisher>Oxford University Press</publisher>
  <pubPlace>Oxford</pubPlace>
  <date>1989</date>
  <idno  type="ISBN"> 0-19-254705-5</idno>
  <availability>
    <p>Copyright 1989, Oxford University
        Press</p>
  </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:

<bibl> (bibliographic citation) contains a loosely-structured bibliographic citation of which the sub-components may or may not be explicitly tagged.

<biblFull> (fully-structured bibliographic citation) contains a fully-structured bibliographic citation, in which all components of the TEI file description are present.

<listBibl> (citation list) contains a list of bibliographic citations of any kind.

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:

<projectDesc> (project description) describes in detail the aim or purpose for which an electronic file was encoded, together with any other relevant information concerning the process by which it was assembled or collected.

<samplingDecl> (sampling declaration) contains a prose description of the rationale and methods used in sampling texts in the creation of a corpus or collection.

<editorialDecl> (editorial practice declaration) provides details of editorial principles and practices applied during the encoding of a text.

<refsDecl> (references declaration) specifies how canonical references are constructed for this text.

<classDecl> (classification declarations) contains one or more taxonomies defining any classificatory codes used elsewhere in the text.

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>
  <p>All quotation marks converted to entity
    references odq and cdq.</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 such division in the form
     XX.yyy where XX is the book number in roman numeral and
     yyy is the section number in arabic.</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:

<taxonomy> defines a typology used to classify texts either implicitly, by means of a bibliographic citation, or explicitly by a structured taxonomy.

<bibl> (bibliographic citation) contains a loosely-structured bibliographic citation of which the sub-components may or may not be explicitly tagged.

<category> contains an individual descriptive category, possibly nested within a superordinate category, within a user-defined taxonomy.

<catDesc> (category description) describes some category within a taxonomy or text typology, either in the form of a brief prose description or in terms of the situational parameters used by the TEI formal textDesc.

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 further described 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:

<creation> contains information about the creation of a text.

<langUsage> (language usage) describes the languages, sublanguages, registers, dialects, etc. represented within a text.

<textClass> (text classification) groups information which describes the nature or topic of a text in terms of a standard classification scheme, thesaurus, etc.

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:

<language> characterizes a single language or sublanguage used within a text.

an example is needed.

The <textClass> element classifies a text by reference to the system or systems defined by the <classDecl> element, and contains one or more of the following elements:

<keywords> contains a list of keywords or phrases identifying the topic or nature of a text.

<classCode> (classification code) contains the classification code used for this text in some standard classification system.

<catRef/> (category reference) specifies one or more defined categories within some taxonomy or text typology.

The element <keywords> contains a list of keywords or phrases identifying the topic or nature of a text. The attribute scheme links these to the classification system defined in <taxonomy>.

```
<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>
```

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 date 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>
```

20 List of Elements Described

The following list shows all the elements defined for the TEI Lite schema, with a brief description of each, and a link to its full specification in the Appendix.

<abbr> (abbreviation) contains an abbreviation of any sort.

<add> (addition) contains letters, words, or phrases inserted in the text by an author, scribe, annotator, or corrector.

<address> contains a postal address, for example of a publisher, an organization, or an individual.

<addrLine> (address line) contains one line of a postal address.

<anchor/> (anchor point) attaches an identifier to a point within a text, whether or not it corresponds with a textual element.

<argument> A formal list or prose description of the topics addressed by a subdivision of a text.

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

<authority> (release authority) supplies the name of a person or other agency responsible for making an electronic file available, other than a publisher or distributor.

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

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

<bibl> (bibliographic citation) contains a loosely-structured bibliographic citation of which the sub-components may or may not be explicitly tagged.

<biblFull> (fully-structured bibliographic citation) contains a fully-structured bibliographic citation, in which all components of the TEI file description are present.

<biblScope> (scope of citation) defines the scope of a bibliographic reference, for example as a list of page numbers, or a named subdivision of a larger work.

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

<byline> contains the primary statement of responsibility given for a work on its title page or at the head or end of the work.

<catDesc> (category description) describes some category within a taxonomy or text typology, either in the form of a brief prose description or in terms of the situational parameters used by the TEI formal textDesc.

<category> contains an individual descriptive category, possibly nested within a superordinate category, within a user-defined taxonomy.

<catRef/> (category reference) specifies one or more defined categories within some taxonomy or text typology.

<cell> contains one cell of a table.

<change> summarizes a particular change or correction made to a particular version of an electronic text which is shared between several researchers.

<choice> groups a number of alternative encodings for the same point in a text.

<cit> (cited quotation) contains a quotation from some other document, together with a bibliographic reference to its source. In a dictionary it may contain an example text with at least one occurrence of the word form, used in the sense being described, or a translation of the headword, or an example.

<classCode> (classification code) contains the classification code used for this text in some standard classification system.

<classDecl> (classification declarations) contains one or more taxonomies defining any classificatory codes used elsewhere in the text.

<closer> groups together salutations, datelines, and similar phrases appearing as a final group at the end of a division, especially of a letter.

<code> contains literal code from some formal language such as a programming language.

<corr> (correction) contains the correct form of a passage apparently erroneous in the copy text.

<creation> contains information about the creation of a text.

<date> contains a date in any format.

<dateline> contains a brief description of the place, date, time, etc. of production of a letter, newspaper story, or other work, prefixed or suffixed to it as a kind of heading or trailer.

<del> (deletion) contains a letter, word, or passage deleted, marked as deleted, or otherwise indicated as superfluous or spurious in the copy text by an author, scribe, annotator, or corrector.

<distributor> supplies the name of a person or other agency responsible for the distribution of a text.

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

<divGen> (automatically generated text division) indicates the location at which a textual division generated automatically by a text-processing application is to appear.

<docAuthor> (document author) contains the name of the author of the document, as given on the title page (often but not always contained in a byline).

<docDate> (document date) contains the date of a document, as given (usually) on a title page.

<docEdition> (document edition) contains an edition statement as presented on a title page of a document.

<docImprint> (document imprint) contains the imprint statement (place and date of publication, publisher name), as given (usually) at the foot of a title page.

<docTitle> (document title) contains the title of a document, including all its constituents, as given on a title page.

<edition> (edition) describes the particularities of one edition of a text.

<editionStmt> (edition statement) groups information relating to one edition of a text.

<editor> secondary statement of responsibility for a bibliographic item, for example the name of an individual, institution or organization, (or of several such) acting as editor, compiler, translator, etc.

<editorialDecl> (editorial practice declaration) provides details of editorial principles and practices applied during the encoding of a text.

<eg> (example) contains any kind of illustrative example.

<emph> (emphasized) marks words or phrases which are stressed or emphasized for linguistic or rhetorical effect.

<encodingDesc> (encoding description) documents the relationship between an electronic text and the source or sources from which it was derived.

<epigraph> contains a quotation, anonymous or attributed, appearing at the start of a section or chapter, or on a title page.

<extent> describes the approximate size of a text as stored on some carrier medium, whether digital or non-digital, specified in any convenient units.

<figure> groups elements representing or containing graphic information such as an illustration or figure.

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

<foreign> (foreign) identifies a word or phrase as belonging to some language other than that of the surrounding text.

<formula> contains a mathematical or other formula.

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

<funder> (funding body) specifies the name of an individual, institution, or organization responsible for the funding of a project or text.

<gap> (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.

<gi> (element name) contains the name (generic identifier) of an element.

<gloss> identifies a phrase or word used to provide a gloss or definition for some other word or phrase.

<group> contains the body of a composite text, grouping together a sequence of distinct texts (or groups of such texts) which are regarded as a unit for some purpose, for example the collected works of an author, a sequence of prose essays, etc.

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

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

<ident> (identifier) contains an identifier or name for an object of some kind in a formal language.

<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.

<index> (index entry) marks a location to be indexed for whatever purpose.

<interp> (interpretation) summarizes a specific interpretative annotation which can be linked to a span of text.

<interpGrp> (interpretation group) collects together a set of related interpretations which share responsibility or type.

<item> contains one component of a list.

<keywords> contains a list of keywords or phrases identifying the topic or nature of a text.

<l> (verse line) contains a single, possibly incomplete, line of verse.

<label> contains the label associated with an item in a list; in glossaries, marks the term being defined.

<language> characterizes a single language or sublanguage used within a text.

<langUsage> (language usage) describes the languages, sublanguages, registers, dialects, etc. represented within a text.

<lb/> (line break) marks the start of a new (typographic) line in some edition or version of a text.

<lg> (line group) contains a group of verse lines functioning as a formal unit, e.g. a stanza, refrain, verse paragraph, etc.

<list> (list) contains any sequence of items organized as a list.

<listBibl> (citation list) contains a list of bibliographic citations of any kind.

<mentioned> marks words or phrases mentioned, not used.

<milestone/> marks a boundary point separating any kind of section of a text, typically but not necessarily indicating a point at which some part of a standard reference system changes, where the change is not represented by a structural element.

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

<note> contains a note or annotation.

<notesStmt> (notes statement) collects together any notes providing information about a text additional to that recorded in other parts of the bibliographic description.

<num> (number) contains a number, written in any form.

<opener> groups together dateline, byline, salutation, and similar phrases appearing as a preliminary group at the start of a division, especially of a letter.

<orig> (original form) contains a reading which is marked as following the original, rather than being normalized or corrected.

<p> (paragraph) marks paragraphs in prose.

<pb/> (page break) marks the boundary between one page of a text and the next in a standard reference system.

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

<profileDesc> (text-profile description) provides a detailed description of non-bibliographic aspects of a text, specifically the languages and sublanguages used, the situation in which it was produced, the participants and their setting.

<projectDesc> (project description) describes in detail the aim or purpose for which an electronic file was encoded, together with any other relevant information concerning the process by which it was assembled or collected.

<ptr/> (pointer) defines a pointer to another location.

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

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

<**pubPlace**> (publication place) contains the name of the place where a bibliographic item was published.

<**q**> (separated from the surrounding text with quotation marks) contains material which is marked as (ostensibly) being somehow different than the surrounding text, 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.

<**ref**> (reference) defines a reference to another location, possibly modified by additional text or comment.

<**refsDecl**> (references declaration) specifies how canonical references are constructed for this text.

<**reg**> (regularization) contains a reading which has been regularized or normalized in some sense.

<**resp**> (responsibility) contains a phrase describing the nature of a person's intellectual responsibility.

<**respStmt**> (statement of responsibility) supplies a statement of responsibility for the intellectual content of a text, edition, recording, or series, where the specialized elements for authors, editors, etc. do not suffice or do not apply.

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

<**row**> contains one row of a table.

<**rs**> (referencing string) contains a general purpose name or referring string.

<**s**> (s-unit) contains a sentence-like division of a text.

<**salute**> (salutation) contains a salutation or greeting prefixed to a foreword, dedicatory epistle, or other division of a text, or the salutation in the closing of a letter, preface, etc.

<**samplingDecl**> (sampling declaration) contains a prose description of the rationale and methods used in sampling texts in the creation of a corpus or collection.

<**seg**> (arbitrary segment) represents any segmentation of text below the "chunk" level.

<**seriesStmt**> (series statement) groups information about the series, if any, to which a publication belongs.

<**sic**> (latin for thus or so) contains text reproduced although apparently incorrect or inaccurate.

<**signed**> (signature) contains the closing salutation, etc., appended to a foreword, dedicatory epistle, or other division of a text.

<**soCalled**> contains a word or phrase for which the author or narrator indicates a disclaiming of responsibility, for example by the use of scare quotes or italics.

<**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.

<**sp**> (speech) An individual speech in a performance text, or a passage presented as such in a prose or verse text.

<**speaker**> A specialized form of heading or label, giving the name of one or more speakers in a dramatic text or fragment.

<sponsor> specifies the name of a sponsoring organization or institution.

<stage> (stage direction) contains any kind of stage direction within a dramatic text or fragment.

<table> contains text displayed in tabular form, in rows and columns.

<taxonomy> defines a typology used to classify texts either implicitly, by means of a bibliographic citation, or explicitly by a structured taxonomy.

<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 prefixed to every TEI-conformant 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.

<term> contains a single-word, multi-word, or symbolic designation which is regarded as a technical term.

<textClass> (text classification) groups information which describes the nature or topic of a text in terms of a standard classification scheme, thesaurus, etc.

<time> contains a phrase defining a time of day in any format.

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

<titlePage> (title page) contains the title page of a text, appearing within the front or back matter.

<titlePart> contains a subsection or division of the title of a work, as indicated on a title page.

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

<trailer> contains a closing title or footer appearing at the end of a division of a text.

<unclear> contains a word, phrase, or passage which cannot be transcribed with certainty because it is illegible or inaudible in the source.

Appendixes

Substantive changes from the P4 version

This revision of the TEI Lite schema conforms to the TEI P5 Guidelines, which makes a number of changes from the TEI P4 Guidelines underlying earlier versions of TEI Lite. The following brief list indicates some of the major changes which will be needed in existing TEI P4-conformant documents before they can be used with the new schema. A fuller list is in preparation for publication as a part of TEI P5: the items listed here relate specifically to changes in TEI Lite only.

- At P5, a TEI document must declare a namespace of `http://www.tei-c.org/ns/1.0`
- The attributes id and lang are replaced by the attributes xml:id and xml:lang respectively. Values for the latter attribute must conform to RFC 3066
- The element <choice> must be used to wrap <reg> and <orig> if both are supplied. Similarly for <sic> and <corr>, and for <abbr> and <expan>.
- "numbered divs" (<div1>, <div2>, etc.) are not supported in this version of TEI Lite
- all pointing and linking mechanisms now use the same W3C-defined mechanism: there is no longer any distinction between internal and external pointing elements
- the content model of <change> has changed significantly
- **hic desunt multa**

Formal specification

The TEI Lite is a pure subset of the TEI. All of the elements defined in it are taken from the following standard TEI modules:

tei, core, header, textstructure, figures, linking, analysis, and tagdocs.

The following elements from those modules are excluded from the schema: <ab>, <alt>, <altGrp>, <altIdent>, <analytic>, <attDef>, <attList>, <attRef>, <biblItem>, <biblStruct>, <binaryObject>, <broadcast>, <c>, <cb>, <cl>, <classSpec>, <classes>, <content>, <correction>, <datatype>, <defaultVal>, <desc>, <distinct>, <div1>, <div2>, <div3>, <div4>, <div5>, <div6>, <div7>, <egXML>, <elementSpec>, <equipment>, <equiv>, <exemplum>, <fsdDecl>, <floatingText>, <headItem>, <headLabel>, <hyphenation>, <imprimatur>, <interpretation>, <join>, <joinGrp>, <link>, <linkGrp>, <listRef>, <m>, <macroSpec>, <measure>, <meeting>, <memberOf>, <metDecl>, <metSym>, <moduleRef>, <moduleSpec>, <monogr>, <normalization>, <phr>, <postBox>, <postCode>, <quotation>,

<recording>, <recordingStmt>, <remarks>, <schemaSpec>, <scriptStmt>, <segmentation>, <series>, , <spanGrp>, <specDesc>, <specGrp>, <specGrpRef>, <specList>, <state>, <stdVals>, <street>, <stringVal>, <tag>, <timeline>, <valDesc>, <valItem>, <valList>, <variantEncoding>, <w>, <when>

Here is the TEI Lite schema itself :

Schema teilite: changed components

---

att.global provides attributes common to all elements in the TEI encoding scheme.

Module tei

Members

Attributes att.global.linking (*@corresp*, *@synch*, *@sameAs*, *@copyOf*, *@next*, *@prev*, *@exclude*, *@select*) att.global.analytic (*@ana*)

`@xml:id` (identifier) provides a unique identifier for the element bearing the attribute.

Status Optional

Datatype `xsd:ID`

Values any valid XML identifier.

Note The xml:id attribute may be used to specify a canonical reference for an element; see section CORS.

`@n` (number) gives a number (or other label) for an element, which is not necessarily unique within the document.

Status Optional

Datatype 1–∞ occurrences of `token { pattern = "(\p{L}|\p{N}|\p{P}|\p{S})+" }` separated by whitespace

Values the value may contain only letters, digits, punctuation characters, or symbols: it may not contain whitespace or word separating characters. It need not be restricted to numbers.

Note The n attribute may be 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 `xsd:language`

Values The value must conform to BCP 47. If the value is a private use code (i.e., starts with `x-` or contains `-x-`) it should, and if not it may, match the value of an ident attribute of a <language> element supplied in the TEI Header of the current document.

Note If no value is specified for xml:lang, the xml:lang value for the immediately enclosing element is inherited; for this reason, a value should always be specified on the outermost element (<TEI>).

`@rend` (rendition) indicates how the element in question was rendered or presented in the source text.

Status Optional

Datatype 1–∞ occurrences of `token { pattern = "(\p{L}|\p{N}|\p{P}|\p{S})+" }` separated by whitespace

Values may contain any number of tokens, each of which may contain letters, punctuation marks, or symbols, but not word-separating characters.

```
<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.

@xml:base provides a base URI reference with which applications can resolve relative URI references into absolute URI references.

Status Optional

Datatype xsd:anyURI

Values any syntactically valid URI reference.

```
<div
   type="bibl">
  <head>Bibliography</head>
  <listBibl
     xml:base="http://www.lib.ucdavis.edu/BWRP/Works/">
    <bibl
       n="1">
      <author>
        <name>Landon, Letitia Elizabeth</name>
      </author>
      <ref
         target="LandLVowOf.sgm">
        <title>The Vow of the Peacock</title>
      </ref>
    </bibl>
    <bibl
       n="2">
      <author>
        <name>Compton, Margaret Clephane</name>
      </author>
      <ref
         target="NortMIrene.sgm">
        <title>Irene, a Poem in Six Cantos</title>
      </ref>
    </bibl>
    <bibl
       n="3">
      <author>
        <name>Taylor, Jane</name>
      </author>
      <ref
         target="TaylJEssay.sgm">
        <title>Essays in Rhyme on Morals and Manners</title>
      </ref>
    </bibl>
  </listBibl>
</div>
```

@xml:space signals an intention about how white space should be managed by applications.

Status Optional

Legal values are: **default** the processor should treat white space according to the default XML white space handling rules

preserve the processor should preserve unchanged any and all white space in the source

Note The XML specification provides further guidance on the use of this attribute.

Schema teilite: unchanged components

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.
abbr: (abbreviation) contains an abbreviation of any sort.
add: (addition) contains letters, words, or phrases inserted in the text by an author, scribe, annotator, or corrector.
addrLine: (address line) contains one line of a postal address.
address: contains a postal address, for example of a publisher, an organization, or an individual.
anchor: (anchor point) attaches an identifier to a point within a text, whether or not it corresponds with a textual element.
appInfo: (application information) records information about an application which has edited the TEI file.
application: provides information about an application which has acted upon the document.
argument: A formal list or prose description of the topics addressed by a subdivision of a text.
att: (attribute) contains the name of an attribute appearing within running text.
att.ascribed: provides attributes for elements representing speech or action that can be ascribed to a specific individual.
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.
att.datable: provides attributes for normalization of elements that contain dates, times, or datable events.
att.datable.w3c: provides attributes for normalization of elements that contain datable events using the W3C datatypes.
att.declarable: provides attributes for those elements in the TEI Header which may be independently selected by means of the special purpose decls attribute.
att.declaring: provides attributes for elements which may be independently associated with a particular declarable element within the header, thus overriding the inherited default for that element.
att.dimensions: provides attributes for describing the size of physical objects.
att.divLike: provides attributes common to all elements which behave in the same way as divisions.
att.docStatus: provides attributes for use on metadata elements describing the status of a document.
att.editLike: provides attributes describing the nature of a encoded scholarly intervention or interpretation of any kind.
att.global.analytic: provides additional global attributes for associating specific analyses or interpretations with appropriate portions of a text.
att.global.linking: defines a set of attributes for hypertext and other linking, which are enabled for all elements when the additional tag set for linking is selected.
att.handFeatures: provides attributes describing aspects of the hand in which a manuscript is written.
att.internetMedia: provides attributes for specifying the type of a computer resource using a standard taxonomy.

att.interpLike: provides attributes for elements which represent a formal analysis or interpretation.
att.measurement: provides attributes to represent a regularized or normalized measurement.
att.naming: provides attributes common to elements which refer to named persons, places, organizations etc.
att.placement: provides attributes for describing where on the source page or object a textual element appears.
att.pointing: defines a set of attributes used by all elements which point to other elements by means of one or more URI references.
att.ranging: provides attributes for describing numerical ranges.
att.responsibility: provides attributes indicating who is responsible for something asserted by the markup and the degree of certainty associated with it.
att.segLike: provides attributes for elements used for arbitrary segmentation.
att.sourced: provides attributes identifying the source edition from which some encoded feature derives.
att.spanning: provides attributes for elements which delimit a span of text by pointing mechanisms rather than by enclosing it.
att.tableDecoration: provides attributes used to decorate rows or cells of a table.
att.transcriptional: provides attributes specific to elements encoding authorial or scribal intervention in a text when transcribing manuscript or similar sources.
att.translatable: provides attributes used to indicate the status of a translatable portion of an ODD document.
att.typed: provides attributes which can be used to classify or subclassify elements in any way.
author: in a bibliographic reference, contains the name(s) of the author(s), personal or corporate, of a work; for example in the same form as that provided by a recognized bibliographic name authority.
authority: (release authority) supplies the name of a person or other agency responsible for making an electronic file available, other than a publisher or distributor.
availability: supplies information about the availability of a text, for example any restrictions on its use or distribution, its copyright status, etc.
back: (back matter) contains any appendixes, etc. following the main part of a text.
bibl: (bibliographic citation) contains a loosely-structured bibliographic citation of which the sub-components may or may not be explicitly tagged.
biblFull: (fully-structured bibliographic citation) contains a fully-structured bibliographic citation, in which all components of the TEI file description are present.
biblScope: (scope of citation) defines the scope of a bibliographic reference, for example as a list of page numbers, or a named subdivision of a larger work.
body: (text body) contains the whole body of a single unitary text, excluding any front or back matter.
byline: contains the primary statement of responsibility given for a work on its title page or at the head or end of the work.
cRefPattern: (canonical reference pattern) specifies an expression and replacement pattern for transforming a canonical reference into a URI.

catDesc: (category description) describes some category within a taxonomy or text typology, either in the form of a brief prose description or in terms of the situational parameters used by the TEI formal textDesc.
catRef: (category reference) specifies one or more defined categories within some taxonomy or text typology.
category: contains an individual descriptive category, possibly nested within a superordinate category, within a user-defined taxonomy.
cell: contains one cell of a table.
change: summarizes a particular change or correction made to a particular version of an electronic text which is shared between several researchers.
choice: groups a number of alternative encodings for the same point in a text.
cit: (cited quotation) contains a quotation from some other document, together with a bibliographic reference to its source. In a dictionary it may contain an example text with at least one occurrence of the word form, used in the sense being described, or a translation of the headword, or an example.
classCode: (classification code) contains the classification code used for this text in some standard classification system.
classDecl: (classification declarations) contains one or more taxonomies defining any classificatory codes used elsewhere in the text.
closer: groups together salutations, datelines, and similar phrases appearing as a final group at the end of a division, especially of a letter.
code: contains literal code from some formal language such as a programming language.
corr: (correction) contains the correct form of a passage apparently erroneous in the copy text.
creation: contains information about the creation of a text.
date: contains a date in any format.
dateline: contains a brief description of the place, date, time, etc. of production of a letter, newspaper story, or other work, prefixed or suffixed to it as a kind of heading or trailer.
del: (deletion) contains a letter, word, or passage deleted, marked as deleted, or otherwise indicated as superfluous or spurious in the copy text by an author, scribe, annotator, or corrector.
desc: (description) contains a brief description of the object documented by its parent element, including its intended usage, purpose, or application where this is appropriate.
distributor: supplies the name of a person or other agency responsible for the distribution of a text.
div: (text division) contains a subdivision of the front, body, or back of a text.
divGen: (automatically generated text division) indicates the location at which a textual division generated automatically by a text-processing application is to appear.
docAuthor: (document author) contains the name of the author of the document, as given on the title page (often but not always contained in a byline).
docDate: (document date) contains the date of a document, as given (usually) on a title page.
docEdition: (document edition) contains an edition statement as presented on a title page of a document.
docImprint: (document imprint) contains the imprint statement (place and date of publication, publisher name), as given (usually) at the foot of a title page.

docTitle: (document title) contains the title of a document, including all its constituents, as given on a title page.
edition: (edition) describes the particularities of one edition of a text.
editionStmt: (edition statement) groups information relating to one edition of a text.
editor: secondary statement of responsibility for a bibliographic item, for example the name of an individual, institution or organization, (or of several such) acting as editor, compiler, translator, etc.
editorialDecl: (editorial practice declaration) provides details of editorial principles and practices applied during the encoding of a text.
eg: (example) contains any kind of illustrative example.
emph: (emphasized) marks words or phrases which are stressed or emphasized for linguistic or rhetorical effect.
encodingDesc: (encoding description) documents the relationship between an electronic text and the source or sources from which it was derived.
epigraph: contains a quotation, anonymous or attributed, appearing at the start of a section or chapter, or on a title page.
expan: (expansion) contains the expansion of an abbreviation.
extent: describes the approximate size of a text as stored on some carrier medium, whether digital or non-digital, specified in any convenient units.
figDesc: (description of figure) contains a brief prose description of the appearance or content of a graphic figure, for use when documenting an image without displaying it.
figure: groups elements representing or containing graphic information such as an illustration or figure.
fileDesc: (file description) contains a full bibliographic description of an electronic file.
foreign: (foreign) identifies a word or phrase as belonging to some language other than that of the surrounding text.
formula: contains a mathematical or other formula.
front: (front matter) contains any prefatory matter (headers, title page, prefaces, dedications, etc.) found at the start of a document, before the main body.
funder: (funding body) specifies the name of an individual, institution, or organization responsible for the funding of a project or text.
gap: (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.
geoDecl: (geographic coordinates declaration) documents the notation and the datum used for geographic coordinates expressed as content of the <geo> element elsewhere within the document.
gi: (element name) contains the name (generic identifier) of an element.
gloss: identifies a phrase or word used to provide a gloss or definition for some other word or phrase.
graphic: indicates the location of an inline graphic, illustration, or figure.
group: contains the body of a composite text, grouping together a sequence of distinct texts (or groups of such texts) which are regarded as a unit for some purpose, for example the collected works of an author, a sequence of prose essays, etc.
head: (heading) contains any type of heading, for example the title of a section, or the heading of a list, glossary, manuscript description, etc.

hi: (highlighted) marks a word or phrase as graphically distinct from the surrounding text, for reasons concerning which no claim is made.
ident: (identifier) contains an identifier or name for an object of some kind in a formal language.
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.
index: (index entry) marks a location to be indexed for whatever purpose.
interp: (interpretation) summarizes a specific interpretative annotation which can be linked to a span of text.
interpGrp: (interpretation group) collects together a set of related interpretations which share responsibility or type.
item: contains one component of a list.
keywords: contains a list of keywords or phrases identifying the topic or nature of a text.
l: (verse line) contains a single, possibly incomplete, line of verse.
label: contains the label associated with an item in a list; in glossaries, marks the term being defined.
langUsage: (language usage) describes the languages, sublanguages, registers, dialects, etc. represented within a text.
language: characterizes a single language or sublanguage used within a text.
lb: (line break) marks the start of a new (typographic) line in some edition or version of a text.
lg: (line group) contains a group of verse lines functioning as a formal unit, e.g. a stanza, refrain, verse paragraph, etc.
list: (list) contains any sequence of items organized as a list.
listBibl: (citation list) contains a list of bibliographic citations of any kind.
macro.anyXML: defines a content model within which any XML elements are permitted
macro.limitedContent: (paragraph content) defines the content of prose elements that are not used for transcription of extant materials.
macro.paraContent: (paragraph content) defines the content of paragraphs and similar elements.
macro.phraseSeq: (phrase sequence) defines a sequence of character data and phrase-level elements.
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.
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.
macro.xtext: (extended text) defines a sequence of character data and gaiji elements.
measureGrp: (measure group) contains a group of dimensional specifications which relate to the same object, for example the height and width of a manuscript page.
mentioned: marks words or phrases mentioned, not used.

milestone: marks a boundary point separating any kind of section of a text, typically but not necessarily indicating a point at which some part of a standard reference system changes, where the change is not represented by a structural element.
model.addrPart: groups elements such as names or postal codes which may appear as part of a postal address.
model.addressLike: groups elements used to represent a postal or e-mail address.
model.applicationLike: groups elements used to record application-specific information about a document in its header.
model.biblLike: groups elements containing a bibliographic description.
model.biblPart: groups elements which represent components of a bibliographic description.
model.catDescPart: groups component elements of the TEI Header Category Description.
model.choicePart: groups elements (other than <choice> itself) which can be used within a <choice> alternation.
model.common: groups common chunk- and inter-level elements.
model.dateLike: groups elements containing temporal expressions.
model.div1Like: groups top-level structural divisions.
model.divBottom: groups elements appearing at the end of a text division.
model.divBottomPart: groups elements which can occur only at the end of a text division.
model.divGenLike: groups elements used to represent a structural division which is generated rather than explicitly present in the source.
model.divLike: groups elements used to represent un-numbered generic structural divisions.
model.divPart: groups paragraph-level elements appearing directly within divisions.
model.divTop: groups elements appearing at the beginning of a text division.
model.divTopPart: groups elements which can occur only at the beginning of a text division.
model.divWrapper: groups elements which can appear at either top or bottom of a textual division.
model.editorialDeclPart: groups elements which may be used inside <editorialDecl> and appear multiple times.
model.egLike: groups elements containing examples or illustrations.
model.emphLike: groups phrase-level elements which are typographically distinct and to which a specific function can be attributed.
model.encodingDescPart: groups elements which may be used inside <encodingDesc> and appear multiple times.
model.entryPart: groups elements appearing at any level within a dictionary entry.
model.entryPart.top: groups high level elements within a structured dictionary entry
model.frontPart: groups elements which appear at the level of divisions within front or back matter.
model.gLike: groups elements used to represent individual non-Unicode characters or glyphs.
model.global: groups elements which may appear at any point within a TEI text.

model.global.edit: groups globally available elements which perform a specifically editorial function.
model.global.meta: groups globally available elements which describe the status of other elements.
model.glossLike: groups elements which provide an alternative name, explanation, or description for any markup construct.
model.graphicLike: groups elements containing images, formulae, and similar objects.
model.headLike: groups elements used to provide a title or heading at the start of a text division.
model.hiLike: groups phrase-level elements which are typographically distinct but to which no specific function can be attributed.
model.highlighted: groups phrase-level elements which are typographically distinct.
model.imprintPart: groups the bibliographic elements which occur inside imprints.
model.inter: groups elements which can appear either within or between paragraph-like elements.
model.lLike: groups elements representing metrical components such as verse lines.
model.labelLike: groups elements used to gloss or explain other parts of a document.
model.limitedPhrase: groups phrase-level elements excluding those elements primarily intended for transcription of existing sources.
model.listLike: groups list-like elements.
model.measureLike: groups elements which denote a number, a quantity, a measurement, or similar piece of text that conveys some numerical meaning.
model.milestoneLike: groups milestone-style elements used to represent reference systems.
model.msItemPart: groups elements which can appear within a manuscript item description.
model.msQuoteLike: groups elements which represent passages such as titles quoted from a manuscript as a part of its description.
model.nameLike: groups elements which name or refer to a person, place, or organization.
model.nameLike.agent: groups elements which contain names of individuals or corporate bodies.
model.noteLike: groups globally-available note-like elements.
model.pLike: groups paragraph-like elements.
model.pLike.front: groups paragraph-like elements which can occur as direct constituents of front matter.
model.pPart.data: groups phrase-level elements containing names, dates, numbers, measures, and similar data.
model.pPart.edit: groups phrase-level elements for simple editorial correction and transcription.
model.pPart.editorial: groups phrase-level elements for simple editorial interventions that may be useful both in transcribing and in authoring.
model.pPart.transcriptional: groups phrase-level elements used for editorial transcription of pre-existing source materials.
model.personPart: groups elements which form part of the description of a person.

model.phrase: groups elements which can occur at the level of individual words or phrases.
model.phrase.xml: groups phrase-level elements used to encode XML constructs such as element names, attribute names, and attribute values
model.profileDescPart: groups elements which may be used inside <profileDesc> and appear multiple times.
model.ptrLike: groups elements used for purposes of location and reference.
model.publicationStmtPart: groups elements which may appear within the <publicationStmt> element of the TEI Header.
model.qLike: groups elements related to highlighting which can appear either within or between chunk-level elements.
model.quoteLike: groups elements used to directly contain quotations.
model.resourceLike: groups non-textual elements which may appear together with a header and a text to constitute a TEI document.
model.respLike: groups elements which are used to indicate intellectual or other significant responsibility, for example within a bibliographic element.
model.segLike: groups elements used for arbitrary segmentation.
model.sourceDescPart: groups elements which may be used inside <sourceDesc> and appear multiple times.
model.stageLike: groups elements containing stage directions or similar things defined by the module for performance texts.
model.teiHeaderPart: groups high level elements which may appear more than once in a TEI Header.
model.titlepagePart: groups elements which can occur as direct constituents of a title page, such as <docTitle>, <docAuthor>, <docImprint>, or <epigraph>.
name: (name, proper noun) contains a proper noun or noun phrase.
note: contains a note or annotation.
notesStmt: (notes statement) collects together any notes providing information about a text additional to that recorded in other parts of the bibliographic description.
num: (number) contains a number, written in any form.
opener: groups together dateline, byline, salutation, and similar phrases appearing as a preliminary group at the start of a division, especially of a letter.
orig: (original form) contains a reading which is marked as following the original, rather than being normalized or corrected.
p: (paragraph) marks paragraphs in prose.
pb: (page break) marks the boundary between one page of a text and the next in a standard reference system.
pc: (punctuation character) a character or string of characters regarded as constituting a single punctuation mark.
postscript: contains a postscript, e.g. to a letter.
principal: (principal researcher) supplies the name of the principal researcher responsible for the creation of an electronic text.
profileDesc: (text-profile description) provides a detailed description of non-bibliographic aspects of a text, specifically the languages and sublanguages used, the situation in which it was produced, the participants and their setting.
projectDesc: (project description) describes in detail the aim or purpose for which an electronic file was encoded, together with any other relevant information concerning the process by which it was assembled or collected.
ptr: (pointer) defines a pointer to another location.

pubPlace: (publication place) contains the name of the place where a bibliographic item was published.
publicationStmt: (publication statement) groups information concerning the publication or distribution of an electronic or other text.
publisher: provides the name of the organization responsible for the publication or distribution of a bibliographic item.
q: (separated from the surrounding text with quotation marks) contains material which is marked as (ostensibly) being somehow different than the surrounding text, 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.
quote: (quotation) contains a phrase or passage attributed by the narrator or author to some agency external to the text.
ref: (reference) defines a reference to another location, possibly modified by additional text or comment.
refState: (reference state) specifies one component of a canonical reference defined by the milestone method.
refsDecl: (references declaration) specifies how canonical references are constructed for this text.
reg: (regularization) contains a reading which has been regularized or normalized in some sense.
relatedItem: contains or references some other bibliographic item which is related to the present one in some specified manner, for example as a constituent or alternative version of it.
resp: (responsibility) contains a phrase describing the nature of a person's intellectual responsibility.
respStmt: (statement of responsibility) supplies a statement of responsibility for the intellectual content of a text, edition, recording, or series, where the specialized elements for authors, editors, etc. do not suffice or do not apply.
revisionDesc: (revision description) summarizes the revision history for a file.
row: contains one row of a table.
rs: (referencing string) contains a general purpose name or referring string.
s: (s-unit) contains a sentence-like division of a text.
said: (speech or thought) indicates passages thought or spoken aloud, whether explicitly indicated in the source or not, whether directly or indirectly reported, whether by real people or fictional characters.
salute: (salutation) contains a salutation or greeting prefixed to a foreword, dedicatory epistle, or other division of a text, or the salutation in the closing of a letter, preface, etc.
samplingDecl: (sampling declaration) contains a prose description of the rationale and methods used in sampling texts in the creation of a corpus or collection.
scriptNote: describes a particular script distinguished within the description of a manuscript or similar resource.
seg: (arbitrary segment) represents any segmentation of text below the "chunk" level.
seriesStmt: (series statement) groups information about the series, if any, to which a publication belongs.
sic: (latin for thus or so) contains text reproduced although apparently incorrect or inaccurate.

signed: (signature) contains the closing salutation, etc., appended to a foreword, dedicatory epistle, or other division of a text.
soCalled: contains a word or phrase for which the author or narrator indicates a disclaiming of responsibility, for example by the use of scare quotes or italics.
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.
sp: (speech) An individual speech in a performance text, or a passage presented as such in a prose or verse text.
speaker: A specialized form of heading or label, giving the name of one or more speakers in a dramatic text or fragment.
sponsor: specifies the name of a sponsoring organization or institution.
stage: (stage direction) contains any kind of stage direction within a dramatic text or fragment.
table: contains text displayed in tabular form, in rows and columns.
taxonomy: defines a typology used to classify texts either implicitly, by means of a bibliographic citation, or explicitly by a structured taxonomy.
teiCorpus: contains the whole of a TEI encoded corpus, comprising a single corpus header and one or more TEI elements, each containing a single text header and a text.
teiHeader: (TEI Header) supplies the descriptive and declarative information making up an electronic title page prefixed to every TEI-conformant text.
term: contains a single-word, multi-word, or symbolic designation which is regarded as a technical term.
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.
textClass: (text classification) groups information which describes the nature or topic of a text in terms of a standard classification scheme, thesaurus, etc.
time: contains a phrase defining a time of day in any format.
title: contains a title for any kind of work.
titlePage: (title page) contains the title page of a text, appearing within the front or back matter.
titlePart: contains a subsection or division of the title of a work, as indicated on a title page.
titleStmt: (title statement) groups information about the title of a work and those responsible for its intellectual content.
trailer: contains a closing title or footer appearing at the end of a division of a text.
typeNote: describes a particular font or other significant typographic feature distinguished within the description of a printed resource.
unclear: contains a word, phrase, or passage which cannot be transcribed with certainty because it is illegible or inaudible in the source.
val: (value) contains a single attribute value.