The Association for Computers and the Humanities (ACH)

The Association for Computational Linguistics (ACL)

The Association for Literary and Linguistic Computing (ALLC)

Guidelines for Electronic Text Encoding and Interchange

Edited by

C. M. Sperberg-McQueen and Lou Burnard

TEI P2, Chapter 6
Text Encoding Initiative
Chicago, Oxford

December 7, 1992

Draft Version 2, December 7, 1992

Note

This is the current draft (as of December 7, 1992) of one section of the *Guidelines for Electronic Text Encoding and Interchange*, version 2 (also known as TEI P2), produced by the Text Encoding Initiative. This fascicle contains:

- a table of contents for TEI P2, indicating the status of each section as of the release date of this fascicle
- a draft for one section of TEI P2
- entries derived from this section for the reference list of tags of TEI P2 (if any)
- a User Response and Comment Form

Please feel free to send us your comments on the contents of this draft and their applicability to your needs. Comments may be sent to the TEI-L list server for general discussion, or directly to either of the TEI editors (adresses on the User Response and Comment form). Bear in mind in reading, however, that TEI P2 is a reference manual, not a tutorial in the use of electronic text, and unavoidably assumes either a certain amount of prior knowledge or patience with unfamiliar technical material. Note also that some matters which may appear obscure in one section may well be treated in more detail elsewhere in the full text of TEI P2.

TEI P2, like P1, is a working draft. The observant will find several typographic and presentational flaws in the current text: for example, cross-references are not properly resolved in the LaTeX versions and some parts of the reference material are not yet complete. We hope to rectify as many as possible of these flaws in the coming months, but have decided not to allow them to delay any further the release of the draft. We hope that those interested in the Guidelines will find the current state of the text readable and usable enough for the purpose of substantive comment. As far as possible, all comments received on this section will be taken into account in the preparation of the next version of the Guidelines, which is due by the end of 1992.

Updates on the status of the draft as a whole will be distributed automatically to all subscribers of the Listserv list TEI-L. To subscribe to TEI-L, send electronic mail to the address LISTSERV@UICVM (or Listserv@uicvm.uic.edu) containing the single line

```
subscribe tei-l J. Smith
```

(substituting your name for “J. Smith”). TEI-L is also the appropriate place to pose questions or offer public comments on the TEI guidelines and other relevant issues. The TEI-L file server contains all the sections of TEI P2 thus far released, as well as other TEI materials. Its contents are shadowed at various other sites around the world; for more information consult one of the following documents:

- TEI ED J8, “Obtaining the Second Version of the TEI Guidelines (TEI P2)” (describes how to retrieve electronic copies of TEI P2 and the various formats they are available in)

- TEI ED J9, “Obtaining Paper Copies of the Second Version of the TEI Guidelines (TEI P2)” (describes how to request paper copies of TEI P2, for those without electronic mail access)

These documents are available from the TEI-L file server or through the editors at the addresses given on the User Response and Comment Form.

This document is made possible in part by financial support from the U.S. National Endowment for the Humanities, an independent federal agency; 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. Direct and indirect support has also been received from the University of Illinois at Chicago, the Oxford University Computing Service, the University of Arizona, the University of Oslo, Queen's University (Kingston, Ont.), and Ohio State University. We acknowledge assistance with software acquisition from the following firms: SEMA Group and Yard Software (for Mark-It and Write-It), Software Exoterica (for CheckMark and Xtran), SoftQuad, Inc., (for Author/Editor and RulesBuilder), Electronic Book Technology (for demonstration copies of DynaText), Xerox Corporation (for Ventura Publisher).

Overall Table of Contents for TEI P2

The following list shows the expected contents of TEI P2 as a whole and indicates the current status of each chapter.

Contents

Part 2

Core Tags and General Rules

Chapter 6

Elements Available in All TEI DTDs

This chapter describes elements which may appear in any kind of text and the tags used to mark them in all TEI document types. Most of these elements are freely floating phrases, which can appear at any point within the textual structure, although most of them must be contained by a higher-level element of some kind (such as a paragraph). A few of the elements described in this chapter (for example, bibliographic citations and lists) have a comparatively well-defined internal structure, but most of them have no consistent inner structure of their own. In the general case, they contain only a few words, and are often identifiable in a conventionally printed text by the use of typographic conventions such as shifts of font, use of quotation or other punctuation marks, or other changes in layout.

To use the terminology introduced in section 3.4, “Element Classes and Other Parameter Entities,” [in separate fascicle], most of the elements described in this chapter are members of the class *phrase*, and a small number are members of the class *chunk*, or *inter*.

This chapter begins by describing the **<p>** tag used to mark paragraphs, which serve as the fundamental formal unit for running text in all base tag sets. This is followed, in section 6.2, “Ambiguous Punctuation,” on page 9, by a discussion of some specific problems associated with the interpretation of conventional punctuation, and the methods proposed by the current Guidelines for resolving ambiguities therein.

The next section (section 6.3, “Highlighting and Quotation,” on page 10) describes a number of phrase-level elements commonly marked by typographic features (and thus well-represented in conventional markup languages). These include features commonly marked by font shifts (section 6.3.2, “Emphasis, Foreign Words, and Unusual Language,” on page 12) and features commonly marked by quotation marks (section 6.3.3, “Quotation,” on page 16) as well as such features as terms, cited words, and glosses (section 6.3.4, “Terms, Glosses, and Cited Words,” on page 21).

The next section (section 6.4, “Names, Numbers, Dates, Abbreviations, and Addresses,” on page 24) describes several phrase-level and inter-level elements which, although often of interest for analysis or processing, are rarely explicitly identified in conventional printing. These include names (section 6.4.1, “Names,” on page 24), numbers and measures (section 6.4.2, “Numbers and Measures,” on page 26), dates and times (section 6.4.3, “Dates and times,” on page 28), abbreviations (section 6.4.4, “Abbreviations and Their Expansions,” on page 30), and addresses (section 6.4.5, “Addresses,” on page 32).

Section 6.5, “Simple Editorial Changes,” on page 33 introduces some phrase-level elements which may be used to record simple editorial emendation or correction of the encoded text. The tags described here constitute a simple subset of the full mechanism for encoding textual variation (described in full in chapter 21, “Text Criticism and Apparatus,” [in separate fascicle]) which should be adequate to most commonly encountered situations.

In the same way, the following section (section 6.6, “Simple Links and Cross References,” on page 41) presents only a subset of the facilities available for the encoding of cross-references or text-linkage. The full story may be found in section 16, “Segmentation and Alignment,” [in separate fascicle]; the tags presented here are intended to be usable for a wide variety of simple applications.

Sections 6.7, “Lists,” on page 44, and 6.8, “Notes, Annotation, and Indexing,” on page 48, describe two kinds of quasi-structural elements, lists and notes, which may appear either within chunk-level elements such as paragraphs, or between them. Several kinds of lists are catered for, of an arbitrary complexity. The section on notes discusses both notes found in the source and simple mechanisms for adding annotations of an interpretive nature during the encoding; again, only a subset of the facilities described in full elsewhere (specifically, in section 17, “Simple Analytic Mechanisms,” [in separate fascicle]) is discussed.

Next, section 6.9, “Reference Systems,” on page 52, describes methods of encoding within a text the conventional system or systems used when making references to it. Some reference systems have attained canonical authority and must be recorded to make the text useable in normal work; in other cases, a convenient reference system must be created by the creator or analyst of an electronic text.

Like lists and notes, the bibliographic citations discussed in section 6.10, “Bibliographic Citations and References,” on page 63, may be regarded as structural elements in their own right. A range of possibilities is presented for the encoding of bibliographic citations or references, which may be treated as simple phrases within a running text, or as highly-structured components suitable for inclusion in a bibliographic database.

Tags for passages of verse or drama (whether prose or verse) are discussed in section 6.11, “Passages of Verse or Drama,” on page 81.

The chapter concludes with a discussion of ways in which text may be segmented for arbitrary analytic purposes, in section 6.12, “Arbitrary Segments,” on page 87.

Except for those tags designed to be used in concurrent markup streams, all the elements described in this chapter occur in the *core* of TEI tags, defined by the following DTD fragments. First, the DTD defines a number of element classes used in the content models:

```
<!-- 6:  Elements available in all TEI dtds                   -->
<!-- TEIcore2:  Core Elements Available in All TEI DTDs       -->

<!-- Text Encoding Initiative: Guidelines for Electronic      -->
<!-- Text Encoding and Interchange                            -->

<!-- DRAFT version 2                                          -->


<!-- Definition of useful element classes, for use in         -->
<!-- defining content models.                                 -->
```

```
<!ENTITY % x.hqinter '' >
<!ENTITY % m.hqinter '%x.hqinter q | quote' >
<!ENTITY % x.hqphrase '' >
<!ENTITY % m.hqphrase '%x.hqphrase distinct | emph | foreign | gloss |
hi | mentioned | soCalled | term' >
<!ENTITY % x.data '' >
<!ENTITY % m.data '%x.data abbr | address | date | dateRange | expan |
measure | name | num | org | organization | person | place | time |
timeRange' >
<!ENTITY % x.edit '' >
<!ENTITY % m.edit '%x.edit add | corr | del | omit | orig | reg | sic' >
<!ENTITY % x.loc '' >
<!ENTITY % m.loc '%x.loc lb | pb | ptr | ref' >
<!ENTITY % x.refsys '' >
<!ENTITY % m.refsys '%x.refsys milestone' >
<!ENTITY % x.lists '' >
<!ENTITY % m.lists '%x.lists label | list | list.bibl' >
<!ENTITY % x.notes '' >
<!ENTITY % m.notes '%x.notes note | span' >
<!ENTITY % x.bibl '' >
<!ENTITY % m.bibl '%x.bibl ' >
<!ENTITY % x.seg '' >
<!ENTITY % m.seg '%x.seg s' >

<!ENTITY % x.phrase '' >
<!ENTITY % m.phrase '%x.phrase %m.data; | %m.edit; | %m.hqphrase; |
%m.loc; | %m.seg;' >
<!ENTITY % x.inter '' >
<!ENTITY % m.inter '%x.inter %m.bibl; | %m.hqinter; | %m.lists; |
%m.notes;' >
<!ENTITY % x.chunk '' >
<!ENTITY % m.chunk '%x.chunk l | lg | p | sp | stage' >
<!-- (continued in sec. 6, 6)                                  -->
```

Next, some useful basic content models are defined; these will be used in the formal declaration of many elements:

```
<!-- 6:  Elements available in all TEI dtds (cont'd)          -->
<!-- (continuation of sec. )                                   -->
<!-- Definition of useful content models.                      -->


<!ENTITY % phrase.seq '(#PCDATA | %m.phrase;)*'  >
<!ENTITY % chunk.seq '(%m.inter; | %m.chunk;)*' >
<!ENTITY % paraContent '(#PCDATA | %m.inter; | %m.phrase;)*'  >
<!ENTITY % specialPara '(((%m.chunk;), (%m.chunk; | %m.inter;)*) |
(%paraContent;))' >
```

Finally, the elements of the TEI core are themselves defined:

```
<!-- 6:  Elements available in all TEI dtds (cont'd)          -->
<!-- (continuation of sec. )                                  -->
<!-- Definition of elements, sub-group by sub-group.          -->

<!-- ... declarations from section 6.1                        -->
<!--     (Paragraph)                                          -->
<!--     go here ...                                          -->
<!-- ... declarations from section 6.3.2.1                    -->
<!--     (Highlighted phrases)                                -->
<!--     go here ...                                          -->
<!-- ... declarations from section 6.4.1                      -->
<!--     (Proper Nouns)                                       -->
<!--     go here ...                                          -->
<!-- ... declarations from section 6.4.2                      -->
<!--     (Numbers and measures)                               -->
<!--     go here ...                                          -->
<!-- ... declarations from section 6.4.3                      -->
<!--     (Dates and times)                                    -->
<!--     go here ...                                          -->
<!-- ... declarations from section 6.5.1                      -->
<!--     (Editorial tags for correction)                      -->
<!--     go here ...                                          -->
<!-- ... declarations from section 6.5.2                      -->
<!--     (Editorial tags for regularization)                  -->
<!--     go here ...                                          -->
<!-- ... declarations from section 6.5.3                      -->
<!--     (Other editorial tags)                               -->
<!--     go here ...                                          -->
<!-- ... declarations from section 6.4.5                      -->
<!--     (Addresses and their components)                     -->
<!--     go here ...                                          -->
<!-- ... declarations from section 6.6                        -->
<!--     (Simple cross references)                            -->
<!--     go here ...                                          -->
<!-- ... declarations from section 6.7                        -->
<!--     (Lists and List Items)                               -->
<!--     go here ...                                          -->
<!-- ... declarations from section 6.8.1                      -->
<!--     (Annotation)                                         -->
<!--     go here ...                                          -->
<!-- ... declarations from section 6.9.5                      -->
<!--     (Milestone tag)                                      -->
<!--     go here ...                                          -->
<!-- ... declarations from section 6.10.1                     -->
<!--     (Tags for Bibliographic Citations)                   -->
<!--     go here ...                                          -->
<!-- ... declarations from section 6.11.1                     -->
<!--     (Verse)                                              -->
<!--     go here ...                                          -->
<!-- ... declarations from section 6.11.2                     -->
<!--     (Drama)                                              -->
<!--     go here ...                                          -->
<!-- ... declarations from section 6.12                       -->
```

```
<!--     (Arbitrary Segments)                                      -->
<!--     go here ...                                               -->
```

6.1 Paragraphs

The paragraph is the fundamental organizational unit for all prose texts, being the smallest regular unit into which prose can be divided. Prose can appear in all TEI texts, not simply in those using the prose base (section 7, “Base Tag Set for Prose,” [in separate fascicle]); the paragraph is therefore described here, as an element which can appear in any kind of text.

Paragraphs can contain any of the other elements described within this chapter, as well as some other elements which are specific to individual text types. We distinguish *phrase-level* elements, which must be entirely contained within a paragraph and cannot appear except within one, from *chunks*, which can appear between, but not within, paragraphs, and from *inter-level* elements, which can appear either within a single paragraph or between paragraphs. The class of phrases includes emphasized or quoted phrases, names, dates, etc. The class of inter-level elements includes bibliographic citations, notes, lists, etc. The class of chunks includes the paragraph, itself.

Because paragraphs may appear in different base or additional tagsets, their possible contents may differ in different kinds of document. In particular, additional elements not listed in this chapter may appear in paragraphs in certain kinds of text. However, the elements described in this chapter are always by default available in all kinds of text.

The paragraph is marked using the **<p>** element:

<p> : marks paragraphs in prose.

If a consistent internal subdivision of paragraphs is desired, the **<s>** (“segment”) element should be used, as discussed in section 6.12, “Arbitrary Segments,” on page 87. More usually however, paragraphs have no firm internal structure, but contain prose encoded as a mix of characters, entity references, phrases marked as described in the rest of this chapter, and embedded elements like lists, figures, or tables.

Since paragraphs are usually explicitly marked in Western texts, typically by indentation, the application of the **<p>** tag usually presents few problems.

In some cases, the body of a text may comprise but a single paragraph:

```
<body>
<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.
</body>
```

This news story shows typically short journalistic paragraphs:

```
<head>SARAJEVO, Bosnia and Herzegovina, April 19</head>
<p>Serbs seized more territory in this struggling new
country today as the United States Air Force ended a
two-day airlift of humanitarian aid into the capital,
Sarajevo.
<p>International relief workers called on European
Community nations to step up their humanitarian aid to
the former Yugoslav republic, in conjunction with new
American aid flights if necessary.
<p>A special envoy from the European Community, Colin
Doyle, harshly condemned the decision by Serbs to shell
Sarajevo on Saturday night during a visit to the Bosnian
capital by a senior American official, Deputy Assistant
Secretary of State Ralph R. Johnson.
<p>...
```

The following extract from a Russian fairy tale demonstrates how other phrase level elements (in this case **<q>** elements representing direct speech; see section 6.3.3, “Quotation,” on page 16) may be nested within, but not across, paragraphs:

```
<p>A fly built a castle, a tall and mighty castle.
There came to the castle the Crawling Louse.  <q>Who,
who's in the castle?  Who, who's in your house?</q>
said the Crawling Louse.  <q>I, I, the Languishing Fly.
And who art thou?</q> <q>I'm the Crawling Louse.</q>

<p>Then came to the castle the Leaping Flea.  <q>Who,
who's in the castle?</q> said the Leaping Flea.  <q>I,
I, the Languishing Fly, and I, the Crawling Louse.  And
who art thou?</q> <q>I'm the Leaping Flea.</q>

<p>Then came to the castle the Mischievous Mosquito.
<q>Who, who's in the castle?</q> said the Mischievous
Mosquito.  <q>I, I, the Languishing Fly, and I, the
Crawling Louse, and I, the Leaping Flea.  And who art
thou?</q> <q>I'm the Mischievous Mosquito.</q>
```

The **<p>** element is formally declared as follows:

```
<!-- 6.1:  Paragraph                                         -->
<!ELEMENT p             - O  (%paraContent;)                   >
<!ATTLIST p                  %a.global;                        >
<!-- This fragment is used in sec. 6                         -->
```

6.2 Ambiguous Punctuation

Punctuation marks cause problems for text markup because they may not be available in the character set used and because they are often ambiguous. In the former case entity names should be used to render the punctuation mark (see 4, “Characters and Character Sets,” [in separate fascicle]). In the latter case, ambiguous punctuation may be treated as described below.

Full stop (period) may mark (orthographic) sentence boundaries, abbreviations, decimal points, or serve as a visual aid in printing numbers. These usages can be distinguished by tagging S-units, abbreviations, and numbers, as described in sections 6.12, “Arbitrary Segments,” on page 87, 6.4.4, “Abbreviations and Their Expansions,” on page 30, and 6.4.2, “Numbers and Measures,” on page 26. There are independent reasons for tagging these, whether or not they are marked by full stops. Alternatively, the following TEI-specific entity names may be used to distinguish stops (and other characters) used for these purposes:

stop.abbr: a stop used to end an abbreviation
stop.sent: a stop used to end a sentence
stop.abse: a stop used both to end an abbreviation and to end a sentence
stop.dec: a stop used as a decimal point
comma.dec: a comma used as a decimal point
middot.dec: a midline dot used as a decimal point
stop.space: a stop used as a numeric space character
comma.space: a comma used as a numeric space character

These entities are defined in the file *teipunc2*, which is documented in chapter 39, “Standard Writing System Declarations,” [in separate fascicle].

Question mark and *exclamation mark* typically mark the end of orthographic sentences, but may also be used as a mid-sentence comment by the author (! to express surprise or some other strong feeling, ? to query a word or expression or mark a sentence as dubious in linguistic discussion). These uses may be distinguished by marking S-units, in which case the mid-sentence uses of these punctuation marks may be left unmarked.

Hyphens at line-end may or may not indicate permanent (“hard”) hyphens in the word. Where the lineation of the machine-readable text differs from the original, the editor may eliminate soft (line-end) hyphens or replace them by a reference to the entity *SHY* (“soft hyphen”), which is defined in the standard public entity set *ISOnum* (which should be invoked in the DTD subset if the entity *SHY* is to be used). The solution chosen should be reported in the **<hyphenation>** tag of the encoding declarations in the TEI header. See chapter 5, “The TEI Header,” [in separate fascicle] for discussion of the TEI header and encoding declarations.

Creators of machine-readable texts are recommended to avoid soft hyphens, as one cannot tell whether the hyphens are soft or hard in the case of compounds or prefixed words which might or might not be hyphenated in mid-sentence.

Dashes are best distinguished in form by using the entity names provided in the public entity set *ISOpub*, defined in ISO 8879: *mdash*, *ndash*, and *dash* (the "true" hyphen). Dashes are used for a variety of purposes: insertion, interruption, new speaker (in dialogue), list item. In the latter two cases it is preferable to mark the underlying feature using the elements **<q>** or **<item>**, on which see section 6.3.3, "Quotation," on page 16, and section 6.7, "Lists," on page 44, respectively.

Quotation marks should generally be replaced by the tags **<q>** or **<quote>**, especially as quotations are not always marked by quotation marks (notably long quotations) or may be marked in a variety of ways; see the discussion of quotation and related features in section 6.3.3, "Quotation," on page 16.

Apostrophes must be distinguished from single quote marks. This is best done by tagging quotations or other uses of quotation marks (see above). However, apostrophes have a variety of uses. In English they mark contractions, genitive forms, and (occasionally) plural forms. Full disambiguation of these uses belongs to the level of linguistic analysis and interpretation.

Where punctuation marks are disambiguated by tagging the underlying feature they signal, it may be debated whether they should be excluded or left as part of the text. In the case of quotation marks, it may sometimes be more convenient to distinguish opening from closing marks simply by using the appropriate entity reference, rather than to use the **<q>** element, with or without a *rend* attribute. The solution chosen will vary depending upon the feature and depending upon the purpose of the project.

6.3 Highlighting and Quotation

This section deals with a variety of textual features, all of which have in common that they are frequently realized in conventional printing practice by the use of such features as underlining, italic fonts, or quotation marks, collectively referred to here as *highlighting*. After an initial discussion of this phenomenon and alternate approaches to encoding it, this section describes ways of encoding the following textual features, all of which are conventionally rendered using some kind of highlighting:

- emphasis, foreign words and other linguistically distinct uses of highlighting
- representation of speech and thought, quotation, etc.
- technical terms, glosses, etc.

6.3.1 What Is Highlighting?

By *highlighting* we mean the use of any combination of typographic features (font, size, hue, etc.) in a printed or written text in order to distinguish some passage of a text from its surroundings.[1] The purpose of highlighting is generally to

[1] Although the way in which a spoken text is performed, (for example, the voice quality, loudness, etc.) might be regarded as analogous to "highlighting" in this

draw the reader's attention to some feature or characteristic of the passage highlighted; this section describes the elements recommended by these Guidelines for the encoding of such textual features.

In conventionally printed modern texts, highlighting is often employed to identify words or phrases which are:

- *distinct* in some way — as foreign, dialectal, archaic, technical, etc.
- emphatic, and which would for example be stressed when spoken
- not part of the body of the text, for example cross references, titles, headings, labels, etc.
- identified with a distinct narrative stream, for example an internal monologue or commentary.
- attributed by the narrator to some other agency, either within the text or outside it: for example direct speech or quotation
- set apart from the text in some other way, for example proverbial phrases, words mentioned but not used, names of persons and places in older texts, editorial corrections or additions, etc.

The textual functions signalled by highlighting may not be rendered consistently in different parts of a text or in different texts. (For example, a foreign word may appear in italics if the surrounding text is in roman, but in roman if the surrounding text is in italics). For this reason, these Guidelines distinguish between the encoding of rendering itself and the encoding of the underlying feature expressed by it.

Highlighting as such may be encoded by using the global *rend* attribute which can be specified for any element in the TEI scheme. This allows the encoder both to specify the function of a highlighted phrase or word, by selecting the appropriate element described here or elsewhere in the Guidelines, and to further describe the way in which it is highlighted, by means of the *rend* attribute. If the encoder wishes to offer no interpretation of the feature underlying the use of highlighting in the source text, then the **<hi>** element may be used, which indicates only that the text so tagged was highlighted in some way.

The possible values carried by the *rend* attribute are not formally defined in this version of the Guidelines. Since the *rend* attribute may be used to document any peculiarity of the way a given segment of text was rendered in the original source text, it may need to express a very large range of typographic features, by no means restricted to type face, type size, etc. For more discussion see the chapter on physical rendition (section 20, “Manuscripts, Analytic Bibliography, and Physical Description of the Source Text,” [in separate fascicle]).

Where it is both appropriate and feasible, these Guidelines recommend that the textual feature marked by the highlighting should be encoded, rather than just the simple fact of the highlighting. This is for the following reasons:

---

sense, these Guidelines recommend distinct elements for the encoding of such “highlighting” in spoken texts. See further section 10, “Base Tag Set for Transcriptions of Spoken Texts,” [in separate fascicle].

- the same kind of highlighting may be used for different purposes in different contexts
- the same textual function may be highlighted in different ways in different contexts
- for analytic purposes, it is in general more useful to know the intended function of a highlighted phrase than simply that it is distinct.

In many, if not most, cases the underlying function of a highlighted phrase will be obvious and non-controversial, since the distinctions indicated by a change of highlighting correspond with distinctions discussed elsewhere in these Guidelines. It should be recognized, however, that cases do exist in which is it not economically feasible to mark the underlying function of highlighting (e.g. in the preparation of large text corpora), as well as cases in which it is not intellectually appropriate (as in the transcription of some older materials, or in the preparation of material for the study of typographic practice). In such cases, the **<hi>** element should be used, as further discussed below.

Elements which are sometimes realized by typographic distinction but which are not discussed in this section include **<title>** (discussed in section 6.10, "Bibliographic Citations and References," on page 63) and **<name>** (discussed in section 6.4.1, "Names," on page 24).

6.3.2 Emphasis, Foreign Words, and Unusual Language

This subsection discusses the following elements:

<foreign> : identifies a word or phrase as belonging to some language other than that of the surrounding text. Attributes include:
  lang : identifies the language of the word or phrase marked.
<emph> : marks words or phrases which are stressed or emphasized for linguistic or rhetorical effect.
<hi> : marks a word or phrase as typographically distinct from the surrounding text, for reasons concerning which no claim is made. Attributes include:
  rend : describes the rendition or presentation of the word or phrase highlighted.
<distinct> : identifies any word or phrase which is regarded as linguistically distinct, for example as archaic, technical, dialectal, non-preferred, etc., or as forming part of a sublanguage. Attributes include:
  type : specifies the sublanguage or register to which the word or phrase is being assigned
  time : specifies how the phrase is distinct diachronically
  space : specifies how the phrase is distinct diatopically
  social : specifies how the phrase is distinct diastatically

6.3.2.1 Foreign Words or Expressions

Words or phrases which are not in the main language of the text should be tagged as such, at least where the fact is indicated in the text. Where the word or phrase concerned is already distinguished from the rest of the text by virtue of its function (for example, because it is a name, a technical term, a quotation, a mentioned word, etc.) then the global *lang* attribute should be used to specify additionally that its language distinguishes it from the surrounding text. Any element in the TEI scheme may take a *lang* attribute, which specifies both the writing system and the language used by its content (see section 4, “Characters and Character Sets,” [in separate fascicle] for discussion of this attribute). Where there is no other applicable element, the tag **<foreign>** may be used to provide a peg onto which the *lang* may be attached.

```
<q>Aren't you confusing <foreign lang=la>post hoc</foreign>
with <foreign lang=la>propter hoc</foreign>?</q> said the
Bee Master.  <q>Wax-moth only succeed when weak bees let
them in.</q>
```

The **<foreign>** tag should not be used to encode foreign words which are mentioned or glossed within the text: for these use the appropriate element from section 6.3.4, “Terms, Glosses, and Cited Words,” on page 21 below. Compare the following example sentences:

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

<mentioned lang=fr>croissant</mentioned> is difficult to
pronounce with your mouth full.

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

The **<foreign>** element is formally defined as follows:

```
<!-- 6.3.2.1:  Highlighted phrases                          -->
<!ELEMENT foreign       - -  (%phrase.seq;)                   >
<!ATTLIST foreign            %a.global;
          lang               IDREF              #IMPLIED      >
<!-- (continued in sec. 6.3.2.2, 6.3.2.3, 6.3.3, 6.3.4)      -->
<!-- This fragment is used in sec. 6                        -->
```

6.3.2.2 Emphatic Words and Phrases

The **<emph>** element is provided to mark words or phrases which are *linguistically* emphatic or stressed. Text which is only typographically “emphasized” falls into the class of highlighted text, and may be tagged with the **<hi>** element. In printed works, emphasis is generally indicated by devices such as the use of an italic font, a large typeface or extra wide letter spacing; in manuscripts and typescripts, it is usually indicated by the use of underlining. As the following examples demonstrate, an encoder may choose whether or not to make explicit the particular type of rendition associated with the emphasis, by use of the *rend* attribute. If a source text consistently renders a particular feature (e.g. emphasis or words in foreign languages) in a particular way, the rendering associated with that feature may be described in the TEI header and the *rend* attribute used only to describe examples which deviate from the norm.

```
<q>Sex, sir, is <emph>purely</emph> a question of appetite!</q>
   Tarr exclaimed.

<q>What it all comes to is this,</q> he said.
<q><emph rend=italic>What does Christopher
Robin do in the morning nowadays?</emph></q>

<l>Here Thou, great <name rend=italics>Anna</name>!
      whom three Realms obey,
<l>Doth sometimes Counsel take &mdash;
      and sometimes <emph rend=italic>Tea</emph>.
```

The **<hi>** element is used to mark words or phrases which are highlighted in some way, but for which identification of the intended distinction is difficult, controversial or impossible. It enables an encoder simply to record the fact of highlighting, possibly describing it by the use of a *rend* attribute, as discussed above, without however taking a position as to the function of the highlighting. This may also be useful if the text is to be processed in two stages: representing simply typographic distinctions during a first pass, and then replacing the **<hi>** tags with more specific tags in a second pass.

Some simple examples:

```
<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 ...
```

In this example, the first highlighted phrase uses black letter or gothic print to mimic the appearance of a legal document, and italic to mark “Walter Shandy” as a name. In a second pass, the elements **<head>** or **<label>** might be appropriate for the first use, and the element **<name>** for the second.

```
The heaviest rain, and snow, and hail, and sleet, could
boast of the advantage over him in only one respect.  They
often <hi rend=quoted>came down</hi> handsomely, and
Scrooge never did.
```

In this example, the phrase “came down” uses inverted commas to indicate a play on words.[2] In a second pass, the element **<soCalled>** might be preferred.

The **<emph>** and **<hi>** elements are formally defined as follows:

```
<!-- 6.3.2.2:  Highlighted phrases (cont'd)                   -->
<!-- (continuation of sec. 6.3.2.1)                           -->
<!ELEMENT emph          - -  (%phrase.seq;)                     >
<!ATTLIST emph               %a.global;                         >
<!ELEMENT hi            - -  (%phrase.seq)                      >
```
&@fn@sep

[2] The Oxford English Dictionary documents the phrase “to come down” in the sense “to bring or put down; *esp.* to lay down money; to make a disbursement” as being in use, mostly in colloquial or humorous contexts, from at least 1700 to the latter half of the 19th century.

```
<!ATTLIST hi              %a.global;
          rend            CDATA            #IMPLIED       >
```

6.3.2.3 Other Linguistically Distinct Material

For some kinds of analysis, it may be desirable to encode the linguistic distinctiveness of words and phrases with more delicacy than is allowed by the **<foreign>** element. The **<distinct>** element is provided for this purpose. Its attributes allow for additional information characterizing the nature of the linguistic distinction to be made in two distinct ways: the *type* attribute simply assigns a user-defined code of some kind to the word or phrase which assigns it to some register, sub-language, etc. No recommendations as to the set of values for this attribute are provided at this time, as little consensus exists in the field.

Alternatively, the remaining three attributes may be used in combination to place a word or phrase on a three-dimensional scale sometimes used in descriptive linguistics.[3] The *time* attribute places a word *diachronically*, for example as archaic, old-fashioned, contemporary, futuristic, etc.; the *place* attribute places a word *diatopically*, that is, with respect to a geographical classification, for example as national, regional, international, etc.; the *social* attribute places a word *diastatically*, that is, with respect to a social classification, for example as technical, polite, impolite, restricted, etc. Again, no recommendations are made for the values of these attributes at this time; the encoder should provide a description of the scheme used in the appropriate section of the header (see section 5.3, "The Encoding Description," [in separate fascicle]).

Examples:

```
Next morning a boy in that dormitory confided to his
bosom friend, a <distinct type=psSlang>fag</distinct> of
Macrea's, that there was trouble in their midst which
King <distinct type=archaic>would fain</distinct> keep
secret.

Next morning a boy in that dormitory confided to his
bosom friend, a
<distinct time=1900 social=publicschool place=GB>fag</distinct>
of Macrea's, that there was trouble in their midst which
King <distinct time=archaic>would fain</distinct> keep
secret.
```

Where more complex (or more rigorous) interpretive analyses of the associations of a word are required, the more detailed and general mechanisms described in section 18, "Feature Structure Analysis," [in separate fascicle] should be preferred to these simple characterizations. It may also be preferable to record the kinds of analysis suggested here by means of the simple annotation elements &@fn@sep

[3] See, for example, *Sociolinguistics/Soziolinguistik (An international handbook of the science of language and society. Ein internationales Handbuch zur Wissenschaft von Sprache und Gesellschaft)* (Berlin, New York: De Gruyter, 1988), I, pp. 271 and 274.

<note> or **<span>** described in section 6.8, “Notes, Annotation, and Indexing,” on page 48.

The **<distinct>** element has the following formal definition:

```
<!-- 6.3.2.3:  Highlighted phrases (cont'd)                   -->
<!-- (continuation of sec. 6.3.2.1)                           -->
<!ELEMENT distinct      - -  (%phrase.seq;)                     >
<!ATTLIST distinct           %a.global;
          type               CDATA              #IMPLIED
          time               CDATA              #IMPLIED
          space              CDATA              #IMPLIED
          social             CDATA              #IMPLIED       >
```

6.3.3 Quotation

This section discusses the following elements, all of which are often rendered by the use of quotation marks:

- **<q>** : contains a representation of the speech or thought of a character or speaker in a text. Attributes include:
  - *who* : identifies the speaker of a piece of direct speech
  - *type* : may be used to indicate whether the quoted matter is spoken or thought, or to characterize it more finely. Sample values include:
    - *spoken* : representation of direct speech, usually marked by quotation marks
    - *thought* : representation of thought, e.g. internal monologue
  - *direct* : may be used to indicate whether the quoted matter is regarded as direct or indirect speech. Sample values include:
    - *y* : speech or thought is represented directly
    - *n* : speech or thought is represented indirectly e.g. by use of a marked verbal aspect
    - *unspecified* : no claim is made
- **<quote>** : contains a phrase or passage attributed by the narrator or author to some agency external to the 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

One form of presentational variation found particularly frequently in written and printed texts is the use of quotation marks. As with the typographic variations discussed in the preceding section, it is generally helpful to separate the encoding of the underlying textual feature (for example, a quotation or a piece of direct speech) from the encoding of its rendering (for example, the use of a particular style of quotation marks).

The most common and important use of quotation marks is, of course, to mark *quotation*, by which we mean simply any part of the text attributed by the author or narrator to some agency other than the narrative voice. Typical examples include passages cited from other works, for which the element **<quote>** may be used, and words or phrases attributed to other voices within the current work, for which the element **<q>** may be used. If this distinction between intra-textual and inter-textual voices cannot be made reliably, or is not of interest, then all quoted matter may simply be marked using the **<q>** tag. The editorial policy in this respect should be stated in the encoding description of the TEI Header. The **<soCalled>** element is used for cases where the author or narrator distances him or herself from the words in question without however attributing them to any other voice in particular.

Quotation may be rendered by changes in type face, by special punctuation marks (single or double or angled quotes, dashes, etc.) and by layout (indented paragraphs, etc.). If these characteristics are of interest, an appropriate value for the *rend* attribute should be given, to record how the **<q>** or **<quote>** element is rendered. For discussion of suggested values for this attribute, see below, and also chapter 20, “Manuscripts, Analytic Bibliography, and Physical Description of the Source Text,” [in separate fascicle].

Quotation marks themselves may, like other punctuation marks, be felt for some purposes to be worth retaining within a text, quite independently of their description by the *rend* attribute. Where this is done, an appropriate entity reference should be chosen from the standard entity sets listed in chapter 4, “Characters and Character Sets,” [in separate fascicle]; this has the advantage that the entity may be redefined as null when the punctuation is to be ignored for some analytic purpose. Well-known ambiguities, such as whether the character ' represents an apostrophe or a closing single quotation mark, or whether the character " represents an opening or closing double quotation mark may all be resolved by the use of appropriate entity references, as discussed in section 6.2, “Ambiguous Punctuation,” on page 9.

Alternatively, the encoder may suppress all quotation marks, possibly recording their form using the *rend* attribute. Where this is done, the following list of entity names (taken from the public entity sets *ISOpub* and *ISOnum*) may be found useful to describe quotation-mark styles common in European and American typesetting:

ldquo: double inverted comma (shaped like 66, superscript)
lsquo: single inverted comma (shaped like 6, superscript)
rdquo: double apostrophe (shaped like 99, superscript)
rsquo: single apostrophe (shaped like 9, superscript)
ldquor: double comma (shaped like 99, printed on base line)
lsquor: single comma (shaped like 9, printed on base line)
laquo: double guillemet open to the right
lsaquo: single guillemet open to the right
raquo: double guillemet open to the left
rsaquo: single guillemet open to the left
mdash: dash the width of a lowercase *m*

These may be used in the *rend* attribute to show how the quotation was opened and closed. For example, if the words *pre* and *post* are used to indicate preceding and following punctuation, then the following example would describe a conventional American book printed using single quotation marks:

```
<q rend='PRE lsquo POST rsquo'>Who-e debel you?</q>
&mdash he at last said &mdash <q
rend='PRE lsquo POST rsquo'>you no speak-e,
damme, I kill-e.</q>  And so saying,
the lighted tomahawk began flourishing
about me in the dark.
```

The following example demonstrates alternative policies which may be adopted with respect to encoding of the punctuation used to mark quotation:

```
Adolphe se tourna vers lui :
<q>&mdash Alors, Albert, quoi de neuf?</q>
<q>&mdash Pas grand-chose.</q>
<q>&mdash Il fait beau,</q> dit Robert.
```

```
Adolphe se tourna vers lui :
<q rend='PRE mdash'>Alors, Albert, quoi de neuf ?</q>
<q rend='PRE mdash'>Pas grand-chose.</q>
<q rend='PRE mdash'>Il fait beau,</q> dit Robert.
```

To make explicit who is speaking, which is not always stated in the above example, the *who* attribute should be used:

```
Adolphe se tourna vers lui :
<q who='Adolphe'>&mdash Alors, Albert, quoi de neuf?</q>
<q who='Albert'>&mdash Pas grand-chose.</q>
<q who='Robert'>&mdash Il fait beau,</q> dit Robert.
```

The *who* attribute is also useful as a means of supplying a normalized form of the speaker's name, to facilitate selection of text by particular speakers. As indicated above, it may be supplied whether or not an indication of the speaker is given explicitly in the text.

Where investigation of "narrative voice" is the primary object of the encoding, it may be convenient to identify each speaker as a *participant* in the work, and to associate individual speeches with them by means of SGML's ID/IDREF mechanism. See section 26, "Additional Tags for the TEI Header," [in separate fascicle] for discussion of the *participant description* component of the TEI Header.

For such analyses, it may also be useful to distinguish representations of speech from representations of thought, in modern printed texts often indicated by a change of typeface. The *type* attribute should be used for this purpose, as in this example:

```
<q type=speech>Oh yes,</q> said Henry, <q type=speech>I mean
Gordon Macrae, for example...</q>
<q type=thought>Jungian Analyst with Winebox! That's what you
called him, you callous bastard, didn't you? Eh? Eh?</q>
```

Quoted matter may be embedded within quoted matter, as when one speaker reports the speech of another:

```
<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:&mdash;<q who=Spaulding>I wish to the Lord, Mr. Wilson, that
I was a red-headed man.</q></q>
```

Direct speech nested in this way is treated in the same way as elsewhere: a change of rendition may occur, but the same element should be used. An encoder may however choose to distinguish between direct speech which contains quotations from extra-textual matter and direct speech itself, as in the following example:

```
<p><q>The Lord! The Lord! It is Sakya Muni himself,</q> the
lama half sobbed; and under his breath began the wonderful
Buddhist invocation:-
<quote><l>To Him the Way -- the Law -- Apart --
<l>Whom Maya held beneath her heart
<l>Ananda's Lord -- the Bodhisat
</quote>
And He is here! The Most Excellent Law is here also.  My
pilgrimage is well begun.  And what work! What work!</q>
```

Quotations from other works are often accompanied by a reference to their source, This should be encoded using the elements for bibliographic references discussed in section 6.10, "Bibliographic Citations and References," on page 63, as in the following example.

```
<div id=MM01 type=chapter><head>Chapter 1</head>
<epigraph>
  <quote><l>Since I can do no good because a woman
  <l>Reach constantly at something that is near it.</quote>
  <bibl><title>The Maids Tragedy</title>
        <author>Beaumont and Fletcher</author></bibl>
</epigraph>
<p>Miss Brooke had that kind of beauty which seems to be thrown into
relief by poor dress...
```

Like other bibliographic references, the citation attached to a quotation may be represented simply by a pointer, as in this example:

```
Lexicography has shown little sign of being affected by the
work of followers of J.R. Firth, probably best summarized
in his slogan, <quote>You shall know a word by the company it
keeps.</quote> <ref target=FI57>(Firth, 1957)</ref>
```

Unlike most of the other elements discussed in this chapter, quotations may frequently contain other high-level elements such as paragraphs or verse lines, as well as being themselves contained by such elements. Three possible solutions exist for this well-known structural problem:

- the quotation is broken into segments, each of which is entirely contained within a paragraph
- the quotation is marked up as part of a concurrent but independent hierarchy
- the quotation boundaries are represented by empty milestone tags

The example of nested direct speech given above, for example, includes a paragraph break in the middle of the outermost **<q>**, which could be represented in any of the ways just noted. The outermost speech (Wilson's) can be broken up to fit into the series of paragraphs, using the *rend* attribute to record the absence of closing quotation marks at the end of each paragraph (the inner speeches, being punctuated conventionally, need not carry *rend* values):

```
<p><q who=Wilson rend='pre ldquo'>The first thing that
put us out was that advertisement.  Spaulding, he came
down into the office just this day eight weeks with
this very paper in his hand, and he
says:&mdash;</q></p>
<p><q who=Wilson rend='pre ldquo'><q who=Spaulding>I
wish to the Lord, Mr. Wilson, that I was a red-headed
man.</q></q></p>
<p><q who=Wilson rend='pre ldquo'><q who=Wilson>Why
that?</q> I asks.</q></p>
```

If quotations are marked as part of a distinct markup stream given the name “QD”, the outermost speech need not be broken up into multiple elements:

```
<(QD)q who=Wilson> ...
<(TEI.2)p>The first thing that put us out was that
advertisement.  Spaulding, he came down into the
office just this day eight weeks with this very paper
in his hand, and he says:&mdash;</(TEI.2)p>
<(TEI.2)p><(QD)q who=Spaulding>I wish to the Lord, Mr.
Wilson, that I was a red-headed man.</(QD)q></(TEI.2)p>
<p><(QD)q who=Wilson>Why
that?</(QD)q> I asks.</(TEI.2)p>
 ... </(QD)q>
```

If quotations are marked using (user-defined) empty elements given the names “QB” and “QE”, then no element contains the speeches and they need not be broken up into multiple elements at the paragraph breaks:

```
<qb who=Wilson> ...
<p>The first thing that put us out was that
advertisement.  Spaulding, he came down into the
office just this day eight weeks with this very paper
in his hand, and he says:&mdash;</p>
<p><qb who=Spaulding>I wish to the Lord, Mr.
Wilson, that I was a red-headed man.<qe></p>
<p><qb who=Wilson>Why
that?<qe> I asks.</p>
 ... <qe>
```

For further discussion, and several examples, see chapter 36, “Markup for Non-Hierarchical Phenomena,” [in separate fascicle].

Finally, in this section, the element **<soCalled>** is provided for all cases in which quotation marks are used to distance the quoted text from the narrator or speaker. Common examples include the “scare” quotes often found in newspaper headlines and advertising copy, where the effect is to cast doubts on the veracity of an assertion:

```
<head>PM dodges <soCalled>election threat</soCalled>
in interview</head>
```

The same element should be used to mark a variety of special ironic usages. Some further examples follow:

```
He hated <soCalled>good</soCalled> books.

<soCalled>Croissants</soCalled> indeed! toast not good enough for you?

Although Chomsky's decision that all NL sentences are finite
objects was never justified by arguments from the attested
properties of NLs, it did have a certain
<soCalled>social</soCalled> justification.  It was
commonly assumed in works on logic until fairly recently
that the notion <mentioned>language</mentioned> is
necessarily restricted to finite strings.
```

The elements discussed in this section are formally defined as follows:

```
<!-- 6.3.3:  Highlighted phrases (cont'd)                  -->
<!-- (continuation of sec. 6.3.2.1)                        -->
<!ELEMENT q            - -  (%phrase.seq;)                   >
<!ATTLIST q                 %a.global;
          type              CDATA              #IMPLIED
          direct            (y | n | unspecified)
                                               unspecified
          who               CDATA              #IMPLIED      >
<!ELEMENT quote        - -  (%seq;)                          >
<!ATTLIST quote             %a.global;                       >
<!ELEMENT soCalled     - -  (%phrase.seq;)                   >
<!ATTLIST soCalled          %a.global;                       >
```

6.3.4 Terms, Glosses, and Cited Words

This section describes the following textual elements, all of which have in common that they may be variously realized using italics, quotation marks or other devices:

<term> : contains a word or phrase regarded as a technical term.

<gloss> : identifies a phrase or word used to provide a gloss or definition for some other word or phrase
<mentioned> : marks words or phrases mentioned, not used

Technical terms are often italicized or emboldened upon first mention in printed texts; an explanation or gloss is sometimes given in quotation marks. Linguistic analyses conventionally cite words in languages under discussion in italics, providing a gloss immediately following marked with single quotation marks. Other texts in which individual words or phrases are *mentioned* (for example, as examples) rather than *used* may mark them either with italics or with quotation marks, and will gloss them less regularly.

A **<term>** may appear with or without a gloss, as may a **<mentioned>** element. Where the **<gloss>** is present, it may be linked to the term it is glossing by means of SGML's ID/IDREF mechanism. To establish such a link, the encoder should give an *id* value to the **<term>** or **<mentioned>** element and provide that id as the value of the *target* attribute on the **<gloss>** element. The following examples demonstrate this facility: for more discussion of this and other kinds of linkage within TEI documents, see chapter 16, “Segmentation and Alignment,” [in separate fascicle].

Examples:

```
We may define <term rend=sc id=tdpv>discoursal point of view</term>
as <gloss target=tdpv>the relationship, expressed through discourse
structure, between the implied author or some other addresser,
and the fiction.</gloss>

<gloss target=T1 rend=unmarked>A computational device
that infers structure from grammatical strings of words</gloss>
is known as a <term id=T1>parser</term>, and much of the
history of NLP over the last 20 years has been occupied with
the design of parsers.

There is thus a striking accentual difference between a
verbal form like <mentioned lang=Greek
id=cw234>eluthemen</mentioned> <gloss termid=cw234>we
were released,</gloss> accented on the second
syllable of the word, and its participial derivative
<mentioned id=cw235 lang=grc>lutheis</mentioned>
<gloss termid=cw235>released,</gloss> accented on the last.
```

The elements discussed in this section have the following formal definitions:

```
<!-- 6.3.4:  Highlighted phrases (cont'd)                    -->
<!-- (continuation of sec. 6.3.2.1)                          -->
<!ELEMENT term          - -  (%phrase.seq;)                    >
<!ATTLIST term               %a.global;                        >
<!ELEMENT mentioned     - -  (%phrase.seq;)                    >
<!ATTLIST mentioned          %a.global;                        >
<!ELEMENT gloss         - -  (%phrase.seq;)                    >
<!ATTLIST gloss              %a.global;
          target             IDREF              #IMPLIED       >
```

6.3.5 Some Further Examples

As a simple example of the elements discussed 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.

A first approximation to the encoding of this sentence, might be simply to record the fact that the phrases printed above in italics are highlighted, as follows:

```
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 lang=fr>romans
d'antiquit&eacute;</hi>, the romances of Chr&eacute;tien de
Troyes, ...
```

This encoding would however lose the important distinction between an italicized title and an italicized foreign phrase. Many other phrases might also be italicized in the text, and a retrieval program seeking to identify foreign terms (for example) would not be able to produce reliable results by simply looking for italicized words. Where economic and intellectual constraints permit, therefore, it would be preferable to encode both the function of the highlighted phrases and their appearance, as follows:

```
On the one hand the <title
rend=italic>Nibelungenlied</title> is associated with the
new rise of romance of twelfth-century France, the <foreign
rend=italic>romans d'antiquit&eacute;</foreign>, the
romances of Chr&eacute;tien de Troyes, ...
```

In this example, the decision as to which textual features are distinguished by the highlighting is relatively uncontroversial. As a less straightforward example, consider the use of italic font in the following passage from Samuel Richardson's *Clarissa* (1747).

> A pretty common case, I believe; in all *vehement* debatings. She says I am *too witty*; Anglicé, *too pert*; I, that she is *too wise*; that is to say, being likewise put into English, *not so young as she has been*: in short, she is grown so much into a *mother*, that she had forgotten she ever was a *daughter*. ...

Clearly, the word “vehement” is not italicized for the same reason as the phrase “not so young as she has been”; the former is emphasized, while the latter is proverbial. It also provides an ironic gloss for the words “too wise”, in the same way as “too pert” glosses “too witty”. The glossed phrases are not however technical terms or cited words, but quoted phrases, as if Clarissa were putting words into her own and her mother's mouths. Finally, the words “mother” and “daughter” are apparently italicized simply to oppose them in the sentence; certainly they do not fit into any of the categories so far proposed as reasons for italicizing. Note also that the word “Anglicé” is not italicized although it is not generally considered an English word.

The following sample encoding for the above passage attempts to take into account all the above points:

```
A pretty common case, I believe; in all
<hi rend=italic>vehement</emph> debatings.  She says I am
<term rend=italic>too witty</term>; <foreign lang=la
rend=roman>Anglic&egrave;</foreign>, <gloss rend=italic>too
pert</gloss>; I, that she is <term rend=italic>too wise</q>;
that is to say, being likewise put into English, <gloss
rend=italic>not so young as she has been</gloss>:  in short,
she is grown so much into a <hi rend=italic>mother</hi>,
that she had forgotten she ever was a
<hi rend=italic>daughter</hi>.
```

6.4 Names, Numbers, Dates, Abbreviations, and Addresses

This section describes a number of textual features which it is often convenient to distinguish from their surrounding text. Names, dates and numbers are likely to be of particular importance to the scholar treating a text as source for a database; distinguishing such items from the surrounding text is however equally important to the scholar primarily interested in the vocabulary.

The treatment of these textual features proposed here is not intended to be exhaustive: fuller treatments for names and dates are provided in section 22, “Additional Tags for Names and Dates,” [in separate fascicle].

6.4.1 Names

Any proper noun or noun-phrase may be marked as such using the general purpose **<name>** element. Where it is thought useful to do so, the kind of object referred to by the name may also be specified, using the *type* attribute. To simplify the encoding of very frequently-encountered types of name (specifically, names for persons, places and organizations), three additional elements are provided for these.

<name> : contains a proper noun or noun phrase. Attributes include:
  type : indicates the type of the object which is being named by the phrase.
<person> : contains a name of a person, or a reference to a person.
<place> : contains the name of, or a reference to, a place.
<org> : contains the name of an organization, or a reference to an organization.

All of these tags are members of the element class *names*, and share the following attributes, which are discussed below:

key : provides an alternative identifier for the object being named, such as a database record key.
reg : gives a normalized or regularized form of the proper name

A proper name may consist of more than one word:

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

Where the distinctions between names of places and names of persons can be made with confidence, an alternative way of encoding the above example might be:

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

The element **<org>** may be used in the same way to identify names or other references to institutions such as government departments or commercial companies:

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

If a more detailed analysis of names and places is being undertaken, in which all references to persons and places are to be associated with a database, the elements **<person>**, **<place>** and **<org>** have an additional attribute, *key*, which may be used as in the following example:

```
<q>My dear <person key=BENM1>Mr. Bennet</person>,</q>
said <person key=BENM2>his lady</person> to him one day,
<q>have you heard that
<place key=NETP1>Netherfield Park</place> is let at last?</q>
```

The *key* attribute provides an external identifier, such as a database key, as a means of specifying the referent of a **<person>**, **<place>** or **<org>** element. As this example shows, it may be used for any reference to a person or place, not necessarily one in the form of a proper noun or noun phrase.

Simply tagging something as a name is generally not 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. No special element is provided to mark the standard form of a name to be used for reference purposes; instead, it may be specified as the value of a *reg* (regularization) attribute:

```
  <person reg='Volanges, Mme de'>Mme. de Volanges</person>
  marie sa fille: c'est encore un secret; mais elle m'en
  a fait part hier.

  <person reg='de la Mare, Walter'>Walter de la Mare</person>
  was born at <place>Charlton</place>, in <name type=county>Kent
  </name>, in <date>1873</date>.

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

This method is adequate for many simple applications. For more complex applications, such as onomastics, or wherever a detailed analysis of the component parts of a name is needed, the specialized elements described in chapter 22, “Additional Tags for Names and Dates,” [in separate fascicle] should be used. That chapter also discusses more tags for providing more systematic information about people and organizations.

These elements are formally declared as follows:

```
<!-- 6.4.1:  Proper Nouns                                     -->
<!ENTITY % a.names '
          key                CDATA             #IMPLIED
          reg                CDATA             #IMPLIED       ' >
<!ELEMENT name          - -  (%phrase.seq;)                    >
<!ATTLIST name               %a.global;
          type               CDATA             #IMPLIED        >
<!ELEMENT person        - -  (%phrase.seq;)                    >
<!ATTLIST person             %a.global;                        >
<!ELEMENT place         - -  (%phrase.seq;)                    >
<!ATTLIST place              %a.global;                        >
<!ELEMENT org           - -  (%phrase.seq;)                    >
<!ATTLIST org                %a.global;                        >
<!-- This fragment is used in sec. 6                          -->
```

6.4.2 Numbers and Measures

Like names or abbreviations, numbers can occur virtually anywhere in a text. Numbers are special in that they 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”). They also frequently appear in association with specialized semi-lexical tokens referring to units of *measure*. Like names and dates, numbers and measures may in some cases require the use of the analytic tools discussed in chapter 18, “Feature Structure Analysis,” [in separate fascicle]. For simple encoding of commonly occurring phrases however, it may be enough to mark numbers and units of measure as such, and provide simple information as to their type and value.

The elements **<num>** and **<measure>** are provided for these purposes:

<num> : contains a number, written in any form. Attributes include:
type : indicates the type of numeric value. Sample values include:
cardinal : absolute number, e.g. “21”, “21.5”
ordinal : ordinal number, e.g. “21st”
fraction : fraction, e.g. “one half” or “three halves”
percentage : e.g. “ten percent”
value : supplies the value of the number in an application-dependent standard form.
<measure> : contains a word or phrase referring to a unit of measurement of any kind, typically associated with a number. Attributes include:
type : specifies the type of units being measured. Sample values include:
weight : measure of weight e.g. kg, pound.
count : unit of count, e.g. dozen, score.
length : measure of length, e.g. pole, mm.
area : measure of area e.g. acre, hectare.
volume : measure of volume e.g. litre, gallon.
currency : unit of currency e.g. ecu, escudo, mark.
reg : gives a standardized form for the name of the measure concerned

For many kinds of application, e.g. natural-language processing or machine translation, numbers are not regarded as “lexical” in the same way as other parts of a text. For these and other applications, the **<num>** element provides a convenient method of distinguishing numbers from the surrounding text. For other kinds of application, numbers are only useful if normalized: here the **<num>** element is useful precisely because it provides a standardized way of representing a numerical value.

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>

<num type=fraction value='0,5'>one half</num>
<num type=fraction value='0,5'>1/2</num>

<l>I've measured it from side to side
<l>'Tis <num value=3>three</num><measure type=length reg='ft'>feet
</measure> long, and <num value=2>two</num>
<measure type=length reg='ft'>feet</measure> wide.

<list>
<item><num value=2>Ii</><measure type=volume>bags</>hops
<item><num value=6>six</><measure type=volume>trusses</>Woolen and linen
<item><num value=5>5</><measure type=weight>tonnes</>coale
    ...
</list>
```

These elements are formally defined as follows:

```
<!-- 6.4.2:  Numbers and measures                              -->
<!ELEMENT num           - -  (%phrase.seq;)                      >
<!ATTLIST num                %a.global;
          type               CDATA               #IMPLIED
          value              CDATA               #IMPLIED        >
<!ELEMENT measure       - -  (%phrase.seq;)                      >
<!ATTLIST measure            %a.global;
          type               CDATA               #IMPLIED
          reg                CDATA               #IMPLIED        >
<!-- This fragment is used in sec. 6                           -->
```

6.4.3 Dates and times

Dates and times, like numbers, can appear in widely varying culture- and language-dependent forms, and can pose similar problems in automatic language processing. The following elements are provided to identify them:

<date> : contains a date in any format. Attributes include:
calendar : indicates the system or calendar to which the date belongs.
value : gives the value of the date in some standard form, usually yyyy-mm-dd.
<time> : contains a phrase defining a time of day in any format. Attributes include:
value : gives the value of the time in a standard form.
<dateRange> : contains two dates or another phrase delimiting a time period. Attributes include:
calendar : indicates the system or calendar to which the date belongs.
from : indicates the starting point of the period in standard form.
to : indicates the ending point of the period in standard form.
exact : indicates the precision to be attached to either or both dates specified. Sample values include:
to : the *to* date is exact
from : the *from* date is exact
both : both dates are exact
none : both dates are approximate or unspecified
<timeRange> : contains two times or another phrase indicating a time period. Attributes include:
from : indicates the starting point of the time period in standard form.
to : indicates the ending point of the time period in standard form.
exact : indicates the precision to be attached to either or both times specified. Sample values include:
to : the *to* time is exact
from : the *from* time is exact
both : both times are exact
none : both times are approximate or unspecified

Dates can occur virtually anywhere in a text, but in some contexts (e.g. bibliographic citations) their encoding is recommended or required rather than optional. Times can also appear anywhere but are generally optional.

Partial dates or times (e.g. “1990,”“September 1990”, “twelvish”) can be expressed in the *value* attribute by simply omitting a part of the value supplied. Imprecise dates or times (for example “early August”, “some time after ten and before twelve”) may be expressed as date or time ranges. If either end of the date or time range is known to be accurate, (for example, “at some time before 1230”, “a few days after Hallowe'en”) the *exact* attribute may be used to specify this.

Where the certainty (i.e. reliability) of the date or time itself is in question, rather than its precision, the encoder should record this fact using the mechanisms discussed in section 19, “Certainty,” [in separate fascicle].

These mechanisms are useful primarily for fully specified dates or times known with certainty. If component parts of dates or times are to be marked up, or if a more complex analysis of the meaning of a temporal expression is required, the techniques described in chapter 22, “Additional Tags for Names and Dates,” [in separate fascicle] should be used in preference to the simple method outlined here.

Examples:

```
<date value='1980-02-21'>21 Feb 1980</date>

Given on the <date value='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>

<date value='1990'>1990</date>
<date value='1990-09'>September 1990</date>

Those five years &mdash; <dateRange from=1918 to=1923>
1918 to 1923</dateRange> &mdash had been, he suspected,
somehow very important.

The Eddic poems are preserved in a unique
manuscript (Codex Regius 2365) from
<dateRange from=1250 to=1300>
the second half of the thirteenth
century</dateRange>, and <title>Hervarar
saga</title> dates from <date value=1300>
around 1300</date>.
```

These elements are formally defined as follows:

```
<!-- 6.4.3:  Dates and times                                  -->
<!ELEMENT date         - -  (%phrase.seq;)                      >
<!ATTLIST date              %a.global;
          calendar          CDATA             #IMPLIED
          value             CDATA             #IMPLIED
          certainty         CDATA             #IMPLIED          >
<!ELEMENT dateRange    - O  (%phrase.seq;)                      >
<!ATTLIST dateRange         %a.global;
          calendar          CDATA             #IMPLIED
```

```
          from              CDATA              #IMPLIED
          to                CDATA              #IMPLIED
          exact             (to|from|both|none)
                                               #IMPLIED       >
<!ELEMENT time          - -  (%phrase.seq;)                   >
<!ATTLIST time              %a.global;
          value             CDATA              #IMPLIED       >
<!ELEMENT timeRange     - -  (%phrase.seq;)                   >
<!ATTLIST timeRange         %a.global;
          from              CDATA              #IMPLIED
          to                CDATA              #IMPLIED
          exact             (to|from|both|none)
                                               #IMPLIED       >
<!-- This fragment is used in sec. 6                         -->
```

6.4.4 Abbreviations and Their Expansions

It is sometimes desirable to mark abbreviations in the copy text, whether to trigger special processing for them, to provide the full form of the word or phrase abbreviated, or to allow for different possible expansions of the abbreviation. Abbreviations may be transcribed as they stand, or expanded; they may be left unmarked, or marked using these tags:

<abbr> : contains an abbreviation of any sort. Attributes include:
- *expan* : gives an expansion of the abbreviation.
- *resp* : identifies the individual responsible for the interpretation given the abbreviation.
- *type* : allows the encoder to classify the abbreviation according to some convenient typology. Sample values include:
  - *suspension* : the abbreviation provides the first letter(s) of the word or phrase, omitting the remainder.
  - *contraction* : the abbreviation omits some letter(s) in the middle.
  - *brevigraph* : the abbreviation comprises a special symbol or mark.
  - *superscription* : the abbreviation includes writing above the line.
  - *acronym* : the abbreviation comprises the initial letters of the words of a phrase.
  - *title* : the abbreviation is for a title of address (Dr, Ms, Mr,
  - *organization* : the abbreviation is for the name of an organization.
  - *geographic* : the abbreviation is for a geographic name.

<expan> : contains the expansion of an abbreviation. Attributes include:
- *abbr* : gives the abbreviation in its unexpanded form.
- *resp* : identifies the individual responsible for the interpretation given the abbreviation.
- *type* : allows the encoder to classify the abbreviation according to some convenient typology.

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, and the *expan* attribute may be used to supply an expansion:

```
<abbr type=title>Dr.</abbr> <abbr type=initial>M.</abbr> Deegan
is the Director of the <abbr expan='Computers in Teaching Initiative'
type=acronym>CTI</abbr> Centre for Textual Studies.
```

Abbreviations such as “Dr M” above may be treated as two abbreviations, as above, or as one:

```
<abbr>Dr. M.</abbr> Deegan is the Director of the
<abbr>CTI</abbr> Centre for Textual Studies.
```

This element is particularly useful where manuscript materials in which abbreviation is very frequent are being transcribed. For example:

```
<l>Ex<abbr type=brevigraph expan='per'
           Resp=PG>&per;</abbr>ience, thogh noon auctoritee
<l>Were in this world, is right ynogh for me
<l>To speke of wo that is in mariage;
```

Here an entity reference “per” has been used to represent the common manuscript symbol “crossed-p”, and its expansion supplied in the associated **<abbr>** tag. The same lines might be transcribed, expanded, as follows:

```
<l>Ex<expan type=brevigraph abbrev='&per;'
          Resp=PG>per</expan>ience, thogh noon auctoritee
<l>Were in this world, is right ynogh for me
<l>To speke of wo that is in mariage;
```

In practice, it may be most convenient to transcribe the abbreviation as an entity reference; this allows the entity reference itself to be expanded either as an **<abbr>** or as an **<expan>** element, depending on the processing to be done at the moment. The text shown here:

```
<l>Ex&per;ience, thogh noon auctoritee
<l>Were in this world, is right ynogh for me
<l>To speke of wo that is in mariage;
```

may be expanded as desired by providing the appropriate choice between the two entity declarations:

```
<!ENTITY per "<abbr type=brevigraph expan='per'
              Resp=PG>&p.crossed;</abbr>" >
<!ENTITY per "<expan type=brevigraph abbrev='&p.crossed;'
              Resp=PG>per</expan>" >
```

For further discussion of manuscript abbreviations, see section 20, “Manuscripts, Analytic Bibliography, and Physical Description of the Source Text,” [in separate fascicle].

These elements are formally defined as follows:

```
<!-- 6.4.4:  Abbreviations                                    -->
<!ELEMENT abbr         - -  (%phrase.seq;)                      >
<!ATTLIST abbr              %a.global;
          expan             CDATA              #IMPLIED
          resp              CDATA              #IMPLIED
          type              CDATA              #IMPLIED       >
<!ELEMENT expan        - -  (%phrase.seq;)                      >
<!ATTLIST expan             %a.global;
          abbr              CDATA              #IMPLIED
          resp              CDATA              #IMPLIED
          type              CDATA              #IMPLIED       >
```

6.4.5 Addresses

The **<address>** element is used to specify a postal address, either as running text, or as a structure composed only of the following elements:

- **<street>** : a full street address including any name or number identifying a building as well as the name of the street or route on which it is located.
- **<city>** : contains the name of a town, city or village occurring within an address or imprint.
- **<region>** : contains the name of any regional or administrative subdivision, smaller than a country but larger than a city, occurring within an address.
- **<country>** : contains the largest geographic, administrative or political unit identified within an address.
- **<postcode>** : contains a numerical or alphanumeric code used as part of a postal address to simplify sorting or delivery of mail.
- **<postbox>** : contains a number or other identifier for some postal delivery point other than a street address.

These elements may each appear once only in any address, but may be given in any order. None of them is required. Where code letters are commonly used in addresses (for example, to identify regions or countries) a useful practice is to supply the full name of the region or country as the content of the element, but to supply the abbreviatory code as the value of the global *n* attribute, so that (for example) an application preparing formatted labels can readily find the required information.

For further discussion of ways of regularizing the names of places, see section 6.4, “Names, Numbers, Dates, Abbreviations, and Addresses,” on page 24. A full postal address will also, of course, include the name of the addressee, tagged either using the appropriate name element discussed in that section, or using a more specific element such as **<publisher>** or **<organization>**. Other elements commonly included with postal addresses include **<phone>** **<fax>** or **<email>** addresses; for more details on these see chapter 22, “Additional Tags for Names and Dates,” [in separate fascicle].

The **<address>** element and its components are formally described as follows:

```
<!-- 6.4.5:  Addresses and their components                   -->
<!ELEMENT address       - o  (org? & street? & city? & region? &
                              country? & postcode? & postbox?)  >
<!ATTLIST address            %a.global;                         >
<!ELEMENT street        - o  (%phrase.seq)                      >
<!ATTLIST street             %a.global;                         >
<!ELEMENT city          - o  (%phrase.seq)                      >
<!ATTLIST city               %a.global;                         >
<!ELEMENT region        - o  (%phrase.seq)                      >
<!ATTLIST region             %a.global;                         >
<!ELEMENT country       - o  (%soup)                            >
<!ATTLIST country            %a.global;                         >
<!ELEMENT postcode      - o  (#PCDATA)                          >
<!ATTLIST postcode           %a.global;                         >
<!ELEMENT postbox       - o  (#PCDATA)                          >
<!ATTLIST postbox            %a.global;                         >
<!-- This fragment is used in sec. 6                          -->
```

6.5 Simple Editorial Changes

As in editing a printed text, so in encoding a text in electronic form, it may be necessary to accommodate editorial comment on the text and to render account of any changes made to the text in preparing it. The tags described in this section may be used to record such editorial interventions, whether made by the encoder, by the editor of a printed edition used as a copy text, by earlier editors, or by the copyists of manuscripts.

The tags described here handle most common types of editorial intervention and stereotyped comment; where less structured commentary of other types is to be included, it should be marked using the **<note>** element described in section 6.8, “Notes, Annotation, and Indexing,” on page 48. Systematic interpretive annotation is also possible using the various methods described in chapter 16, “Segmentation and Alignment,” [in separate fascicle]. The examples given here illustrate only simple cases of editorial intervention; in particular, they permit economical encoding of two alternative readings of a text only. To encode more than two views of any one segment of text, the mechanisms described in chapters 16, “Segmentation and Alignment,” [in separate fascicle]and 21, “Text Criticism and Apparatus,” [in separate fascicle] must be used.

The first two pairs of elements here discussed (**<sic>** and **<corr>**, **<reg>** and **<orig>**) may both be used to record simultaneously a text in its "original", uncorrected and unaltered form and also in an "edited" form. In this way they resemble the pair **<abbr>** and **<expan>**, described in section 6.4.4, "Abbreviations and Their Expansions," on page 30. Such paired elements enable software to move automatically from one "view" of the text to the other.

Three categories of editorial intervention are discussed in this section:

- correction (or non-correction) of apparent errors
- regularization (or non-regularization) of variant, irregular, non-standard, or eccentric forms
- editorial additions, suppressions, and omissions

6.5.1 Correction of Apparent Errors

When the copy text is manifestly faulty, an encoder or transcriber may elect simply to correct it without comment. For scholarly purposes, it will often be more generally useful to record both the correction and the original state of the text. The elements described here enable this to be done is such a way as not to distract the reader.

<sic> : contains text reproduced although apparently incorrect or inaccurate. Attributes include:
  corr : gives a correction for the apparent error in the copy text.
  resp : identifies the individual responsible for the correct form given.
<corr> : contains the correct form of a passage apparently erroneous in the copy text. Attributes include:
  sic : gives the original form of the apparent error in the copy text.
  resp : identifies the individual responsible for the correct form given.

The following examples show alternative treatment of the same material. The copy text reads:

> Another property of computer-assisted historical research is that data modelling must permit any one textual feature or part of a textual feature to be a part of more than one information model and to allow the researcher to draw on several such models simultaneously, for example, to select from a machine-readable text those marginal comments which indicate that the date's mentioned in the main body of the text are incorrect.

An encoder may choose to correct the typographic error, either silently or with an indication that a correction has been made, as follows:

```
... marginal comments which indicate that the
<corr>dates</corr> mentioned in
the main body of the text are incorrect.
```

Alternatively, the encoder may simply record the typographic error without correcting it, either without comment or with a **<sic>** element to indicate the error is not a transcription error in the encoding:

```
... marginal comments which indicate that the
<sic>date's</sic> mentioned in
the main body of the text are incorrect.
```

If the encoder elects both to record the original source text and to provide a correction for the sake of word-search and other programs, either **<sic>** or **<corr>** may be used with the appropriate attribute:

```
... marginal comments which indicate that the
<sic corr='dates' resp=MSM>date's</sic> mentioned in
the main body of the text are incorrect.

... marginal comments which indicate that the
<corr sic="date's" resp='MSM'>dates</corr> mentioned in
the main body of the text are incorrect.
```

If both readings are given, the choice between **<sic>** and **<corr>** is largely a question of individual preference; since both record the same information, either may be mechanically transformed into the other. If the original reading contains SGML tags, it will prove more convenient to use **<sic>** than **<corr>** (and vice versa if there are tags within the corrected reading), since SGML tags are not recognized in attribute values. If both readings contain subordinate tags, then recourse may be had to the methods described in section 21, "Text Criticism and Apparatus," [in separate fascicle].

The **<sic>** and **<corr>** tags imply a degree of editorial certainty about error and its correction and should be strictly limited in use to cases where such certainty exists. They are not intended for use in cases where emendation is conjectural; for these cases, the mechanisms proposed in chapter 21, "Text Criticism and Apparatus," [in separate fascicle] are recommended.

The formal definition of these elements is as follows:

```
<!-- 6.5.1:  Editorial tags for correction                    -->
<!ELEMENT sic          - -  (%specialPara;)                     >
<!ATTLIST sic               %a.global;
          corr              CDATA               #IMPLIED
          resp              CDATA               #IMPLIED        >
<!ELEMENT corr         - -  (%specialPara;)                     >
<!ATTLIST corr              %a.global;
          sic               CDATA               #IMPLIED
          resp              CDATA               #IMPLIED        >
<!-- This fragment is used in sec. 6                           -->
```

6.5.2 Regularization and Normalization

When the source text makes extensive use of variant forms or non-standard spellings, it may be desirable for a number of reasons to *regularize* it: that is, to provide “standard” or “regularized” forms equivalent to the non-standard forms.[4]

As with other such changes to the copy text, the changes may be made silently (in which case the TEI header should specify the types of silent changes made) or may be explicitly marked using the following elements:

<reg> : contains a reading which has been regularized or normalized in some sense. Attributes include:
- *orig* : gives the unregularized form of the text as found in the source copy.
- *resp* : identifies the individual responsible for the regularization of the word or phrase.

<orig> : contains the original form of a reading, for which a regularized form is given in an attribute value. Attributes include:
- *reg* : gives a regularized (normalized) form of the text.
- *resp* : identifies the individual responsible for the regularization of the word or phrase.

Typical applications for these elements include the production of editions intended for student or lay readers, linguistic research in which spelling or usage variation is not the main question at issue, production of spelling dictionaries, etc.

Consider this 16th-century text:

> how godly a dede it is to overthrowe so wicked a race the world may judge: for my part I thinke there canot be a greater sacryfice to God.

An encoder may choose to preserve the original spelling of this text, but simply flag it as nonstandard by using the **<orig>** element with no attributes specified, as follows:

```
how godly a <orig>dede</orig> it is to
<orig>overthrowe</orig> so wicked a race the
world may judge:  for my part I <orig>thinke</orig>
there <orig>canot</orig> be a greater
<orig>sacryfice</orig> to God.
```

Alternatively, the encoder may simply indicate that certain words have been modernized by using the **<reg>** element with no attributes specified, as follows:

```
how godly a <reg>deed</reg> it is to
<reg>overthrow</reg> so wicked a race the
world may judge:  for my part I <reg>think</reg>
there <reg>cannot</reg> be a greater
<reg>sacrifice</reg> to God.
```

[4] In some contexts, the term *regularization* has a narrower and more specific significance than that proposed here: the **<reg>** element may be used for any kind of regularization, including normalization, standardization, and modernization.

More usefully, the encoder may elect to record both old and new spellings, so that (for example) the same electronic text may serve as the basis of an old- or new-spelling edition:

```
how godly a <reg orig='dede'>deed</reg> it is to
<reg orig='overthrowe'>overthrow</reg> so wicked a race the
world may judge:  for my part I <reg orig='thinke'>think</reg>
there <reg orig='canot'>cannot</reg> be a greater
<reg orig='sacryfice'>sacrifice</reg> to God.
```

Or the **<orig>** tag might be preferred

```
how godly a <orig reg='deed'>dede</orig> it is to
<orig reg='overthrow'>overthrowe</orig> so wicked a race the
world may judge:  for my part I <orig reg='think'>thinke</orig>
there <orig reg='cannot'>canot</orig> be a greater
<orig reg='sacrifice'>sacryfice</orig> to God.
```

The *resp* attribute should be used to specify the agency responsible for the regularization. This may be an identifiable individual, for example an editor, or a descriptive phrase such as “copyist”. For example, in the first stanza of the Old Norse poem *Grógaldr*, the manuscript form *dura* is usually regularized in modern editions to *dyra* “doors”. The manuscript's “vek ek <thorn>ik dau<eth>ra dura” might thus be recorded together with its regularization in two ways, as follows:

```
vek ek &thorn;ik dau&eth;ra <reg orig='dura' resp=ed>dyra</reg>
```

or:

```
vek ek &thorn;ik dau&eth;ra <orig reg='dyra' rep=ed>dura</orig>
```

These elements are formally defined as follows:

```
<!-- 6.5.2:  Editorial tags for regularization               -->
<!ELEMENT reg          - -  (%phrase.seq;)                     >
<!ATTLIST reg               %a.global;
          orig              CDATA             #IMPLIED
          resp              CDATA             #IMPLIED        >
<!ELEMENT orig         - -  (%phrase.seq;)                     >
<!ATTLIST orig              %a.global;
          reg               CDATA             #IMPLIED
          resp              CDATA             #IMPLIED        >
<!-- This fragment is used in sec. 6                          -->
```

6.5.3 Additions, Deletions and Omissions

The following elements are used to indicate when words or phrases has been omitted from, added to, or marked for deletion from, a text. Like the other editorial elements, they allow for a wide range of editorial practices:

<omit> : 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 or inaudible. Attributes include:
- *desc* : gives a description of the omitted text.
- *reason* : gives the reason for omission. Sample values include "sampling", "illegible", "inaudible", "irrelevant", "canceled", "canceled and illegible".
- *extent* : indicates approximately how much text has been lost, in letters, minims, inches, or any appropriate unit, for cases where a deletion renders the text quite illegible.
- *resp* : indicates the individual responsible for the omission (in the case of canceled material, the individual responsible for the cancellation)

<add> : contains letters, words, or phrases inserted in the text by a scribe, annotator, corrector, editor, or encoder. Attributes include:
- *place* : if the the addition is written into the copy text, indicates where the additional text is written. Sample values include:
  - *inline* : addition is made in a space left in the witness by an earlier scribe
  - *supralinear* : addition is made above the line
  - *infralinear* : addition is made below the line
  - *left* : addition is made in left margin
  - *right* : addition is made in right margin
  - *top* : addition is made in top margin
  - *bottom* : addition is made in bottom margin
  - *opposite* : addition is made on opposite page
  - *verso* : addition is made on verso of sheet
  - *mixed* : addition is made somewhere, one or more of other values
- *resp* : identifies the individual responsible for the addition.

<del> : contains a word or passage deleted, marked as deleted, or identified as superfluous or spurious in the copy text. Attributes include:
- *type* : classifies the type of deletion using any convenient typology.
- *status* : may be used to indicate faulty deletions, e.g. strikeouts which include too much or too little text.
- *resp* : identifies the individual responsible for making the deletion.

Encoders may choose to omit parts of the copy text for reasons ranging from illegibility of the source or impossibility of transcribing it, to editorial policy, e.g. a systematic exclusion of poetry or prose from an encoding. The full details of the policy decisions concerned should be documented in the TEI Header (see section 5.3, "The Encoding Description," [in separate fascicle]). Each place in the text at which omission has taken place should be marked with an **<omit>** element, with optionally further information about the reason for the omission, its extent and the person or agency responsible for it, as in the following examples:

```
<omit desc="Prose commentary" reason="sampling"
      extent="120 lines" resp=PR>

... Their arrangement with respect to Jupiter and to each
other was as follows:
<omit desc="diagram" reason="sampling" extent="2 cm x 1 col">
That is, there were two starts on the easterly side and one
to the west; ...
```

```
<omit reason="illegible" desc="ink blot" extent="two words">

<omit reason='overwriting, illegible'
      resp='hand 1'
      extent='8 chars'>
```

The **<add>** and **<del>** elements may be used to record where words or phrases have been added or deleted in the copy text. They are not appropriate where longer passages have been added or deleted, which span several SGML elements; for these, the elements **<addSpan>** and **<delSpan>**, or other mechanisms described in section 20, “Manuscripts, Analytic Bibliography, and Physical Description of the Source Text,” [in separate fascicle] must be used.

Additions to a text may be recorded for a number of reasons. Sometimes they are marked in a distinctive way in the source text, for example by bracketting or insertion above the line (*supralinear* insertion), as in the following example, taken from a 19th century manuscript:

```
The story I am going to relate is true as to its main
facts, and as to the consequences <add
place='supralinear' resp='auth'>of these facts</add>
from which this tale takes its title.
```

The **<add>** element may also be used to mark simple editorial changes, such as supplying a word omitted by mistake from the source text:

```
... and a very great <add resp=DW>slaughter</add>
of the Danes was made, and the part that escaped
were saved by flight.
```

or a passage omitted from one source but present in another:

```
... they went both together up along the Thames,
and a great reinforcement came to them both from
the East Angles and the Northumbrians.
<add resp=DW>[They then went up along the Thames
until they reached the Severn, then up along
the Severn.]</add>
<note resp=DW place=foot n=1>Supplied from A. ...</note>
Then ...
```

The **<add>** element is suitable for simple cases of this nature only; where several sources are involved, the mechanisms described in section 21, “Text Criticism and Apparatus,” [in separate fascicle] should be used. Where the addition is simply to correct an obvious error (for example, a word accidentally left out during the transcription), the **<corr>** or **<sic>** elements may be used in preference.

The **<del>** element is used to mark material which is deleted in the source (as opposed to material which has been omitted by the encoder or transcriber). This is of particular importance in transcribing manuscript material, though deletion is also found in printed texts, sometimes for humorous purposes:

```
<l>One day I will sojourn to your shores
<l>I live in the middle of England
<l>But!
<l>Norway! My soul resides in your watery
   <del type='overstrike'>fiords fyords fiiords</del>
<l>Inlets.
```

The *type* attribute may be used to distinguish different methods of deletion in manuscript or typescript material, as in this line from the typescript of Eliot's Wasteland:

```
<l><del type=overtyped>Mein</del> Frisch
   <del type=overstrike>schwebt</del> weht der Wind
```

Deletion in manuscript or typescript is often associated with addition:

```
<l><del type=overstrike>Inviolable</del>
   <add type=sublinear>Inexplicable</add>
   splendour of Corinthian white and gold
```

For any detailed transcription of a manuscript or typescript with more than trivial amounts of alteration, the reader should consult chapter 21, “Text Criticism and Apparatus,” [in separate fascicle], and chapter 20, “Manuscripts, Analytic Bibliography, and Physical Description of the Source Text,” [in separate fascicle].

These elements are formally defined as follows:

```
<!-- 6.5.3:  Other editorial tags                              -->
<!ELEMENT omit         - O  EMPTY                                >
<!ATTLIST omit              %a.global;
          desc              CDATA              #IMPLIED
          reason            CDATA              #IMPLIED
          resp              CDATA              #IMPLIED
          extent            CDATA              #IMPLIED          >
<!ELEMENT add          - -  (%specialPara;)                      >
<!ATTLIST add               %a.global;
          place             CDATA              #IMPLIED
          resp              CDATA              #IMPLIED          >
<!ELEMENT del          - -  (%phrase.seq;)                       >
<!ATTLIST del               %a.global;
          rend              CDATA              #IMPLIED
          type              CDATA              #IMPLIED
          status            CDATA              'unremarkable'

          resp              CDATA              #IMPLIED          >
<!-- This fragment is used in sec. 6                           -->
```

6.6 Simple Links and Cross References

Cross-references or links between one location in a document and another, or between one location and several others, may be encoded using the elements **<ptr>** and **<ref>**, as discussed in this section. These elements both "point" from one location in a document, the place that the element itself appears, to another (or to several), specified by the *target* attribute. Linkages of several other kinds are also provided for in these guidelines; see further chapter 16, "Segmentation and Alignment," [in separate fascicle].

The pointing facility of these elements depends on the ability to supply a unique identifier for any element in the TEI scheme, using the global *id* attribute. Where the object or objects of a cross-reference are not identifiable in this way, either because they are located in a distinct SGML document or because no *id* attribute is available, the elements **<xptr>** or **<xref>** may be used instead.[5] Alternatively, if no explicit link is to be encoded, but it is simply required to mark the phrase as a cross-reference, the **<ref>** element may be used without a *target* attribute.

<ptr> : defines a pointer to another location in the current document. Attributes include:
- *target* : specifies the destination of the pointer as one or more SGML identifiers

<ref> : defines a reference to another location in the current document, in terms of one or more identifiable elements, possibly modified by additional text or comment. Attributes include:
- *target* : specifies the destination of the reference as one or more SGML identifiers

The elements **<ptr>** and **<ref>** share, as members of the element class *pointer*, the following attributes:

type : categorizes the pointer in some respect, using any convenient set of categories.
resp : specifies the creator of the pointer.
crdate : specifies when the pointer was created.
targType : specifies the kinds of elements to which this pointer may point.
targOrder : where more than one identifier is supplied as the value of the *target* attribute, this attribute specifies whether the order in which they are supplied is significant. Sample values include:
- *Y* : Yes: the order in which IDREFs are specified as the value of a *target* attribute should be followed when combining the targeted elements.
- *N* : No: the order in which IDREFs are specified as the value of a *target* attribute has no significance when combining the pointered elements.
- *U* : Unspecified: the order in which IDREFs are specified as the value of a *target* attribute may or may not be significant.

[5] See chapter 16, "Segmentation and Alignment," [in separate fascicle], for a discussion of these elements and the extended syntax they provide for "hypertext" links.

evaluate : specifies the intended meaning when the target of a pointer is itself a pointer. Sample values include:
  all : if the element pointed to is itself a pointer, then the target of that pointer will be taken, and so on, until an element is found which is not a pointer.
  one : if the element pointed to is itself a pointer, then its target (whether a pointer or not) is taken as the target of this pointer.
  none : no further evaluation of targets is carried out beyond that needed to find the element specified in the pointer's target.

The shared attributes of the two elements may be used in the same way; the difference between the elements is that while the **<ptr>** element is empty, the **<ref>** element may contain phrases specifying, or defining more exactly, the target of a cross reference, which form the content of the element. Since its content thus serves as a human-readable pointer, in the simplest case a **<ref>** element need not identify its target in any other way. For example:

```
See <ref>section 12 on page 34</ref>.
```

More usually, it will be desirable to identify the target of the cross-reference using the *target* attribute, so that processing software can access it directly, for example to implement a linkage or to generate an appropriate reference. Assuming that section 12 in the previous example has been tagged **<div1 id=SEC12>**, the same cross reference might more exactly be encoded as

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

If the text for the cross reference is to be generated according to a fixed pattern, or if no text is to appear in the body of the cross reference, the **<ptr>** element would be used as follows:

```
See in particular <ptr target=sec12>.
```

A cross-reference may point to any number of locations simultaneously, simply by giving more than one identifier as the value of its *target* attribute. This may be particularly useful where an analytic index is to be encoded, as in the following example:

```
<list>
<item>Saints aid rejected in mel. <ptr target=p299>
<item>Sallets censured <ptr target="p143 p144">
<item>Sanguine mel. signs <ptr target="p263 p312 p332">
<item>Scilla or sea onyon, a purger of mel. <ptr target=p442>
...
```

Here the targets of the cross references are simply page numbers; it is assumed that corresponding elements with identifiers “p299” “p143”, etc. have been provided in the body of the text. If it is desired to check that the targetted elements are of a particular type, the *targType* (target type) attribute may be specified:

```
<list>
<item>Saints aid rejected in mel <ptr target=p299 targType=pb>
<item>Sallets censured <ptr target="p143 p144" targType=pb>
...
```

Here, a processing application can check that the elements with identifiers “p299”, “p143”, and “p144” are all **<pb>** (page-break) elements. It is a semantic error in a text if the targets given do not match the values specified on a *targType* attribute.

The *type* and *resp* attributes may be used, as elsewhere, to categorize the cross reference according to any system of importance to the encoder and to supply a code identifying the person or agency responsible for the cross reference. If bibliographic references require special processing (e.g. in order to provide a consistent short-form reference), they might be tagged thus:

```
Similar forms, often called <term rend='ldquo
rdquo'>rewriting systems</term>, have a long
history among mathematicians, but the specific form
of <ptr targType=fig target=Fig22> was first
studied extensively by Chomsky <ptr type=bibliog
targType='bibl bibl.struct bibl.full' target=Chom59>.
```

Here “type=bibliog” signals for the processing appropriate to a bibliographic reference, while “targType='bibl bibl.struct bibl.full'” restricts the legal targets to bibliographic elements, and “target=Chom59” indicates which bibliographic element actually is being referred to. For further discussion of bibliographic references, see section 6.10.3, “Citation References,” on page 79.

If the order in which the objects of a multi-headed cross reference are specified is of importance, the *targOrder* (target order) attribute should be specified.

```
<p>The following discussions of this topic should
be consulted for further information:
<ptr target='ch3 sec332 sec45 sec722' targOrder=Y>
```

The **<ptr>** and **<ref>** tags have many applications in addition to the simple cross-referencing facilities illustrated in this section. In conjunction with the analytic tools discussed in sections 16, “Segmentation and Alignment,” [in separate fascicle], 17, “Simple Analytic Mechanisms,” [in separate fascicle], and 18, “Feature Structure Analysis,” [in separate fascicle], they may be used to link analyses of a text to their object, to combine corresponding segments of a text, or to align segments of a text with a temporal or other axis or with each other.

These elements are formally defined as follows:

```
<!-- 6.6:  Simple cross references                          -->
<!ENTITY % a.pointer '
          type              CDATA               #IMPLIED
          resp              CDATA               #IMPLIED
          crdate            CDATA               #IMPLIED
          targType          NAMES               #IMPLIED
          targOrder         (Y | N | U)         U
```

```
          evaluate          (all | one | none)  #IMPLIED       ' >
<!ELEMENT ptr           - o  EMPTY                             >
<!ATTLIST ptr                %a.global;
          target             IDREFS             #REQUIRED      >
<!ELEMENT ref           - -  (%phrase.seq)                     >
<!ATTLIST ref                %a.global;
          target             IDREFS             #IMPLIED       >
<!-- This fragment is used in sec. 6                          -->
```

6.7 Lists

The following elements are provided for the encoding of lists, their constituent items, and the labels or headings associated with them:

<list> : contains any sequence of items organized as a list. Attributes include:
- *type* : describes the form of the list. Sample values include:
  - *ordered* : list items are numbered or lettered.
  - *bulleted* : list items are marked with a “bullet” or other typographic device.
  - *simple* : list items are not numbered or bulleted.
  - *gloss* : each list item glosses some term or concept, which is given by a **<label>** element preceding the list item.

<item> : contains one component of a list. Attributes include:
- *n* : indicates the number (or letter, or other enumerator) borne by this list item in the copy text.

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

<head> : contains any heading, for example, the title of a section, or the heading of a list or glossary.

<head.label> : contains the heading for the label or term column in a glossary list or similar structured list.

<head.item> : contains the heading for the item or gloss column in a glossary list or similar structured list.

The **<list>** element should be used to mark any kind of *list*: numbered, lettered, bulleted, or unmarked. Lists formatted as such in the copy text should in general be encoded using this element, with an appropriate value for the *type* attribute. Lists given as run-on text may also be encoded using this element, where this is felt to be appropriate.

Each distinct item in the list should be encoded as a distinct **<item>** element. If the numbering or other identification for the items in a list is unremarkable and may be reconstructed by any processing program, no enumerator need be specified. If however an enumerator is retained in the encoded text, it may be supplied either by using the *n* attribute on the **<item>** element, or by using a **<label>** element. The following examples are thus equivalent:

```
I will add two facts, which have seldom occurred in
the composition of six, or at least of five quartos.
<list type=ordered rend=runon>
<label>(1) <item>My first rough manuscript, without any
intermediate copy, has been sent to the press.
<label>(2) <item>Not a sheet has been seen by any human
eyes, excepting those of the author and the printer:
the faults and the merits are exclusively my own.
</list>

I will add two facts, which have seldom occurred in
the composition of six, or at least of five quartos.
<list type=ordered rend=runon>
<item n='1'>My first rough manuscript, without any
intermediate copy, has been sent to the press.
<item n='2'>Not a sheet has been seen by any human
eyes, excepting those of the author and the printer:
the faults and the merits are exclusively my own.
</list>
```

The two styles may not be mixed in the same list: if one item is preceded by a label, all must be.

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

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

A list may be given a heading or title, for which the **<head>** element should be used, as in the next example, which also demonstrates simple use of the **<label>** element to mark a tabular or glossary list in which each item is associated with a word or phrase rather than a numeric or alphabetic enumerator:

```
<list type=gloss>
<head>Report of the conduct and progress of Ernest Pontifex.  Upper Vth
form &dash half term ending Midsummer 1851</head>
<label>Classics<item>Idle listless and unimproving
<label>Mathematics<item>ditto
<label>Divinity<item>ditto
<label>Conduct in house<item>Orderly
<label>General conduct<item>Not satisfactory, on account of his great
unpunctuality and inattention to duties
</list>
```

In such a list, the individual items have internal structure. In complex cases, where list items contain many components, the list is better treated as a *table*, on which see section 25, “Formulae and Tables,” [in separate fascicle]. A particularly important instance of the simple two-column table is the “glossary list”, which should be marked by the tag **<list type=gloss>**. In such lists, each **<label>** element contains a term and each **<item>** its gloss; it is a semantic error for a list tagged with “type=gloss” not to have labels. For example:

```
<list type=gloss>
<head>Unit Three -- Vocabulary</head>
<label lang=la>acerbus, -a, -m       <item>bitter, harsh
<label lang=la>ager, agr&imacr;, M.  <item>field
<label lang=la>audi&omacr;, &imacr;re,
       &imacr;v&imacr;, &imacr;tus   <item>hear, listen (to)
<label lang=la>bellum, -&imacr;, N.  <item>war
<label lang=la>bonus, -a, -um        <item>good
<!-- etc. -->
</list>
```

Additionally, the **<term>** and **<gloss>** elements discussed in section 6.3.4, “Terms, Glosses, and Cited Words,” on page 21 might be used to make explicit the role that each column in the glossary list has, as follows:

```
<list type=gloss>
<head>Unit Three -- Vocabulary</head>
<label><term lang=la>acerbus, -a, -m</term>
<item><gloss>bitter, harsh</gloss></item>
<label><term lang=la>ager, agr&imacr;, M. </term>
<item><gloss>field</gloss></item>
<label><term lang=la>audi&omacr;, &imacr;re,
       &imacr;v&imacr;, &imacr;tus  </term>
<item><gloss>hear, listen (to)</gloss></item>
<label><term lang=la>bellum, -&imacr;, N. </term>
<item><gloss>war</gloss></item>
<label><term lang=la>bonus, -a, -um</term>
<item><gloss>good</gloss></item>
<!-- etc. -->
</list>
```

Note in the above examples the use of the global *lang* attribute to specify on the **<label>** (or **<term>**) element what language the term is from. For further discussion of the *lang* attribute see section 3.3, “Global Attributes,” [in separate fascicle], and section 4.2, “Shifting among Character Sets,” [in separate fascicle]. A more elaborate markup for this glossary would distinguish the headword forms from the grammatical information (principal parts and gender), using tags described more fully in chapters 13, “Base Tag Set for Terminological Data,” [in separate fascicle] or 12, “Base Tag Set for Printed Dictionaries,” [in separate fascicle].

In addition to the **<head>** element used to supply a title or heading for the whole list, headings for the two columns of a glossary-style list may be specified using the two special elements **<head.lab>** and **<head.it>**:

```
The simple, straightforward statement of an idea is
preferable to the use of a worn-out expression.
<list type=gloss>
<head.label>TRITE
<head.item>SIMPLE, STRAIGHTFORWARD
<label>bury the hatchet  <item>stop fighting, make peace
<label>at loose ends     <item>disorganized
<label>on speaking terms <item>friendly
<label>fair and square   <item>completely honest
<label>at death's door   <item>near death
</list>
```

The elements **<label>**, **<head>**, **<head.lab>**, and **<head.it>** may contain only phrase-level elements. The **<item>** element however may contain paragraphs or other “chunks”, including other lists. In this 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 &dash a solitaire; one
           banished from human society.</item>
     </list> <!-- end of first nested 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><!-- end of second nested list --></item>
</list><!-- end of glossary list -->
```

Lists of different types may be nested to arbitrary depths in this way.

The formal declarations for lists and list items are as follows.

```
<!-- 6.7:  Lists and List Items                               -->
<!ELEMENT list          - -  (head?, ( (item*) | (head.label?,
                             head.item?, (label, item)+)))      >
<!ATTLIST list               %a.global;
          type               CDATA               simple         >
<!ELEMENT item          - O  (%specialPara;)                    >
<!ATTLIST item               %a.global;
          n                  CDATA               #IMPLIED       >
<!ELEMENT label         - o  (%phrase.seq;)                     >
<!ATTLIST label              %a.global;                         >
<!ELEMENT head          - -  (%phrase.seq;)                     >
<!ATTLIST head               %a.global;                         >
```

```
<!ELEMENT head.label    - O  (%phrase.seq;)                   >
<!ATTLIST head.label         %a.global;                       >
<!ELEMENT head.item     - o  (%phrase.seq;)                   >
<!ATTLIST head.item          %a.global;                       >
<!-- This fragment is used in sec. 6                         -->
```

6.8 Notes, Annotation, and Indexing

6.8.1 Notes and Simple Annotation

The following elements are provided for the encoding of discursive notes, either already present in the copy text or supplied by the encoder:

<note> : contains a note or annotation. Attributes include:
- *n* : indicates the symbol or number used to mark the note's point of attachment to the main text.
- *type* : describes the type of note.
- *resp* : indicates who is responsible for the annotation: author, editor, translator, etc. Sample values include:
  - *auth[or]* : note originated with the author of the text.
  - *ed[itor]* : note added by the editor of the text.
  - *comp[iler]* : note added by the compiler of a collection.
  - *tr[anslator]* : note added by the translator of a text.
  - *transcr[iber]* : note added by the transcriber of a text into electronic form.
  - *(initials)* : note added by the individual indicated by the initials.
- *place* : indicates where the note appears in the source text. Sample values include:
  - *foot* : note appears at foot of page.
  - *end* : note appears at end of chapter or volume.
  - *inline* : note appears as a marked paragraph in the body of the text.
  - *left* : note appears in left margin.
  - *right* : note appears in right margin.
  - *interlinear* : note appears between lines of the text.
  - *app[aratus]* : note appears in the apparatus at the foot of the page.
- *anchored* : indicates whether the copy text shows the exact place of reference for the note. Sample values include:
  - *yes* : copy text indicates the place of attachment for the note.
  - *no* : copy text indicates no place of attachment for the note.
- *target* : indicates the point of attachment of a note, or the beginning of the span to which the note is attached.
- *target.end* : points to the end of the span to which the note is attached, if the note is not embedded in the text at that point.

<span> : attaches simple analytic categories to an arbitrary passage of text. Attributes include:
- *resp* : indicates who is responsible for the annotation being performed.
- *type* : indicates what type of phenomenon is being noted in the passage. Sample values include:
  - *image* : an image in the passage is being identified.
  - *character* : the span is associated with a particular character.

theme : the **<span>** tag identifies a theme in the passage.
allusion : the **<span>** tag identifies an allusion to another text.
discourse type : the **<span>** tag identifies a passage marked by a particular discourse type.
value : identifies the specific phenomenon being annotated.
target : indicates the beginning of the span.
target.end : points to the end of the span being annotated.

A *note* is any additional comment found in a text, marked in some way as being out of the main textual stream. We propose that all notes should be marked using the same tag, **<note>**, whether they appear as block notes in the main text area, at the foot of the page, at the end of the chapter or volume, in the margin, or in some other place.

Notes may be in a different hand or typeface, may be authorial or editorial, and may have been added later. Attributes may be used to specify these and other characteristics of notes, as detailed below.

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 marginal notes, 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. In some cases, however, it may be desirable to transcribe notes not at their point of attachment to the text but at their point of appearance (at the end of the volume, or the end of the chapter — not, in general, when the notes appear at the foot of the page); in this case the *target* and *target.end* attributes should be used to specify the point of attachment. In some cases, the note is explicitly attached not to a point but to a span of text; for a full discussion of pointing to points and spans in the text, see chapter 6.6, “Simple Links and Cross References,” on page 41.

In addition to transcribing notes from the copy text, researchers may wish to annotate the electronic text itself, by attaching analytic notes in some structured vocabulary to particular passages of text, e.g. to specify the topics or themes of a text. The empty **<span>** element is provided for such applications.

Examples:

```
<l>The self-same moment I could pray
<l>And from my neck so free
<l>The albatross fell off, and sank
<l>Like lead into the sea.
<note type=auth place=margin>The spell begins to break</note>

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 <cited>collection</cited>
instead of the expected <cited>set</cited>.
Our usage corresponds to the <cited>aggregate</cited> of many
mathematical writings and to the sense of <cited>class</cited>
found in older logical writings.
</note>
The elements ...
```

The **<span>** tag may be used to mark any span of text; for example, the structural divisions assigned to a narrative by an interpreter:

```
<p id=P1><anchor id=A1>Sigmund, the son of Volsung, was a king in
Frankish country.  Sinfiotli was the eldest of his sons ...  <anchor
id=A2> <s id=S4a>But Sinfiotli ... wooed the same woman</s> <s id=s4b>
and Sinfiotli killed him over it.</s> <anchor id=A4>And when he came
home, Borghild asked him to go away, but Sigmund offered her weregild,
and she was obliged to accept it.  At the funeral feast Borghild was
serving beer.  She took poison, a big drinking horn full, and brought it
to Sinfiotli.  ...  Sinfiotli drank it off and at once fell dead.
<anchor=A5>

<p id=P2>Sigmund carried him a long way in his arms and came to a long,
narrow fjord, and there was a small boat there and a man in it.  ...

<p id=P3>King Sigmund lived a long time in Denmark ...

<p id=P4>Sigmund and all his sons were tall ...

<s id=nil></s>

<span resp=TMA type='structural unit' value='introduction'
        target=A1 targend=A2>
<span resp=TMA type='structural unit' value='conflict'
        target=s4a>
<span resp=TMA type='structural unit' value='climax'
        target=s4b>
<span resp=TMA type='structural unit' value='revenge'
        target=A4 targend=A5>
<span resp=TMA type='structural unit' value='reconciliation'
        target=nil>
<span resp=TMA type='structural unit' value='aftermath'
        target=P2 targend=P4>
```

The formal declarations for the **<note>** and **<span>** elements are these:

```
<!-- 6.8.1:  Annotation                                       -->
<!ELEMENT note          - O  (%specialPara;)                    >
<!ATTLIST note               %a.global;
          n                  CDATA              #IMPLIED
          type               CDATA              #IMPLIED
          resp               CDATA              #IMPLIED
          place              CDATA              'unspecified'
          anchored           (yes | no)         yes
          target             IDREF              #IMPLIED
          target.end         IDREF              #IMPLIED        >
<!ELEMENT span          - O  EMPTY                              >
<!ATTLIST span               %a.global;
          resp               CDATA              #IMPLIED
          type               CDATA              #IMPLIED
          value              CDATA              #IMPLIED
```

```
          target            IDREF             #IMPLIED
          target.end        IDREF             #IMPLIED      >
<!-- ... declarations from section 6.8.2                    -->
<!--     (Index Entries)                                    -->
<!--     go here ...                                        -->
<!-- This fragment is used in sec. 6                        -->
```

6.8.2 Index Entries

Machine-readable versions of existing texts rarely reproduce any index published with the copy text. Should a printed index be transcribed, the **<div1>** tag or a **<div>** tag at an appropriate level should be used to demarcate the index, and the index itself may be transcribed as a structured list or table.

It is convenient, however, to be able to generate a new index from a machine-readable text, whether the text is being written for the first time with the tags here defined or was transcribed from some other source. The **<index>** tag is provided for this purpose; it may be useful for marking points of particular interest for whatever reason, and not merely for generating printed indexes for a printed version of the text. The **<div.gen>** element indicates the point at which an index, or any other generated text (e.g. a table of contents), is to appear in the output of a text production process.

<index> : marks a location to be indexed for whatever purpose. Attributes include:
- *level1* : gives the form under which the index entry is to be made.
- *level2* : gives the second-level form, if any.
- *level3* : gives the third-level form, if any.
- *level4* : gives the fourth-level form, if any.
- *index* : indicates which index (of several) the index entry belongs to.

<div.gen> : indicates the location at which an **<div>** generated automatically by a text-processing application is to appear. Attributes include:
- *type* : specifies what type of generated text division (e.g. index, table of contents, etc.) is to appear. Sample values include:
  - *index* : an index is to be generated and inserted at this point.
  - *toc* : a table of contents
  - *figlist* : a list of figures
  - *tablist* : a list of tables

The tag **<index>** associates up to four levels of index terms with a specific point in the text. The index terms are supplied in attributes named *level1*, *level2*, *level3*, and *level4*. An *index* attribute associates the entry with a particular index, so multiple indices are possible.

All index terms must be supplied as attribute values; no part of the text itself is taken as a term. This may require words or phrases to be repeated, as illustrated below; it also allows spelling to be normalized, as the example shows:

```
The students understand procedures for Arabic lemmatisation
<index level1='Arabic lemmatization'>and are beginning
to build parsers.
```

The **<div.gen>** element marks the place at which an index generated from the **<index>** elements should be inserted into the output of a processing program; typically, this will be at some point within the back matter of the document; its *type* attribute should be used to specify which index is to be generated, and its *n* attribute to specify a name for the index:

```
<back>
<div><head>Examples</head>
<p> ...
</div>
<div><head>Bibliography</head>
<list.bibl> ... </list.bibl>
</div>
<div.gen type='index 1' n='Index Nominum'>
<div.gen type='index 2' n='Index Rerum'  >
</back>
```

The formal declaration for these elements is as follows.

```
<!-- 6.8.2:  Index Entries                                  -->
<!ELEMENT index         - O  EMPTY                           >
<!ATTLIST index              %a.global;
          index              CDATA              #IMPLIED
          level1             CDATA              #REQUIRED
          level2             any string of characters.
                                                #IMPLIED
          level3             any string of characters.
                                                #IMPLIED
          level4             any string of characters.
                                                #IMPLIED       >
<!ELEMENT div.gen       - O  EMPTY                           >
<!ATTLIST div.gen            %a.global;
          type               CDATA              #IMPLIED       >
<!-- This fragment is used in sec. 6.8.1                    -->
```

6.9 Reference Systems

By *reference system* we mean the system by which names or references are associated with particular passages of a text (e.g. “Ps. 23:3” for the third verse of Psalm 23 or “Amores 2.10.7” for Ovid's *Amores*, book 2, poem 10, line 7). Such names make it possible to mark a place within a text and enable other readers to find it again. A reference system may be based on structural units (chapters, paragraphs, sentences; stanza and verse), typographic units (page and line numbers), or divisions created specifically for reference purposes (chapter and verse in Biblical texts). Where one exists, the traditional reference system for a text should be preserved in an electronic transcript of it, if only to make it easier to compare electronic and non-electronic versions of the text.

Reference systems may be recorded in TEI-encoded texts in any of the following ways:

- where a reference system exists, and is based on the same logical structure as that of the text's SGML markup, the reference for a passage may be recorded as the value of the global *id* or *n* attribute on an appropriate tag, or may be constructed by combining attribute values from several levels of tags, as described below in section 6.9.1, “Using the ID and N Attributes,” on page 54
- where there is no pre-existing reference system, the global *id* or *n* attributes may be used to construct one (e.g. collections and corpora created in electronic form), as described below in section 6.9.2, “Creating New Reference Systems,” on page 55
- where a reference system exists which is not based on the same logical structure as that of the text's SGML markup, (for example, one based on the page and line numbers of particular editions of the text rather than on the structural divisions of it), then a distinct logical structure representing the reference system may be encoded as a concurrent markup stream, as described below in sections 6.9.3, “Concurrent Markup for Pages and Lines,” on page 56 and 6.9.4, “Concurrent Markup for Other Hierarchies,” on page 58
- where a reference system exists which does not correspond to any particular logical structure, or where the logical structure concerned is of no interest to the encoder except as a means of supporting the referencing system, then references may be encoded by means of **<milestone>** elements, which simply mark points in the text at which values in the reference system change, as described below in section 6.9.5, “Milestone Tags,” on page 59

The specific method used to record traditional or new reference systems for a text should be declared in the TEI header, as further described in section 6.9.6, “Declaring Reference Systems,” on page 61 and in chapter 37, “Algorithm for Recognizing Canonical References,” [in separate fascicle].

When a text has no pre-existing associated reference system of any kind, these Guidelines recommend as a minimum that at least the page boundaries of the source text be marked using one of the methods outlined in this section. Retaining page breaks in the markup is also recommended for texts which have a detailed reference system of their own. Line breaks in prose texts may be, but need not be, tagged.[6]

[6] Many encoders find it convenient to retain the line breaks of the original during data entry, to simplify proof-reading, but this may be done without inserting a tag for each line break of the original. For further discussion of tagging for pagination, lineation, and other physical characteristics of the copy text, see chapter 20, “Manuscripts, Analytic Bibliography, and Physical Description of the Source Text,” [in separate fascicle].

6.9.1 Using the ID and N Attributes

When traditional reference schemes represent a hierarchical structuring of the text which mirrors that of the SGML document, the *n* attribute defined for all elements may be used to indicate the traditional identifier of the relevant structural units. The *n* attribute may also be used to record the numbering of sections or list items in the copy text if the copy-text numbering is important for some reason, for example because the numbers are out of sequence.

For example, a traditional reference to Ovid's *Amores* might be “Amores 2.10.7”—book 2, poem 10, line 7. Book, poem, and line are structural units of the work and will therefore be tagged in any case. (See chapter 8, “Base Tag Set for Verse,” [in separate fascicle] for a discussion of structural units in verse collections.) In such cases, it is convenient to record traditional reference numbers of the structural units using the *n* attribute. The relevant tags for our example would be:

```
<div0 n=Amores> ...
    <div1 n='2' type=book> ...
         <div2 n='10' type=poem> ...
              <l n='7'> ...
```

One may also place the entire standard reference for each portion of the text into the appropriate value for the *n* attribute, though for obvious reasons this takes more space in the file:

```
<div0 n=Amores> ...
    <div1 n='Amores 2' type=book> ...
         <div2 n='Amores 2.10' type=poem> ...
              <l n='Amores 2.10.7'> ...
```

If the names used by the traditional reference system can be formulated as SGML identifiers, then the references can be given as values for the *id* attribute; this requires that the reference be given without internal spaces, begin with a letter, and contain no characters other than letters, digits, hyphens, and full stops.[7] Unlike values for the *n* attribute, values for the *id* attribute must be unique throughout the document. Our example then looks like this:

```
<div0 id=Amores> ...
    <div1 id='Am.2' type=book> ...
         <div2 id='Am.2.10' type=poem> ...
              <l id='Am.2.10.7'> ...
```

To document the usage and to allow automatic processing of these standard references, it is recommended that the TEI header be used to declare whether standard references are recorded in the *n* or *id* attributes and which elements may carry standard references or portions of them. See below, section 6.9.6,

[7] The legal form of identifiers depends in part on the SGML declaration; with appropriate modifications in the declaration, other characters may be made legal in identifiers; this is allowed though not encouraged in TEI-conformant documents.

"Declaring Reference Systems," on page 61 for examples of declarations for the reference systems just shown.

Using the *n* attribute one can specify only a single standard referencing system, a limitation not without problems, since some editions may define structural units differently and thus create alternative reference systems. For example, another edition of the *Amores* considers poem 10 a continuation of poem 9, and therefore would specify the same line as "Amores 2.9.31". In order to record both of these reference systems, one must either use the *id* or *n* attribute for one, with competing systems in concurrent markup hierarchies, as discussed below in section 6.9.3, "Concurrent Markup for Pages and Lines," on page 56, or else use the **<milestone>** tags described in section 6.9.5, "Milestone Tags," on page 59.

6.9.2 Creating New Reference Systems

If a text has no canonical reference system of its own, a reference system, if needed, may be derived from the structure of the electronic text, specifically from the SGML markup of the text. As with any reference system intended for long-term use, it is important to see the reference as an established, unchanging point in the text. Should the text be revised or rearranged, the reference-system identifiers associated with any bit of text must stay with that bit of text, even if it means the reference numbers fall out of sequence. (A new reference system may always be created beside the old one if out-of-sequence numbers must be avoided.)

The global attributes *n* and *id* may be used to assign reference identifiers to segments of the text. Identifiers specified by either attribute apply to the entire element for which they are given. SGML enforces uniqueness on ID attributes within a single document, and ID values must begin with a letter. No such restrictions are made on the values of *n* attributes.

A convenient method of mechanically generating unique values for *id* or *n* attributes, based on the SGML structure of the document is to use the "typed path" or "untyped path" methods described in section 16, "Segmentation and Alignment," [in separate fascicle] to identify elements within the **<text>** segment of a TEI document. Identifiers generated with these methods should use the **<text>** segment as their starting point, rather than the **<TEI.2>** or **<body>** elements, since the **<TEI.2>** element need be taken as a starting point only if identifiers need to be generated for the **<TEI.header>**, which is not usually the case, and using the **<body>** element as a root would prevent assignment of identifiers for the front and back matter. The typed and untyped path methods are convenient, but are in no way required for anyone creating a reference system.

If the *id* attribute is used to record the reference identifiers generated, each value should record the entire path. If the *n* attribute is used, each value may record either the entire path or only the subpath from the SGML parent element. The attribute used, the elements which can bear standard reference identifiers, and the method for constructing standard reference identifiers, should all be declared in the header as described in chapter 5.3.4, "The Reference System Declaration," [in separate fascicle].

6.9.3 Concurrent Markup for Pages and Lines

When the hierarchy of the SGML-encoded document and that of the reference system differ (e.g. for reference systems based on page and line numbers) or when more than one reference system is to be encoded, the encoder may choose to represent the alternative reference system(s) as elements in one or more concurrent document hierarchies. See the discussion of the CONCUR feature in section 2, “Concise Summary of SGML,” [in separate fascicle]for an introduction to the concept of concurrent hierarchies. For further discussion of this and other mechanisms, see chapter 36, “Markup for Non-Hierarchical Phenomena,” [in separate fascicle].

One common form of traditional reference system specifies the page and line, or page, column, and line of a passage as it appears in some standard edition. Such references may be specified using a concurrent markup hierarchy which divides the body of a text into pages and lines or into pages, columns, and lines. Volumes may also need to be identified. The document type name should be a short identifier for the edition cited. The following tags may be used:

<vol> : marks the individual volumes of a reference edition.
<page> : marks pages in a reference edition.
<col> : contains one column of a multi-column reference edition.
<line> : contains one line of a reference edition.

Page and line numbers for an edition by Lachmann, for example, might be specified thus:

```
<(La)page n=37> <!-- Text from Lachmann, p. 37 -->
                ...
<(La)line n=32> <!-- Text from Lachmann, p. 37, line 32 -->
<(La)line n=33> <!-- Text from Lachmann, p. 37, line 33 -->
<(La)line n=34> <!-- Text from Lachmann, p. 37, line 34 -->
<(La)page n=38> <!-- Text from Lachmann, p. 38 -->
<(La)line n=1>  <!-- Text from Lachmann, p. 38, line 1 -->
<(La)line n=2>  <!-- Text from Lachmann, p. 38, line 2 -->
<(La)line n=3>  <!-- Text from Lachmann, p. 38, line 3 -->

                <!-- etc. -->
<(La)page n=39> <!-- Text from Lachmann, p. 39 --> ...
<(La)line n=18> <!-- Text from Lachmann, p. 39, line 18 -->
<(La)line n=21> <!-- Text from Lachmann, p. 39, line 21 -->
                <!-- etc. -->
<(La)page n=40> <!-- Text from Lachmann, p. 40 --> ...
                <!-- etc. -->
<(La)page n=41> <!-- Text from Lachmann, p. 41 --> ...
                <!-- etc. -->
```

The markup shown above would be interleaved with the normal markup for the document. Since SGML requires tags in concurrent markup streams to be labeled with their document type, however, the “normal” markup would need to have the notation “TEI.2” inserted before each tag's generic identifier. The combined markup might look something like this:

```
<(TEI.2)TEI.2>
<(TEI.2)TEI.Header> ... </(TEI.2)TEI.Header>
<(TEI.2)text> ...
<(TEI.2)div0><(TEI.2)head> ... </(TEI.2)head>
...
<(TEI.2)div1>
<(TEI.2)div2> ...
<(La)page n=37> <!-- Text from Lachmann, p. 37 -->
                ...
<(La)line n=32> <!-- Text from Lachmann, p. 37, line 32 -->
<(La)line n=33> <!-- Text from Lachmann, p. 37, line 33 -->
<(La)line n=34> <!-- Text from Lachmann, p. 37, line 34 -->
<(La)page n=38> <!-- Text from Lachmann, p. 38 -->
<(La)line n=1>  <!-- Text from Lachmann, p. 38, line 1 -->
<(La)line n=2>  <!-- Text from Lachmann, p. 38, line 2 -->
<(La)line n=3>  <!-- Text from Lachmann, p. 38, line 3 -->
</(TEI.2)div2>
<(TEI.2)div2>
<(La)line n=4>  <!-- Text from Lachmann, p. 38, line 4 -->
                <!-- etc. -->
</(TEI.2)div2>
</(TEI.2)div1>
<(TEI.2)div1><head>... </head>
                ...
<!-- Text from Lachmann, p. 39 --> ...
<(La)page n=39> <!-- Text from Lachmann, p. 39 --> ...
<(La)line n=18> <!-- Text from Lachmann, p. 39, line 18 -->
<(La)line n=19> <!-- Text from Lachmann, p. 39, line 19 -->
<(La)line n=20> <!-- Text from Lachmann, p. 39, line 20 -->
<(La)line n=21> <!-- Text from Lachmann, p. 39, line 21 -->
                ... <!-- etc. -->
<(TEI.2)div1>
<(TEI.2)div2>
</(TEI.2)div2>
<!-- etc. -->
<!-- Text from Lachmann, p. 40 --> ...
<!-- etc. -->
</(TEI.2)text>
</(TEI.2)TEI.2>
```

The following SGML declarations give the formal specification for the standard pre-defined document type designed for recording page and line numbers of a reference edition in a concurrent markup stream.

```
<!-- 6.9.3: Concurrent Document Type for Page and Line        -->
<!-- References                                                -->
<!ENTITY % version 'ref' >
<!ELEMENT %version      - -  (#PCDATA | page | vol)*           >
<!ATTLIST %version           %a.global;                        >
<!ELEMENT vol           - -  (#PCDATA | page)*                 >
<!ATTLIST vol                %a.global;                        >
<!ELEMENT page          - O  (#PCDATA | line | col)*           >
<!ATTLIST page               %a.global;                        >
```

```
<!ELEMENT col          - O  (#PCDATA | line)*              >
<!ATTLIST col               %a.global;                     >
<!ELEMENT line         - -  (#PCDATA)                      >
<!ATTLIST line              %a.global;                     >
```

This concurrent hierarchy is enabled as shown below: after the document type declaration for the TEI.2 document type, the document should contain the sequence of lines:

```
<!DOCTYPE La system 'teipl2.dtd' [
<!ENTITY % version "La" >
]>
```

which call the document type for page and line references and give it the name “La”. If page and line numbers from more than one standard edition are to be marked, then the relevant lines may be repeated, each time using a different value for the document type and entity definition (where the example has “La”). For example, to show page and line numbers from the editions of Lachmann (La), Kraus (Kr), and Moser/Tervooren (MT) at the same time, one might use declarations like the following:

```
<!DOCTYPE La system 'teipl2.dtd' [
     <!ENTITY % version "La" >
]>
<!DOCTYPE Kr system 'teipl2.dtd' [
     <!ENTITY % version "Kr" >
]>
<!DOCTYPE MT system 'teipl2.dtd' [
     <!ENTITY % version "MT" >
]>
```

See below, section 6.9.6, “Declaring Reference Systems,” on page 61for examples of declarations for the reference systems just shown.

6.9.4 Concurrent Markup for Other Hierarchies

Hierarchies similar to that defined above can be provided for most common hierarchical reference systems. Hierarchies such as act / scene / line, for conventional dramatic structure, book / canto / stanza / line, for longer verse texts, or book / poem / stanza / line, for collections of verse, may be readily expressed with concurrent SGML markup. Since these hierarchical structures can readily be represented using the base tag sets described in part III of these Guidelines, however, reference systems with such structures may most readily be expressed using the *n* ore *id* attributes, as described above in section 6.9.1, “Using the ID and N Attributes,” on page 54.

Any text with an idiosyncratic standard reference system will require its own DTD, so that appropriately named tags can be created for the reference units. Such DTDs may follow the pattern of those described in the preceding section; they should also be documented in an auxiliary tag set description file, using the tags described in chapter 30, “Tag Set Declaration,” [in separate fascicle].

6.9.5 Milestone Tags

If concurrent markup is not desired (e.g. because the available SGML parser does not support the CONCUR feature), or if the desired reference system does not correspond to any particular structural hierarchy, it may be more convenient to mark up changes in the reference system by using empty **<milestone>** elements.

<milestone> : marks the boundary between sections of a text, as indicated by changes in a standard reference system. Attributes include:
- *ed* : indicates which edition or version the milestone applies to.
- *n* : indicates the new number or other value for the unit which changes at this milestone.
- *unit* : indicates what kind of section is changing at this milestone. Sample values include:
  - *page* : page breaks in the reference edition.
  - *column* : column breaks.
  - *line* : line breaks.
  - *book* : any units termed *book*, *liber*, etc.
  - *poem* : individual poems in a collection.
  - *canto* : cantos or other major sections of a poem.
  - *stanza* : stanzas within a poem, book, or canto.
  - *act* : acts within a play.
  - *scene* : scenes within a play or act.
  - *section* : sections of any kind.
  - *absent* : passages not present in the reference edition.

<pb> : marks the boundary between one page of a text and the next in a standard reference system.

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

These elements simply mark the points in a text at which some category in a reference system changes. They have no content,but subdivide the text into regions just as milestones divide a road into segments. The elements **<pb>** and **<lb>** are provided to mark specific types of milestone, namely page and line boundaries, as further described in chapter 20, “Manuscripts, Analytic Bibliography, and Physical Description of the Source Text,” [in separate fascicle]. No SGML validation of a reference system based on **<milestone>** tags is possible, so it will be the responsibility of the encoder or the application software to ensure that milestone tags occur in a correct order.

Using these tags, the reference system of any one edition can be recreated from a text in which all are marked by simply ignoring all tags that do not specify that edition on the *ed* attribute.

For example, to mark the breaks between books and poems in an edition other than the source text, tags like the following may be interspersed in the document:

```
<milestone ed=N unit=work n='Amores'>
<milestone ed=N unit=book>
<milestone ed=N unit=poem>
...
<milestone ed=N unit=poem>
...
<milestone ed=N unit=book>
...
<milestone ed=N unit=book>
...
<milestone ed=N unit=book>
<milestone ed=N unit=poem>
...
<milestone ed=N unit=poem>
...
<milestone ed=N unit=book>
...
```

In this case no *n* value is specified, since the numbers rise predictably and the application can keep a count from the start of the document, if desired.

The value of the *n* attribute may but need not include the identifiers used for any larger sections. That is, either of the following styles is legitimate:

```
<milestone ed=N unit=work n='Amores'>
<milestone ed=N unit=book n=1>
<milestone ed=N unit=poem n=1>
... <!-- text of Amores 1.1 -->
<milestone ed=N unit=poem n=2>
... <!-- text of Amores 1.2 -->
<milestone ed=N unit=book n=3>
...
```

or

```
<milestone ed=N unit=work n='Amores'>
<milestone ed=N unit=book n=1>
<milestone ed=N unit=poem n='1.1'>
... <!-- text of Amores 1.1 -->
<milestone ed=N unit=poem n='1.2'>
... <!-- text of Amores 1.2 -->
<milestone ed=N unit=book n='1.3'>
...
```

When using **<milestone>** tags, line numbers may be supplied for every line or only periodically (every fifth, every tenth line). The latter may be simpler; the former is more reliable.

The style of numbering used in the values of *n* is unrestricted: for the example above, “I.i”, “I.ii”, and “I.iii” could have been used equally well if preferred. The special value *unnumbered* should be reserved for marking sections of text which fall outside the normal numbering system (e.g. chapter heads, poem numbers, titles, or speaker attributions in a verse drama).

No automatic validation of any hierarchical ordering can be defined for the various types of **<milestone>** tag; it is the encoder's responsibility to ensure that milestone values are valid.

Because the *ed* attribute is unrestricted, no change need be made to the document type declaration of a file before adding tags to describe a new reference system. (The value of *ed* may be restricted to a defined set of edition symbols by using the techniques described in chapter 32, “Modifying TEI DTDs,” [in separate fascicle].)

See below, section 6.9.6, “Declaring Reference Systems,”, for examples of declarations for the reference systems just shown.

The milestone tag has the following formal declaration:

```
<!-- 6.9.5:  Milestone tag                                    -->
<!ELEMENT milestone     - O  EMPTY                             >
<!ATTLIST milestone          %a.global;
          ed                 CDATA               #REQUIRED
          n                  CDATA               #IMPLIED
          unit               CDATA               #REQUIRED      >
<!-- This fragment is used in sec. 6                          -->
```

6.9.6 Declaring Reference Systems

Whatever kind of reference system is used in an electronic text, it is recommended that the TEI header contain a description of its construction in the **<refsDecl>** element described in section 5.3.4, “The Reference System Declaration,” [in separate fascicle]. As described there, the declaration may consist either of a formal declaration using the **<step>** tag or an informal description in prose. The former is recommended because unlike prose it can be processed by software.

The three examples given in section 6.9.1, “Using the ID and N Attributes,” on page 54would be declared as follows. The first example encodes the standard references for Ovid's *Amores* one level at a time, using the *n* attribute on the **<div0>**, **<div1>**, **<div2>**, and **<l>** tags. The header for such an encoding should look something like this:

```
<TEI.Header>
    <fileDesc> ...
    </fileDesc>
    <encodingDesc>
         ...
         <refsDecl>
              <step gi=div0 delim=' ' att=n>
              <step gi=div1 delim='.' att=n>
              <step gi=div2 delim='.' att=n>
              <step gi=L>
         </refsDecl>
         ...
    </encodingDesc>
</TEI.Header>
```

The second example encodes the same reference system, again using the *n* attribute on the **<div0>**, **<div1>**, **<div2>**, and **<l>** tags, but giving the reference string in full on each tag. The **<refsDecl>** element for such an encoding should look something like this:

```
<refsDecl>
    <step gi='div0 div1 div2 l' att='n'>
</refsDecl>
```

The third example encodes the same reference system, this time giving the entire reference string as the value of the *id* attribute on the relevant tags. The reference system declaration for such an encoding should look something like this:

```
<refsDecl>
    <step gi='div0 div1 div2 l' att='id'>
</refsDecl>
```

Declarations for reference system encoded in concurrent markup hierarchies take the following form. The example in section 6.9.3, “Concurrent Markup for Pages and Lines,” on page 56 showing page and line numbers from editions edited by Lachmann, von Kraus, and Moser and Tervooren would require a formal declaration something like this:

```
<refsDecl n=La>
    <step dtd=La gi='page' att='n'>
    <step dtd=La gi='line' att='n'>
</refsDecl>
<refsDecl n=Kr>
    <step dtd=Kr gi='page' att='n'>
    <step dtd=Kr gi='line' att='n'>
</refsDecl>
<refsDecl n=MT>
    <step dtd=MT gi='page' att='n'>
    <step dtd=MT gi='line' att='n'>
</refsDecl>
```

Reference systems recorded by means of milestone tags can also be declared; the example given in section 6.9.5, “Milestone Tags,” on page 59could be declared in this way, using a prose description:

```
<refsDecl>
    <p>Standard references to work, book, poem, and line may be
    constructed from the <gi>milestone</gi> tags in the text.
</refsDecl>
```

Or in this way, using a formal declaration:

```
<refsDecl>
    <state unit=work>
    <state unit=book>
    <state unit=poem>
    <state unit=line>
```

```
</refsDecl>
```

6.10 Bibliographic Citations and References

Bibliographic citations (that is, full descriptions of bibliographic items such as books, articles, films, broadcasts, songs, etc.) or pointers to them may appear at various places in a TEI text. They are required at several points within the TEI Header's source description, as discussed in section 5.4, “The Profile Description,” [in separate fascicle]; they may also appear within the body of a text, either singly, (for example within a footnote), or collected together in a list as a distinct part of a text.

In printed texts, the individual constituents of a citation are conventionally marked off from each other and from the flow of text by such features as bracketing, italics, special punctuation conventions, underlining, etc. In electronic texts, such distinctions are also important, whether in order to produce acceptably formatted output or to facilitate intelligent retrieval processing,[8] quite apart from the need to distinguish the citation itself as a textual object with particular linguistic properties.

It should be emphasized that for citations as for other textual features, the primary or sole consideration is not how the text should be formatted when it is printed. The distinctions permitted by the scheme outlined here may not necessarily be all that particular formatters or bibliographic styles require, although they should prove adequate to the needs of many such commonly used software systems.[9] The features distinguished and described in section 6.10.2, “Components of Bibliographic Citations,” on page 67constitute a set which has been useful for a wide range of bibliographic purposes and in many applications, and which moreover corresponds to a great extent with existing bibliographic and library cataloguing practice. For a fuller account of that practice as applied to electronic texts see section 5.2.7, “The Source Description,” [in separate fascicle]; for a brief mention of related library standards see section 5.7, “Note for Library Cataloguers,” [in separate fascicle].

6.10.1 Bibliographic Citation Elements

The following elements are used to mark individual bibliographic citations as wholes, or groups of bibliographic citations:

&@fn@sep

[8] For example, to distinguish “London” as an author's name from “London” as a place of publication or as a component of a title.

[9] Among the bibliographic software systems and subsystems consulted in the design of the **<bibl.struct>** structure were BibTeX, Scribe, and ProCite. The distinctions made by all three may be preserved in **<bibl.struct>** structures, though the nature of their design prevents a simple one-to-one mapping from their data elements to TEI elements. For further information, see section 6.10.4, “Relationship to Other Bibliographic Schemes,” on page 80.

<bibl> : contains a loosely-structured bibliographic citation of which the sub-components may or may not be explicitly tagged.
<bibl.struct> : contains a structured bibliographic citation, in which only bibliographic subelements appear and in a specified order.
<bibl.full> : contains a fully-structured bibliographic citation, in which all components of the TEI file description save the source description are present.
<list.bibl> : contains a list of bibliographic citations of any kind.

These elements all share a number of possible component sub-elements. For the **<bibl>** and **<bibl.struct>** elements, exactly the same sub-elements are concerned, and they are described together in section 6.10.2, “Components of Bibliographic Citations,” on page 67; for the **<bibl.full>** element, the sub-elements concerned are fully described in section 5.2.7, “The Source Description,” [in separate fascicle].

Different levels of specific tagging may be appropriate in different situations. In some cases, it may be felt necessary to mark just the extent of the citation itself, with perhaps a few distinctions being made within it (for example, between the part of the citation which identifies a title or author and the rest). Such citations, containing a mixture of text with specialized bibliographic elements, are regarded as **<bibl>** elements, and tagged accordingly. For example:

```
<p>A book which had a great influence on him
was <bibl>Tufte's <title>Envisioning
Information</title></bibl>, although he may
never have actually read it.
```

Indeed, some encoders may find it unnecessary to mark the bibliographic citation at all:

```
<p>A book which had a great influence on him
was Tufte's <title>Envisioning Information</title>,
although he may never have actually read it.
```

Some bibliographic citations are extremely elliptical, often only a string of the form “Baxter, 1983”. If no further details of Baxter's book are given in the source text and none are supplied by the encoder, then the citation thus given should be tagged as a **<bibl>**:

```
All of this is of course much more fully treated
in <bibl>Baxter, 1983</bibl>.
```

In general, however, normal modern bibliographic practice, and these Guidelines, distinguish between a bibliographic *citation*, which is a self-sufficient description of a bibliographic item, and a bibliographic *reference*, which is a short-form citation (such as “Baxter, 1983”) which serves usually as a place-holder or pointer to a full long-form citation found elsewhere in the text. The usual encoding of short-form citations such as “Baxter, 1983” is not as **<bibl>** elements but as cross-references to such elements; see section 6.10.3, “Citation References,” on page 79 below.

In cases where the encoder wishes to impose more structure on the bibliographic information, for example to make sure it conforms to a particular stylesheet or retrieval processor, the **<bibl.struct>** element should be used. Note that several of the features in this and later examples are explained later in the current section.

```
<bibl.struct>
    <author>Edward R. Tufte</>
    <title>Envisioning Information</>
    <imprint>
         <city>Cheshire, Conn.</>
         <publisher>Graphics Press</>
         <date>1990</>
    </imprint>
</bibl.struct>
```

The highest level of detail and the most complex structure supported by the current proposals is provided by the **<bibl.full>** element, which closely resembles the **<file.desc>** element of the TEI Header (section 5.2.7, "The Source Description," [in separate fascicle]).

```
<bibl.full>
    <titleStmt>
         <author>Tufte, Edward R[olf]</>
         <title>Envisioning Information</>
    </titleStmt>
    <extent>126 pp.</>
    <publicationStmt>
         <creation.date>1990</>
         <publisher>Graphics Press</>
         <city>Cheshire, Conn. USA</>
         <date>1990</>
    </publicationStmt>
</bibl.full>
```

A list of bibliographic items, of whatever kind, may be treated in the same way as any other list (see section 6.7, "Lists," on page 44). Alternatively, the specialized **<list.bibl>** element may be used. The difference between the two is that a **<list>** contains **<items>**, within which citation elements (**<bibl>**, **<bibl.struct>** or **<bibl.full>**) may appear, as well as other phrase- and paragraph-level elements, whereas the **<list.bibl>** may contain only citation elements, optionally preceded by a heading and a series of introductory paragraphs. The former would be appropriate for a list of bibliographic elements in which descriptive prose predominated, and the latter for a more formal bibliography. The following are thus both legal encodings of a list of bibliographic entries: a **<list.bibl>**:[10]

---

[10] In these examples, white space has been liberally used to help indicate the structure of the entries; it is not required and is not necessarily used in practice.

```
<list.bibl>
<head>Bibliography</head>
<bibl.struct id=NEL80>
    <analytic>
         <author>Nelson, T. H.</>
         <title>Replacing the printed word:
                a complete literary system.</>
    </analytic>
    <monogr>
         <title>Information Processing '80:  Proceedings of the IFIPS
                Congress, October 1980</>
         <editor>Simon H. Lavington</>
         <imprint>
              <publisher>North-Holland</>
              <city>Amsterdam</>
              <date>1980</>
         </imprint>
         <biblScope>pp 1013-23
    </monogr>
    <note>Apparently a draft of section 4 of <title>Literary
          Machines</title>.</note>
</bibl.struct>
<bibl id=NEL88>Ted Nelson:  <title>Literary Machines</title>
    (privately published, 1987)</bibl>
<bibl id=BAX88>
    <author>Baxter, Glen</author>
    <title>Glen Baxter His Life: the years of struggle</title>
    London: Thames and Hudson, 1988.
    </bibl>
</list.bibl>
```

or a simple **<list>**:

```
<list><head>Bibliography</>
<item><bibl id=NEL80>
        <author>Nelson, T. H.</>
        <title level=a>Replacing the printed word:
             a complete literary system.</>
        <title level=m>Information Processing '80:
             Proceedings of the IFIPS Congress, October 1980</>
        <editor>Simon H. Lavington</>
        <publisher>North-Holland</>
        <city>Amsterdam</>
        <date>1980</>
        <biblScope>pp 1013-23
        <note>Apparently a draft of section 4 of <title>Literary
         Machines</title>.</note>
    </bibl></item>
<item><bibl id=NEL88>Ted Nelson:  <title>Literary Machines</title>
    (privately published, 1987)</bibl></item>
<item><bibl id=BAX88>
    <author>Baxter, Glen</author>
    <title>Glen Baxter His Life: the years of struggle</title>
```

```
    London: Thames and Hudson, 1988.
    </bibl></item>
</list>
```

The formal declarations for these elements are as follows:

```
<!-- 6.10.1:  Tags for Bibliographic Citations              -->
<!ELEMENT bibl          - o  (#PCDATA | %m.phrase; |
                             %m.bibl.part;)*                  >
<!ATTLIST bibl               %a.global;                       >
<!ELEMENT bibl.struct   - o  (analytic?, (monogr, series*)+,
                             (note | idno)*)                  >
<!ATTLIST bibl.struct        %a.global;                       >
<!ELEMENT bibl.full     - o  (title.stmt, edition.stmt?,
                             extent?, publication.stmt,
                             series.stmt?, notes.stmt?,
                             source.desc*)                    >
<!ATTLIST bibl.full          %a.global;                       >
<!ELEMENT list.bibl     - -  (head?, (bibl | bibl.struct |
                             bibl.full)+, trailer?)           >
<!ATTLIST list.bibl          %a.global;                       >
<!-- (continued in sec. 6.10.2.1, 6.10.2.2, 6.10.2.3)       -->
<!-- This fragment is used in sec. 6                        -->
```

6.10.2 Components of Bibliographic Citations

This section discusses a number of very commonly occurring component elements of bibliographic citations. They fall into four groups:

- elements for grouping components of the *analytic*, *monographic*, and *series* levels in a structured citation
- titles of various kinds, and statements of intellectual responsibility (authorship, etc.)
- information relating to the publication, pagination, etc. of an item
- annotation, commentary and further detail

The following sections describe the elements which may be used to represent such information within a **<bibl>** or **<bibl.struct>** element. Within the former, any or all of these may be used and in any order. Within the latter, such of these elements as exist for a given citation must be distinguished, and must also be presented in a specific order, discussed further below (section 6.10.2.6, “Order of Components within Citations,” on page 79).

6.10.2.1 Analytic, Monographic, and Series Levels

In common library practice a clear distinction is made between an individual item within a larger collection and a free-standing book, journal, or collection. Similarly a book in a series is distinguished sharply from the series within which it appears. An article forming part of a collection which itself appears in a series thus has a bibliographic description with three quite distinct levels of information:

1. the *analytic* level, giving the title, author, etc., of the article
2. the *monographic* level, giving the title, editor, etc., of the collection
3. the *series* level, giving the title of the series, possibly the names of its editors, etc., and the number of the volume within that series.

In the same way, an article in a journal requires at least two levels of information: the analytic level describing the article itself, and the monographic level describing the journal.

These three levels may be distinguished within a **<bibl>** element, and must be distinguished within a **<bibl.struct>** element if present, by means of the following tags:

- **<analytic>** : contains bibliographic elements describing an item (e.g. an article or poem) published within a monograph or journal and not as an independent publication.
- **<monogr>** : contains bibliographic elements describing an item (e.g. a book or journal) published as an independent item (i.e. as a separate physical object).
- **<series>** : contains information about the series in which a book or other bibliographic item has appeared.

For purposes of TEI encoding, journals and anthologies are both treated as monographs; a journal title will thus be tagged **<title level=j>** or **<monogr><title>** ... **</title>** ... **</monogr>**. Individual articles in the journal or collected texts should be treated at the “analytic” level. When an article has been printed in more than one journal or collection, the bibliographic citation may have more than one **<monogr>** element, each possibly followed by one or more **<series>** elements. A **<series>** element always relates to the most recently preceding **<monogr>** element. (Whether reprints of an article are treated in the same bibliographic citation or a separate one varies among different styles. Library lists typically use a different entry for each publication, while academic footnoting practice typically treats all publications of the same article in a single entry.)

For example, the article cited in this example has been published twice, once in a journal and once in a collection which appeared in a series:

```
<bibl.struct>
     <analytic>
          <author>Thaller, Manfred</author>
          <title level=a>A Draft Proposal for a Standard for the
                 Coding of Machine Readable Sources</>
     </analytic>
     <monogr>
          <!-- In -->
          <title level=j>Historical Social Research</title>
          <imprint>
                <biblScope type=vol>40</>
                <date>October 1986</>
                <biblScope type=pages>3-46</>
          </imprint>
     </monogr>
```

```
        <monogr>
             <!-- Rpt. in -->
             <title level=m>Modelling Historical Data:
                    Towards a Standard for Encoding and Exchanging
                    Machine-Readable Texts</title>
             <editor>Daniel I. Greenstein</editor>
             <imprint>
                  <city>St. Katharinen</city>
                  <publ>Max-Planck-Institut f&uuml;r Geschichte
                        In Kommission bei
                        Scripta Mercaturae Verlag</publ>
                  <date>1991</date>
             </imprint>
        <monogr>
        <series>
             <title level=s>Halbgraue Reihe
                  zur Historischen Fachinformatik</title>
             <resp><role>Herausgegeben von</role>
                   <individual>
                   <name type=person>Manfred Thaller</>
                   <org>Max-Planck-Institut f&uuml;r
                       Geschichte</></individual>
             </resp>
             <title level=s>Serie A:  Historische Quellenkunden</>
             <biblScope>Band 11</biblScope>
        </series>
   </bibl.struct>
```

Punctuation may not appear between the elements within a structured bibliographic entry; if punctuation is to be given explicitly in the encoding, it must be contained within the elements it delimits. As the example shows, it is possible to encode the entry without any inter-element punctuation: this facilitates use of the **<bibl.struct>** element in systems which can render bibliographic citations in any of several styles.

The formal declarations for the elements defined in this section are as follows:

```
<!-- 6.10.2.1:  Tags for Bibliographic Citations (cont'd)     -->
<!-- (continuation of sec. )                                  -->
<!ELEMENT analytic      - O  (author | editor | resp | title)*  >
<!ATTLIST analytic           %a.global;                         >
<!ELEMENT monogr        - O  ( ( ((author | editor | resp)+,
                             title+, (editor | resp)*) |
                             (title+, (author | editor |
                             resp)*)), (note | meeting)*,
                             (edition, (editor | resp)*),
                             imprint+, biblScope* )            >
<!ATTLIST monogr             %a.global;                        >
<!ELEMENT series        - O  (title | editor | resp |
                             biblScope)*                       >
<!ATTLIST series             %a.global;                        >
```

6.10.2.2 Authors, Titles, and Editors

Bibliographic citations typically begin with a statement of the title being cited and the names of those intellectually responsible for it. For articles in journals or collections, such statements should appear both for the analytic and for the monographic level. The following elements are provided for tagging such elements:

- **<title>** : contains the title of a work, whether article, book, journal, or series, including any alternative titles or subtitles. Attributes include:
  - *level* : indicates whether this is the title of an article, book, journal, series, or unpublished material. Sample values include:
    - *a* : analytic title (article, poem, or other item published as part of a larger item)
    - *m* : monographic title (book, collection, or other item published as a distinct item, including single volumes of multi-volume works)
    - *j* : journal title
    - *s* : series title
    - *u* : title of unpublished material (including theses and dissertations unless published by a commercial press)
  - *type* : classifies the title according to some convenient typology. Sample values include:
    - *main* : main title
    - *subordinate* : subtitle, title of part
    - *parallel* : alternate title, often in another language, by which the work is also known
    - *abbreviated* : abbreviated form of title
- **<author>** : primary *statement of responsibility* for a bibliographic item, for example the name of a single author, institution or organization, or of several such.
- **<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.
- **<resp>** : supplies a statement of responsibility for someone responsible for the intellectual content of a text, edition, recording, or series, where the **<author>**, **<editor>** and other specialized elements do not suffice.
- **<role>** : contains a phrase describing the nature of a person's intellectual responsibility.
- **<name>** : contains a proper noun or noun phrase.
- **<person>** : contains a name of a person, or a reference to a person.
- **<indiv>** : contains structured information about an individual.
- **<org>** : contains the name of an organization, or a reference to an organization.
- **<organization>** : contains structured information about an organization: its name and address.

In bibliographic citations, all titles should be tagged as such, whether analytic, monographic, or series titles. The single element **<title>** is used for all these cases. When it appears directly within an **<analytic>**, **<monogr>**, or **<series>** element, **<title>** is interpreted as belonging to the appropriate level. When it appears elsewhere, its *level* attribute should be used to signal its bibliographic level. It is a semantic error to give a value for the *level* attribute which is inconsistent with the context; such values may be ignored. The *level* value “a” implies the analytic level; the values “m”, “j”, and “u” imply the monographic level; the value “s” implies the series level. Note, however, that the semantic error occurs only if the nested title is directly enclosed by the **<analytic>**, **<monogr>**, or **<series>** element; if it is enclosed only indirectly, no semantic error need be present. For example, the analytic title may contain a monographic title:

```
<bibl.struct>
    <analytic><author>Lucy Allen Paton</>
              <title>Notes on Manuscripts
                   of the <title level=m>Proph&eacute;cies
                   de Merlin</title></title>
    </analytic>
    <monogr><title level=j>PMLA</>
            <biblscope type=vol>8</>
            <date>1913</date>
            <biblscope type=pages>122</>
    </monogr>
</bibl.struct>
```

In this case, the analytic title “Notes on Manuscripts of the *Prophécies de Merlin*” needs no *level* attribute because it is directly contained by the **<analytic>** level; the monographic title contained within it, “Prophécies de Merlin,” does not create a semantic error because it is not directly contained by the **<analytic>** element.

In some bibliographic applications, it may prove useful to distinguish main titles from subordinate titles, parallel titles, etc. The *type* attribute is provided to allow this distinction to be recorded.

The following citation, from a national standard for bibliographic references,[11] illustrates this type of analysis with its distinction between main and subordinate titles.

```
<bibl>Saarikoski, Pirkko-Liisa, and Paavo Suomalainen,
<title level=a type=main>Studies on the physiology of
the hibernating hedgehog, 15</title>
<title level=a type=subordinate>Effects of seasonal
and temperature changes on the in vitro
glycerol release from brown adipose tissue</>
<title level=j>Ann. Acad. Sci. Fenn., Ser. A4</>
<date>1972</> <biblScope type='vol: pp'>187: 1-4</>
</bibl>
```

Slightly more complex is the distinction made below among main, subordinate, and parallel titles, in an example from the same source (p. 63). The punctuation and the bibliographic analysis are those given in ANSI Z39.29-1977; the punctuation is in the style prescribed by the International Standard Bibliograph-

[11] American National Standard for Bibliographic References, ANSI Z39.29-1977 (New York: American National Standards Institute, 1977), p. 34 (sec. A.2.2.1).

ic Description (ISBD).[12]

```
<bibl>Tchaikovsky, Peter Ilich.
<title level=m type=main>The swan lake ballet</>
= <title level=m type=parallel>Le lac des cygnes</>
: <title level=m type=subordinate>grand ballet en 4 actes</>
: <title level=m type=subordinate>op. 20</>
[Score].
New York:  Broude Brothers; [1951] (B.B. 59). vi, 685 p.
</bibl>
```

The elements **<author>** and **<editor>** have, for printed books and articles, a fairly obvious significance; for other kinds of bibliographic items their proper usage may be less obvious. The **<author>** element should be used for the person or agency with primary responsibility for a work's intellectual content, and the element **<editor>** for an editor of the work. Thus an organization such as a radio or television station is usually accounted "author" of a broadcast, for example, while the author of a Government report will usually be the agency which produced it.

For anyone else with responsibility for the work, the **<resp>** element may or must be used; this element contains both a **<role>** and the name of the responsible party (tagged as **<name>**, **<person>**, **<individual>**, **<org>**, or **<organization>**). Examples of secondary responsibility of this kind include the roles of illustrator, translator, editor, annotator. **<resp>** may also be used for editors, if it is desired to record the specific terms in which their role is described.

Examples of **<author>** and **<editor>** may be found in sections 6.10.1, "Bibliographic Citation Elements," on page 63, and 6.10.2.1, "Analytic, Monographic, and Series Levels," on page 67; wherever **<author>** and **<editor>** may occur, the **<resp>** element may also occur. When one of these elements precedes or immediately follows a title, it applies to that title; when it follows an **<edition>** element or occurs within an edition statement, it applies to the edition in question.

In this example, the **<resp>** elements apply to the work as a whole, not merely to the first edition:

```
<bibl>
    <author>Lominadze, D. G.</>
    <title level=m>Cyclotron waves in plasma.</title>
    <resp><role>translated by</> <name>A. N. Dellis;</name></resp>
    <resp><role>edited by</> <name>S. M. Hamberger.</name></resp>
    <edition>1st ed.</>
    <imprint><place>Oxford:</> <publisher>Pergamon Press,</>
       <date>1981.</></imprint>
    <extent>206 p.</>
```

[12] The analysis is not wholly unproblematic: as the text of the standard points out, the first subordinate title is subordinate only to the parallel title in French, while the second is subordinate to both the English main title and the French parallel title, without this relationship being made clear, either in the markup given below or in the reference structure offered by the standard.

```
    <title level=s>International series in natural philosophy.</>
    <note place=inline>Translation of:  <title level=m>Ciklotronnye
    volny v plazme.</></>
</bibl>
```

In this example, by contrast, the **<resp>** element applies to the edition, and not to the collection per se (Moser and Tervooren were not responsible for the first thirty-five printings); the elements of the citation have been reordered from their appearance on the title page of the volume in order to ensure the correct relationship of the collection title, the edition statement, and the statement of responsibility.

```
<bibl.struct>
<monogr>
    <title>Des Minnesangs Fr&uuml;hling</title>
    <edition>36., neugestaltete und erweiterte Auflage</edition>
    <note place=inline>Mit 1 Faksimile</note>
    <resp>
         <role>Unter Benutzung der Ausgaben
         von <person>Karl Lachmann</> und <person>Moriz Haupt</>,
         <person>Friedrich Vogt</> und <person>Carl von Kraus</>
         bearbeitet von
         </role>
         <person>Hugo Moser</>
         <!-- und -->
         <person>Helmut Tervooren</>
    </resp>
    <biblScope type=volume>I</>
    <title level=m type='volume title'>Texte</>
    <imprint>
         <city>Stuttgart</>
         <publisher>S. Hirzel Verlag</>
         <date>1977</>
    </imprint>
    <extent>459 pp.</>
</monogr>
</bibl.struct>
```

The formal definition of the **<title>** element may be found in section 6.3.2, “Emphasis, Foreign Words, and Unusual Language,” on page 12. The formal definition of the **<name>**, **<person>**, **<individual>**, **<org>**, and **<organization>** elements may be found in section 6.4, “Names, Numbers, Dates, Abbreviations, and Addresses,” on page 24. The formal definitions of the remaining elements described in this section are as follows:

```
<!-- 6.10.2.2:  Tags for Bibliographic Citations (cont'd)     -->
<!-- (continuation of sec. )                                  -->
<!ELEMENT author        - o  (%phrase.seq)                      >
<!ATTLIST author             %a.global;                         >
<!ELEMENT editor        - o  (%phrase.seq)                      >
<!ATTLIST editor             %a.global;
          role               CDATA              editor          >
<!ELEMENT resp          - o  ((role & (%m.agent)), (role |
```

```
                              (%m.agent))*+)                   >
<!ATTLIST resp                %a.global;                       >
<!ELEMENT role          - o  (%phrase.seq;)                    >
<!ATTLIST role                %a.global;                       >
```

6.10.2.3 Imprint, Pagination, and Other Details

By *imprint* is meant all the information relating to the publication of a work: the person or organization by whose authority and in whose name a bibliographic entity such as a book is made public or distributed (whether a commercial publisher or some other organization), the place of publication, and a date. It may also include a full address for the publisher or organization. Full bibliographic citations usually specify either the number of pages in a print publication (or equivalent information for non-print materials), or the specific location of the material being cited within its containing publication. The following elements are provided to hold this information:

<imprint> : groups information relating to the publication or distribution of a bibliographic item.
<city> : contains the name of a town, city or village occurring within an address or imprint.
<address> : contains a postal or other address, for example of a publisher, an organization, or an individual
<place> : contains the name of, or a reference to, a place.
<publisher> : provides the name of the organization responsible for the publication or distribution of a bibliographic item.
<org> : contains the name of an organization, or a reference to an organization.
<organization> : contains structured information about an organization: its name and address.
<date> : contains a date in any format.
<idno> : supplies any standard or non-standard number used to identify a bibliographic item. Attributes include:
  type : categorizes the number, for example as an ISBN or other standard series.
<extent> : describes the approximate size of the electronic text as stored on some carrier medium, specified in any convenient units.
<biblScope> : defines the scope of a bibliographic reference, for example as a list of pagenumbers, or a named subdivision of a larger work. Attributes include:
  type : identifies the type of information conveyed by the element, e.g. “pages”, “volume”. Sample values include:
    volume : the element contains a volume number.
    issue : the element contains an issue number, or volume and issue numbers.
    pages : the element contains a page number or page range.
    chapter : the element contains a chapter indication (number and/or title)
    part : the element identifies a part of a book or collection.

The **<address>** element is as described in section 6.4.5, "Addresses," on page 32. For bibliographic purposes, usually only the city (or cities) of publication are required, possibly including the name of the country; for this reason, the element **<city>** may be supplied on its own (outside of an **<address>** element) in this context. Where the full postal address is likely to be of importance in identifying or locating the bibliographic item concerned, it should if known be supplied and tagged using the **<address>** element. If desired, the **<place>** element described in section 6.4.1, "Names," on page 24 may be used instead; this involves no claim that the information given is either a full address or the name of a city.

The name of the publisher of an item may be marked using the **<publisher>** tag; if the item is made public ("published") by an organization other than a conventional publisher, it may seem preferable to identify that organization using the **<org>** or **<organization>** element. This is frequently the case with technical reports:

```
<bibl.struct>
<monogr>
    <author>Nicholas, Charles K.</>
    <author>Welsch, Lawrence A.</>
    <title>On the interchangeability of SGML and ODA</>
    <imprint>
         <city>Gaithersburg, MD</city>
         <org>National Institute of Standards and Technology</>
         <date value='1992-01'>January 1992</>
         <stdnum type='NIST'>NISTIR 4681</>
    </imprint>
    <extent>19 pp.</>
</monogr>
</bibl.struct>
```

and with dissertations:

```
<bibl.struct>
<monogr>
    <author>Hansen, W.</>
    <title level=u>Creation of hierarchic text
         with a computer display</>
    <note place=inline>Ph.D. dissertation</>
    <imprint>
         <org>Dept. of Computer Science, Stanford Univ.</>
         <place>Stanford, CA</>
         <date value='1971-06'>June 1971</>
    </imprint>
</monogr>
</bibl.struct>
```

When an item has been reprinted, especially reprinted without change from a specific earlier edition, the reprint may appear in a **<monogr>** element with only the **<imprint>** and other details of the reprint. In the following example, a microform reprint has been issued without any change in the title or authorship. The series statement here applies only to the second **<monogr>** element.

```
<bibl.struct>
<monogr>
    <author>Shirley, James</>
    <title type=main>The gentlemen of Venice</>
    <title type=subordinate>a tragi-comedie presented at
         the private house in Salisbury Court by
         Her Majesties servants</>
    <note place=inline>[Microform]</>
    <imprint>
         <city>London</>
         <publisher>H. Moseley</>
         <date>1655</>
    </imprint>
    <extent>78 p.</>
</monogr>
<monogr>
    <imprint>
         <city>New York</>
         <publisher>Readex Microprint</>
         <date>1953</>
    </imprint>
    <extent>1 microprint card, 23 x 15 cm.</>
</monogr>
<series>
    <title>Three centuries of drama:  English, 1642-1700</>
</series>
</bibl>
```

A bibliographic description, particularly for an analytic title, will often include some additional information specifying its location, for example as a volume number, page number, range of page numbers, or name or number of a subdivision of the host work. The element **<biblScope>** may be used to identify such information if it is present. Where it is desired to distinguish different class of such information (volume number, page number, chapter number, etc.), the *type* attribute may be used with any convenient typology.

When the item being cited is a journal article, the **<imprint>** element describing the issue in which it appeared will typically contain **<biblScope>** elements for volume and page numbers, together with a **<date>** element.

For example:

```
<bibl.struct>
<analytic>
    <author>Wrigley, E. A.</>
    <title>Parish registers and the historian</>
</analytic>
<!-- in -->
<monogr>
    <editor>Steel, D. J.</>
    <title>National index of parish registers</>
    <imprint>
         <city>London</city>
```

```
          <publisher>Society of Genealogists</>
          <date value='1968'>1968</>
     </imprint>
     <biblScope type=volume>vol. 1</>
     <biblScope type=pages>pp. 155-167.</>
</monogr>
</bibl.struct>
```

The *type* attribute on **<biblScope>** is optional: both the following are legal examples:

```
<bibl.struct>
<analytic>
    <author>Boguraev, Branimir</>
    <author>Neff, Mary</>
    <title>Text Representation, Dictionary Structure,
         and Lexical Knowledge</>
</analytic>
<!-- in -->
<monogr>
    <title level=j>Literary & Linguistic Computing</>
    <imprint>
         <biblScope type=volume>7</>
         <biblScope type=issue>2</>
         <date>1992</>
         <biblScope type=pages>110-112</>
    </imprint>
</monogr>
</bibl.struct>

<bibl.struct>
<analytic>
    <author>Chesnutt, David</>
    <title>Historical Editions in the States</>
</analytic>
<!-- in -->
<monogr>
    <title level=j>Computers and the Humanities</>
    <imprint>
         <biblScope type=volume>25</>
         <biblScope type=issue>6</>
         <date value='1991-12'>December, 1991</>
         <biblScope type=pages>377-380</>
    </imprint>
</monogr>
</bibl.struct>
```

Formal definitions for the elements described in this section are as follows:

```
<!-- 6.10.2.3:  Tags for Bibliographic Citations (cont'd)     -->
<!-- (continuation of sec. )                                  -->
<!ELEMENT title         - o  (%soup;)                          >
<!ATTLIST title              %a.global;
          level              (a | m | j | s | u)
                                                  #IMPLIED
          type               CDATA                #IMPLIED      >
<!ELEMENT imprint       - o  (%m.place; | publisher | %m.date |
                             biblScope)*                       >
<!ATTLIST imprint            %a.global;                        >
<!ELEMENT publisher     - o  (%phrase.seq)                     >
<!ATTLIST publisher          %a.global;                        >
<!ELEMENT idno          - o  (#PCDATA)                         >
<!ATTLIST idno               %a.global;
          type               CDATA                #IMPLIED      >
<!ELEMENT extent        - o  (%phrase.seq; )                   >
<!ATTLIST extent             %a.global;                        >
<!ELEMENT biblScope     - o  (%phrase.seq)                     >
<!ATTLIST biblScope          %a.global;
          type               CDATA                #IMPLIED      >
<!-- Note and dates are defined elsewhere
                   -->
```

6.10.2.4 Series Information

Series information may (in **<bibl>** elements) or must (in **<bibl.struct>** elements) be enclosed in a **<series>** element or (in a **<bibl.full>** element) a **<seriesStmt>** element. The title of the series may be tagged **<title level=s>**, the volume number **<biblScope type=volume>**, and responsibility statements for the series (e.g. the name and affiliation of the editor, as in the example in section 6.10.2.1, "Analytic, Monographic, and Series Levels," on page 67) may be tagged **<editor>** or **<resp>**.

6.10.2.5 Notes and Other Additional Information

Explanatory notes about the publication of unusual items, the form of an item (e.g. "[Score]" or "[Microform]"), or its provenance (e.g. "translation of ...") may be tagged using the **<note>** element. The same element may be used for annotation of a bibliographic entry in a database.

<note> : contains a note or annotation. Attributes include:
- *type* : describes the type of note.
- *place* : indicates where the note appears in the source text. Sample values include:
  - *foot* : note appears at foot of page.
  - *end* : note appears at end of chapter or volume.
  - *inline* : note appears as a marked paragraph in the body of the text.
  - *left* : note appears in left margin.
  - *right* : note appears in right margin.

interlinear : note appears between lines of the text.
app[aratus] : note appears in the apparatus at the foot of the page.

```
<bibl><author>Coombs, James H., Allen H. Renear,
       and Steven J. DeRose.</author>
       <title level=a>Markup Systems and the Future of Scholarly
       Text Processing.</title>
       <title level=j>Communications of the ACM</title>
       <biblScope>30.11 (November 1987):  933-947.</biblScope>
       <note>Classic polemic supporting descriptive over procedural
            markup in scholarly work.</note>
```

6.10.2.6 Order of Components within Citations

The order of elements in **<bibl>** elements is not constrained.

In a **<bibl.struct>** elements, the **<analytic>** element, if it occurs, must come first, followed by one or more **<monogr>** and **<series>** elements, which may appear intermingled (as long as a **<monogr>** element comes first). Within **<analytic>**, the title(s), author(s), editor(s), and other statements of responsibility may appear in any order; it is recommended that all forms of the title be given together. Within **<monogr>**, the author, editor, and statements of responsibility may either come first or else follow the monographic title(s). Following these, the elements must appear in the following order:

- **<note>**s on the publication (and **<meeting>** elements describing the conference, in the case of a proceedings volume)
- **<edition>** elements, each followed by any related **<editor>** or **<resp>** elements
- **<imprint>**
- **<biblScope>**

Within **<imprint>**, the elements allowed may appear in any order.

Finally, within the **<series>** information in a **<bibl.struct>**, the sequence of elements is not constrained.

If more detailed structuring of a bibliographic description is required, the **<bibl.full>** element should be used. This is not further described here, as its contents are essentially equivalent to those of the **<fileDesc>** element in the **<TeiHeader>**, which is fully described in section 5.2.7, “The Source Description,” [in separate fascicle].

6.10.3 Citation References

References which are pointers to bibliographic items, of whatever kind, should be treated in the same way as other cross-references (see section 6.6, “Simple Links and Cross References,” on page 41). As discussed in that section, cross referencing within TEI texts is in general represented by means of **<ptr>** or **<ref>** elements. A *target* attribute on these elements is used to supply an identifying value for the target of the cross reference, which should be, in the case of bibliographic elements, a citation element of some kind. Where the form of the reference itself is unimportant, or may be reconstructed mechanically, or is not to be encoded, the **<ptr>** element is used, as in the following example:

```
As shown above (<ptr target=NEL80>)...
```

Where the form of the reference is important, or contains additional qualifying information which is to be kept but distinguished from the surrounding text, the **<ref>** element should be used, as in the following example:

```
Nelson claims <ref target=NEL80>(ibid, passim)</ref> ...
```

When it is desired to distinguish the short form of the reference itself from the additional qualifying information added to it, the additional information may be marked as a **<biblScope>** element:

```
Nelson claims (<ref target=NEL80>Nelson [1980], <biblScope>especially
pages 13-37</ref>) ...
```

(This distinction between the short form and the qualifying information may be useful in the context of document production or normalization of bibliographic references.)

6.10.4 Relationship to Other Bibliographic Schemes

The bibliographic tagging defined here can capture the distinctions required by most bibliographic encoding systems; for the benefit of users of some commonly used systems, the following lists of equivalences are offered, showing the relationship of the markup defined here to the fields defined for bibliographic records in the Scribe, BibTeX, and ProCite systems.

The various bibliographic fields defined for use in the Scribe and BibTeX systems of bibliographic databases have the following equivalents in the scheme presented here:[13] Elements and structures available in the tag set defined here which have no analogues in Scribe and BibTeX are not noted.

address: tag as **<city>**, **<place>**, or **<address>**
annote: tag as **<note>**
author: tag as **<author>**
booktitle: tag as **<title level=m>** or **<title>** within **<monogr>**
chapter: tag as **<biblScope type=chapter>**
date: used only to record date entry was made in the bibliographic database; not supported
edition: tag as **<edition>**
editor: tag as **<editor>** or **<resp>**

[13] The BibTeX scheme is intentionally compatible with that of Scribe, although it omits some fields used by Scribe. Hence only one list of fields is given here.

editors: tag as multiple **<editor>** or **<resp>** elements
fullauthor: use the *reg* attribute on **<author>** or **<name>**
fullorganization: use the *reg* attribute on **<org>** or **<name>**
howpublished: tag as **<note>**, possibly using the form **<note place=inline>**
institution: used only for issuer of technical reports; tag as **<publisher>**, **<org>**, or **<organization>**
journal: tag as **<title level=j>** or **<title>** within **<monogr>**
key: used to specify an alternate sort key for the bibliographic item, for use instead of author's or editor's name; not supported
meeting: tag as **<meeting>** or as **<note>**
month: use **<date>**; if the date is not in a trivially parseable form, use the *value* attribute to provide an ISO-format equivalent
note: tag as **<note>**
number: tag as **<biblScope type=issue>** or **<biblScope type=number>**; for technical report numbers, use **<idno type=docno>**
organization: used only for sponsor of conference; use **<org>** or **<organization>** within **<resp>** within **<meeting>** element
pages: tag as **<biblScope type=pages>**
publisher: tag as **<publisher>**
school: used only for institutions at which thesis work is done; tag as **<org>** or **<organization>**
series: tag as **<title level=s>** or **<title>** within **<series>**
title: tag as **<title>** in appropriate context or with appropriate *level* value
volume: tag as **<biblScope type=volume>**
year: tag as **<date>**; if the date is not in a trivially parseable form, use the *value* attribute to provide an ISO-format equivalent

6.11 Passages of Verse or Drama

The following elements are included in the core tag set for the convenience of those encoding texts which include mixtures of prose, verse and drama.

- **<l>** : contains a single, possibly incomplete, line of verse. Attributes include:
  - *part* : specifies whether or not the line is metrically complete. Sample values include:
    - *Y* : the line is metrically incomplete
    - *N* : either the line is complete, or no claim is made as to its completeness
    - *I* : the initial part of an incomplete line
    - *M* : a medial part of an incomplete line
    - *F* : the final part of an incomplete line
  - *met* : contains a user-specified encoding for the metrical structure of the line.
- **<lg>** : contains a group of verse lines functioning as a formal unit e.g. a stanza, refrain, verse paragraph, etc. Attributes include:
  - *part* : specifies whether or not the line is metrically complete. Sample values include:
    - *y* : the line group is metrically incomplete
    - *n* : either the line group is complete, or no claim is made as to its completeness.

- *i* : the initial part of an incomplete line group
- *m* : a medial part of an incomplete line group
- *f* : the final part of an incomplete line group

type : characterizes the line group in some respect, e.g. as a verse paragraph, stanza, verse, etc. Sample values include:

- *stanza* : stanzas of regular number of lines
- *para* : blank verse paragraph
- *free* : arbitrary grouping e.g. in free verse
- *refrain* : a chorus or refrain

rhyme : contains a user-specified encoding for the rhyme scheme of this line group.

<sp> : 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> : contains any kind of stage direction within a dramatic text or fragment. Attributes include:

type : indicates the kind of stage direction. Sample values include:

- *setting* : describes a setting.
- *entrance* : describes an entrance.
- *exit* : describes an exit.
- *business* : describes stage business.
- *novelistic* : is a narrative, motivating stage direction.
- *delivery* : describes how a character speaks.
- *modifier* : gives some detail about a character.
- *location* : describes a location.
- *mixed* : more than one of the above

Full details of other, more specialized, elements for the encoding of texts which are predominantly verse or drama are described in the appropriate chapter of part three (for verse, see the verse base described in chapter 8, “Base Tag Set for Verse,” [in separate fascicle]; for performance texts, see the drama base described in chapter 9, “Base Tag Set for Drama,” [in separate fascicle]). In this section, we describe only the elements listed above, all of which can appear in any text, whichever of the three modes prose, verse, or drama may predominate in it.

6.11.1 Verse

Like other written texts, verse texts or poems may be hierarchically subdivided, for example into books or cantos. These structural subdivisions should be encoded using the general purpose **<div>** or **<div1>** (etc.) elements described below in chapters 7, “Base Tag Set for Prose,” [in separate fascicle] and 8, “Base Tag Set for Verse,” [in separate fascicle]. The fundamental unit of a verse text is the verse line rather than the paragraph, however.

The **<l>** element is used to mark up verse lines, that is metrical rather than typographic lines. Where a metrical line is interrupted by a typographic line break, the encoder may choose to ignore the fact entirely or to use the empty **<lb>** (line break) element discussed in 6.9, “Reference Systems,” on page 52. In the copy text, the following example is printed on four typographic lines, beginning with the words “There”, “From”, “The”, and “the”.

```
<l>There they lie, in the largest, in an
   open space in the woods,
<l>From 500 to 600 poor fellows &mdash; the groans
   and screams &mdash;
<l>The odor of blood, mixed with the fresh scent
   of the night, <lb>the grass, the trees &mdash;
   that Slaughter-house!
```

Where verse lines are not properly nested within the enclosing hierarchy (for example where verse lines cross larger boundaries such as verse paragraphs or speeches) the encoder may choose to use one of the techniques discussed in chapter 36, “Markup for Non-Hierarchical Phenomena,” [in separate fascicle], or to use the *part* attribute to indicate that the verse line is incomplete, as in the following example:

```
<l>Thou fumblest <name>Eros</>, and my Queenes a Squire
<l>More tight at this, then thou:  Dispatch. O Loue,
<l>That thou couldst see my Warres to day, and knew'st
<l>The Royall Occupation, thou should'st see
<l part=i>A Workeman in't.
<stage>Enter an Armed Soldier.</stage>
<l part=f>Good morrow to thee, welcome. ...
```

In some verse forms, regular groupings of lines are regarded as units of some kind, often identified by a regular verse scheme. In stichic verse and couplets, groups of lines analogous to paragraphs are often indicated by indentation. In other verse forms, lines are grouped into irregular sequences indicated simply by white space. The neutral **<lg>** or line group element may be used to mark any such grouping of lines; the *type* is available to further categorize the line group where this is felt desirable, as in the following example. This example also demonstrates the use of the *rhyme* attribute as a means of encoding the rhyme scheme used by a stanza, and the use of the *rend* attribute to indicate whether or not a line is indented.

```
<lg type=stanza rhyme='AABCBC'>
<l>Come fill up the Glass,
<l rend=indent>Round, round let it pass,
<l>'Till our Reason be lost in our Wine:
<l rend=indent>Leave Conscience's Rules
<l rend=indent>To Women and Fools,
<l>This only can make us divine.
</lg>
<lg type=refrain rhyme'ABBA'>
<head>Chorus.</head>
<l>Then a Mohock, a Mohock I'll be,
<l>No Laws shall restrain
<l>Our Libertine Reign,
<l>We'll riot, drink on, and be free.
</lg>
```

For some kinds of analysis, it may be useful to identify different kinds of line group within the same piece of verse. Such line groups may self-nest, in much the same way as the un-numbered **<div>** element described in chapter 7, "Base Tag Set for Prose," [in separate fascicle]. For example:

The *part* attribute may also be attached to a **<lg>** element to indicate that it is incomplete, for example because it forms part of a group that is divided between two speakers, as in the following example:

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

For alternative methods of aligning groups of lines which do not form simple hierarchic groups, or which are discontinuous, see the more detailed discussion in chapter 16, "Segmentation and Alignment," [in separate fascicle]. For discussion of other elements specific to the encoding of verse, see chapter 8, "Base Tag Set for Verse," [in separate fascicle].

These elements are defined as follows:

```
<!-- 6.11.1:  Verse                                          -->
<!ELEMENT l             - o  (%phrase.seq)                    >
<!ATTLIST l                  %a.global;
          part               (Y | N | I | M | F)
                                                 N
          met                CDATA               #IMPLIED     >
<!ELEMENT lg            - o  (head?,(l|lg)+)                  >
<!ATTLIST lg                 %a.global;
          type               CDATA               #IMPLIED
          rhyme              CDATA               '*INHERITED'
          part               (y | n | i | m | f)
                                                 N            >
<!-- This fragment is used in sec. 6                         -->
```

6.11.2 Drama

Like other written texts, dramatic and other *performance texts* such as cinema or TV scripts are often hierarchically organized, for example into acts and scenes. These structural subdivisions should be encoded using the general purpose **<div>** or **<div1>** (etc.) elements described below in chapters 7, "Base Tag Set for Prose," [in separate fascicle] and 9, "Base Tag Set for Drama," [in separate fascicle]. Within these divisions, the body of a performance text typically consists of *speeches*, often prefixed by a phrase indicating who is speaking, and occasionally interspersed with stage directions of various kinds.

In the following simple example, each speech consists of a single paragraph:

```
<div2 type=scene n='I.2'>
<head>Scene 2.</head>
<stage type=setting>Peachum, Filch.</stage>
<sp><speaker>FILCH.</speaker><p>Sir, Black Moll hath sent word her
Trial comes on in the Afternoon, and she hopes you will order Matters
so as to bring her off.
<sp><speaker>PEACHUM.</speaker><p>Why, she may plead her Belly
at worst; to my Knowledge she hath taken care of that Security.
But, as the Wench is very active and industrious, you may satisfy
her that I'll soften the Evidence.
<sp><speaker>FILCH.</speaker><p>Tom Gagg, sir, is found guilty.
```

In the following example, each speech consists of a sequence of verse lines, some of them being marked as metrically incomplete:

```
<div1 name='Act' n='I'><head>ACT I</>
<div2 name='Scene' n='1'><head>SCENE I</head>
<stage rend=italic>
Enter Barnardo and Francisco, two Sentinels, at several doors</stage>
<sp><speaker>Barn</><l part=i>Who's there?
<sp><speaker>Fran</><l>Nay, answer me.  Stand and unfold yourself.
<sp><speaker>Barn</><l part=i>Long live the King!
<sp><speaker>Fran</><l part=m>Barnardo?
<sp><speaker>Barn</><l part=f>He.
<sp><speaker>Fran</><l>You come most carefully upon your hour.
<sp><speaker>Barn</><l>'Tis now struck twelve.  Get thee to bed,
 Francisco.
<sp><speaker>Fran</><l>For this relief much thanks.'Tis bitter cold,
<l part=i>And I am sick at heart.
```

In some cases, as here in the First Quarto of *Hamlet*, the printed speaker attributions need to be supplemented by use of the *who* attribute; again, the lines are marked as complete or incomplete:

```
<stage>Enter two Centinels.
<add place=right resp=ms>
Now call'd Bernardo & Francesco.
</add></stage>
<sp who='Francisco'><speaker>1.</>
    <l part=y>STand:  who is that?</sp>
<sp who='Barnardo'><speaker>2.</>
    <l part=y>Tis I.</sp>
<sp who='Francisco'><speaker>1.</>
    <l>O you come most carefully vpon your
       watch,</sp>
<sp who='Barnardo'><speaker>2.</>
    <l>And if you meete Marcellus and Horatio,
    <l>The partners of my watch, bid them make haste.</sp>
<sp who='Francisco'><speaker>1.</>
    <l part=y>I will:  See who goes there.</sp>
<stage>Enter Horatio and Marcellus.</>
```

```
<sp who='Horatio'><speaker>Hor.</>
    <l part=i>Friends to this ground.</sp>
<sp who='Marcellus'><speaker>Mar.</>
    <l part=f>And leegemen to the Dane,
    <l>O farewell honest souldier, who hath
    releeued you?</sp>
<sp who='Francisco'><speaker>1.</>
    <l>Barnardo hath my place, giue you good night.</sp>
```

By contrast with the preceding examples, the following encodes an early printed edition without making any assumption about which parts are prose or verse:

```
<div1 type=act n='I'>
<div2 type=scene n='1'>
<head rend=italic>Actus primus, Scena prima.</head>
<stage type=setting rend=italic>
A tempestuous noise of Thunder and Lightning heard:  Enter
a Ship-master, and a Boteswaine.</stage>
<sp><speaker>Master.</speaker> Bote-swaine.
<sp><speaker>Botes.</speaker> Heere Master: What cheere?
<sp><speaker>Mast.</speaker> Good: Speake to th' Mariners: fall
too't, yarely, or we run our selues #a ground,
bestirre, bestirre.  <stage type=move>Exit.</stage>
<stage type=move>Enter Mariners.</stage>
<sp><speaker>Botes.</speaker>
Heigh my hearts, cheerely, cheerely my harts:
yare, yare: Take in the toppe-sale: Tend to th' Masters
whistle: Blow till thou burst thy winde, if roome e-nough.
```

The **<sp>** and **<stage>** elements should also be used to mark parts of a text otherwise in prose which are presented as if they were dialogue in a play. For example:

```
<sp><speaker>The reverend Doctor Opimiam</speaker>
<p>I do not think I have named a single unpresentable
fish. </sp>
<sp><speaker>Mr Gryll</speaker><p>Bream, Doctor: there
is not much to be said for bream.</sp>
<sp><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>Fish, Miss Gryll -- I could discourse to you on
fish by the hour:  but for the present I will
forbear...
</sp>

<sp><speaker>Lord Curryfin</speaker>
<stage>(after a pause).</stage>
<p><q>Mass</q> as the second grave-digger
says in <title>Hamlet</title>,
<q>I cannot tell.</q>
```

```
</sp>
<p>A chorus of laughter dissolved the sitting.
```

These elements are defined as follows:

```
<!-- 6.11.2:  Drama                                             -->
<!ELEMENT sp           - -  (speaker?, (%specialPara) )         >
<!ATTLIST sp                %a.global;
          who               IDREFS              #IMPLIED        >
<!ELEMENT speaker      - -  (%phrase.seq)        -(speaker)     >
<!ATTLIST speaker           %a.global;                          >
<!ELEMENT stage        - -  (%para.seq)          -(stage)       >
<!ATTLIST stage             %a.global;
          type              CDATA               mix             >
<!-- This fragment is used in sec. 6                            -->
```

delete part on ambiguous punctuation-->

6.12 Arbitrary Segments

For many purposes, one would like to be able to segment text systematically into pieces smaller than the paragraph. The element **<s>** (segment) is provided for this purpose. Segments of any size or type may be marked using this tag.

<s> : contains any arbitrary phrase-level unit of text. Attributes include:
type : indicates what sort of segment is marked.

The global *id* and *n* attributes may be used to associate standard reference identifiers with the segment for purposes of citation or retrieval.

The *type* attribute may be used to categorize the segment, for example as an *orthographic sentence*, or as a *phrase* of some specific type. For some purposes, it may be desired to characterize segments as *orthographic words*, or even explicitly to mark sequences of *punctuation*.

To illustrate this flexibility we first consider the simplest common requirement: the ability to mark orthographic sentences. These we term *S-units* to prevent confusion with the linguistic notion of “sentence”, which is commonly taken to depend implicitly or explicitly on a formal grammatical analysis. S-units are signalled by initial capitalization and particular punctuation marks (e.g. full stop, exclamation mark, question mark), for example:

```
<s>Literate and illiterate speech in a language like English
   are plainly different.</s>
<s>This in fact is the layman's chief interest in linguistics.</s>
```

These S-units are both full sentences, but consideration of any real life text will rapidly demonstrate the need to mark as S-units phrases or even single words. For example:

```
<s>I pause to record that I feel in extraordinary form.</s>
<s>Delirium perhaps.</s>
```

S-units are commonly used in this way to demarcate manageable units of text, for example as input to an automatic word-class tagger. The global *id* attribute may be used to associate an identifier with each such segment. This identifier may then be used to link the segment with syntactic or other kinds of analysis.

```
<s id=BL0034>Literate and illiterate speech in a
   language like English are plainly different.</s>
<s id=BL0035>This in fact is the layman's chief
   interest in linguistics.</s>
```

The **<s>** element may also be used to mark constituent elements of any kind, such as are typically output by an automatic parsing system:

```
<s id=BL0034>
    <s type=noun-phrase> Literate and illiterate speech </s>
    <s type=prep-phrase> in a language like English </s>
    <s type=complement> are plainly different.     </s>
</s>
```

As indicated above, the **<s>** element may nest within itself, and thus provides a simple but powerful mechanism for representing segmentation down to an arbitrary level of complexity, provided that the structure can be represented by a single tree. Mechanisms for representing segmentations of a text which do not resolve to a single tree, or in which constituent features are discontinuous, are discussed in section 16, “Segmentation and Alignment,” [in separate fascicle].

As a final example, we present the same two sentences segmented down to the level of words and punctuation. Each word or punctuation sequence is given a unique identifier.[14]

```
<p>
<s type=orth-sent id=s1>
   <s type=phrase rendition='small caps' id=phr1>
      <s type=orth-word id=s1.1>    Literate     </s>
      <s type=orth-word id=s1.2>    and          </s>
      <s type=orth-word id=s1.3>    illiterate   </s>
      <s type=orth-word id=s1.4>    speech       </s>
   </s>
   <s type=orth-word id=s1.5>       in           </s>
   <s type=orth-word id=s1.6>       a            </s>
   <s type=orth-word id=s1.7>       language     </s>
   <s type=orth-word id=s1.8>       like         </s>
   <s type=orth-word id=s1.9>       English      </s>
   <s type=orth-word id=s1.10>      are          </s>
```
&@fn@sep

14 White space and self-explanatory code names have been liberally introduced in this example, in order to aid legibility: a real life example would, of course, be considerably more terse.

```
   <s type=orth-word id=s1.11>      plainly      </s>
   <s type=orth-word id=s1.12>      different    </s>
   <s type=punct     id=punc1>      .            </s>
</s>

<s type=orth-sent id=s2>
   <s type=orth-word id=s2.1>      This         </s>
   <s type=punct     id=punc16>    ,            </s>
   <s type=orth-word id=s2.2>      in           </s>
   <s type=orth-word id=s2.3>      fact         </s>
   <s type=punct     id=punc17>    ,            </s>
   <s type=orth-word id=s2.4>      is           </s>
   <s type=orth-word id=s2.5>      the          </s>
   <s type=orth-word id=s2.6>      layman
       <s type=punct id=punc18>    '        </s>
                                   s            </s>
   <s type=orth-word id=s2.7>      chief        </s>
   <s type=orth-word id=s2.8>      interest     </s>
   <s type=orth-word id=s2.9>      in           </s>
   <s type=orth-word id=s2.10>     linguistics  </s>
   <s type=punct     id=punc19>    .            </s>
   </s>
```

In the first sentence, the phrase “literate and illiterate speech”, identified by its distinctive rendition, has been marked as a segment of a different type from the remainder. In the second sentence, the “orthographic word” with identifier s2.6 (“layman's”) contains a nested **<s>** element of type “punct”. Each s-unit has an unique identifier which may be used to associate it with a more detailed linguistic analysis, as further discussed in sections 16, “Segmentation and Alignment,” [in separate fascicle] and 18, “Feature Structure Analysis,” [in separate fascicle].

The formal declaration for the **<s>** element is as follows:

```
<!-- 6.12:  Arbitrary Segments                                 -->
<!ELEMENT s             - -  (%phrase.seq)                       >
<!ATTLIST s                  %a.global;
          type               CDATA               #IMPLIED        >
<!-- This fragment is used in sec. 6                           -->
```

Part 7

Alphabetic Reference List of Tags

<abbr> (i.e. abbreviation)

Description: contains an abbreviation of any sort.
Attributes:

expan (i.e. expansion): gives an expansion of the abbreviation.
Data type: CDATA
Value: any string of characters
Default value: #IMPLIED
Example:

```
The address is Southmoor <abbr expan='road'>Rd</>.
```

Remarks: Only one expansion may be given for an abbreviation; if different expansions are to be proposed, the tags for critical apparatus should be used.
This attribute is optional.

resp (i.e. responsibility): identifies the individual responsible for the interpretation given the abbreviation.
Data type: CDATA
Value: any string of characters, usually the initials of the individual, if known, or a descriptive phrase (e.g. “16c hand”) if not.
Default value: #IMPLIED
Example:

```
The address is Southmoor <abbr resp=LB>Rd</>.
```

Remarks: If no expansion is given, the *resp* attribute has no meaning.
This attribute is optional.

type: allows the encoder to classify the abbreviation according to some convenient typology.
Data type: CDATA
Sample values include:
- suspension : the abbreviation provides the first letter(s) of the word or phrase, omitting the remainder.
- contraction : the abbreviation omits some letter(s) in the middle.
- brevigraph : the abbreviation comprises a special symbol or mark.
- superscription : the abbreviation includes writing above the line.
- acronym : the abbreviation comprises the initial letters of the words of a phrase.

title : the abbreviation is for a title of address (Dr, Ms, Mr, ...)
organization : the abbreviation is for the name of an organization.
geographic : the abbreviation is for a geographic name.

Default value: #IMPLIED
Example:

```
<abbrev type=acronym>RSVP</abbrev>
<abbrev type=organization>SPQR</abbrev>
<abbrev type=brevigraph>&per;</abbrev>
<abbrev type=contraction>yr hbl servt</abbrev>
```

Remarks: The *type* attribute is provided for the sake of those who wish to classify abbreviations at their point of occurrence; this may be useful in some circumstances, though usually the same abbreviation will have the same type in all occurrences. As the sample values make clear, abbreviations may be classified by the method used to construct them, the method of writing them, or the referent of the term abbreviated; the typology used is up to the encoder and should be carefully planned to meet the needs of the expected use.

This attribute is optional.

Example:

```
<abbrev type=organization
  expan='senatus populusque romanorum'
  resp=author>SPQR</abbrev>
```

Remarks: This tag is the mirror image of the **<expan>** tag; both allow the encoder to transcribe both an abbreviation and its expansion. In **<abbr>**, however, the original is transcribed as the content of the element and the expansion as an attribute value; **<expan>** reverses this. The choice between the two is up to the user.

The **<abbr>** tag is not required; if appropriate, the encoder may transcribe abbreviations in the source text silently, without tagging them. If abbreviations are not transcribed directly but *expanded* silently, then the TEI header should so indicate.

For a typology of ME abbreviations, see A. G. Petty, *English literary hands from Chaucer to Dryden* (London: Edward Arnold, 1977), pp. 22-25.

Part: additional tag set for common core features.
Member of classes: data
DTD file:
Data description: May contain character data and phrase-level elements.
May occur within:
May contain: %phrase.seq;
Element Declaration:

```
<!ELEMENT abbr          - -  (%phrase.seq;)                    >
```

Attribute List Declaration:

```
<!ATTLIST abbr               %a.global;
          expan              CDATA              #IMPLIED
          resp               CDATA              #IMPLIED
          type               CDATA              #IMPLIED       >
```

Discussed in 6.4.4, “Abbreviations and Their Expansions,” on page 30.
This tag is optional.

<add> (i.e. addition)

Description: contains letters, words, or phrases inserted in the text by a scribe, annotator, corrector, editor, or encoder.
Attributes:

place: if the the addition is written into the copy text, indicates where the additional text is written.
Data type: CDATA
Suggested values include:
inline : addition is made in a space left in the witness by an earlier scribe
supralinear : addition is made above the line
infralinear : addition is made below the line
left : addition is made in left margin
right : addition is made in right margin
top : addition is made in top margin
bottom : addition is made in bottom margin
opposite : addition is made on opposite page
verso : addition is made on verso of sheet
mixed : addition is made somewhere, one or more of other values
Default value: #IMPLIED
Example:

Remarks: For additions made by editors or encoders, no location need be specified.
This attribute is recommended.

resp (i.e. responsible): identifies the individual responsible for the addition.
Data type: CDATA
Value: Any string of characters, e.g. “hand 1”, “hand 2”, “author”, etc., or the initials of individuals known by name.
Default value: #IMPLIED
Example:

This attribute is optional.

Example:

Remarks:
Part: additional tag set for common core features.
Member of classes: edit
DTD file:
Data description: May contain character data and phrase-level elements.
May occur within:
May contain: %specialPara;
Element Declaration:

```
<!ELEMENT add           - -  (%specialPara;)                   >
```

Attribute List Declaration:

```
<!ATTLIST add                %a.global;
          place              CDATA              #IMPLIED
          resp               CDATA              #IMPLIED       >
```

Discussed in 6.5.3, “Additions, Deletions and Omissions,” on page 37.
This tag is optional.

<address>

Description: contains a postal or other address, for example of a publisher, an organization, or an individual
Attributes: [None: global and inherited attributes only.]
Example:

```
<address><street>via Marsala 24</>
        <postcode>40126</>
        <city>Bologna</>
        <country n=I>Italy</>
</address>

<address>
        <postbox>P.O. Box 6998</>
        <city>Chicago</>
        <region n=IL>Illinois</>
        <postcode>60680</>
        <country>USA</>
</address>
```

Remarks: This is a general purpose crystal: it may be given only using the defined substructural elements. To tag addresses in running text without this substructure, use the **<place>** element.

Part: base tag set for common core features.
Member of classes: data
DTD file:
Data description: If given as running prose, use a consistent format wherever possible, for example separating lines of the address by commas, and including any postal code in the standard form.
May occur within: indiv organization
May contain: city country org postbox postcode region street
Element Declaration:

```
<!ELEMENT address       - o  (org? & street? & city? & region? &
                              country? & postcode? & postbox?)  >
```

Attribute List Declaration:

```
<!ATTLIST address            %a.global;                         >
```

Discussed in 5.2.4, “Publication, Distribution, etc.,” [in separate fascicle].
Discussed in 6.10, “Bibliographic Citations and References,” on page 63.
This tag is optional.

Class Name: **agent** (i.e. agent (individual or corporate body))
Description: defines a group of elements which contain names of individuals or corporate bodies.
Attributes: [None: global and inherited attributes only.]
Remarks: This class is used in the **<resp>** element, to allow a statement of responsibility to apply to an individual or a body.
Part: base tag set for common core features.
Member of classes:
DTD file:
Discussed in 6.4, “Names, Numbers, Dates, Abbreviations, and Addresses,” on page 24.

<analytic> (i.e. analytic level)

Description: contains bibliographic elements describing an item (e.g. an article or poem) published within a monograph or journal and not as an independent publication.
Attributes: [None: global and inherited attributes only.]
Example:

Remarks: The **<analytic>** element may occur only within bibliographic citation or reference elements; it is mandatory for description of the analytic level of **<bibl.struct>** elements.
Part: base tag set for common core features.
Member of classes: biblpart
DTD file:
Data description: May contain titles and statements of responsibility (author, editor, or other), in any order.
May occur within: bibl.struct
May contain: author editor resp title
Element Declaration:

```
<!ELEMENT analytic      - O  (author | editor | resp | title)*  >
```

Attribute List Declaration:

```
<!ATTLIST analytic           %a.global;                         >
```

Discussed in 6.10.2.1, “Analytic, Monographic, and Series Levels,” on page 67.
This tag is mandatory when applicable.

<author>

Description: primary *statement of responsibility* for a bibliographic item, for example the name of a single author, institution or organization, or of several such.
Attributes: [None: global and inherited attributes only.]
Example:

```
<author>British Broadcasting Corporation
<author>La Fayette, Marie Madeleine Pioche de la Vergne,
        comtesse de (1634-1693)
```

Remarks: Particularly where cataloguing is likely to be based on the content of the header, it is advisable to use generally recognized authority lists for the exact form of personal names.

In the case of a broadcast, use this element for the name of the company or network which broadcasts the program.
Part: base tag set for common core features.
Member of classes: biblpart
DTD file:
Data description: Any string of characters and phrase-level tags.
May occur within: analytic monogr
May contain: %phrase.seq
Element Declaration:

```
<!ELEMENT author        - o  (%phrase.seq)                    >
```

Attribute List Declaration:

```
<!ATTLIST author             %a.global;                       >
```

Discussed in 5.2.1, “The Title Statement,” [in separate fascicle].
Discussed in 6.10, “Bibliographic Citations and References,” on page 63.
This tag is optional.

<bibl> (i.e. bibliographic citation)

Description: contains a loosely-structured bibliographic citation of which the sub-components may or may not be explicitly tagged.
Attributes: [None: global and inherited attributes only.]
Example:

```
<bibl>Blain, Clements and Grundy: Feminist Companion to Literature in
English (Yale, 1990)
</bibl>

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

Remarks:
Part: base tag set for common core features.
Member of classes: clbibl
DTD file:
Data description: Contains either just prose, or any combination of elements from the *citation* class
May occur within: list.bibl
May contain: %m.bibl.part; %m.phrase; #PCDATA
Element Declaration:

```
<!ELEMENT bibl          - o  (#PCDATA | %m.phrase; |
                             %m.bibl.part;)*                  >
```

Attribute List Declaration:

```
<!ATTLIST bibl              %a.global;                         >
```

Discussed in 6.10, “Bibliographic Citations and References,” on page 63.
This tag is optional.

Class Name: **bibl**
Description: bibliographic elements.
Attributes: [None: global and inherited attributes only.]
Remarks:
Part: base tag set for common core features.
Member of classes: inter
DTD file:
Discussed in 6.10, “Bibliographic Citations and References,” on page 63.

Class Name: **bibl.part** (i.e. bibliographic citation part)
Description: elements which can appear within bibliographic citations, but cannot appear freely in running text outside bibliographic citations.
Attributes: [None: global and inherited attributes only.]
Remarks: This class is used in defining the content model of **<bibl>**.
Part: base tag set for common core features.
Member of classes:
DTD file:
Discussed in 6.10, “Bibliographic Citations and References,” on page 63.

<bibl.full>

Description: contains a fully-structured bibliographic citation, in which all components of the TEI file description save the source description are present.
Attributes: [None: global and inherited attributes only.]
Example:

```
<bibl.full>
        <title.stmt>
           <author>Blain, Virginia
           <author>Clements, Patricia
           <author>Grundy, Isobel
           <title>The Feminist Companion to Literature in English:
               women writers from the middle ages to the present
        </title.stmt>
        <edition.stmt>
           <edition>UK edition</edition>
```

```
            <extent>1231 pp</extent>
            <publication.stmt>
              <creation.date>1990
              <publisher>Yale University Press
              <place>New Haven and London
              <date>1990
            </publication.stmt>
            <source.desc>No source: this is an original work</source.desc>
    </bibl.full>
```

Remarks:
Part: base tag set for common core features.
Member of classes: clbibl
DTD file:
Data description:
May occur within: list.bibl
May contain: edition.stmt extent notes.stmt publication.stmt series.stmt source.desc title.stmt
Element Declaration:

```
<!ELEMENT bibl.full     - o  (title.stmt, edition.stmt?,
                             extent?, publication.stmt,
                             series.stmt?, notes.stmt?,
                             source.desc*)                 >
```

Attribute List Declaration:

```
<!ATTLIST bibl.full          %a.global;                    >
```

Discussed in 6.10, “Bibliographic Citations and References,” on page 63.
Discussed in 5.2.7, “The Source Description,” [in separate fascicle].
This tag is optional.

<biblScope> (i.e. scope of citation)

Description: defines the scope of a bibliographic reference, for example as a list of pagenumbers, or a named subdivision of a larger work.
Attributes:

type: identifies the type of information conveyed by the element, e.g. “pages”, “volume”.
Data type: CDATA
Suggested values include:
volume : the element contains a volume number.
issue : the element contains an issue number, or volume and issue numbers.
pages : the element contains a page number or page range.

chapter : the element contains a chapter indication (number and/or title)
part : the element identifies a part of a book or collection.
Default value: #IMPLIED
Example:

Remarks:
This attribute is optional.

Example:

```
<biblScope>pp 12-34

<biblScope>Volume 2 "My apprenticeship"
```

Remarks:
Part: base tag set for common core features.
Member of classes: biblpart
DTD file:
May occur within: imprint monogr series
May contain: %phrase.seq
Element Declaration:

```
<!ELEMENT biblScope     - o  (%phrase.seq)                    >
```

Attribute List Declaration:

```
<!ATTLIST biblScope          %a.global;
          type               CDATA              #IMPLIED       >
```

Discussed in 6.10, “Bibliographic Citations and References,” on page 63.
This tag is optional.

<bibl.struct> (i.e. structured bibliographic citation)

Description: contains a structured bibliographic citation, in which only bibliographic subelements appear and in a specified order.
Attributes: [None: global and inherited attributes only.]
Example:

```
<bibl.struct>
    <monogr>
         <author>Blain, Virginia
         <author>Clements, Patricia
         <author>Grundy, Isobel
         <title>The Feminist Companion to Literature in
                English: women writers from the middle ages
```

```
                              to the present</>
                       <imprint><publisher>Yale University Press</>
                                <city>New Haven and London</>
                                <date>1990</>
              </monogr>
              </bibl.struct>
```

Remarks:
Part: base tag set for common core features.
Member of classes: clbibl
DTD file:
Data description:
May occur within: list.bibl
May contain: analytic idno monogr note series
Element Declaration:

```
<!ELEMENT bibl.struct   - o  (analytic?, (monogr, series*)+,
                             (note | idno)*)                    >
```

Attribute List Declaration:

```
<!ATTLIST bibl.struct        %a.global;                         >
```

Discussed in 6.10, “Bibliographic Citations and References,” on page 63.
This tag is optional.

Class Name: **chunk**

Description: includes all elements which can occur between, but not within, paragraphs and other chunks.
Attributes: [None: global and inherited attributes only.]
Remarks: This element class does not contain all those elements which can appear between chunks: the class *inter* contains a set of elements which can appear either within or between chunks. Unlike elements of that class, chunks cannot occur within chunks. In prose, this means the elements in this class can appear between but not within paragraphs.
Part: base tag set for common core features.
Member of classes:
DTD file: teicore2
Discussed in 3.4, “Element Classes and Other Parameter Entities,” [in separate fascicle].

Entity Name: **chunk.seq** (i.e. chunk sequence)

Description: defines a sequence of *chunks*, i.e. of paragraph-level elements of the class *chunk* or of the class *inter*.
Remarks:
Expansion: '(%m.inter | %m.chunk;)*' 3.4, “Element Classes and Other Parameter Entities,” [in separate fascicle]

<city> (i.e. city)

Description: contains the name of a town, city or village occurring within an address or imprint.
Attributes: [None: global and inherited attributes only.]
Example:

```
<city>London
<city>Truth or Consequences
```

Remarks:
Part: base tag set for common core features.
Member of classes: biblpart
DTD file:
May occur within: address
May contain: %phrase.seq
Element Declaration:

```
<!ELEMENT city          - o  (%phrase.seq)                     >
```

Attribute List Declaration:

```
<!ATTLIST city              %a.global;                         >
```

Discussed in 6.10, “Bibliographic Citations and References,” on page 63.
This tag is optional.

<col> (i.e. column)

Description: contains one column of a multi-column reference edition.
Attributes: [None: global and inherited attributes only.]
Example:

```
<(La)page n=32>
<(La)col n=a> ...
<(La)col n=b> ...
<(La)page n=33> ...
```

Remarks: The **<col>** tag should be used wherever a standard reference scheme uses references to individual columns of a reference edition; otherwise it need not be used.
Part: base tag set for physical description of the copy text.
Member of classes:
DTD file:
Data description: May contain **<line>** elements, or character data.
May occur within: page
May contain: #PCDATA line
Element Declaration:

```
<!ELEMENT col          - O  (#PCDATA | line)*                 >
```

Attribute List Declaration:

```
<!ATTLIST col              %a.global;                         >
```

Discussed in 6.9.3, “Concurrent Markup for Pages and Lines,” on page 56.
This tag is optional.

<corr> (i.e. correction)

Description: contains the correct form of a passage apparently erroneous in the copy text.
Attributes:

sic: gives the original form of the apparent error in the copy text.
Data type: CDATA
Value: any string of characters
Default value: #IMPLIED
Example:

```
for his nose was as sharp as a pen, and
<corr sic="a Table"> a' babbled</corr>
of green fields.
```

This attribute is optional.

resp (i.e. responsibility): identifies the individual responsible for the correct form given.
Data type: CDATA
Value: Sample values include “ed”, “scribe”, or the initials of the individual in question.
Default value: #IMPLIED
Example:

This attribute is optional.

Example:

If all that is desired is to call attention to the fact that the copy text has been corrected, no attributes are required:

```
I don't know, Juan.  It's so far in the past now &mdash;
how <corr>can we</corr> prove or disprove anyone's theories?
```

It is also possible to provide a correct reading and to identify the individual responsible for the correction:

```
I don't know, Juan.  It's so far in the past now &mdash;
how <corr sic='we can' resp='MSM'>can we</corr> prove or
disprove anyone's theories?
```

Remarks: The **<corr>** tag is a mirror of **<sic>**: the latter leaves the original text untouched, giving the correction as an attribute value; the former substitutes the correction, leaving the original reading as an attribute value. The choice between them is up to the encoder.
Part: additional tag set for common core features.
Member of classes: edit
DTD file:
Data description: May contain character data and phrase-level elements.
May occur within:
May contain: %specialPara;
Element Declaration:

```
<!ELEMENT corr          - -  (%specialPara;)                   >
```

Attribute List Declaration:

```
<!ATTLIST corr               %a.global;
          sic                CDATA               #IMPLIED
          resp               CDATA               #IMPLIED       >
```

Discussed in 6.5.1, “Correction of Apparent Errors,” on page 34.
This tag is optional.

<country> (i.e. country)

Description: contains the largest geographic, administrative or political unit identified within an address.
Attributes: [None: global and inherited attributes only.]
Example:

```
<country>Bosnia-Herzegovina

<country n=DK>Denmark
```

Remarks: Values for identifying attributes should be taken from recognised lists of abbreviations, e.g. ISO 3166.
Part: base tag set for common core features.
Member of classes:
DTD file:
May occur within: address
May contain: %soup
Element Declaration:

```
<!ELEMENT country       - o  (%soup)                          >
```

Attribute List Declaration:

```
<!ATTLIST country            %a.global;                       >
```

Discussed in 6.10, “Bibliographic Citations and References,” on page 63.
This tag is optional.

Class Name: **data**
Description: phrase-level elements containing names, dates, numbers, measures, and similar data.
Attributes: [None: global and inherited attributes only.]
Remarks:
Part: base tag set for common core features.
Member of classes: phrase
DTD file: teicore2
Discussed in 6.4, “Names, Numbers, Dates, Abbreviations, and Addresses,” on page 24.

<date>

Description: contains a date in any format.
Attributes:

calendar: indicates the system or calendar to which the date belongs.
Data type: CDATA
Value: Recommended values include: *Gregorian*, *Julian*, *Roman*, *Mosaic*, *Revolutionary*, *Islamic*.
Default value: #IMPLIED
Example:

```
He was born on <date calendar=Gregorian>Feb. 22, 1732</date>
(<date calendar=Julian value='1732-2-22'>Feb. 11, 1731/32,
O.S.</date>).
```

This attribute is optional.

value: gives the value of the date in some standard form, usually yyyy-mm-dd.
Data type: CDATA
Value: Any string representing a date in standard format; recommended form is "yyyy-mm-dd", as defined by ISO 8601: 1988, *Data elements and interchange formats — Information interchange — Representation of dates and times.*
Default value: #IMPLIED
Example:

```
This list begins in the year 1632, more precisely
on Trinity Sunday, i.e. the Sunday after Pentecost,
in that year the <date calendar=Julian value=1632-6-6>
27th of May (old style)</date>.
```

Remarks: For simple dates, the value should give the Gregorian date in the form (yyyy-mm-dd) specified by ISO 2014. More complicated dates or special applications may require another calendar or another form; these should be documented in the **<std.vals>** element in the TEI Header.
This attribute is optional.

certainty: indicates the degree of certainty attributed to the date.
Data type: CDATA
Value: Any appropriate value, e.g. *ca.*, *approx*, *after*, *before.*
Default value: #IMPLIED
This attribute is optional.

Example:

```
<date value='1980-02-21'>21 Feb 1980</date>

Given on the <date value='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>

<date value='1990-09'>September 1990</date>
```

Remarks:
Part: additional tag set for common core features.
Member of classes: data date
DTD file:
Data description: May contain character data and phrase-level elements.
May occur within:
May contain: %phrase.seq;
Element Declaration:

```
<!ELEMENT date          - -  (%phrase.seq;)                    >
```

Attribute List Declaration:

```
<!ATTLIST date               %a.global;
          calendar           CDATA              #IMPLIED
          value              CDATA              #IMPLIED
          certainty          CDATA              #IMPLIED       >
```

Discussed in 6.4.3, “Dates and times,” on page 28.
This tag is optional.

Class Name: **date** (i.e. dates and date ranges)
Description: defines a class of elements containing date specifications.
Attributes: [None: global and inherited attributes only.]
Remarks: This class allows certain content models to allow either a single date or a date-range element.
Part: base tag set for common core features.
Member of classes:
DTD file:
Discussed in 6.4, “Names, Numbers, Dates, Abbreviations, and Addresses,” on page 24.

<dateRange> (i.e. date range)

Description: contains two dates or another phrase delimiting a time period.
Attributes:

calendar: indicates the system or calendar to which the date belongs.
Data type: CDATA
Value: Recommended values include: *Gregorian*, *Julian*, *Roman*, *Mosaic*, *Revolutionary*, *Islamic*.
Default value: #IMPLIED
This attribute is optional.

from: indicates the starting point of the period in standard form.
Data type: CDATA
Value: any date in a standard form; recommended form is yyyy-mm-dd.
Default value: #IMPLIED
Remarks: The value should conform to the standard form declared in the **<std.vals>** element in the TEI header.
This attribute is optional.

to: indicates the ending point of the period in standard form.
Data type: CDATA

Value: any date in a standard form; recommended form is yyyy-mm-dd.
Default value: #IMPLIED
Remarks: The value should conform to the standard form declared in the **<stdVals>** element in the TEI header.
This attribute is optional.

exact: indicates the precision to be attached to either or both dates specified.
Data type: (to|from|both|none)

to : the *to* date is exact
from : the *from* date is exact
both : both dates are exact
none : both dates are approximate or unspecified

Default value: #IMPLIED
Example:

```
<dateRange from=1100 to=1120 exact=from>
     Early 12th century</dateRange>
```

Remarks: date values should conform to the standard form declared in the **<stdVals>** element in the TEI header.
This attribute is optional.

Example:

```
He edited (<dateRange from=1846 to=1848>1846-48</dateRange>)
the first collection of Hungarian folk poetry.
```

Part: additional tag set for common core features.
Member of classes: data date
DTD file:
Data description: May contain character data and phrase-level elements.
May occur within:
May contain: %phrase.seq;
Element Declaration:

```
<!ELEMENT dateRange     - O  (%phrase.seq;)                    >
```

Attribute List Declaration:

```
<!ATTLIST dateRange          %a.global;
          calendar           CDATA               #IMPLIED
          from               CDATA               #IMPLIED
          to                 CDATA               #IMPLIED
          exact              (to|from|both|none)
                                                 #IMPLIED       >
```

Discussed in 6.4.3, “Dates and times,” on page 28.
This tag is optional.

<del> (i.e. deletion)

Description: contains a word or passage deleted, marked as deleted, or identified as superfluous or spurious in the copy text.

Attributes:

rend (i.e. rendition): indicates how the deletion was indicated in the copy text.
Data type: CDATA
Sample values include:
subpunction : dots below the line indicate matter to be deleted.
overstrike : lines through the text indicated matter to be deleted.
erasure : material to be deleted has been erased (but remains legible enough to transcribe).
bracketed : brackets around the material indicate that it is spurious or superfluous.
Default value: #IMPLIED
Example:

This attribute is optional.

type: classifies the type of deletion using any convenient typology.
Data type: CDATA
Value: any string identifying the class of deletion.
Default value: #IMPLIED
Example:

Remarks: No recommendation of any particular typology is made here; to record the manner in which the deletion is signaled, use *rend*, not *type*.
This attribute is optional.

status: may be used to indicate faulty deletions, e.g. strikeouts which include too much or too little text.
Data type: CDATA
Value: any description of flaws in the marking of a deletion, e.g. “excess left”, “excess right”, “short left”, “short right”.
Default value: 'unremarkable'
Example:

Remarks: Status information on each deletion is needed rather rarely except in critical editions from authorial manuscripts.
This attribute is optional.

resp (i.e. responsible): identifies the individual responsible for making the deletion.
Data type: CDATA

Value: Any string of characters, e.g. “hand 1”, “hand 2”, etc., or the initials of individuals known by name.
Default value: #IMPLIED
Example:

This attribute is optional.

Example:

```
<l><del type=overtyped>Mein</del> Frisch
   <del type=overstrike>schwebt</del> weht der Wind
```

Remarks: Cf. **<omit>**.

Degrees of uncertainty over what (if anything) can still be read may be indicated by use of the **<certainty>** element (see 19, “Certainty,” [in separate fascicle]).

This element should be used for deletion of shorter sequences of text, typically single words or phrases. The **<delSpan>** element should be used for longer sequences of text, for those containing structural subdivisions, and for those containing overlapping additions and deletions.

Part: additional tag set for common core features.
Member of classes: edit
DTD file:
Data description: May contain character data and phrase-level elements.
May occur within:
May contain: %phrase.seq;
Element Declaration:

```
<!ELEMENT del           - -  (%phrase.seq;)                    >
```

Attribute List Declaration:

```
<!ATTLIST del                %a.global;
          rend               CDATA               #IMPLIED
          type               CDATA               #IMPLIED
          status             CDATA               'unremarkable'

          resp               CDATA               #IMPLIED       >
```

Discussed in 6.5.3, “Additions, Deletions and Omissions,” on page 37.
Discussed in 21, “Text Criticism and Apparatus,” [in separate fascicle].
This tag is optional.

<distinct> (i.e. distinct)

Description: identifies any word or phrase which is regarded as linguistically distinct, for example as archaic, technical, dialectal, non-preferred, etc., or as forming part of a sublanguage.

Attributes:

type: specifies the sublanguage or register to which the word or phrase is being assigned
Data type: CDATA
Value: a semi-open user-defined list
Default value: #IMPLIED

time: specifies how the phrase is distinct diachronically
Data type: CDATA
Value: a semi-open user-defined list
Default value: #IMPLIED

space: specifies how the phrase is distinct diatopically
Data type: CDATA
Value: a semi-open user-defined list
Default value: #IMPLIED

social: specifies how the phrase is distinct diastatically
Data type: CDATA
Value: a semi-open user-defined list
Default value: #IMPLIED

Example:

```
Next morning a boy in that dormitory confided to his
bosom friend, a <distinct type=ps_slang>fag</distinct> of
Macrea's, that there was trouble in their midst which
King <distinct type=archaic>would fain</distinct> keep
secret.

     <!-- need another example -->
```

Remarks:
Part: base tag set for common core features.
Member of classes: hqphrase
DTD file:
Data description:
May occur within:
May contain: %phrase.seq;
Element Declaration:

```
<!ELEMENT distinct      - -  (%phrase.seq;)                    >
```

Attribute List Declaration:

```
<!ATTLIST distinct          %a.global;
          type              CDATA              #IMPLIED
          time              CDATA              #IMPLIED
          space             CDATA              #IMPLIED
          social            CDATA              #IMPLIED       >
```

Discussed in 6.3.2.3, “Other Linguistically Distinct Material,” on page 15.
This tag is optional.

<div.gen> (i.e. text division generated automatically)

Description: indicates the location at which an **<div>** generated automatically by a text-processing application is to appear.

Attributes:

type: specifies what type of generated text division (e.g. index, table of contents, etc.) is to appear.

Data type: CDATA

Sample values include:

index : an index is to be generated and inserted at this point.

toc : a table of contents

figlist : a list of figures

tablist : a list of tables

Default value: #IMPLIED

Example:

Remarks: Valid values are application-dependent; those shown are of obvious utility in document production, but are by no means exhaustive.

This attribute is optional.

Example:

One use for this element is to allow document preparation software to generate an index and insert it in the appropriate place in the output. The example below assumes that the *index* attribute on the **<index>** element has been used to specify index entries for multiple indices, 1 and 2.

```
<back>
<div1><head>Bibliography</head>
<list.bibl> ... </list.bibl>
<divGen type='index 1' n='Index Nominum'>
<divGen type='index 2' n='Index Rerum'>
</back>
```

Another use for **<div.gen>** is to specify the location of an automatically produced table of contents:

```
<front>
<titlePage> ... </titlePage>
<divGen type='toc'>
<div><head>Preface</head>
<p> ...
</front>
```

Remarks: This element is intended primarily for use in document production or manipulation, rather than in the transcription of pre-existing materials; it makes it easier to specify the location of indices, tables of contents, etc., to be generated by text preparation or word processing software. The *n* attribute should be used to give a title for the text division being generated.
Part: additional tag set for common core features.
Member of classes:
DTD file:
Data description: Empty.
May occur within:
May contain: EMPTY
Element Declaration:

```
<!ELEMENT div.gen       - O  EMPTY                              >
```

Attribute List Declaration:

```
<!ATTLIST div.gen            %a.global;
          type               CDATA              #IMPLIED       >
```

Discussed in 6.8.2, “Index Entries,” on page 51.
This tag is optional.

Class Name: **edit**
Description: phrase-level elements for simple editorial correction and transcription.
Attributes: [None: global and inherited attributes only.]
Remarks:
Part: base tag set for common core features.
Member of classes: phrase
DTD file: teicore2
Discussed in 6.5, “Simple Editorial Changes,” on page 33.

<edition> (i.e. Edition)
Description: describes the particularities of one edition of a text.
Attributes: [None: global and inherited attributes only.]

Example:

```
<edition>First edition <date>Oct 1990</date>
<edition n='S2'>Students' edition
```

Remarks:
Part: auxiliary tag set for TEI headers.
Member of classes: biblpart
DTD file:
Data description:
May occur within: monogr
May contain: %phrase.seq;
Element Declaration:

```
<!ELEMENT edition       - o  (%phrase.seq;)                    >
```

Attribute List Declaration:

```
<!ATTLIST edition            %a.global;                        >
```

Discussed in 5.2.2, “The Edition Statement,” [in separate fascicle].
This tag is optional.

<editor> (i.e. editor)

Description: 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.
Attributes:

role: specifies the nature of the intellectual responsibility
Data type: CDATA
Value: semi-open list (examples might include: translator, editor, compiler, illustrator, etc.)
Default value: editor

Example:

```
<editor>Eric Johnson
<editor role=illustrator>John Tenniel
```

Remarks: Particularly where cataloguing is likely to be based on the content of the header, it is advisable to use generally recognized authority lists for the exact form of personal names.
Part: base tag set for common core features.
Member of classes: biblpart
DTD file:
Data description: A consistent format should be adopted
May occur within: analytic monogr series
May contain: %phrase.seq

Element Declaration:

```
<!ELEMENT editor        - o  (%phrase.seq)                    >
```

Attribute List Declaration:

```
<!ATTLIST editor             %a.global;
          role               CDATA              editor        >
```

Discussed in 5.2.9, “Computer Files Composed of Transcribed Speech,” [in separate fascicle].
Discussed in 6.10, “Bibliographic Citations and References,” on page 63.
This tag is optional.

<emph> (i.e. emphasized)

Description: marks words or phrases which are stressed or emphasized for linguistic or rhetorical effect.
Attributes: [None: global and inherited attributes only.]
Example:

```
You took the car and did <emph>what</emph>?!!

<q>What it all comes to is this,</q> he said. <q><emph>What
does Christopher Robin do in the morning nowadays?</emph></q>
```

Remarks:
Part: base tag set for common core features.
Member of classes: hqphrase
DTD file:
Data description: free prose
May occur within:
May contain: %phrase.seq;
Element Declaration:

```
<!ELEMENT emph          - -  (%phrase.seq;)                   >
```

Attribute List Declaration:

```
<!ATTLIST emph               %a.global;                       >
```

Discussed in 6.3.2.2, “Emphatic Words and Phrases,” on page 13.
This tag is recommended when applicable.

<expan> (i.e. expansion)

Description: contains the expansion of an abbreviation.
Attributes:

abbr (i.e. abbreviation): gives the abbreviation in its unexpanded form.
Data type: CDATA
Value: any string of characters
Default value: #IMPLIED
Example:

```
The address is Southmoor <expan abbr='Rd'>Road</>.
```

This attribute is optional.

resp (i.e. responsibility): identifies the individual responsible for the interpretation given the abbreviation.
Data type: CDATA
Value: any string of characters, usually the initials of the individual, if known, or a descriptive phrase (e.g. “16c hand”) if not.
Default value: #IMPLIED
Example:

This attribute is optional.

type: allows the encoder to classify the abbreviation according to some convenient typology.
Data type: CDATA
Value: any useful classification name, e.g. “suspension”, “contraction”, “brevigraph”, “title”, “organization”, “geographic”, etc.
Default value: #IMPLIED
Example:

Remarks: The *type* attribute is provided for the sake of those who wish to classify abbreviations at their point of occurrence; this may be useful in some circumstances, though usually the same abbreviation will have the same type in all occurrences. As the sample values make clear, abbreviations may be classified by the method used to construct them, the method of writing them, or the referent of the term abbreviated; the typology used is up to the encoder and should be carefully planned to meet the needs of the expected use.
This attribute is optional.

Example:

Remarks: This tag is the mirror image of the **<abbr>** tag; both allow the encoder to transcribe both an abbreviation and its expansion. In **<abbr>**, however, the original is transcribed as the content of the element and the expansion as an attribute value; **<expan>** reverses this. The choice between the two is up to the user.

The **<expan>** tag is not required; if appropriate, the encoder may expand abbreviations in the source text silently, without tagging them. If this is done, the TEI header should so indicate.

Part: additional tag set for common core features.
Member of classes: data
DTD file:
Data description: May contain character data and phrase-level elements.
May occur within:
May contain: %phrase.seq;
Element Declaration:

```
<!ELEMENT expan         - -  (%phrase.seq;)                     >
```

Attribute List Declaration:

```
<!ATTLIST expan              %a.global;
          abbr               CDATA              #IMPLIED
          resp               CDATA              #IMPLIED
          type               CDATA              #IMPLIED       >
```

Discussed in 6.4.4, “Abbreviations and Their Expansions,” on page 30.
Discussed in 21, “Text Criticism and Apparatus,” [in separate fascicle].
This tag is optional.

<extent>

Description: describes the approximate size of the electronic text as stored on some carrier medium, specified in any convenient units.
Attributes: [None: global and inherited attributes only.]
Example:

```
<extent>3200 sentences
<extent>between 10 and 20 Mb
<extent>ten 3.5 inch high density diskettes
```

Remarks:
Part: auxiliary tag set for TEI headers.
Member of classes: biblpart
DTD file:
Data description:
May occur within: bibl.full
May contain: %phrase.seq;
Element Declaration:

```
<!ELEMENT extent        - o  (%phrase.seq; )                    >
```

Attribute List Declaration:

```
<!ATTLIST extent             %a.global;                          >
```

Discussed in 5.2.3, “Type and Extent of File,” [in separate fascicle].
This tag is optional.

<foreign> (i.e. foreign)

Description: identifies a word or phrase as belonging to some language other than that of the surrounding text.
Attributes:

lang (i.e. language): identifies the language of the word or phrase marked.
Data type: IDREF
Value: contains a language code (associated with some writing system declaration) for the language in question; where applicable, the codes of ISO 639 should be used.
Default value: #IMPLIED
Example:

Remarks: It is strongly recommended that the *lang* attribute be consistently specified on all **<foreign>** elements.
This attribute is recommended.

Example:

```
This is heathen Greek to you still?
Your <foreign lang=lat>lapis philosophicus</foreign>?
```

Remarks: This element is intended for use only where no other element is available to mark the phrase or words concerned. The global *lang* attribute should be used in preference to this element where it is intended to mark the language of the whole of some text element.

The **<distinct>** element may be used to identify phrases belonging to sublanguages or registers not generally regarded as true languages.
Part: base tag set for common core features.
Member of classes: hqphrase
DTD file:
Data description: free prose
May occur within:
May contain: %phrase.seq;
Element Declaration:

```
<!ELEMENT foreign       - -  (%phrase.seq;)                    >
```

Attribute List Declaration:

```
<!ATTLIST foreign            %a.global;
          lang               IDREF               #IMPLIED       >
```

Discussed in 6.3.2.1, “Foreign Words or Expressions,” on page 12.
This tag is optional.

<gloss>

Description: identifies a phrase or word used to provide a gloss or definition for some other word or phrase

Attributes:

target: identifies the associated term element
Data type: IDREF
Value: must be a valid identifier for some **<term>** element in the current document
Default value: #IMPLIED
This attribute is optional.

Example:

```
We may define <term rend=sc id=tdpv>discoursal point of view</term>
as <gloss target=tdpv>the relationship, expressed through discourse
structure, between the implied author or some other addresser,
and the fiction.</gloss>
```

Remarks:
Part: base tag set for common core features.
Member of classes: hqphrase
DTD file:
Data description: free prose
May occur within:
May contain: %phrase.seq;
Element Declaration:

```
<!ELEMENT gloss         - -  (%phrase.seq;)                    >
```

Attribute List Declaration:

```
<!ATTLIST gloss              %a.global;
          target             IDREF               #IMPLIED       >
```

Discussed in 6.3.4, “Terms, Glosses, and Cited Words,” on page 21.
This tag is recommended when applicable.

<head> (i.e. heading)

Description: contains any heading, for example, the title of a section, or the heading of a list or glossary.
Attributes: [None: global and inherited attributes only.]
Example: The most common use for the **<head>** element is to mark the headings of sections. In older writings, the headings or *incipits* may be rather longer than usual in modern works. If a section has an explicit ending as well as a heading, it should be marked as a **<trailer>**, as in this example:

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

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

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

Remarks: The **<head>** tag is used for headings at all levels; processing programs which treat (e.g.) chapter headings, section headings, and list titles differently must determine the proper processing of a **<head>** element based on its structural position. A **<head>** occurring as the first element of a list is the title of that list; one occurring as the first element of a **<div1>** is the title of that chapter or section.
Part: additional tag set for common core features.
Member of classes:
DTD file:
Data description: May contain character data and phrase-level elements.
May occur within: lg list list.bibl
May contain: %phrase.seq;

Element Declaration:

```
<!ELEMENT head          - -  (%phrase.seq;)                    >
```

Attribute List Declaration:

```
<!ATTLIST head              %a.global;                          >
```

Discussed in 6.7, “Lists,” on page 44.
Discussed in 7, “Base Tag Set for Prose,” [in separate fascicle].
This tag is recommended when applicable.

<head.item> (i.e. heading for list items)

Description: contains the heading for the item or gloss column in a glossary list or similar structured list.
Attributes: [None: global and inherited attributes only.]
Example:

```
The simple, straightforward statement of an idea is
preferable to the use of a worn-out expression.
<list type=gloss>
<head.label rend='small caps'>TRITE
<head.item  rend='small caps'>SIMPLE, STRAIGHTFORWARD
<label>bury the hatchet  <item>stop fighting, make peace
<label>at loose
ends  <item>disorganized
<label>on speaking terms <item>friendly
<label>fair and
square      <item>completely honest
<label>at death's door   <item>near death
</list>
```

Remarks: The **<head.item>** element may appear only if each item in the list is preceded by a **<label>**.
Part: additional tag set for common core features.
Member of classes:
DTD file:
Data description: May contain character data and phrase-level elements.
May occur within: list
May contain: %phrase.seq;
Element Declaration:

```
<!ELEMENT head.item     - o  (%phrase.seq;)                    >
```

Attribute List Declaration:

```
<!ATTLIST head.item          %a.global;                         >
```

Discussed in 6.7, “Lists,” on page 44.
This tag is recommended when applicable.

<head.label> (i.e. heading for list labels)

Description: contains the heading for the label or term column in a glossary list or similar structured list.
Attributes: [None: global and inherited attributes only.]
Example:

```
The simple, straightforward statement of an idea is
preferable to the use of a worn-out expression.
<list type=gloss>
<head.label rend='small caps'>TRITE
<head.item  rend='small caps'>SIMPLE, STRAIGHTFORWARD
<label>bury the hatchet  <item>stop fighting, make peace
<label>at loose
ends  <item>disorganized
<label>on speaking terms <item>friendly
<label>fair and
square      <item>completely honest
<label>at death's door   <item>near death
</list>
```

Remarks: The **<head.label>** element may appear only if each item in the list is preceded by a **<label>**.
Part: additional tag set for common core features.
Member of classes:
DTD file:
Data description: May contain character data and phrase-level elements.
May occur within: list
May contain: %phrase.seq;
Element Declaration:

```
<!ELEMENT head.label    - O  (%phrase.seq;)                     >
```

Attribute List Declaration:

```
<!ATTLIST head.label         %a.global;                         >
```

Discussed in 6.7, “Lists,” on page 44.
This tag is recommended when applicable.

<hi> (i.e. highlighted)

Description: marks a word or phrase as typographically distinct from the surrounding text, for reasons concerning which no claim is made.
Attributes:

rend (i.e. rendition): describes the rendition or presentation of the word or phrase highlighted.
Data type: CDATA
Value: a set of keywords from some suitable vocabulary; no systematic recommendations for such a vocabulary are made by these Guidelines.
Default value: #IMPLIED
Example:

Remarks: The *rend* attribute, though optional in general, is recommended on the **<hi>** element.
This attribute is recommended.

Example:

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

Remarks:
Part: auxiliary tag set for TEI headers.
Member of classes: hqphrase
DTD file:
Data description: free prose
May occur within:
May contain: %phrase.seq
Element Declaration:

```
<!ELEMENT hi            - -  (%phrase.seq)                      >
```

Attribute List Declaration:

```
<!ATTLIST hi                 %a.global;
          rend               CDATA              #IMPLIED       >
```

Discussed in 6.3.2.2, “Emphatic Words and Phrases,” on page 13.
This tag is optional.

Class Name: hqphrase

Description: phrase-level elements related to highlighting.
Attributes: [None: global and inherited attributes only.]

Remarks:
Part: base tag set for common core features.
Member of classes: phrase
DTD file: teicore2
Discussed in 6.3, “Highlighting and Quotation,” on page 10.

Class Name: **hqinter**
Description: intermediate-level elements related to highlighting.
Attributes: [None: global and inherited attributes only.]
Remarks:
Part: base tag set for common core features.
Member of classes: inter
DTD file: teicore2
Discussed in 6.3, “Highlighting and Quotation,” on page 10.

<idno> (i.e. identifying number)

Description: supplies any standard or non-standard number used to identify a bibliographic item.
Attributes:

type: categorizes the number, for example as an ISBN or other standard series.
Data type: CDATA
Value: A name or abbreviation indicating what type of identifying number is given (e.g. ISBN, LCCN).
Default value: #IMPLIED

Example:

```
<idno type=ISSN>0143-3385
<idno type=OTA>116
```

Remarks:
Part: auxiliary tag set for TEI headers.
Member of classes: biblpart
DTD file:
Data description:
May occur within: bibl.struct
May contain: #PCDATA
Element Declaration:

```
<!ELEMENT idno          - o  (#PCDATA)                          >
```

Attribute List Declaration:

```
<!ATTLIST idno              %a.global;
          type              CDATA              #IMPLIED       >
```

Discussed in 5.2.4, “Publication, Distribution, etc.,” [in separate fascicle].
This tag is optional.

<imprint>

Description: groups information relating to the publication or distribution of a bibliographic item.
Attributes: [None: global and inherited attributes only.]
Example:

```
<imprint>Oxford, Clarendon Press, 1987
<imprint><place>Oxford<publisher>Clarendon Press<date>1987
```

Remarks:
Part: base tag set for common core features.
Member of classes: biblpart
DTD file:
Data description:
May occur within: monogr
May contain: %m.date %m.place; biblScope publisher
Element Declaration:

```
<!ELEMENT imprint       - o  (%m.place; | publisher | %m.date |
                             biblScope)*                      >
```

Attribute List Declaration:

```
<!ATTLIST imprint           %a.global;                        >
```

Discussed in 6.10, “Bibliographic Citations and References,” on page 63.
This tag is optional.

<index> (i.e. index entry)

Description: marks a location to be indexed for whatever purpose.
Attributes:

index (i.e. index number): indicates which index (of several) the index entry belongs to.
Data type: CDATA
Value: any string of characters; valid values are application-dependent.
Default value: #IMPLIED

Remarks: This attribute makes it possible to create multiple indices for a text.
This attribute is optional.

level1 (i.e. first-level index entry): gives the form under which the index entry is to be made.
Data type: CDATA
Value: any string of characters.
Default value: #REQUIRED
Remarks: At least one level of entry is required.
This attribute is required.

level2 (i.e. second-level index entry): gives the second-level form, if any.
Data type: any string of characters.
Default value: #IMPLIED
This attribute is optional.

level3 (i.e. third-level index entry): gives the third-level form, if any.
Data type: any string of characters.
Default value: #IMPLIED
This attribute is optional.

level4 (i.e. fourth-level index entry): gives the fourth-level form, if any.
Data type: any string of characters.
Default value: #IMPLIED
This attribute is optional.

Example:

```
David's other principal backer, Josiah ha-Kohen
<index level1='Josiah ha-Kohen b. Azarya'
       level2='cousin and backer of David b. Daniel'>
b. Azarya, son of one of the last gaons of Sura,
<index level1='Azarya' level2='gaon of Sura'>
<index level1='Sura'>
was David's own first cousin.
```

Part: additional tag set for common core features.
Member of classes:
DTD file:
Data description: Empty.
May occur within:
May contain: EMPTY
Element Declaration:

```
<!ELEMENT index         - O  EMPTY                              >
```

Attribute List Declaration:

```
<!ATTLIST index              %a.global;
          index              CDATA             #IMPLIED
          level1             CDATA             #REQUIRED
          level2             any string of characters.
                                               #IMPLIED
          level3             any string of characters.
                                               #IMPLIED
          level4             any string of characters.
                                               #IMPLIED       >
```

Discussed in 6.8.2, “Index Entries,” on page 51.
This tag is optional.

<indiv> (i.e. individual)

Description: contains structured information about an individual.
Attributes: [None: global and inherited attributes only.]
Example:

Remarks: To tag just the name of the individual, use the **<name>** or **<person>** tag.
Part: base tag set for common core features.
Member of classes: agent
DTD file:
Data description: May contain the name of the individual, zero or more addresses, and zero or more affiliations.
May occur within:
May contain: address affil name person
Element Declaration:

```
<!ELEMENT indiv         - O  ((person | name), address*,
                             affil*)                          >
```

Attribute List Declaration:

```
<!ATTLIST indiv              %a.global;                       >
```

Discussed in 6.4, “Names, Numbers, Dates, Abbreviations, and Addresses,” on page 24.
This tag is optional.

Class Name: inter

Description: elements of the intermediate (inter-level) class, which can occur both within paragraphs and between paragraphs.
Attributes: [None: global and inherited attributes only.]
Remarks: This element class contains a subset of those elements which can appear in the unstructured "soup" with which paragraph and other elements at the lowest level of crystal structures are filled: specifically all the elements which can also occur as structural elements in their own right. In prose, this means the elements in this class can appear both within and between paragraphs. This class is thus distinct from the purely phrase-level elements which can appear only within soup, and not on their own; the latter class, in keeping with this metaphor, is called "broth"; it is represented by the class *phrase*. Cf. also the class *chunks*.
Part: base tag set for common core features.
Member of classes:
DTD file: teicore2
Discussed in 3.4, "Element Classes and Other Parameter Entities," [in separate fascicle].

<item>

Description: contains one component of a list.
Attributes:

n (i.e. name or number): indicates the number (or letter, or other enumerator) borne by this list item in the copy text.
Data type: CDATA
Value: Whatever string of characters is used to label a list item in the copy text may be used as the value of *n*, but it is not required that numbering be recorded explicitly.
Default value: #IMPLIED
Remarks: In ordered lists, the *n* attribute on the **<item>** element is by definition synonymous with the use of the **<label>** element to record the enumerator of the list item. In glossary lists, however, the term being defined should be given with the **<label>** element, not *n*.
This attribute is optional.

Example:

```
<list type=ordered>
<head>Here begin the chapter headings of Book IV</head>
<item n='4.1'>The death of Queen Clotild.
<item n='4.2'>How King Lothar wanted to appropriate
     one third of the Church revenues.
<item n='4.3'>The wives and children of Lothar.
<item n='4.4'>The Counts of the Bretons.
<item n='4.5'>Saint Gall the Bishop.
<item n='4.6'>The priest Cato.
<item> ...
</list>
```

Part: additional tag set for common core features.
Member of classes:
DTD file:
Data description: May contain simple prose or a sequence of chunks.
May occur within: list
May contain: %specialPara;
Element Declaration:

```
<!ELEMENT item          - O  (%specialPara;)                    >
```

Attribute List Declaration:

```
<!ATTLIST item               %a.global;
          n                  CDATA               #IMPLIED       >
```

Discussed in 6.7, “Lists,” on page 44.
This tag is recommended when applicable.

<l> (i.e. verse line)

Description: contains a single, possibly incomplete, line of verse.
Attributes:

part: specifies whether or not the line is metrically complete.
Data type: (Y | N | I | M | F)
Legal values are:
- Y : the line is metrically incomplete
- N : either the line is complete, or no claim is made as to its completeness
- I : the initial part of an incomplete line
- M : a medial part of an incomplete line
- F : the final part of an incomplete line

Default value: N
Remarks: The values I, M, or F should be used only where it is clear how the line is to be reconstituted.
This attribute is mandatory when applicable.

met: contains a user-specified encoding for the metrical structure of the line.
Data type: CDATA
Value: No recommendations are made at this time; some commonly used example formats are described in section 8, “Base Tag Set for Verse,” [in separate fascicle].
Default value: #IMPLIED
Remarks: The conventions used for encoding metrical patterns must be documented in the TEI Header.
This attribute is optional.

Example:

```
<l part=Y met='-/-/-/-/-/'>
```

Part: base tag set for common core features.
Member of classes: chunk
DTD file:
Data description: contains character data or phrase level elements only
May occur within: lg
May contain: %phrase.seq
Element Declaration:

```
<!ELEMENT l             - o  (%phrase.seq)                      >
```

Attribute List Declaration:

```
<!ATTLIST l                  %a.global;
          part               (Y | N | I | M | F)
                                                N
          met                CDATA              #IMPLIED       >
```

Discussed in 6.11.1, “Verse,” on page 82.
This tag is mandatory when applicable.

<label>

Description: contains the label associated with an item in a list glossaries, marks the term being defined. in
Attributes: [None: global and inherited attributes only.]
Example:

Labels are most commonly used for the headwords in glossary lists; note the use of the global *lang* attribute to set the default language of the glossary list to Middle English, and identify the glosses and headings as modern English or Latin.

```
<list type=gloss lang=ME>
<head        lang=EN>Vocabulary</head>
<head.label lang=EN>Middle English</head.label>
<head.item  lang=EN>New English</head.item>
<label>nu   <item lang=EN>now
<label>lhude      <item lang=EN>loudly
<label>bloweth    <item lang=EN>blooms
<label>med  <item lang=EN>meadow
<label>wude <item lang=EN>wood
<label>awe  <item lang=EN>ewe
<label>lhouth     <item lang=EN>lows
<label>sterteth  <item lang=EN>bounds, frisks
                  (cf. <citation>Chaucer, <title>K.T.</title>
                  644</citation>:  <q>a courser,
                  <egwd>sterting</egwd> as the fyr</q>
```

```
<label>verteth     <item lang=LA>pedit
<label>murie       <item lang=EN>merrily
<label>swik <item lang=EN>cease
<label>naver       <item lang=EN>never
</list>
```

Labels may also be used to record explicitly the numbers or letters which mark list items in ordered lists, as in this extract from Gibbon's *Autobiography*. In this usage the **<label>** element is synonymous with the *n* attribute on the **<item>** element.

```
I will add two facts, which have seldom occurred in
the composition of six, or at least of five quartos.
<list type=ordered rend=runon>
<label>(1) <item>My first rough manuscript, without any
intermediate copy, has been sent to the press.
<label>(2) <item>Not a sheet has been seen by any human
eyes, excepting those of the author and the printer:
the faults and the merits are exclusively my own.
</list>
```

Labels may also be used for other structured list items, as in this extract from the journal of Edward Gibbon:

```
<list type=gloss>
<label>March 1757. <item>I wrote some critical observations
upon Plautus.
<label>March 8th.  <item>I wrote a long dissertation upon
some lines of Virgil.
<label>June.        <item>I saw Mademoiselle Curchod --
<q lang=LA>Omnia vincit amor, et nos cedamus amori.</q>
<label>August.      <item>I went to Crassy, and staid two days.
</list>

<list type=gloss>
<label>The fourth volume of the <title>History of the
Decline and Fall of the Roman Empire</title>
<item>begun March 1, 1782 -- ended June, 1784.
<label>The fifth volume
<item>begun July, 1784 -- ended May 1, 1786.
<label>The sixth volume
<item>begun May 18, 1786 -- ended June 27, 1787.
</list>
```

Part: additional tag set for common core features.
Member of classes: lists
DTD file:
Data description: May contain character data and phrase-level elements.
May occur within: list
May contain: %phrase.seq;
Element Declaration:

```
<!ELEMENT label         - o  (%phrase.seq;)                    >
```

Attribute List Declaration:

```
<!ATTLIST label             %a.global;                         >
```

Discussed in 6.7, “Lists,” on page 44.
This tag is recommended when applicable.

<lg> (i.e. line group)

Description: contains a group of verse lines functioning as a formal unit e.g. a stanza, refrain, verse paragraph, etc.

Attributes:

type: characterizes the line group in some respect, e.g. as a verse paragraph, stanza, verse, etc.
Data type: CDATA
Sample values include:
stanza : stanzas of regular number of lines
para : blank verse paragraph
free : arbitrary grouping e.g. in free verse
refrain : a chorus or refrain
Default value: #IMPLIED
This attribute is recommended.

rhyme: contains a user-specified encoding for the rhyme scheme of this line group.
Data type: CDATA
Value: No recommendations are made at this time; some commonly used example formats are described in section 8, “Base Tag Set for Verse,” [in separate fascicle].
Default value: '*INHERITED'
Remarks: The conventions used for encoding rhyme schemes must be documented in the TEI Header.

part: specifies whether or not the line is metrically complete.
Data type: (y | n | i | m | f)
Legal values are:
y : the line group is metrically incomplete
n : either the line group is complete, or no claim is made as to its completeness.
i : the initial part of an incomplete line group
m : a medial part of an incomplete line group
f : the final part of an incomplete line group
Default value: N
Remarks: The values I, M, or F should be used only where it is clear how the line group is to be reconstituted.
This attribute is mandatory when applicable.

Example:

```
<lg type=free>
<l>
</lg>
```

Remarks:
Part: base tag set for common core features.
Member of classes: chunk
DTD file:
Data description: contains verse lines only, possibly prefixed by a heading.
May occur within: lg
May contain: head l lg
Element Declaration:

```
<!ELEMENT lg            - o  (head?,(l|lg)+)                    >
```

Attribute List Declaration:

```
<!ATTLIST lg                 %a.global;
          type               CDATA               #IMPLIED
          rhyme              CDATA               '*INHERITED'
          part               (y | n | i | m | f)
                                                 N              >
```

Discussed in 6.11.1, “Verse,” on page 82.
This tag is optional.

<lb> (i.e. line break)

Description: marks the start of a new (typographic) line in some edition or version of a text.
Attributes:

ed (i.e. edition): indicates the edition or version in which the line break is located at this point
Data type: CDATA
Value: Any string of characters; usually a siglum conventionally used for the edition.
Default value: #REQUIRED
Example:

```
<lb ed=Riverside n=123>
```

This attribute is required.

n (i.e. number or name): indicates the number or other value associated with the line which follows the point of insertion of this **<lb>**.
Data type: CDATA

Value: Any string of characters.
Default value: #IMPLIED
Remarks: Encoders should adopt a clear and consistent policy as to whether the numbers associated with line breaks relate to the physical sequence number of the line within the page, or to some aspect of the logical structure of the text. By convention, **<lb>** elements should appear at the start of the line to which they refer.
This attribute is optional.

Example:

Remarks: Like other forms of milestone tag, **<lb>** tags cannot be automatically verified by SGML; for better validation, a concurrent markup stream should be used.

The **<lb>** tag is intended for making typographic line breaks in prose. It should be carefully distinguished from the **<l>** element, used to mark lines of verse.
Part: additional tag set for common core features.
Member of classes: loc
DTD file:
May occur within:
May contain: EMPTY
Element Declaration:

```
<!ELEMENT lb            - O  EMPTY                              >
```

Attribute List Declaration:

```
<!ATTLIST lb                 %a.global;
          ed                 CDATA              #REQUIRED
          n                  CDATA              #IMPLIED       >
```

Discussed in 6.9.3, “Concurrent Markup for Pages and Lines,” on page 56.
This tag is optional.

<line>

Description: contains one line of a reference edition.
Attributes: [None: global and inherited attributes only.]
Example:

```
<(La)page n=32>
<(La)line n=1> ... [text of edition La, p. 32, l. 1]
<(La)line n=2> ... [text of edition La, p. 32, l. 2]
<(La)line n=3> ... [text of edition La, p. 32, l. 3]
<(La)line n=4> ... [text of edition La, p. 32, l. 4]
```

```
<(La)page n=33> ...
```

Remarks: <part>base tag set for physical description of the copy text</part> (optional)
Member of classes:
DTD file:
Data description: May contain character data only.
May occur within: col page
May contain: #PCDATA
Element Declaration:

```
<!ELEMENT line          - -  (#PCDATA)                          >
```

Attribute List Declaration:

```
<!ATTLIST line               %a.global;                         >
```

Discussed in 6.9.3, “Concurrent Markup for Pages and Lines,” on page 56.
This tag is optional.

<list>

Description: contains any sequence of items organized as a list.
Attributes:

type: describes the form of the list.
Data type: CDATA
Suggested values include:
ordered : list items are numbered or lettered.
bulleted : list items are marked with a “bullet” or other typographic device.
simple : list items are not numbered or bulleted.
gloss : each list item glosses some term or concept, which is given by a **<label>** element preceding the list item.
Default value: simple
Example:

Remarks: The formal syntax of the element declarations allows **<label>** tags to be omitted from lists tagged **<list type=gloss>**; this is however a semantic error.
This attribute is optional.

Example:

```
<list type=ordered>
<item>a butcher
<item>a baker
<item>a candlestick maker, with
<list type=bullets>
<item>rings on his fingers
<item>bells on his toes
</list>
</list>
```

The following example treats the short numbered clauses of Anglo-Saxon legal codes as lists of items. The text is from an ordinance of King Athelstan (924-939).

```
<div1><head>Athelstan's Ordinance</head>
<list type=ordered>
<item n='1'>Concerning thieves.  First, that no thief is to be
spared who is caught with the stolen goods, [if he is] over
twelve years and [if the value of the goods is] over eightpence.
<list type=ordered>
<item n='1.1'>And if anyone does spare one, he is to pay for the
thief with his wergild -- and the thief is to be no nearer a
settlement on that account -- or to clear himself by an oath of
that amount.
<item n=1.2>If, however, he [the thief] wishes to defend himself
or to escape, he is not to be spared [whether younger or older
than twelve].
<item n=1.3>If a thief is put into prison, he is to be in prison
40 days, and he may then be redeemed with 120 shillings; and the
kindred are to stand surety for him that he will desist for ever.
<item n=1.4>And if he steals after that, they are to pay for him
with his wergild, or to bring him back there.
<item n=1.5>And if
he steals after that, they are to pay for him with his wergild,
whether to the king or to him to whom it rightly belongs; and
everyone of those who supported him is to pay 120 shillings to
the king as a fine.
</list>
<item n=2>Concerning lordless men.  And we pronounced about these
lordless men, from whom no justice can be obtained, that one
should order their kindred to fetch back such a person to justice
and to find him a lord in public meeting.
<list type=ordered>
<item n=2.1>And if they then will not, or cannot, produce him on
that appointed day, he is then to be a fugitive afterwards, and
he who encounters him is to strike him down as a thief.
<item n=2.2>And he who harbours him after that, is to pay for him
with his wergild or to clear himself by an oath of that amount.
</list>
<item n=3>Concerning the refusal of justice.  The lord who
refuses justice and upholds his guilty man, so that the king is
appealed to, is to repay the value of the goods and 120 shillings
to the king; and he who appeals to the king before he demands
```

```
justice as often as he ought, is to pay the same fine as the
other would have done, if he had refused him justice.
<list type=ordered>
<item n=3.1>And the lord who is an accessory to a theft by his
slave, and it becomes known about him, is to forfeit the slave
and be liable to his wergild on the first occasionp if he does it
more often, he is to be liable to pay all that he owns.
<item n=3.2>And likewise any of the king's treasurers or of our
reeves, who has been an accessory of thieves who have committed
theft, is to liable to the same.
</list>
<item n=4>Concerning treachery to a lord.  And we have pronounced
concerning treachery to a lord, that he [who is accused] is to
forfeit his life if he cannot deny it or is afterwards convicted
at the three-fold ordeal.
</list>
```

Note that nested lists have been used so the tagging mirrors the structure indicated by the two-level numbering of the clauses. The clauses could have been treated as a one-level list with irregular numbering, if desired.

```
<p>These decrees, most blessed Pope Hadrian, we propounded in the
public council ... and they confirmed them in our hand in your
stead with the sign of the Holy Cross, and afterwards inscribed
with a careful pen on the paper of this page, affixing thus the
sign of the Holy Cross.
<list type=simple>
<item>I, Eanbald, by the grace of God archbishop of the holy
church of York, have subscribed to the pious and catholic
validity of this document with the sign of the Holy Cross.
<item>I, &AElig;lfwold, king of the people across the Humber,
consenting have subscribed with the sign of the Holy Cross.
<item>I, Tilberht, prelate of the church of Hexham, rejoicing
have subscribed with the sign of the Holy Cross.
<item>I, Higbald, bishop of the church of Lindisfarne, obeying
have subscribed with the sign of the Holy Cross.
<item>I, Ethelbert, bishop of Candida Casa, suppliant, have
subscribed with thef sign of the Holy Cross.
<item>I, Ealdwulf, bishop of the church of Mayo, have subscribed
with devout will.
<item>I, &AElig;thelwine, bishop, have subscribed through
delegates.
<item>I, Sicga,  patrician', have subscribed with serene mind
with the sign of the Holy Cross.
</list>
```

Part: additional tag set for common core features.
Member of classes: lists
DTD file:
Data description: May contain an optional heading followed by a series of items, or a series of label and item pairs, the latter being optionally preceded by one or two specialized headings.

May occur within:
May contain: head head.item head.label item label
Element Declaration:

```
<!ELEMENT list          - -  (head?, ( (item*) | (head.label?,
                             head.item?, (label, item)+)))      >
```

Attribute List Declaration:

```
<!ATTLIST list               %a.global;
          type               CDATA              simple         >
```

Discussed in 6.7, “Lists,” on page 44.
This tag is recommended when applicable.

<list.bibl> (i.e. citation list)

Description: contains a list of bibliographic citations of any kind.
Attributes: [None: global and inherited attributes only.]
Example:

```
<list.bibl>
    <head>Works consulted</head>
    <bibl>Blain, Clements and Grundy: Feminist Companion to Literature in
          English (Yale, 1990)
    </bibl>
    <bibl.struct><title.piece>The Interesting story of the Children in the
Wood
      <title>The Penny Histories</title>
      <author>Victor E Neuberg</author>
      <imprint><publisher>OUP<date>1968
    </bibl.struct>
</list.bibl>
```

Remarks:
Part: base tag set for common core features.
Member of classes: lists
DTD file:
Data description:
May occur within:
May contain: bibl bibl.full bibl.struct head trailer
Element Declaration:

```
<!ELEMENT list.bibl     - -  (head?, (bibl | bibl.struct |
                             bibl.full)+, trailer?)            >
```

Attribute List Declaration:

```
<!ATTLIST list.bibl         %a.global;                    >
```

Discussed in 6.10, “Bibliographic Citations and References,” on page 63.
This tag is optional.

Class Name: **lists**
Description: list-like elements.
Attributes: [None: global and inherited attributes only.]
Remarks:
Part: base tag set for common core features.
Member of classes: inter
DTD file: teicore2
Discussed in 6.7, “Lists,” on page 44.

Class Name: **loc**
Description: elements used for purposes of location and reference
Attributes: [None: global and inherited attributes only.]
Remarks:
Part: base tag set for common core features.
Member of classes: phrase
DTD file: teicore2
Discussed in 6.6, “Simple Links and Cross References,” on page 41.

<measure>

Description: contains a word or phrase referring to a unit of measurement of any kind, typically associated with a number.
Attributes:

type: specifies the type of units being measured.
Data type: CDATA
Sample values include:
weight : measure of weight e.g. kg, pound.
count : unit of count, e.g. dozen, score.
length : measure of length, e.g. pole, mm.
area : measure of area e.g. acre, hectare.
volume : measure of volume e.g. litre, gallon.
currency : unit of currency e.g. ecu, escudo, mark.
Default value: #IMPLIED
Remarks:
This attribute is optional.

reg (i.e. regularized): gives a standardized form for the name of the measure concerned
Data type: CDATA
Value: Any string of characters.
Default value: #IMPLIED
Remarks: In providing a “regularized” form, no claim is made that the form in the source text is incorrect; the regularized form is simply that chosen as the main form for purposes of unifying variant forms under a single heading.
This attribute is optional.

Example:

```
<measure type=currency reg='pound sterling'>quid</measure>
<measure type=area reg=old-merks>merks of old extent</measure>
```

Remarks:
Part: additional tag set for common core features.
Member of classes: data
DTD file:
Data description: May contain character data and phrase-level elements.
May occur within:
May contain: %phrase.seq;
Element Declaration:

```
<!ELEMENT measure       - -  (%phrase.seq;)                    >
```

Attribute List Declaration:

```
<!ATTLIST measure            %a.global;
          type               CDATA              #IMPLIED
          reg                CDATA              #IMPLIED       >
```

Discussed in 6.4.2, “Numbers and Measures,” on page 26.
This tag is optional.

<mentioned>

Description: marks words or phrases mentioned, not used
Attributes: [None: global and inherited attributes only.]
Example:

```
There is thus a striking accentual difference between a verbal
form like <mentioned lang=EL id=c234>eluthemen</mentioned>
<gloss target=c234>we were released,</mentioned> accented on
the second syllable of the word, and its
participial derivative <mentioned lang=EL id=c235>lutheis</mentioned>
<gloss target=c235>released,</mentioned> accented on the last.
```

Remarks:
Part: base tag set for common core features.
Member of classes: hqphrase
DTD file:
Data description: free prose
May occur within:
May contain: %phrase.seq;
Element Declaration:

```
<!ELEMENT mentioned     - -  (%phrase.seq;)                    >
```

Attribute List Declaration:

```
<!ATTLIST mentioned          %a.global;                        >
```

Discussed in 6.3.4, "Terms, Glosses, and Cited Words," on page 21.
This tag is recommended when applicable.

<milestone>

Description: marks the boundary between sections of a text, as indicated by changes in a standard reference system.
Attributes:

ed (i.e. edition): indicates which edition or version the milestone applies to.
Data type: CDATA
Value: Any string of characters; usually a siglum conventionally used for the edition.
Default value: #REQUIRED
Example:

This attribute is required.

n (i.e. number or name): indicates the new number or other value for the unit which changes at this milestone.
Data type: CDATA
Value: Any string of characters.
Default value: #IMPLIED
Remarks: The special value *unnumbered* should be used in passages which fall outside the normal numbering scheme (e.g. chapter heads, poem numbers or titles, or speaker attributions in verse drama).
This attribute is optional.

unit: indicates what kind of section is changing at this milestone.
Data type: CDATA
Suggested values include:

page : page breaks in the reference edition.
column : column breaks.
line : line breaks.
book : any units termed *book*, *liber*, etc.
poem : individual poems in a collection.
canto : cantos or other major sections of a poem.
stanza : stanzas within a poem, book, or canto.
act : acts within a play.
scene : scenes within a play or act.
section : sections of any kind.
absent : passages not present in the reference edition.

Default value: #REQUIRED

Example:

Remarks: If the milestone marks the beginning of a piece of text not present in the reference edition, the special value *absent* may be used as the value of *unit*. The normal interpretation is that the reference edition does not contain the text which follows, until the next **<milestone>** tag for the edition in question is encountered.

In addition to the values suggested, other terms may be appropriate (e.g. *Stephanus* for the Stephanus numbers in Plato).

This attribute is required.

Example:

```
<milestone ed=La name=Dreissiger n=23>
...
<milestone ed=La name=Dreissiger n=24>
...
```

Remarks: Milestones for page and column should precede milestones for line numbers, but this and other logical requirements cannot be enforced automatically by SGML; for better validation, a concurrent markup stream should be used.

Part: additional tag set for common core features.
Member of classes: refsys
DTD file:
May occur within:
May contain: EMPTY
Element Declaration:

```
<!ELEMENT milestone     - O  EMPTY                              >
```

Attribute List Declaration:

```
<!ATTLIST milestone          %a.global;
          ed                 CDATA              #REQUIRED
          n                  CDATA              #IMPLIED
          unit               CDATA              #REQUIRED      >
```

Discussed in 6.9, “Reference Systems,” on page 52.
This tag is optional.

<monogr> (i.e. monographic level)

Description: contains bibliographic elements describing an item (e.g. a book or journal) published as an independent item (i.e. as a separate physical object).
Attributes: [None: global and inherited attributes only.]
Example:

Remarks: The **<monogr>** element may occur only within bibliographic citation or reference elements; it is mandatory for description of the monographic level of **<bibl.struct>** elements.
Part: base tag set for common core features.
Member of classes: biblpart
DTD file:
Data description: May contain specialized bibliographic elements, in a prescribed order.
May occur within: bibl.struct
May contain: author biblScope edition editor imprint meeting note resp title
Element Declaration:

```
<!ELEMENT monogr        - O  ( ( ((author | editor | resp)+,
                             title+, (editor | resp)*) |
                             (title+, (author | editor |
                             resp)*)), (note | meeting)*,
                             (edition, (editor | resp)*),
                             imprint+, biblScope* )            >
```

Attribute List Declaration:

```
<!ATTLIST monogr             %a.global;                        >
```

Discussed in 6.10.2.1, “Analytic, Monographic, and Series Levels,” on page 67.
This tag is mandatory when applicable.

<name> (i.e. name, proper noun)

Description: contains a proper noun or noun phrase.
Attributes:

type: indicates the type of the object which is being named by the phrase.
Data type: CDATA
Value: Values such as person, place, institution, product, acronym.
Default value: #IMPLIED
This attribute is recommended.

Example:

Remarks: Proper nouns referring to people, places, and organizations may be tagged instead with **<person>**, **<place>**, or **<org>**.
Part: additional tag set for common core features.
Member of classes: data agent
DTD file:
Data description: May contain character data and phrase-level elements.
May occur within: indiv organization
May contain: %phrase.seq;
Element Declaration:

```
<!ELEMENT name          - -  (%phrase.seq;)                     >
```

Attribute List Declaration:

```
<!ATTLIST name               %a.global;
          type               CDATA              #IMPLIED       >
```

Discussed in 6.4.1, “Names,” on page 24.
This tag is optional.

Class Name: # names

Description: elements with proper nouns as content.
Attributes:

key: provides an alternative identifier for the object being named, such as a database record key.
Data type: CDATA
Value: any string
Default value: #IMPLIED
Example:

```
<name type='place'>Montaillou</name>
is not a large parish.
At the time of the events which led to <name type='person' key=BXII>
Fournier's</name> investigations,
the local population consisted of between 200 and
250 inhabitants.
```

Remarks: The value may be a unique identifier from a database, or simply a more explicit name for the referent. Its purpose is only to record an identification; if the analysis leading to the identification is to be recorded as well, the analytic tags described in chapter 18, “Feature Structure Analysis,” [in separate fascicle] should be used in addition or instead.
This attribute is optional.

reg (i.e. regularization): gives a normalized or regularized form of the proper name
Data type: CDATA
Value: Any string of characters.
Default value: #IMPLIED
Example:

```
At the time of the events which led to
<name type=person reg='Benedict XII, Pope of Avignon
(Jacques Fournier)'>Fournier's</name> investigations,
the local population consisted of between 200 and
250 inhabitants.
```

Remarks: In providing a “regularized” form, no claim is made that the form in the source text is incorrect; the regularized form is simply that chosen as the main form for purposes of unifying variant forms under a single heading.
This attribute is optional.

Remarks:
Part: base tag set for common core features.
Member of classes:
DTD file:
Discussed in 6.4.1, “Names,” on page 24.

<note>

Description: contains a note or annotation.
Attributes:

n (i.e. number or symbol): indicates the symbol or number used to mark the note's point of attachment to the main text.
Data type: CDATA
Value: any string of characters
Default value: #IMPLIED

Example:

```
Mevorakh b. Saadya's mother, the matriarch of the family
during the second half of the eleventh century,
   <note n=126>
        The alleged mention of Judah Nagid's mother
        in a letter from 1071 is, in fact, a
        reference to Judah's children; cf. above,
        nn. 111 and 54.
   </note>
is well known from Geniza documents published by Jacob Mann.
```

Remarks: If notes are numbered in sequence and their numbering can be reconstructed automatically by processing software, it may not be considered necessary to record the note numbers.
This attribute is optional.

type: describes the type of note.
Data type: CDATA
Value: Values can be taken from any convenient typology of annotation suitable to the work in hand; e.g. annotation, gloss, citation, digression, preliminary, temporary, ...
Default value: #IMPLIED
This attribute is optional.

resp (i.e. responsible): indicates who is responsible for the annotation: author, editor, translator, etc.
Data type: CDATA
Sample values include:
auth[or]: note originated with the author of the text.
ed[itor]: note added by the editor of the text.
comp[iler]: note added by the compiler of a collection.
tr[anslator]: note added by the translator of a text.
transcr[iber]: note added by the transcriber of a text into electronic form.
(initials): note added by the individual indicated by the initials.
Default value: #IMPLIED
Remarks: For specialized types of editorial annotation (e.g. for marking corrections, normalizations, cruxes, etc.), see chapter 21, “Text Criticism and Apparatus,” [in separate fascicle].
This attribute is recommended when applicable.

place: indicates where the note appears in the source text.
Data type: CDATA
Sample values include:
foot: note appears at foot of page.
end: note appears at end of chapter or volume.
inline: note appears as a marked paragraph in the body of the text.
left: note appears in left margin.
right: note appears in right margin.
interlinear: note appears between lines of the text.

app[aratus]: note appears in the apparatus at the foot of the page.
Default value: 'unspecified'
Example:

Remarks: For pages with multiple apparatus, values such as *app1* and *app2* can be used.

The *place* attribute can be used to indicate to text formatting software where a note should be printed. If the locations indicated do not agree with those in the copy text, that fact should be indicated in the TEI header.
This attribute is optional.

anchored: indicates whether the copy text shows the exact place of reference for the note.
Data type: (yes | no)
Legal values are:
yes: copy text indicates the place of attachment for the note.
no: copy text indicates no place of attachment for the note.
Default value: yes
Remarks: In modern texts, notes are usually anchored by means of explicit footnote or endnote symbols. An explicit indication of the phrase or line annotated may however be used instead (e.g. "page 218, lines 3-4"). The *anchored* attribute indicates whether any explicit location is given, whether by symbol or by prose cross-reference. If the specific symbols used are to be recorded, use the *n* attribute.
This attribute is optional.

target: indicates the point of attachment of a note, or the beginning of the span to which the note is attached.
Data type: IDREF
Value: reference to the *id* of an element which begins at the location in question (e.g. the *id* of an **<anchor>** element).
Default value: #IMPLIED
Example:

Remarks: If *target* and *target.end* are to be used to indicate where notes attach to the text, then elements at the appropriate locations (**<anchor>** elements if necessary) must be given *id* values to be pointed at.
This attribute is recommended when applicable.

target.end: points to the end of the span to which the note is attached, if the note is not embedded in the text at that point.
Data type: IDREF
Value: reference to the *id* of an element which *ends* at the location in question, or to an empty element at the point in question.
Default value: #IMPLIED
Example:

Remarks:
This attribute is recommended when applicable.

Example:

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

Remarks:
Part: additional tag set for common core features.
Member of classes: notes biblpart
DTD file:
Data description: May contain character data and phrase-level elements.
May occur within: bibl.struct monogr
May contain: %specialPara;
Element Declaration:

```
<!ELEMENT note          - O  (%specialPara;)                    >
```

Attribute List Declaration:

```
<!ATTLIST note               %a.global;
          n                  CDATA               #IMPLIED
          type               CDATA               #IMPLIED
          resp               CDATA               #IMPLIED
          place              CDATA               'unspecified'
          anchored           (yes | no)          yes
          target             IDREF               #IMPLIED
          target.end         IDREF               #IMPLIED       >
```

Discussed in 6.8, “Notes, Annotation, and Indexing,” on page 48.
This tag is recommended.

Class Name: **notes**
Description: note-like elements.
Attributes: [None: global and inherited attributes only.]
Remarks:
Part: base tag set for common core features.
Member of classes: inter

DTD file: teicore2
Discussed in 6.8, “Notes, Annotation, and Indexing,” on page 48.

<num> (i.e. number)

Description: contains a number, written in any form.
Attributes:

type: indicates the type of numeric value.
Data type: CDATA
Suggested values include:
cardinal : absolute number, e.g. “21”, “21.5”
ordinal : ordinal number, e.g. “21st”
fraction : fraction, e.g. “one half” or “three halves”
percentage : e.g. “ten percent”
Default value: #IMPLIED
Remarks: If a different typology is desired, other values can be used for this attribute.
This attribute is optional.

value: supplies the value of the number in an application-dependent standard form.
Data type: CDATA
Value: any numeric value in the chosen standard form.
Default value: #IMPLIED
Remarks: The standard form used should be described in the **<std.vals>** element in the TEI header. Standard forms may be defined from scratch, or borrowed from existing practice (e.g. “Standard values are given in the notation defined for numeric constants in the C language.”)
This attribute is optional.

Example:

```
<num type=cardinal value='21'>twenty-one</num>
<num type=cardinal value='1.5'>1.5</num>
He stands <num type=cardinal value='1.9'>1 &middot; 90</num>m. high.
```

Remarks: Detailed analyses of quantities and units of measure in historical documents should use the feature structure mechanism described in chapter 18, “Feature Structure Analysis,” [in separate fascicle]. The **<num>** element is intended for use in simple applications.
Part: additional tag set for common core features.
Member of classes: data
DTD file:
Data description: May contain character data and phrase-level elements.
May occur within:
May contain: %phrase.seq;
Element Declaration:

```
<!ELEMENT num          - -  (%phrase.seq;)                    >
```

Attribute List Declaration:

```
<!ATTLIST num               %a.global;
          type              CDATA              #IMPLIED
          value             CDATA              #IMPLIED       >
```

Discussed in 6.4.2, “Numbers and Measures,” on page 26.
This tag is optional.

<omit> (i.e. omitted material)

Description: 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 or inaudible.
Attributes:

desc (i.e. description): gives a description of the omitted text.
Data type: CDATA
Value: a prose description of the material omitted.
Default value: #IMPLIED
Example:

This attribute is optional.

reason: gives the reason for omission. Sample values include “sampling”, “illegible”, “inaudible”, “irrelevant”, “canceled”, “canceled and illegible”.
Data type: CDATA
Value: any short indication of the reason for the omission.
Default value: #IMPLIED
Example:

This attribute is optional.

resp (i.e. responsibility): indicates the individual responsible for the omission (in the case of canceled material, the individual responsible for the cancellation)
Data type: CDATA
Value: any string of characters identifying the individual in question, e.g. “2d hand” or the initials of a known individual.
Default value: #IMPLIED
Example:

This attribute is optional.

extent: indicates approximately how much text has been lost, in letters, minims, inches, or any appropriate unit, for cases where a deletion renders the text quite illegible.
Data type: CDATA
Value: any string of characters
Default value: #IMPLIED
Example:

Remarks: It is possible, but not always necessary, to provide measurements precise to the millimeter or even to the printer's point. The degree of precision attempted will vary with the purpose of the encoding and the nature of the material.
This attribute is optional.

Example:

Remarks:
Part: additional tag set for common core features.
Member of classes: edit
DTD file:
Data description: Empty.
May occur within:
May contain: EMPTY
Element Declaration:

```
<!ELEMENT omit          - O  EMPTY                            >
```

Attribute List Declaration:

```
<!ATTLIST omit               %a.global;
          desc               CDATA              #IMPLIED
          reason             CDATA              #IMPLIED
          resp               CDATA              #IMPLIED
          extent             CDATA              #IMPLIED       >
```

Discussed in 6.5.3, “Additions, Deletions and Omissions,” on page 37.
This tag is optional.

<org> (i.e. organization)

Description: contains the name of an organization, or a reference to an organization.
Attributes: [None: global and inherited attributes only.]
Example:

```
<org key='Miniluv'>The Ministry of Love</org>
```

Remarks: This tag may be used to mark either all references to organizations, or only those which use proper nouns; in the latter case it is synonymous with **<name type=org>**. No distinction is made between organizations and institutions; the **<org>** tag may be used to mark references to either, or both. To provide further information about the organization, e.g. an address, use the **<organization>** element.
Part: additional tag set for common core features.
Member of classes: data agent
DTD file:
Data description: May contain character data and phrase-level elements.
May occur within: address organization
May contain: %phrase.seq;
Element Declaration:

```
<!ELEMENT org          - -  (%phrase.seq;)                    >
```

Attribute List Declaration:

```
<!ATTLIST org               %a.global;                        >
```

Discussed in 6.4.1, “Names,” on page 24.
This tag is optional.

<organization> (i.e. organization (structured information))

Description: contains structured information about an organization: its name and address.
Attributes: [None: global and inherited attributes only.]
Example:

```
<organization>
  <org key='TEI'>ACH/ACL/ALLC Text Encoding Intiative</org>
  <address>
      <org>Computer Center (M/C 135), UIC</>
      <street>1940 W. Taylor St. Room 124</>
      <city>Chicago</>
      <region n=IL>Illinois</>
      <postcode>60612-7352</>
      <country>U.S.A.</>
  </address>
  <address>
      <org>Oxford University Computing Service</>
      <street>13 Banbury Rd.</>
      <city>Oxford</>
      <postcode>OX2 6NN</>
      <country>England</>
```

```
            </address>
          </organization>
```

Remarks: This tag may be used to encode a structured description of an organization, providing at least its name and its address. Further elements may be added as they are found useful. To tag only the name of the organization, use the **<org>** or the **<name>** element.
Part: additional tag set for common core features.
Member of classes: data agent
DTD file:
Data description: May contain the organizations name and zero or more addresses.
May occur within:
May contain: address name org
Element Declaration:

```
<!ELEMENT organization  - O  ((org | name), address*)           >
```

Attribute List Declaration:

```
<!ATTLIST organization       %a.global;                          >
```

Discussed in 6.4, "Names, Numbers, Dates, Abbreviations, and Addresses," on page 24.
This tag is optional.

<orig> (i.e. original form)

Description: contains the original form of a reading, for which a regularized form is given in an attribute value.
Attributes:

reg (i.e. original): gives a regularized (normalized) form of the text.
Data type: CDATA
Value: any string of characters
Default value: #IMPLIED
Example:

This attribute is optional.

resp (i.e. responsibility): identifies the individual responsible for the regularization of the word or phrase.
Data type: CDATA
Value: any string of characters, typically the initials of the individual involved, or a role identifier like "editor" if not known by name.
Default value: #IMPLIED
Example:

This attribute is optional.

Example:

Remarks:
Part: additional tag set for common core features.
Member of classes: edit
DTD file:
Data description: May contain character data and phrase-level elements.
May occur within:
May contain: %phrase.seq;
Element Declaration:

```
<!ELEMENT orig          - -  (%phrase.seq;)                    >
```

Attribute List Declaration:

```
<!ATTLIST orig               %a.global;
          reg                CDATA               #IMPLIED
          resp               CDATA               #IMPLIED       >
```

Discussed in 6.5.1, “Correction of Apparent Errors,” on page 34.
Discussed in 21, “Text Criticism and Apparatus,” [in separate fascicle].
This tag is optional.

<p> (i.e. paragraph)

Description: marks paragraphs in prose.
Attributes: [None: global and inherited attributes only.]
Example:

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

Remarks: In some contexts, the paragraph may have a specialized meaning, e.g. in the tag set for dictionaries, **<p>** is used to enclose any running text, and thus does not imply text set off as is conventionally done in running prose.
Part: additional tag set for common core features.
Member of classes: chunk
DTD file:

Data description: May contain character data and phrase-level elements.
May occur within:
May contain: %paraContent;
Element Declaration:

```
<!ELEMENT p             - O  (%paraContent;)                    >
```

Attribute List Declaration:

```
<!ATTLIST p                  %a.global;                         >
```

Discussed in 6.1, “Paragraphs,” on page 7.
This tag is required.

<pb> (i.e. page break)

Description: marks the boundary between one page of a text and the next in a standard reference system.
Attributes:

ed (i.e. edition): indicates the edition or version in which the page break is located at this point
Data type: CDATA
Value: Any string of characters; usually a siglum conventionally used for the edition.
Default value: #REQUIRED
Example:

```
<pb ed=Riverside n=123>
```

This attribute is required.

n (i.e. number or name): indicates the number or other value associated with the page which follows the point of insertion of this **<pb>**.
Data type: CDATA
Value: Any string of characters.
Default value: #IMPLIED
Remarks: Encoders should adopt a clear and consistent policy as to whether the numbers associated with page breaks relate to the physical sequence number of the page or the page number or signature printed on it. By convention, **<pb>** elements should appear at the start of the page to which they refer.
This attribute is optional.

Example:

Remarks: Like other forms of milestone tag, **<pb>** tags cannot be automatically verified by SGML; for better validation, a concurrent markup stream should be used.
Part: additional tag set for common core features.
Member of classes: loc
DTD file:
May occur within:
May contain: EMPTY
Element Declaration:

```
<!ELEMENT pb            - O  EMPTY                              >
```

Attribute List Declaration:

```
<!ATTLIST pb                 %a.global;
          ed                 CDATA              #REQUIRED
          n                  CDATA              #IMPLIED       >
```

Discussed in 6.9.3, “Concurrent Markup for Pages and Lines,” on page 56.
This tag is optional.

<page>

Description: marks pages in a reference edition.
Attributes: [None: global and inherited attributes only.]
Example:

```
<(La)page n=32> ... [text of edition La, page 32]
<(La)page n=32> ... [text of edition La, page 33]
<(La)page n=32> ... [text of edition La, page 34]
<(La)page n=32> ... [text of edition La, page 35]
```

Remarks: The **<page>** element in a concurrent page-reference DTD may be subdivided into columns or lines if desired, or left unanalysed.
Part: base tag set for physical description of the copy text.
Member of classes:
DTD file:
Data description: May contain **<col>** elements, **<line>** elements, or character data.
May occur within: %version vol
May contain: #PCDATA col line
Element Declaration:

```
<!ELEMENT page          - O  (#PCDATA | line | col)*            >
```

Attribute List Declaration:

```
<!ATTLIST page              %a.global;                        >
```

Discussed in 6.9.3, “Concurrent Markup for Pages and Lines,” on page 56.
This tag is optional.

<person>

Description: contains a name of a person, or a reference to a person.
Attributes: [None: global and inherited attributes only.]
Example:

Remarks: This element may be used either for proper nouns only (in which case it is synonymous with **<name type='person'>**) or for any reference to persons it is felt desirable to tag.[1]
Part: additional tag set for common core features.
Member of classes: data agent
DTD file:
Data description: May contain character data and phrase-level elements.
May occur within: indiv
May contain: %phrase.seq;
Element Declaration:

```
<!ELEMENT person        - -  (%phrase.seq;)                   >
```

Attribute List Declaration:

```
<!ATTLIST person            %a.global;                        >
```

Discussed in 6.4.1, “Names,” on page 24.
This tag is optional.

Entity Name: **paraContent** (i.e. paragraph content)
Description: defines the legal contents of paragraphs and similar elements as comprising character data, phrase-level elements which can occur only in paragraph content, and some higher-level elements which can occur either within paragraph content or within sequences of high-level elements (as defined by the parameter entity *seq*).

&@fn@sep

[1] This ambiguity may need to be removed. Use broader definition only and expect **<name>** to nest within **<person>** if the tagger wants it that way? -MSM I think it shd be restricted to names -LB

Expansion: '(#PCDATA | %m.inter | %m.phrase;)*' 3.4, "Element Classes and Other Parameter Entities," [in separate fascicle]

Class Name: phrase

Description: includes elements which can occur at the level of individual words or phrases.
Attributes: [None: global and inherited attributes only.]
Remarks: This class of elements can occur only within larger elements of the class *inter* or *chunk*. In prose, this means these elements can occur within paragraphs, list items, lines of verse, etc.
Part: base tag set for common core features.
Member of classes:
DTD file: teicore2
Discussed in 3.4, "Element Classes and Other Parameter Entities," [in separate fascicle].

Entity Name: phrase.seq (i.e. phrase sequence)

Description: defines a sequence of character data and phrase-level elements.
Expansion: '(#PCDATA | %m.phrase)*' 7.7, "Overall Structure of the Prose DTD," [in separate fascicle]

<place>

Description: contains the name of, or a reference to, a place.
Attributes: [None: global and inherited attributes only.]
Example:

Remarks: This element may be used either for proper nouns only (in which case it is synonymous with **<name type='place'>**) or for any reference to places it is felt desirable to tag.[2]
Part: additional tag set for common core features.
Member of classes: data
DTD file:
Data description: May contain character data and phrase-level elements.
May occur within:
May contain: %phrase.seq;

[2] This ambiguity may need to be removed. Use broader definition only and expect **<name>** to nest within **<place>** if the tagger wants it that way? -MSM I think it shd be restricted to names -LB

Element Declaration:

```
<!ELEMENT place         - -  (%phrase.seq;)                  >
```

Attribute List Declaration:

```
<!ATTLIST place              %a.global;                       >
```

Discussed in 6.4.1, “Names,” on page 24.
This tag is optional.

Class Name: # place

Description: defines a class of elements which contain names of places, addresses, etc.
Attributes: [None: global and inherited attributes only.]
Remarks: This class allows any type of location specification to be used in certain places (e.g. within the **<imprint>** element of a structured bibliographic reference).
Part: base tag set for common core features.
Member of classes:
DTD file:
Discussed in 6.4, “Names, Numbers, Dates, Abbreviations, and Addresses,” on page 24.

Class Name: # pointer

Description: elements which point to other elements, using IDREFs.

Attributes:

type: categorizes the pointer in some respect, using any convenient set of categories.
Data type: CDATA
Value: The type should indicate the intended function of the pointer, or the rhetorical relationship between its source and target.
Default value: #IMPLIED

This attribute is optional.

resp : specifies the creator of the pointer.
Data type: CDATA
Value: any string of characters, usually the initials or name of the creator.
Default value: #IMPLIED

This attribute is optional.

crdate : specifies when the pointer was created.
Data type: CDATA
Value: any string representing a date.
Default value: #IMPLIED
This attribute is optional.

targType : specifies the kinds of elements to which this pointer may point.

Data type: NAMES
Value: A list of names of SGML elements defined in the DTD of the current document.
Default value: #IMPLIED
Remarks: If this attribute is supplied, every element specified as a target must be of one or the other of the types specified. An application may choose whether or not to report failures to satisfy this constraint as errors, but may not access an element of the right identifier but the wrong type.

This attribute is optional.

targOrder : where more than one identifier is supplied as the value of the *target* attribute, this attribute specifies whether the order in which they are supplied is significant.
Data type: (Y | N | U)
Legal values are:
- Y : Yes: the order in which IDREFs are specified as the value of a *target* attribute should be followed when combining the targeted elements.
- N : No: the order in which IDREFs are specified as the value of a *target* attribute has no significance when combining the pointered elements.
- U : Unspecified: the order in which IDREFs are specified as the value of a *target* attribute may or may not be significant.

Default value: U

This attribute is optional.

evaluate : specifies the intended meaning when the target of a pointer is itself a pointer.
Data type: (all | one | none)

- all : if the element pointed to is itself a pointer, then the target of that pointer will be taken, and so on, until an element is found which is not a pointer.

one : if the element pointed to is itself a pointer, then its target (whether a pointer or not) is taken as the target of this pointer.
none : no further evaluation of targets is carried out beyond that needed to find the element specified in the pointer's target.

Default value: #IMPLIED

Remarks: If no value is given, the application program is responsible for deciding (possibly on the basis of user input) how far to trace a chain of pointers.

This attribute is optional.

Remarks:
Part: base tag set for common core features.
Member of classes:
DTD file:
Discussed in 6.6, “Simple Links and Cross References,” on page 41.

<postbox> (i.e. postbox)

Description: contains a number or other identifier for some postal delivery point other than a street address.
Attributes: [None: global and inherited attributes only.]
Example:

```
<postbox>P.O. Box 280

<postbox>Postbus 532
```

Remarks: The position and nature of postal codes is highly country-specific; the conventions appropriate to the country concerned should be used.
Part: base tag set for common core features.
Member of classes:
DTD file:
May occur within: address
May contain: #PCDATA
Element Declaration:

```
<!ELEMENT postbox       - o  (#PCDATA)                        >
```

Attribute List Declaration:

```
<!ATTLIST postbox           %a.global;                      >
```

Discussed in 6.10, “Bibliographic Citations and References,” on page 63.
This tag is optional.

<postcode> (i.e. postcode)

Description: contains a numerical or alphanumeric code used as part of a postal address to simplify sorting or delivery of mail.
Attributes: [None: global and inherited attributes only.]
Example:

```
<postcode>HR1 3LR

<postcode>60142-7
```

Remarks: The position and nature of postal codes is highly country-specific; the conventions appropriate to the country concerned should be used.
Part: base tag set for common core features.
Member of classes:
DTD file:
May occur within: address
May contain: #PCDATA
Element Declaration:

```
<!ELEMENT postcode      - o  (#PCDATA)                      >
```

Attribute List Declaration:

```
<!ATTLIST postcode          %a.global;                      >
```

Discussed in 6.10, “Bibliographic Citations and References,” on page 63.
This tag is optional.

<ptr>

Description: defines a pointer to another location in the current document.

Attributes:

target: specifies the destination of the pointer as one or more SGML identifiers
Data type: IDREFS
Value: Each value specified must be the same as that specified as value for an *ID* attribute for some other element in the current SGML document.

Default value: #REQUIRED

This attribute is required.

Example:

```
<ptr target="P2 P5" resp=LB type=s>
<ptr target='D1' type='defn'>
```

Remarks:
Part: base tag set for common core features.
Member of classes: loc
DTD file:
May occur within:
May contain: EMPTY
Element Declaration:

```
<!ELEMENT ptr           - o  EMPTY                          >
```

Attribute List Declaration:

```
<!ATTLIST ptr                %a.global;
          target             IDREFS             #REQUIRED      >
```

Discussed in 6.6, “Simple Links and Cross References,” on page 41.
This tag is optional.

<publisher>

Description: provides the name of the organization responsible for the publication or distribution of a bibliographic item.
Attributes: [None: global and inherited attributes only.]
Example:

```
<publishr>Faber and Faber Ltd
<imprint><city>Oxford<publisher>Clarendon Press<date>1987
```

Remarks:
Part: base tag set for common core features.
Member of classes: biblpart
DTD file:
Data description: Use the full form of the name by which a company is usually referred to, rather than any abbreviation of it which may appear on a title page
May occur within: imprint
May contain: %phrase.seq
Element Declaration:

```
<!ELEMENT publisher     - o  (%phrase.seq)                    >
```

Attribute List Declaration:

```
<!ATTLIST publisher         %a.global;                        >
```

Discussed in 5.2.4, “Publication, Distribution, etc.,” [in separate fascicle].
Discussed in 6.10, “Bibliographic Citations and References,” on page 63.
This tag is optional.

<q> (i.e. quoted speech or thought)

Description: contains a representation of the speech or thought of a character or speaker in a text.

Attributes:

type: may be used to indicate whether the quoted matter is spoken or thought, or to characterize it more finely.
Data type: CDATA
Sample values include:
spoken : representation of direct speech, usually marked by quotation marks
thought : representation of thought, e.g. internal monologue
Default value: #IMPLIED
This attribute is recommended when applicable.

direct: may be used to indicate whether the quoted matter is regarded as direct or indirect speech
Data type: (y | n | unspecified)
Sample values include:
y : speech or thought is represented directly
n : speech or thought is represented indirectly e.g. by use of a marked verbal aspect
unspecified : no claim is made
Default value: unspecified
This attribute is recommended when applicable.

who: identifies the speaker of a piece of direct speech
Data type: CDATA
Value: may be an idref
Default value: #IMPLIED

Example:

```
And so, as Tiny Tim observed,
<q type=speech>God Bless Us, Every One!</q>
```

Remarks:
Part: base tag set for common core features.
Member of classes: hqinter

DTD file:
May occur within:
May contain: %phrase.seq;
Element Declaration:

```
<!ELEMENT q             - -  (%phrase.seq;)                     >
```

Attribute List Declaration:

```
<!ATTLIST q                  %a.global;
          type               CDATA               #IMPLIED
          direct             (y | n | unspecified)
                                                 unspecified
          who                CDATA               #IMPLIED       >
```

Discussed in 6.3, “Highlighting and Quotation,” on page 10.
This tag is optional.

<quote> (i.e. quotation)

Description: contains a phrase or passage attributed by the narrator or author to some agency external to the text.
Attributes: [None: global and inherited attributes only.]
Example:

```
Lexicography has shown little sign of being affected by the
work of followers of J.R. Firth, probably best summarized
in his slogan, <quote>You shall know a word by the company it
keeps</quote><ref target=FI57>(Firth, 1957)</ref>
```

Remarks: If a citation is supplied for the source of a quotation, this should be included within the **<quote>** element.
Part: base tag set for common core features.
Member of classes: hqinter
DTD file:
May occur within:
May contain: %seq;
Element Declaration:

```
<!ELEMENT quote         - -  (%seq;)                            >
```

Attribute List Declaration:

```
<!ATTLIST quote              %a.global;                         >
```

Discussed in 6.3.3, “Quotation,” on page 16.
This tag is optional.

<ref>

Description: defines a reference to another location in the current document, in terms of one or more identifiable elements, possibly modified by additional text or comment.

Attributes:

target: specifies the destination of the reference as one or more SGML identifiers
Data type: IDREFS
Value: Each value specified must be the same as that specified as value for an *ID* attribute for some other element in the current SGML document.
Default value: #IMPLIED

This attribute is optional.

Example:

```
<ref target="P2 P5" resp=LB type=s>
   See especially the second sentence</ref>
<ref>see also s.v. <term>locution</term></ref>
```

Remarks:
Part: base tag set for common core features.
Member of classes: loc
DTD file:
May occur within:
May contain: %phrase.seq
Element Declaration:

```
<!ELEMENT ref           - -  (%phrase.seq)                      >
```

Attribute List Declaration:

```
<!ATTLIST ref                %a.global;
          target             IDREFS             #IMPLIED       >
```

Discussed in 6.6, “Simple Links and Cross References,” on page 41.
This tag is optional.

Class Name: **refsys** (i.e. reference system elements)
Description: milestone-style elements used in reference systems
Attributes: [None: global and inherited attributes only.]
Remarks:
Part: base tag set for common core features.
Member of classes:
DTD file: teicore2

Discussed in 3.4, “Element Classes and Other Parameter Entities,” [in separate fascicle].

<reg> (i.e. regularization)

Description: contains a reading which has been regularized or normalized in some sense.

Attributes:

orig (i.e. original): gives the unregularized form of the text as found in the source copy.
Data type: CDATA
Value: any string of characters
Default value: #IMPLIED
Example:

```
<reg orig='auctoritee'>Authority</reg>
```

This attribute is optional.

resp (i.e. responsibility): identifies the individual responsible for the regularization of the word or phrase.
Data type: CDATA
Value: any string of characters, typically the initials of the individual involved, or a role identifier like “editor” if not known by name.
Default value: #IMPLIED
Example:

This attribute is optional.

Example:

Remarks: The **<reg>** tag is mirrored by the **<orig>** tag, which allows the unnormalized form of the original to be retained as the content of the element, while still providing the opportunity to record the regularized or normalized form postulated by a researcher. The choice between the two elements is up to the encoder.
Part: additional tag set for common core features.
Member of classes: edit
DTD file:
Data description: May contain character data and phrase-level elements.
May occur within:
May contain: %phrase.seq;
Element Declaration:

```
<!ELEMENT reg           - -  (%phrase.seq;)                     >
```

Attribute List Declaration:

```
<!ATTLIST reg                %a.global;
          orig               CDATA              #IMPLIED
          resp               CDATA              #IMPLIED       >
```

Discussed in 6.5.2, “Regularization and Normalization,” on page 35.
Discussed in 21, “Text Criticism and Apparatus,” [in separate fascicle].
This tag is optional.

<region> (i.e. region)

Description: contains the name of any regional or administrative subdivision, smaller than a country but larger than a city, occurring within an address.

Attributes: [None: global and inherited attributes only.]
Example:

```
<region n=IL>Illinois

<region>Provence
```

Remarks:
Part: base tag set for common core features.
Member of classes:
DTD file:
May occur within: address
May contain: %phrase.seq
Element Declaration:

```
<!ELEMENT region        - o  (%phrase.seq)                      >
```

Attribute List Declaration:

```
<!ATTLIST region             %a.global;                         >
```

Discussed in 6.10, “Bibliographic Citations and References,” on page 63.
This tag is optional.

<resp> (i.e. statement of responsibility)

Description: supplies a statement of responsibility for someone responsible for the intellectual content of a text, edition, recording, or series, where the **<author>**, **<editor>** and other specialized elements do not suffice.
Attributes: [None: global and inherited attributes only.]
Example:

```
<resp><role>transcribed from original ms</role>
      <name>Claus Huitfeldt</name>
</resp>

<resp><role>converted to SGML encoding</role>
      <name>Alan Morrison</name>
</resp>
```

Remarks:
Part: base tag set for common core features.
Member of classes: biblpart
DTD file:
Data description:
May occur within: analytic monogr series
May contain: %m.agent role
Element Declaration:

```
<!ELEMENT resp          - o  ((role & (%m.agent)), (role |
                             (%m.agent))*+)                    >
```

Attribute List Declaration:

```
<!ATTLIST resp               %a.global;                        >
```

Discussed in 5.2.1, “The Title Statement,” [in separate fascicle].
This tag is recommended when applicable.

<role>

Description: contains a phrase describing the nature of a person's intellectual responsibility.
Attributes: [None: global and inherited attributes only.]
Example:

```
<resp><role>compiler</role> <name>Edward Child</name></resp>
```

Remarks:
Part: auxiliary tag set for TEI headers.
Member of classes:
DTD file:
Data description:
May occur within: resp
May contain: %phrase.seq;

Element Declaration:

```
<!ELEMENT role          - o  (%phrase.seq;)                    >
```

Attribute List Declaration:

```
<!ATTLIST role              %a.global;                         >
```

Discussed in 5.2.1, “The Title Statement,” [in separate fascicle].
This tag is optional.

<s> (i.e. arbitrary segment)

Description: contains any arbitrary phrase-level unit of text.
Attributes:

type: indicates what sort of segment is marked.
Data type: CDATA
Value: any string of characters.
Default value: #IMPLIED
Remarks: The *type* attribute may be used to provide any suitable classification for the **<s>** elements tagged in a text. It should not be used to tag features for which other tags are defined in these Guidelines.
This attribute is optional.

Example:

```
<s>When are you leaving?</s>
<s>Tomorrow.</s>
```

Remarks: The **<s>** element may be used to mark orthographic sentences, or other arbitrary low-level segments of the text.
Part: additional tag set for common core features.
Member of classes: seg
DTD file:
Data description: May contain character data and phrase-level elements.
May occur within:
May contain: %phrase.seq
Element Declaration:

```
<!ELEMENT s             - -  (%phrase.seq)                     >
```

Attribute List Declaration:

```
<!ATTLIST s                  %a.global;
          type               CDATA              #IMPLIED       >
```

Discussed in 6.12, “Arbitrary Segments,” on page 87.
Discussed in 10.4.1, “Segments,” [in separate fascicle].
This tag is optional.

Class Name: seg

Description: elements for arbitrary segmentation.
Attributes: [None: global and inherited attributes only.]
Remarks:
Part: base tag set for common core features.
Member of classes: phrase
DTD file: teicore2
Discussed in 6.12, “Arbitrary Segments,” on page 87.

<series> (i.e. series information)

Description: contains information about the series in which a book or other bibliographic item has appeared.
Attributes: [None: global and inherited attributes only.]
Example:

Remarks:
Part: base tag set for common core features.
Member of classes: biblpart
DTD file:
Data description: May contain character data and phrase-level elements.
May occur within: bibl.struct
May contain: biblScope editor resp title
Element Declaration:

```
<!ELEMENT series        - O  (title | editor | resp |
                             biblScope)*                    >
```

Attribute List Declaration:

```
<!ATTLIST series             %a.global;                     >
```

Discussed in 6.10, “Bibliographic Citations and References,” on page 63.
This tag is mandatory when applicable.

<sic>

Description: contains text reproduced although apparently incorrect or inaccurate.

Attributes:

corr (i.e. correction): gives a correction for the apparent error in the copy text.
Data type: CDATA
Value: any string of characters
Default value: #IMPLIED
Example:

```
for his nose was as sharp as a pen, and
<sic corr="a' babbled">a Table</sic>
of green fields.
```

This attribute is optional.

resp (i.e. responsibility): identifies the individual responsible for the correct form given.
Data type: CDATA
Value: Sample values include “author”, “ed”, “scribe”, or the initials of the individual in question.
Default value: #IMPLIED
Example:

```
for his nose was as sharp as a pen, and
<sic corr="a' babbled" resp='Theobald'>a Table</sic>
of green fields.
```

This attribute is optional.

Example:

If all that is desired is to call attention to the apparent problem in the copy text, no attributes are required:

```
I don't know, Juan.  It's so far in the past now &mdash;
how <sic>we can</sic> prove or disprove anyone's theories?
```

It is also possible to provide a correct reading and to identify the individual responsible for the correction:

```
I don't know, Juan.  It's so far in the past now &mdash;
how <sic corr='can we' resp='MSM'>we can</sic> prove or
disprove anyone's theories?
```

Remarks: The **<sic>** tag is a mirror of **<corr>**: the former leaves the original text untouched, giving the correction as an attribute value; the latter substitutes the correction, leaving the original reading as an attribute value. The choice between them is up to the encoder.

Part: additional tag set for common core features.

Member of classes: edit
DTD file:
Data description: May contain character data and phrase-level elements.
May occur within:
May contain: %specialPara;
Element Declaration:

```
<!ELEMENT sic           - -  (%specialPara;)                    >
```

Attribute List Declaration:

```
<!ATTLIST sic               %a.global;
          corr              CDATA               #IMPLIED
          resp              CDATA               #IMPLIED        >
```

Discussed in 6.5.1, “Correction of Apparent Errors,” on page 34.
This tag is optional.

<soCalled> (i.e. so called)

Description: 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
Attributes: [None: global and inherited attributes only.]
Example:

```
To edge his way along the crowded paths of life, warning
all human sympathy to keep its distance, was what the
knowing ones call <soCalled>nuts</soCalled> to Scrooge.
```

Remarks:
Part: base tag set for common core features.
Member of classes: hqphrase
DTD file:
May occur within:
May contain: %phrase.seq;
Element Declaration:

```
<!ELEMENT soCalled      - -  (%phrase.seq;)                     >
```

Attribute List Declaration:

```
<!ATTLIST soCalled          %a.global;                          >
```

Discussed in 6.3, “Highlighting and Quotation,” on page 10.
This tag is optional.

<sp> (i.e. speech)

Description: An individual speech in a performance text, or a passage presented as such in a prose or verse text.
Attributes:

who: identifies the speaker of the part by supplying an ID.
Data type: IDREFS
Value: The IDREFS are derived from the ID attribute on the role elements in the cast list or from a list of the participants.
Default value: #IMPLIED

Example:

```
<sp><speaker>The reverend Doctor Opimiam</speaker>
<p>I do not think I have named a single unpresentable fish.
<sp><speaker>Mr Gryll</speaker>
<p>Bream, Doctor: there is not much to be said for bream.
<sp><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>Fish, Miss Gryll -- I could discourse to you on fish by the hour:
but for the present I will forbear...
</sp>
```

Remarks: The *who* attribute on this element may be used either in addition to the **<speaker>** element or as an alternative.
Part: base tag set for common core features.
Member of classes: chunk
DTD file:
Data description: Lines or paragraphs, stagedirections and phrase level elements.
May occur within:
May contain: %specialPara speaker
Element Declaration:

```
<!ELEMENT sp            - -  (speaker?, (%specialPara) )       >
```

Attribute List Declaration:

```
<!ATTLIST sp                 %a.global;
          who                IDREFS              #IMPLIED      >
```

Discussed in 6.11.2, “Drama,” on page 84.
This tag is optional.

<span>

Description: attaches simple analytic categories to an arbitrary passage of text.
Attributes:

resp (i.e. responsibility): indicates who is responsible for the annotation being performed.
Data type: CDATA
Value: initials of the researcher providing the analysis.
Default value: #IMPLIED
This attribute is recommended.

type: indicates what type of phenomenon is being noted in the passage.
Data type: CDATA
Sample values include:
image: an image in the passage is being identified.
character: the span is associated with a particular character.
theme: the **<span>** tag identifies a theme in the passage.
allusion: the **<span>** tag identifies an allusion to another text.
discourse type: the **<span>** tag identifies a passage marked by a particular discourse type.
Default value: #IMPLIED
Example:

This attribute is recommended.

value: identifies the specific phenomenon being annotated.
Data type: CDATA
Value: A value from any convenient vocabulary for the type of phenomenon involved: names for images, themes, characters, types of discourse, etc.
Default value: #IMPLIED
Remarks: Automatic processing is easier if the values for this attribute come from a controlled vocabulary.
This attribute is recommended.

target: indicates the beginning of the span.
Data type: IDREF
Value: reference to the *id* of an element which begins at the location in question (e.g. the *id* of an **<anchor>** element).
Default value: #IMPLIED
Remarks: If no value is supplied, the passage is assumed to begin at the point where the **<span>** element occurs. If the *id* value given is that of an element with content, and no value is given for *target.end*, then the passage is assumed coterminous with the element pointed at.
This attribute is recommended when applicable.

target.end: points to the end of the span being annotated.
Data type: IDREF
Value: reference to the *id* of an element which *ends* at the location in question, or to an empty element at the point in question.
Default value: #IMPLIED
Remarks: If no value is supplied, the annotation is assumed to apply to the span or single point indicated by the *target* value.
This attribute is recommended when applicable.

Example:

```
<span resp=JBS type=theme value='fire'  target=loc23 target.end=loc25>
<span resp=JBS type=theme value='water' target=loc24 target.end=loc25>
<span resp=DF  type=type-scene value='hero on the beach'
          target=loc19 target.end=loc33>
```

Remarks: The **<span>** element is intended for simple analyses of text, in which specific values (for example names of themes, characters, allusions, etc.) are associated with specific passages or spans of text. The *type* and *value* attributes allow the researcher to use whatever categories and values seem apt to the purpose. For more discursive annotation which requires less restrictive markup, the **<note>** element may be used. For analysis with more internal structure, use the feature structure mechanisms described in chapter 18, “Feature Structure Analysis,” [in separate fascicle].

Note that since the **<span>** tag points both to the beginning and to the ending of the passage it annotates, it need not be located at the passage itself; this allows the base text and annotation to be kept separate if that is desired.

Part: additional tag set for common core features.
Member of classes: notes
DTD file:
Data description: Empty.
May occur within:
May contain: EMPTY
Element Declaration:

```
<!ELEMENT span          - O  EMPTY                             >
```

Attribute List Declaration:

```
<!ATTLIST span               %a.global;
          resp               CDATA               #IMPLIED
          type               CDATA               #IMPLIED
          value              CDATA               #IMPLIED
          target             IDREF               #IMPLIED
          target.end         IDREF               #IMPLIED       >
```

Discussed in 6.8, “Notes, Annotation, and Indexing,” on page 48.
This tag is optional.

<speaker>

Description: A specialized form of heading or label, giving the name of one or more speakers in a dramatic text or fragment.
Attributes: [None: global and inherited attributes only.]
Example:

```
<sp who="NI RSA">
<speaker>Nancy and Robert</speaker>
<stage type=delivery>(speaking simultaneously)</stage>
<p>The future? ...
</sp>
```

Remarks: The *who* attribute on the **<sp>** element may be used either in addition to this element or as an alternative.
Part: base tag set for common core features.
Member of classes:
DTD file:
Data description: Any sequence of phrase level data
May occur within: sp
May contain: %phrase.seq
Element Declaration:

```
<!ELEMENT speaker       - -  (%phrase.seq)       -(speaker)    >
```

Attribute List Declaration:

```
<!ATTLIST speaker           %a.global;                         >
```

Discussed in 6.11.2, “Drama,” on page 84.
This tag is optional.

Entity Name: specialPara (i.e. special paragraph)

Description: defines the behavior of certain “special” paragraph-like elements, which can either behave like a single paragraph (containing character data and phrases) or contain a series of paragraphs and paragraph-level objects.
Remarks: This parameter entity is used to define the contents of several elements, including **<item>**, **<note>**, and **<sp>**.
Expansion: '(((%m.chunk), (%m.chunk; | %m.inter;)*) | (%paraContent;))' 3.4, “Element Classes and Other Parameter Entities,” [in separate fascicle]

<stage> (i.e. stage direction)

Description: contains any kind of stage direction within a dramatic text or fragment.
Attributes:

type: indicates the kind of stage direction.
Data type: CDATA
Suggested values include:
setting : describes a setting.

entrance : describes an entrance.
exit : describes an exit.
business : describes stage business.
novelistic : is a narrative, motivating stage direction.
delivery : describes how a character speaks.
modifier : gives some detail about a character.
location : describes a location.
mixed : more than one of the above
Default value: mix

Example:

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

Remarks:
Part: base tag set for common core features.
Member of classes: chunk
DTD file:
May occur within:
May contain: %para.seq
Element Declaration:

```
<!ELEMENT stage         - -  (%para.seq)         -(stage)      >
```

Attribute List Declaration:

```
<!ATTLIST stage              %a.global;
          type               CDATA              mix            >
```

Discussed in 6.11.2, “Drama,” on page 84.
This tag is mandatory when applicable.

<street> (i.e. street)

Description: a full street address including any name or number identifying a building as well as the name of the street or route on which it is located.
Attributes: [None: global and inherited attributes only.]
Example:

```
<street>via della Faggiola, 36</street>

<street>Duntaggin,
        110 Southmoor Road
</street>
```

Remarks: The order and presentation of house names and numbers and street names, etc., may vary considerably in different countries. The encoding should reflect the order which is appropriate in the country concerned.
Part: base tag set for common core features.
Member of classes:
DTD file:
Data description: Running prose only
May occur within: address
May contain: %phrase.seq
Element Declaration:

```
<!ELEMENT street        - o  (%phrase.seq)                    >
```

Attribute List Declaration:

```
<!ATTLIST street             %a.global;                       >
```

Discussed in 6.10, “Bibliographic Citations and References,” on page 63.
This tag is optional.

<term>

Description: contains a word or phrase regarded as a technical term.
Attributes: [None: global and inherited attributes only.]
Example:

```
A computational device that infers structure from grammatical
strings of words is known as a <term>parser</term>, and much
of the history of NLP over the last 20 years has been occupied
with the design of parsers.

We may define <term rend=sc id=tdpv>discoursal point of view</term>
as <gloss target=tdpv>the relationship, expressed through discourse
structure, between the implied author or some other addresser,
and the fiction.</gloss>
```

Remarks:
Part: base tag set for common core features.
Member of classes: hqphrase
DTD file:
Data description: free prose
May occur within:
May contain: %phrase.seq;

Element Declaration:

```
<!ELEMENT term          - -  (%phrase.seq;)                    >
```

Attribute List Declaration:

```
<!ATTLIST term              %a.global;                         >
```

Discussed in 6.3.4, “Terms, Glosses, and Cited Words,” on page 21.
This tag is recommended when applicable.

<time>

Description: contains a phrase defining a time of day in any format.
Attributes:

value: gives the value of the time in a standard form.
Data type: CDATA
Value: Any string representing a time in standard format; recommended form is “hhmm”, using the 24 hour clock.
Default value: #IMPLIED
Example:

```
<time value=1600>four o' clock</time>
```

Remarks: If the normalized form of time is not given using the twenty-four hour clock, this fact should be documented in the **<stdVals>** element in the TEI Header.
This attribute is optional.

Example:

```
As he sat smiling, the quarter struck &mdash <time value=1145>
the quarter to twelve</time>.
```

```
Given on the <time value='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.</time>

<time value='1990-09'>September 1990</time>
```

Remarks:
Part: additional tag set for common core features.
Member of classes: data
DTD file:
Data description: May contain character data and phrase-level elements.

May occur within:
May contain: %phrase.seq;
Element Declaration:

```
<!ELEMENT time          - -  (%phrase.seq;)                    >
```

Attribute List Declaration:

```
<!ATTLIST time               %a.global;
          value              CDATA              #IMPLIED       >
```

Discussed in 6.4.3, “Dates and times,” on page 28.
This tag is optional.

<timeRange> (i.e. time range)

Description: contains two times or another phrase indicating a time period.
Attributes:

from: indicates the starting point of the time period in standard form.
Data type: CDATA
Value: a time of day; recommended form is to use the 24 hour clock.
Default value: #IMPLIED
Remarks: The value should conform to the standard form declared in the **<stdVals>** element in the TEI header.
This attribute is optional.

to: indicates the ending point of the time period in standard form.
Data type: CDATA
Value: a time of day; recommended form is to use the 24 hour clock.
Default value: #IMPLIED
Remarks: The value should conform to the standard form declared in the **<stdVals>** element in the TEI header.
This attribute is optional.

exact: indicates the precision to be attached to either or both times specified.
Data type: (to|from|both|none)

to : the *to* time is exact
from : the *from* time is exact
both : both times are exact
none : both times are approximate or unspecified

Default value: #IMPLIED
Example:

```
<timeRange from=1600 to=1620 exact=from>
     Just after teatime</timeRange>
```

Remarks: time values should conform to the standard form declared in the **<stdVals>** element in the TEI header.
This attribute is optional.

Example:

Part: additional tag set for common core features.
Member of classes: data
DTD file:
Data description: May contain character data and phrase-level elements.
May occur within:
May contain: %phrase.seq;
Element Declaration:

```
<!ELEMENT timeRange     - -  (%phrase.seq;)                    >
```

Attribute List Declaration:

```
<!ATTLIST timeRange          %a.global;
          from               CDATA              #IMPLIED
          to                 CDATA              #IMPLIED
          exact              (to|from|both|none)
                                                #IMPLIED       >
```

Discussed in 6.4.3, “Dates and times,” on page 28.
This tag is optional.

<title>

Description: contains the title of a work, whether article, book, journal, or series, including any alternative titles or subtitles.
Attributes:

level (i.e. bibliographic level (or class) of title): indicates whether this is the title of an article, book, journal, series, or unpublished material.
Data type: (a | m | j | s | u)
Legal values are:
a: analytic title (article, poem, or other item published as part of a larger item)
m: monographic title (book, collection, or other item published as a distinct item, including single volumes of multi-volume works)
j: journal title
s: series title

u: title of unpublished material (including theses and dissertations unless published by a commercial press)

Default value: #IMPLIED

Example:

Remarks: If the title appears directly enclosed within an **<analytic>** element, the *level*, if given, must be “a”; if it appears directly enclosed within a **<monogr>** element, *level* must be “m”, “j”, or “u”; when **<title>** is directly enclosed by **<series>**, *level* must be “s”.

This attribute is recommended when applicable.

type (i.e. type of title): classifies the title according to some convenient typology.

Data type: CDATA

Sample values include:

main : main title

subordinate : subtitle, title of part

parallel : alternate title, often in another language, by which the work is also known

abbreviated : abbreviated form of title

Default value: #IMPLIED

Example:

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

This attribute is optional.

Example:

```
<title>La vie mode d'emploi. Romans.</title>

<title>Synthese: an international journal for epistemology, methodology
and history of science</title>

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

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

Remarks:

Part: base tag set for common core features.

Member of classes:

DTD file:

Data description:

May occur within: analytic monogr series
May contain: %soup;
Element Declaration:

```
<!ELEMENT title         - o  (%soup;)                          >
```

Attribute List Declaration:

```
<!ATTLIST title              %a.global;
          level              (a | m | j | s | u)
                                                   #IMPLIED
          type               CDATA                 #IMPLIED       >
```

Discussed in 5.2.1, “The Title Statement,” [in separate fascicle].
Discussed in 6.10, “Bibliographic Citations and References,” on page 63.
This tag is optional.

<%version> (i.e. version name)

Description: serves as the root element for a concurrent markup stream which will be used to mark page and line numbers of a reference edition of the text.
Attributes: [None: global and inherited attributes only.]
Example:

```
<!DOCTYPE TEI.2 system 'tei2.dtd' >
<!DOCTYPE La    system 'teip2.dtd' [
       <!ENTITY % version 'La'>
]>
<(TEI.2)TEI.2><(La)La>
<(TEI.2)TEI.Header> ...
</(TEI.2)TEI.Header>
<(TEI.2)text>
<(La)page n='[1]'>
  ...
<(La)page n='4'>
  ...
<(La)page n='5'>
  ...
</(TEI.2)text>
</(La)La>
</(TEI.2)TEI.2>
```

Remarks: The version name should be short (one or two characters), as it must be prefixed to all page and line tags in the concurrent markup stream.
Part: base tag set for physical description of the copy text.
Member of classes:
DTD file:
Data description: May contain character data, **<vol>**, and **<page>** elements.

May occur within:
May contain: #PCDATA page vol
Element Declaration:

```
<!ELEMENT %version      - -  (#PCDATA | page | vol)*           >
```

Attribute List Declaration:

```
<!ATTLIST %version           %a.global;                        >
```

Discussed in 6.9, “Reference Systems,” on page 52.
This tag is mandatory when applicable.

Entity Name: version (i.e. version)

Description: defines the name to be used for the root element of a concurrent markup stream for marking pages and lines of some reference edition.
Remarks: Some name for the edition should be supplied by defining this parameter entity within the appropriate DTD subset. If none is defined, the parameter entity (and thus the document type) default to “ref”.
Expansion: 'ref' 6.9.3, “Concurrent Markup for Pages and Lines,” on page 56

<vol> (i.e. volume)

Description: marks the individual volumes of a reference edition.
Attributes: [None: global and inherited attributes only.]
Example:

```
<(La)vol  n=2>  ... [text of edition La, volume 2]
<(La)page n=32> ... [text of edition La, vol. 2, page 32]
<(La)page n=32> ... [text of edition La, vol. 2, page 33]
<(La)page n=32> ... [text of edition La, vol. 2, page 34]
<(La)page n=32> ... [text of edition La, vol. 2, page 35]
```

Remarks: The **<vol>** may be used if the reference edition has more than one volume; otherwise it need not be used. Any data contained within a **<vol>** element but not within a **<page>** element is assumed not to appear in the edition from which the reference scheme derives.
Part: base tag set for physical description of the copy text.
Member of classes:
DTD file:
Data description: May contain character data and **<page>** elements.
May occur within: %version
May contain: #PCDATA page
Element Declaration:

```
<!ELEMENT vol           - -  (#PCDATA | page)*                 >
```

Attribute List Declaration:

```
<!ATTLIST vol                %a.global;                        >
```

Discussed in 6.9.3, “Concurrent Markup for Pages and Lines,” on page 56.
This tag is optional.

Usability Report and Comment

If you have comments of any kind on the current draft of the TEI Guidelines, please complete this form and return it to one of the following addresses:

C. M. Sperberg-McQueen, Computer Center (M/C 135), University of Illinois at Chicago, 1940 W. Taylor Room 124, Chicago IL 60612-7352, USA. Bitnet: U35395@UICVM, Internet: U35395@uicvm.uic.edu

Lou Burnard, Oxford University Computing Service, 13 Banbury Road, Oxford OX2 6NN, UK. Fax: +44 (865) 273275, Janet: LOU@UK.AC.OXFORD.VAX Internet: LOU@vax.oxford.ac.uk

Further copies of this form may be downloaded from LISTSERV @ UICVM (or listserv @ uicvm.uic.edu). Send a mail message containing the line 'GET TEIURC MEMO TEI-L'

Your name:
Your postal address:

Your e-mail address (if any):
Your occupation:
Your academic background:
Your immediate reactions to the Guidelines (please tick or cross one only):

Relevance to your present concerns or interests: high medium low
Importance to your future research plans: high medium low
Comprehensibility/usability of current draft: high medium low

Detailed technical comment: [Please feel free to comment on any part of the Guidelines or in general. Comments relating to specific parts of the text are most helpful if they include a section or chapter number]

Thank you!

Temporary Index