6 Elements Available in All TEI Documents

This chapter describes elements which may appear in any kind of text and the tags used to mark them in all TEI documents. Most of these elements are freely floating phrases, which can appear at any point within the textual structure, although they must generally 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.

This chapter begins by describing the <p> tag used to mark paragraphs, which serve as the fundamental formal unit for running text in many base tag sets, and are available in all. This is followed, in section 6.2 *Treatment of Punctuation*, 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*) 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*) and features commonly marked by quotation marks (section 6.3.3 *Quotation*) as well as such features as terms, cited words, and glosses (section 6.3.4 *Terms, Glosses, and Cited Words*).

The next section (section 6.4 *Names, Numbers, Dates, Abbreviations, and Addresses*) 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 *Referring Strings*), numbers and measures (section 6.4.3 *Numbers and Measures*), dates and times (section 6.4.4 *Dates and Times*), abbreviations (section 6.4.5 *Abbreviations and Their Expansions*), and addresses (section 6.4.2 *Addresses*).

Section 6.5 *Simple Editorial Changes* 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 mechanisms for encoding such information (described in full in chapter 18 *Transcription of Primary Sources*), which should be adequate to most commonly encountered situations.

In the same way, the following section (section 6.6 *Simple Links and Cross References*) presents only a subset of the facilities available for the encoding of cross-references or text-linkage. The full story may be found in chapter 14 *Linking, Segmentation, and Alignment*; the tags presented here are intended to be usable for a wide variety of simple applications.

Sections 6.7 *Lists*, and 6.8 *Notes, Annotation, and Indexing*, 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 chapter 15 *Simple Analytic Mechanisms*) is discussed.

Next, section 6.9 *Reference Systems*, describes methods of encoding within a text the conventional system or systems used when making references to the text. 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*, 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.

Additional elements for the encoding of passages of verse or drama (whether prose or verse) are discussed in section 6.11 *Passages of Verse or Drama*.

The chapter concludes with a technical overview of the structure and organization of the tag set described here. This should be read in conjunction with chapter 3 *Structure of the TEI Document Type Definition*, describing the structure of the TEI document type definition.

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 8 *Base Tag Set for Prose*); 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, and other elements which have similar structural properties, notably the <ab> (anonymous block) element described in 14.3 *Blocks, Segments and Anchors*) which may be used as an alternative to the paragraph in some kinds of texts.

Because paragraphs may appear in different base or additional tag sets, their possible contents may differ in different kinds of documents. 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> or <seg> ('segment') elements may be used, as discussed in chapters 14 *Linking, Segmentation, and Alignment* and 15 *Simple Analytic Mechanisms* respectively. 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.</p>
</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>
<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>
<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>
<p>...</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*) 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>
<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>
<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>
</p>
```

The <p> element is formally declared as follows:

```
<!-- 6.1: Paragraph-->
<!ELEMENT p %om.RO;  %paraContent;>
<!ATTLIST p
      %a.global;
      TEIform CDATA 'p'  >
<!-- end of 6.1-->
```

6.2 Treatment of 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 *Languages and Character Sets*). 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 14.3 *Blocks, Segments and Anchors*, 6.4.5 *Abbreviations and Their Expansions*, and 6.4.3 *Numbers and Measures*. There are independent reasons for tagging these, whether or not they are marked by full stops. Alternatively, entity names like the following might 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

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 either eliminate non-significant line-end hyphens or replace them by a reference to an appropriate character entity.[70] Whichever method is

[70] shy ('soft hyphen') is defined in the standard public entity set ISOnum; Unicode reserves code point 2010 for the hyphen, and 2011 for the 'non-breaking' hyphen.

adopted, it should be reported using the <hyphenation> element within the encoding declarations in the TEI header. See chapter 5 *The TEI Header* for discussion of the TEI header and encoding declarations.

When creating a machine-readable text from scratch, it is best not to introduce hyphenation simply to make lines of a predefined length, since one cannot then easily tell whether the hyphens are soft or hard. When compounds or prefixed words are hyphenated in mid-sentence, it may be impossible to tell whether the hyphenation is due to formatting or to linguistic concerns.

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). Alternatively, in a standalone XML context, these entities may be represented as Unicode characters —, –, or ‐ respectively. 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*, and section 6.7 *Lists*, 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*.

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.

Parentheses and other marks of suspension such as dashes or ellipses are often used to signal information about the syntactic structure of a text fragment. Full disambiguation of their uses also belongs to the level of linguistic analysis and interpretation, and is therefore discussed in chapter 15 *Simple Analytic Mechanisms*.

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 using 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.[71] The purpose of highlighting is generally to 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 regarded as being one or more of the following:

- 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.

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:

- 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 it is 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*) and <name> (discussed in section 6.4.1 *Referring Strings*).

[71] 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 sense, these Guidelines recommend distinct elements for the encoding of such 'highlighting' in spoken texts. See further section 11.2.6 *Shifts*.

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.

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

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

<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
Values a semi-open user-defined list

time specifies how the phrase is distinct diachronically
Values a semi-open user-defined list

space specifies how the phrase is distinct diatopically
Values a semi-open user-defined list

social specifies how the phrase is distinct diastatically
Values a semi-open user-defined list

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 4.3 *Code shifting* 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* 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 %om.RR; %paraContent;>
<!ATTLIST foreign
      %a.global;
      TEIform CDATA 'foreign'  >
<!--continued in 6.3.2.1: -->
<!--continued in 6.3.2.1: -->
<!--continued in 6.3.2.1: Quotation-->
<!--continued in 6.3.2.1: Terms, glosses, etc.-->
<!-- end of 6.3.2.1-->
```

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>
<l>Doth sometimes Counsel take &mdash;
   and sometimes <emph rend="italic">Tea</emph>.</l>
```

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.[72] In a second pass, the element <soCalled> might be preferred.

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

```
<!-- 6.3.2.2: -->
<!ELEMENT emph %om.RR;  %paraContent;>
<!ATTLIST emph
      %a.global;
      TEIform CDATA 'emph'  >
<!ELEMENT hi %om.RR; %paraContent;>
<!ATTLIST hi
      %a.global;
      TEIform CDATA 'hi'  >
<!-- end of 6.3.2.2-->
```

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

[72] 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.

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.[73] The time attribute places a word *diachronically*, for example as archaic, old-fashioned, contemporary, futuristic, etc.; the space 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*).

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" space="GB" social="publicschool">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 chapter 16 *Feature Structures* 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 element <note> described in section 6.8 *Notes, Annotation, and Indexing*, or the element described in section 15.3 *Spans and Interpretations*.

The <distinct> element has the following formal definition:

```
<!-- 6.3.2.3: -->
<!ELEMENT distinct %om.RR; %phrase.seq;>
<!ATTLIST distinct
      %a.global;
      type CDATA #IMPLIED
      time CDATA #IMPLIED
      space CDATA #IMPLIED
      social CDATA #IMPLIED
      TEIform CDATA 'distinct'  >
<!-- end of 6.3.2.3-->
```

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 quotation or apparent quotation — a representation of speech or thought marked as being quoted from someone else (whether in fact quoted or not); in narrative, the words are usually those of of a character or speaker; in dictionaries, <q> may be used to mark real or contrived examples of usage. Attributes include:

who identifies the speaker of a piece of direct speech.
Values may be an idref

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.

[73] 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.

`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.

Legal values are:

`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.

<cit> A quotation from some other document, together with a bibliographic reference to its source.

<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.

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 37 *Obtaining TEI WSDs*; 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 *Treatment of Punctuation*.

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 the ID/IDREF mechanism. See section 23.2.2 *The Participants Description* 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&hellip;</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:-<q>
    <quote>
      <l>To Him the Way &mdash; the Law &mdash; Apart &mdash;</l>
      <l>Whom Maya held beneath her heart</l>
      <l>Ananda's Lord &mdash; the Bodhisat</l>
    </quote>
  And He is here! The Most Excellent Law is here also. My
  pilgrimage is well begun. And what work! What work!</q>
</p>
```

Quotations from other works are often accompanied by a reference to their source. The <cit> element may be used to group together the quotation and its associated bibliographic reference, which should be encoded using the elements for bibliographic references discussed in section 6.10 *Bibliographic Citations and References*, as in the following example.

```
<div id="mm01" type="chapter">
<head>Chapter 1</head>
<epigraph><cit>
   <quote>
      <l>Since I can do no good because a woman</l>
      <l>Reach constantly at something that is near it.</l>
   </quote>
   <bibl>
      <title>The Maid's Tragedy</title>
      <author>Beaumont and Fletcher</author>
   </bibl>
</cit></epigraph>
<p>Miss Brooke had that kind of beauty which seems to be thrown into
relief by poor dress...</p>
</div>
```

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, <cit>
    <quote>You shall know a word by the company it keeps.</quote>
    <ref target="fi57">(Firth, 1957)</ref>
</cit>
```

Unlike most of the other elements discussed in this chapter, direct speech and 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

For further discussion, and several examples, see chapter 31 *Multiple Hierarchies*.

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: Quotation-->
<!ELEMENT q %om.RR; %specialPara;>
<!ATTLIST q
      %a.global;
      type CDATA #IMPLIED
      direct (y | n | unspecified) "unspecified"
      who CDATA #IMPLIED
      TEIform CDATA 'q'  >
<!ELEMENT quote %om.RR; %specialPara;>
<!ATTLIST quote
      %a.global;
      TEIform CDATA 'quote'  >
<!ELEMENT cit %om.RR; ( (q | quote | %m.bibl; | %m.loc; | %m.Incl; )+)>
<!ATTLIST cit
      %a.global;
      TEIform CDATA 'cit'  >
<!ELEMENT soCalled %om.RR; %phrase.seq;>
<!ATTLIST soCalled
      %a.global;
      TEIform CDATA 'soCalled'  >
<!-- end of 6.3.3-->
```

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 single-word, multi-word, or symbolic designation which is regarded as a technical term. Attributes include:

type classifies the term using some typology.

Values any string of characters; for serious terminological work, values should be taken from the dictionary of data element types specified in ISO WD 12 620.

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

target identifies the associated term element

Values must be a valid identifier for some <term> element in the current document

<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 the 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 14 *Linking, Segmentation, and Alignment*.

Examples:

```
We may define <term id="tdpv" rend="sc">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 rend="unmarked" target="t1">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 id="cw234" lang="grc">eluthemen</mentioned>
<gloss target="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
target="cw235">released,</gloss> accented on the last.
```

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

```
<!-- 6.3.4: Terms, glosses, etc.-->
<!ELEMENT term %om.RR; %phrase.seq;>
<!ATTLIST term
      %a.global;
      type CDATA #IMPLIED
      TEIform CDATA 'term'  >
<!ELEMENT mentioned %om.RR;  %phrase.seq;>
<!ATTLIST mentioned
      %a.global;
      TEIform CDATA 'mentioned'  >
<!ELEMENT gloss %om.RR;  %phrase.seq;>
<!ATTLIST gloss
      %a.global;
      target IDREF #IMPLIED
      TEIform CDATA 'gloss'  >
<!-- end of 6.3.4-->
```

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 lang="fr" rend="italic">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 <emph>vehement</emph>
debatings. She says I am <q rend="italic">too witty</q>;
<foreign lang="la" rend="roman">Anglic&egrave;</foreign>,
<gloss rend="italic">too pert</gloss>; I, that she is
<q 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; di[stinguishing such items from the surrounding text is however equally important to the scholar primarily interested in lexis.

The treatment of these textual features proposed here is not intended to be exhaustive: fuller treatments for names, numbers, measures, and dates are provided in the additional tag set for names and dates (see chapter 20 *Names and Dates*).

6.4.1 Referring Strings

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

<rs> contains a general purpose name or referring string. Attributes include:

- **type** indicates more specifically the object referred to by the referencing string. Values might include “person”, “place”, “ship”, “element” etc.
  Values Any string of characters.

<name> contains a proper noun or noun phrase. Attributes include:

- **type** indicates the type of the object which is being named by the phrase.
  Values Values such as person, place, institution, product, acronym.

Where it is thought useful to do so, the kind of object referred to may be specified using the type attribute.

Examples include:

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

```
Collectors of water-rents were appointed by the
<rs type="organization">Watering Committee</rs>.
They were paid a commission not exceeding four per
cent, and gave bond.
```

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

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

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

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

string contains a mixture of common and proper nouns. The following example shows an alternative way of encoding the short sentence from *Pride and Prejudice* quoted above:

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

The following example shows how a proper name may be nested within a referring string:

```
<rs>His Excellency the Life President, <name>Ngwazi Dr H. Kamuzu Banda</name></rs>
```

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.

The following attributes, common to all members of the names element class, are provided to help overcome these difficulties:

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 name used.

Either or both of these attributes may be specified, as appropriate. The key attribute may be useful as a means of gathering together all references to the same individual or location scattered throughout a document:

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

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

```
My personal life during the administration of
<rs key="POJA1" reg="Polk, James K." type="person">Col. Polk</rs>
has but poorly compensated me for the suspended
enjoyments and pursuits of private and professional spheres

<name key="VOM1" reg="Volanges, Mme de" type="person">Mme. de Volanges</name>
marie sa fille: c'est encore un secret; mais elle m'en a fait part hier.

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

<name type="place">Montaillou</name> is not a large parish.
At the time of the events which led to
<name reg="Benedict XII, Pope of Avignon (Jacques Fournier)" type="person">Fournier's</name>
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 20 *Names and Dates* or the analytical tools described in chapter 16 *Feature Structures* should be used.

These elements are formally declared as follows:

```
<!-- 6.4.1: Proper Nouns-->
<!ELEMENT name %om.RR;  %phrase.seq;>
<!ATTLIST name
      %a.global;
      %a.names;
      type CDATA #IMPLIED
      TEIform CDATA 'name'  >
<!ELEMENT rs %om.RR; %phrase.seq;>
<!ATTLIST rs
      %a.global;
      %a.names;
      type CDATA #IMPLIED
      TEIform CDATA 'rs'  >
<!-- end of 6.4.1-->
```

6.4.2 Addresses

The simplest way of encoding an address is to regard it as a series of distinct lines, just as they might be printed on an envelope. The following elements support this view:

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

<addrLine> contains one line of a postal or other address.

Alternatively, an address may be encoded as a structure composed of the following elements, which constitute the addrPart element class:

<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.

<name> contains a proper noun or noun phrase. Attributes include:

type indicates the type of the object which is being named by the phrase.

Values Values such as person, place, institution, product, acronym.

<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.

Any number of elements from the addrPart class may appear within an address and 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. Other components of addresses should be represented using the general-purpose <name> element.

Some examples follow:

```
<address>
   <addrLine>110 Southmoor Road,</addrLine>
   <addrLine>Oxford OX2 6RB,</addrLine>
   <addrLine>UK</addrLine>
</address>
```

The above address could also be represented as follows:

```
<address>
   <street>110 Southmoor Road</street>
   <name type="city">Oxford</name>
   <postCode>OX2 6RB</postCode>
   <name type="country">United Kingdom</name>
</address>
```

The order of elements within an address is highly culture-specific, and is therefore unconstrained:

```
<address>
   <name type="org">Universit&agrave; di Bologna</name>
   <name type="country">Italy</name>
   <postCode>40126</postCode>
   <name type="city">Bologna</name>
   <street>via Marsala 24</street>
</address>
```

For further discussion of ways of regularizing the names of places, see section 6.4 *Names, Numbers, Dates, Abbreviations, and Addresses*. A full postal address may also include the name of the addressee, tagged as above using the general purpose <name> element. When the additional tag set for names and dates is enabled, more specific elements such as <publisher> or <org> may be used, as further discussed in chapter 20 *Names and Dates*.

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

```
<!-- 6.4.2: Addresses and their components-->
<!ELEMENT address %om.RO; ( (%m.Incl;)*,
                    ( (addrLine, (%m.Incl;)*)+ | ((%m.addrPart;), (%m.Incl;)*)* )
) >
```

```
<!ATTLIST address
      %a.global;
      TEIform CDATA 'address'  >
<!ELEMENT addrLine %om.RO;  %phrase.seq;>
<!ATTLIST addrLine
      %a.global;
      TEIform CDATA 'addrLine'  >
<!ELEMENT street %om.RO; %phrase.seq;>
<!ATTLIST street
      %a.global;
      TEIform CDATA 'street'  >
<!ELEMENT postCode %om.RO; (#PCDATA)>
<!ATTLIST postCode
      %a.global;
      TEIform CDATA 'postCode'  >
<!ELEMENT postBox %om.RO; (#PCDATA)>
<!ATTLIST postBox
      %a.global;
      TEIform CDATA 'postBox'  >
<!--Other components of addresses should be represented
using the general purpose NAME element-->
<!-- end of 6.4.2-->
```

6.4.3 Numbers and Measures

This section describes two elements provided for the simple encoding of numbers and measures and gives some indication of circumstances in which this may usefully be done. The following phrase level elements are provided for this purpose:

<num> contains a number, written in any form. Attributes include:

type indicates the type of numeric value.

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

value supplies the value of the number in an application-dependent standard form.

Values any numeric value in the chosen standard form.

<measure> contains a word or phrase referring to some quantity of an object or commodity, usually comprising a number, a unit, and a commodity name. Attributes include:

type specifies the type of unit in which the measure is expressed.

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.

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’).

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>
```

The word ‘measure’ is used here to refer to a special kind of referring string, the referent of which is a ‘virtual object’. In its fullest form, a measure consists of a number, a phrase expressing units of measure and a phrase expressing the commodity being measured. Not all of these components need be present in every case. For some applications, particularly quantitative ones, the internal components of measure need to be marked so that their values can be calculated. Thus, in order to evaluate a monetary measure according to some standard, it is necessary to mark its currency unit (e.g. US dollars, pounds sterling). Similarly, the expression ‘2 ounces’ will have a different meaning when it is associated with ‘flour’ from that which it has when associated with ‘water’.

Such applications will require the elements discussed in chapter 20 *Names and Dates*, or the more powerful analytical tools discussed in chapter 16 *Feature Structures*. Elsewhere, it may be sufficient simply to encode measures as such, perhaps also indicating their numeric content with the <num> element, as in the following examples:

```
<l>I've measured it from side to side</l>
<l>'Tis
<measure reg="0.924 m" type="length">
<num value="3">three</num> feet</measure>
long, and
<measure reg="0.616 m" type="length">
<num value="2">two</num> feet</measure>
wide.</l>
```

As the above example also demonstrates, the <measure> element is a member of the class names like other referencing strings, and may thus bear a reg attribute to indicate a normalized value. The form of normalization used should conform to a defined standard such as the International System of Units (SI). The <measure> element may also carry a key attribute to indicate a database key value, as in the following example:

```
<list>
  <item><measure key="BH2" type="volume">
        <num value="2">ii</num> bags hops
        </measure>
  </item>
  <item><measure key="TW6" type="volume">
        <num value="6">six</num> trusses Woolen and linen goods
        </measure>
  </item>
  <item><measure key="WC5" type="weight">
        5 tonnes coale
       </measure>
  </item>
  <!-- ... -->
</list>
```

These elements are formally defined as follows:

```
<!-- 6.4.3: Numbers and measures-->
<!ELEMENT num %om.RR;  %phrase.seq;>
<!ATTLIST num
      %a.global;
      type CDATA #IMPLIED
      value CDATA #IMPLIED
      TEIform CDATA 'num'  >
<!ELEMENT measure %om.RR;  %phrase.seq;>
<!ATTLIST measure
      %a.global;
```

```
      %a.names;
      type CDATA #IMPLIED
      TEIform CDATA 'measure'  >
<!-- end of 6.4.3-->
```

6.4.4 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.

Values Recommended values include: *Gregorian*, *Julian*, *Roman*, *Mosaic*, *Revolutionary*, *Islamic*.

value gives the value of the date in some standard form, usually yyyy-mm-dd.

Values Any string representing a date in standard format; recommended form is ISO 8601:2000 5.2.1.1 Complete representation, extended format (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 some standard form, usually hh:mm.

Values Any string representing a time in standard format; recommended forms are the extended formats from ISO 8601:2000 (hh:mm, hh:mmZ, hh:mm±hh)

<dateRange> contains two dates or another phrase delimiting a time period. Attributes include:

calendar indicates the system or calendar to which the date belongs.

Values Recommended values include: *Gregorian*, *Julian*, *Roman*, *Mosaic*, *Revolutionary*, *Islamic*.

from indicates the starting point of the period in standard form.

Values any date in a standard form; recommended form is yyyy-mm-dd.

to indicates the ending point of the period in standard form.

Values any date in a standard form; recommended form is yyyy-mm-dd.

exact indicates the precision to be attached to either or both dates specified.

Legal values are:

`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 a standard form, usually hh:mm.

Values a string representing a time in standard format; recommended forms are the extended formats from ISO 8601.

to indicates the ending point of the time period in standard form, usually hh:mm.

Values a string representing a time in standard format; recommended forms are the extended formats from ISO 8601.

exact indicates the precision to be attached to either or both times specified.

Legal values are:

`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 chapter 17 *Certainty and Responsibility*.

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 20 *Names and Dates* should be used in preference to the simple method outlined here.

The value attribute is a useful way of normalizing or disambiguating dates and times which can appear in many formats, as the following examples show:

```
<date value="1980-02-12">12/2/1980</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="2001">2001</date>
<date value="2001-09">September 2001</date>
<date value="2001-09-11">11 Sept 01</date>
<date value="2001-09-11">9/11</date>, <time value="08:48">8:48</time>
<date value="2001-09-11T12:48Z">Sept 11th, 12 minutes before 9 am</date>
```

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

The following examples demonstrate the use of the <dateRange> element to mark a period of time:

```
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:

```
<!-- 6.4.4: Dates and times-->
<!ELEMENT date %om.RR;  %phrase.seq;>
<!ATTLIST date
      %a.global;
      calendar CDATA #IMPLIED
      value CDATA #IMPLIED
      certainty CDATA #IMPLIED
      TEIform CDATA 'date'  >
<!ELEMENT dateRange %om.RO;  %phrase.seq;>
<!ATTLIST dateRange
      %a.global;
      calendar CDATA #IMPLIED
      from CDATA #IMPLIED
      to CDATA #IMPLIED
      exact (to|from|both|none) #IMPLIED
      TEIform CDATA 'dateRange'  >
<!ELEMENT time %om.RR;  %phrase.seq;>
<!ATTLIST time
      %a.global;
      value CDATA #IMPLIED
      type (am | pm | 24hour | descriptive) #IMPLIED
      zone CDATA #IMPLIED
      TEIform CDATA 'time'  >
```

```
<!ELEMENT timeRange %om.RR;  %phrase.seq;>
<!ATTLIST timeRange
      %a.global;
      from CDATA #IMPLIED
      to CDATA #IMPLIED
      exact (to|from|both|none) #IMPLIED
      TEIform CDATA 'timeRange'  >
<!-- end of 6.4.4-->
```

6.4.5 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 (expansion) gives an expansion of the abbreviation.
Values any string of characters

resp (responsibility) signifies the editor or transcriber responsible for supplying the expansion of the abbreviation held as the value of the expan attribute.
Values must be one of the identifiers declared in the document header, associated with a person asserted as responsible for some aspect of the text's creation, transcription, editing, or encoding (see chapter 17 *Certainty and Responsibility*).

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.

cert (certainty) signifies the degree of certainty ascribed to the expansion of the abbreviation.

<expan> contains the expansion of an abbreviation. Attributes include:

abbr (abbreviation) gives the abbreviation in its unexpanded form.
Values any string of characters

resp (responsibility) signifies the editor or transcriber responsible for supplying the expansion of the abbreviation held as the content of the <expan> element.
Values must be one of the identifiers declared in the document header, associated with a person asserted as responsible for some aspect of the text's creation, transcription, editing, or encoding (see chapter 17 *Certainty and Responsibility*).

type allows the encoder to classify the abbreviation according to some convenient typology.
Values any useful classification name, e.g. 'suspension', 'contraction', 'brevigraph', 'title', 'organization', 'geographic', etc.

cert (certainty) signifies the degree of certainty ascribed to the expansion of the abbreviation.

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

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

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

The type attribute may be used to distinguish types of abbreviation by their function, 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 expan="per" resp="pg" type="brevigraph">&per;</abbr>ience,
thogh noon auctoritee</l>
<l>Were in this world, is right ynogh for me</l>
<l>To speke of wo that is in mariage;</l>
```

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

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. (For further discussion of such documentation, see section 25.4.3 *Documenting Coded Character Sets and Entity Sets*.) The text shown here:

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

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' abbr='&p.crossed;' Resp='PG'>per</expan>">
```

For further discussion of manuscript abbreviations, see chapter 18 *Transcription of Primary Sources*.

These elements are formally defined as follows:

```
<!-- 6.4.5: Abbreviations-->
<!ELEMENT abbr %om.RR;  %phrase.seq;>
<!ATTLIST abbr
      %a.global;
      expan CDATA #IMPLIED
      resp IDREF %INHERITED;
      cert CDATA #IMPLIED
      type CDATA #IMPLIED
      TEIform CDATA 'abbr'  >
<!ELEMENT expan %om.RR; %phrase.seq;>
<!ATTLIST expan
      %a.global;
      abbr CDATA #IMPLIED
      resp IDREF %INHERITED;
      cert CDATA #IMPLIED
      type CDATA #IMPLIED
      TEIform CDATA 'expan'  >
<!-- end of 6.4.5-->
```

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*. Systematic interpretive annotation is also possible using the various methods described in chapter 14 *Linking, Segmentation, and Alignment*. 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 14 *Linking, Segmentation, and Alignment* and 19 *Critical Apparatus* 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 6.4.5 *Abbreviations and Their Expansions*. 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

A more extended treatment of the use of these tags in transcriptional and editorial work is given in chapter 18 *Transcription of Primary Sources*.

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** (correction) gives a correction for the apparent error in the copy text.
  Values any string of characters
- **resp** (responsibility) signifies the editor or transcriber responsible for suggesting the correction held as the value of the corr attribute.
  Values must be one of the identifiers declared in the document header, associated with a person asserted as responsible for some aspect of the text's creation, transcription, editing, or encoding (see chapter 17 *Certainty and Responsibility*).
- **cert** (certainty) signifies the degree of certainty ascribed to the correction held as the value of the corr attribute.

<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.
  Values any string of characters
- **resp** (responsibility) signifies the editor or transcriber responsible for suggesting the correction held as the content of the <corr> element.
  Values must be one of the identifiers declared in the document header, associated with a person asserted as responsible for some aspect of the text's creation, transcription, editing, or encoding (see chapter 17 *Certainty and Responsibility*).

cert (certainty) signifies the degree of certainty ascribed to the correction held as the content of the <corr> element.

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 tags, it will prove more convenient to use <sic> than <corr> (and vice versa if there are tags within the corrected reading), since tags are not recognized in attribute values. If both readings contain subordinate tags, then recourse must be had to the methods described in chapter 19 *Critical Apparatus*.

The cert attribute on the <sic> and <corr> elements permits a statement of the degree of editorial confidence in a particular correction. For example, using a confidence scale of one to ten, an editor may indicate the conjectural status of a correction by assigning a value to this attribute of less than ten. In the following instance, some uncertainty is expressed concerning a commonly-accepted emendation:

```
An <corr sic="Antony" cert="8">Autumn</corr> it was,
That grew the more by reaping
```

See further the discussion in section 18.1.3 *Correction and Conjecture*.

Where the correction takes the form of adding text, the encoder must choose whether to use the <corr> (or <sic>) tag, the <add> tag (see section 6.5.3 *Additions, Deletions, and Omissions* below), or the more detailed facilities provided by the additional tag set for primary source description. The following discussion may be helpful when making this decision:

- where the correction is an addition by a scribe or author in a manuscript or other primary source (typescript, proof or galley, etc.) then either <corr> (or <sic>) or <add> might be appropriate, depending on the circumstances. The <add> tag is more expressive, and may convey information on just how the addition was performed (hand, place, etc.) which the <corr> tag cannot. See further the discussion in section 18.1.5 *Substitutions*.
- where the correction is an addition by a transcriber or editor, correcting a perceived deficiency in the text but in circumstances where there is no clearly assertable reason for the deficiency (as a manuscript lacuna, or damage to the page) the <corr> tag should be used. The <add> tag should not be used in this case.
- where the correction is an addition by a transcriber or editor, correcting a perceived deficiency in the text and where there is a clearly assertable reason for the deficiency (as a manuscript lacuna, or damage to the page) which the encoder wishes to record, or when supplying text from a parallel version of the text, the <supplied> element provided by the additional tag set for primary source description should be used (see section 18.1.3 *Correction and Conjecture*).

The formal definition of these elements is as follows:

```
<!-- 6.5.1: Editorial tags for correction-->
<!ELEMENT sic %om.RR;  %specialPara;>
<!ATTLIST sic
      %a.global;
      corr CDATA #IMPLIED
      resp CDATA %INHERITED;
      cert CDATA #IMPLIED
      TEIform CDATA 'sic'  >
<!ELEMENT corr %om.RR;  %specialPara;>
<!ATTLIST corr
      %a.global;
      sic CDATA #IMPLIED
      resp CDATA %INHERITED;
      cert CDATA #IMPLIED
      TEIform CDATA 'corr'  >
<!-- end of 6.5.1-->
```

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.[74]

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** (original) gives the unregularized form of the text as found in the source copy.
  Values any string of characters
- **resp** (responsibility) identifies the individual responsible for the regularization of the word or phrase.
  Values any string of characters, typically the initials of the individual involved, or a role identifier like 'editor' if not known by name.

<orig> contains the original form of a reading, for which a regularized form is given in an attribute value. Attributes include:

- **reg** (regularization) gives a regularized (normalized) form of the text.
  Values any string of characters
- **resp** (responsibility) identifies the individual responsible for the regularization of the word or phrase.
  Values any string of characters, typically the initials of the individual involved, or a role identifier like 'editor' if not known by name.

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.
```

[74] 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.

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.
```

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 þik dauðra 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" resp="ed">dura</orig>
```

These elements are formally defined as follows:

```
<!-- 6.5.2: Editorial tags for regularization-->
<!ELEMENT reg %om.RR;  %phrase.seq;>
<!ATTLIST reg
      %a.global;
      orig CDATA #IMPLIED
      resp CDATA #IMPLIED
      TEIform CDATA 'reg'  >
<!ELEMENT orig %om.RR;  %phrase.seq;>
<!ATTLIST orig
      %a.global;
      reg CDATA #IMPLIED
      resp CDATA #IMPLIED
      TEIform CDATA 'orig'  >
<!-- end of 6.5.2-->
```

6.5.3 Additions, Deletions, and Omissions

The following elements are used to indicate when words or phrases have 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:

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

desc (description) gives a description of the omitted text.
Values a prose description of the material omitted.

reason gives the reason for omission. Sample values include ‘sampling’, ‘illegible’, ‘inaudible’, ‘irrelevant’, ‘cancelled’, ‘cancelled and illegible’.
Values any short indication of the reason for the omission.

extent indicates approximately how much text has been omitted from the transcription, in letters, minims, inches, or any appropriate unit, either because of editorial policy or because a deletion, damage, or other cause has rendered transcription impossible.
Values any string of characters

resp (responsibility) indicates the editor, transcriber or encoder responsible for the decision not to provide any transcription of the text and hence the application of the `<gap>` tag.
Values must be one of the identifiers declared in the document header, associated with a person asserted as responsible for some aspect of the text's creation, transcription, editing, or encoding (see chapter 17 *Certainty and Responsibility*).

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

reason indicates why the material is hard to transcribe.
Values any phrase describing the difficulty, e.g. 'faded', 'ambient noise', 'passing truck', 'ill formed', 'eccentric ductus'.

resp indicates the individual responsible for the transcription of the word, phrase, or passage contained with the `<unclear>` element.
Values must be one of the identifiers declared in the document header, associated with a person asserted as responsible for some aspect of the text's creation, transcription, editing, or encoding (see chapter 17 *Certainty and Responsibility*).

<add> contains letters, words, or phrases inserted in the text by an author, scribe, annotator, or corrector. Attributes include:

place if the the addition is written into the copy text, indicates where the additional text is written.
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

resp (responsible) signifies the editor or transcriber responsible for identifying the hand of the addition.
Values must be one of the identifiers declared in the document header, associated with a person asserted as responsible for some aspect of the text's creation, transcription, editing, or encoding (see chapter 17 *Certainty and Responsibility*).

hand signifies the hand of the agent which made the addition.
Values must be one of the hand identifiers declared in the document header (see section 18.2.1 *Document Hands*).

cert (certainty) signifies the degree of certainty ascribed to the identification of the hand of the addition.

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

type classifies the type of deletion using any convenient typology.
Values any string identifying the class of deletion.

status may be used to indicate faulty deletions, e.g. strikeouts which include too much or too little text.

Values any description of flaws in the marking of a deletion, e.g. 'excess left', 'excess right', 'short left', 'short right'.

resp (responsible) signifies the editor or transcriber responsible for identifying the hand of the deletion.

Values must be one of the identifiers declared in the document header, associated with a person asserted as responsible for some aspect of the text's creation, transcription, editing, or encoding (see chapter 17 *Certainty and Responsibility*).

hand signifies the hand of the agent which made the deletion.

Values must be one of the hand identifiers declared in the document header (see section 18.2.1 *Document Hands*).

cert (certainty) signifies the degree of certainty ascribed to the identification of the hand of 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*). Each place in the text at which omission has taken place should be marked with a <gap> 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:

```
<gap desc="Prose commentary" reason="sampling" resp="pr" extent="120 lines"/>

... Their arrangement with respect to Jupiter and to each other was as follows:
<gap 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; ...

<gap desc="ink blot" reason="illegible" extent="two words"/>

<gap reason="overwriting, illegible" resp="h1" 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 elements; for these, the elements <addSpan> and <delSpan>, or other mechanisms described in section 18 *Transcription of Primary Sources* 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 brackets 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 should not be used to mark editorial changes, such as supplying a word omitted by mistake from the source text or a passage present in another version. In these cases, either the <corr> or <supplied> tags should be used, as discussed above in section 6.5.1 *Correction of Apparent Errors*, and in section 18.1.3 *Correction and Conjecture*, respectively.

The <unclear> element is used to mark passages in the original which cannot be read with confidence, or about which the transcriber is uncertain for other reasons, as for example when transcribing a partially inaudible or illegible source. Its reason and resp attributes are used, as with the <gap> element, to indicate the cause of uncertainty and the person responsible for the conjectured reading.

For example:

```
<l>And where the sandy mountain Fenwick scald</l>
<l><unclear reason="ink blot" resp="LB">The</unclear> sea between
yet hence his pray'r prevail'd</l>
```

or from a spoken text:

```
and then <unclear reason="passing truck">marbled queen</unclear>
```

Where the material affected is entirely illegible or inaudible, the <gap> element discussed above should be used in preference.

The <del> element is used to mark material which is deleted in the source but which can still be read with some degree of confidence, as opposed to material which has been omitted by the encoder or transcriber either because it is entirely illegible or for some other reason. 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>
<l>I live in the middle of England</l>
<l>But!</l>
<l>Norway! My soul resides in your watery
<del type="overstrike">fiords fyords fiiords</del></l>
<l>Inlets.</l>
```

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 *Waste Land*:

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

Deletion in manuscript or typescript is often associated with addition:

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

The <del> element should not be used where the deletion is such that material cannot be read with confidence, or read at all, or where the material has been omitted by the transcriber or editor for some other reason. Where the material cannot be read with confidence following deletion, the <unclear> tag should be used with the reason attribute indicated that the difficulty of transcription is due to deletion. Where material has been omitted by the transcriber or editor, this may be indicated by use of the <corr> (or <sic>) and <gap> elements. Observe that the distinction between recommended uses of the <del>, <corr>, and <gap> tags parallels the distinction drawn between the <add>, <corr>, and <supplied> tags in section 6.5.1 *Correction of Apparent Errors* and section 18.1.3 *Correction and Conjecture*:

- where the correction is a deletion by a scribe or author in a manuscript or other primary source (typescript, proof, galley, etc.) then either <corr> (or <sic>) or <del> might be appropriate, depending on the circumstances. The <del> tag is more expressive, and may convey information on just how the deletion was performed (hand, place, etc.) which the <corr> tag cannot. See further the discussion in section 18.1.5 *Substitutions*.
- where the correction is a deletion by a transcriber or editor, correcting a perceived superfluity in the text but in circumstances where there is no clearly assertable reason for the superfluity (as a spurious addition) the <corr> tag should be used. The <del> tag should not be used in this case.
- where the correction is a deletion by a transcriber or editor, correcting a perceived superfluity in the text where there is a clearly assertable reason for the superfluity (as a spurious addition) the <gap> tag should be used with the reason attribute carrying the reason for the superfluity and hence the deletion of text. Neither the <del> nor <corr> tag should be used in these cases.

For any detailed transcription of a manuscript or typescript with more than trivial amounts of alteration, the reader should consult chapter 19 *Critical Apparatus*, and chapter 18 *Transcription of Primary Sources*.

These elements are formally defined as follows:

```
<!-- 6.5.3: Other editorial tags-->
<!ELEMENT gap %om.RO;  EMPTY>
<!ATTLIST gap
      %a.global;
      desc CDATA #IMPLIED
      reason CDATA #IMPLIED
      resp IDREF %INHERITED;
      hand IDREF %INHERITED;
      agent CDATA #IMPLIED
```

```
      extent CDATA #IMPLIED
      TEIform CDATA 'gap'  >
<!ELEMENT add %om.RR;  %specialPara;>
<!ATTLIST add
      %a.global;
      place CDATA #IMPLIED
      resp IDREF %INHERITED;
      cert CDATA #IMPLIED
      hand IDREF %INHERITED;
      TEIform CDATA 'add'  >
<!ELEMENT del %om.RR;  %phrase.seq;>
<!ATTLIST del
      %a.global;
      type CDATA #IMPLIED
      status CDATA "unremarkable"
      resp IDREF %INHERITED;
      cert CDATA #IMPLIED
      hand IDREF %INHERITED;
      TEIform CDATA 'del'  >
<!ELEMENT unclear %om.RO;  %paraContent;>
<!ATTLIST unclear
      %a.global;
      reason CDATA #IMPLIED
      resp CDATA %INHERITED;
      cert CDATA #IMPLIED
      hand IDREF %INHERITED;
      agent CDATA #IMPLIED
      TEIform CDATA 'unclear'  >
<!-- end of 6.5.3-->
```

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 14 *Linking, Segmentation, and Alignment*.

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 document or because no id attribute is available, the elements <xptr> or <xref> may be used instead.[75] 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 in terms of one or more identifiable elements. Attributes include:

target specifies the destination of the pointer by supplying the values used on the id attribute of one or more other elements in the current document

Values One or more valid identifiers, separated by white space.

<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 by supplying the value of the id attribute on one or more other elements in the current document.

Values One or more valid identifiers, separated by white space.

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.

[75] See chapter 14 *Linking, Segmentation, and Alignment* for a discussion of these elements and the extended syntax they provide for 'hypertext' links.

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. Legal values are:

- Y Yes: the order in which IDREF values are specified as the value of a target attribute should be followed when combining the targeted elements.
- N No: the order in which IDREF values are specified as the value of a target attribute has no significance when combining the targeted elements.
- U Unspecified: the order in which IDREF values are specified as the value of a target attribute may or may not be significant.

evaluate specifies the intended meaning when the target of a pointer is itself a pointer. Legal values are:

- 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>
   <item>Sallets censured <ptr target="p143 p144"/></item>
   <item>Sanguine mel. signs <ptr target="p263 p312 p332"/></item>
   <item>Scilla or sea onyon, a purger of mel. <ptr target="p442"/></item>
   <!-- ... -->
</list>
```

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 target elements are of a particular type, the targType (target type) attribute may be specified:

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

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="figure" 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 *Bibliographic Pointers*.

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 targOrder="Y" target="ch3 sec332 sec45 sec722"/></p>
```

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 chapters 14 *Linking, Segmentation, and Alignment*, 15 *Simple Analytic Mechanisms*, and 16 *Feature Structures*, 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-->
<!ELEMENT ptr %om.RO; EMPTY>
<!ATTLIST ptr
      %a.global;
      %a.pointer;
      target IDREFS #REQUIRED
      TEIform CDATA 'ptr'  >
<!ELEMENT ref %om.RR; %paraContent;>
<!ATTLIST ref
      %a.global;
      %a.pointer;
      target IDREFS #IMPLIED
      TEIform CDATA 'ref'  >
<!-- end of 6.6-->
```

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.

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.

<item> contains one component of a list.

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

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

type categorizes the heading in some way meaningful to the encoder.

Values A set of user-defined keywords may be employed. Their significance should be documented in the header.

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

<headItem> 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 rend="runon" type="ordered">
      <label>(1)</label>
      <item>My first rough manuscript, without any
intermediate copy, has been sent to the press.</item>
      <label>(2)</label>
      <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.</item>
   </list>
```

```
I will add two facts, which have seldom occurred in
the composition of six, or at least of five quartos.
<list rend="runon" type="ordered">
      <item n="1">My first rough manuscript, without any
intermediate copy, has been sent to the press.</item>
      <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.</item>
   </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, the following is a reasonable encoding of a list which (in the original) is simply printed as a single paragraph:

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

```
  <item n="n">those that resemble flies from a distance. </item>
  </list>
```

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 &mdash; half term ending Midsummer 1851</head>
 <label>Classics</label>    <item>Idle listless and unimproving</item>
 <label>Mathematics</label> <item>ditto</item>
 <label>Divinity</label>    <item>ditto</item>
 <label>Conduct in house</label> <item>Orderly</item>
 <label>General conduct</label>
 <item>Not satisfactory, on account of his great
    unpunctuality and inattention to duties</item>
</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 chapter 22 *Tables, Formulae, and Graphics*. 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 &mdash; Vocabulary</head>
  <label lang="la">acerbus, -a, -um     </label> <item>bitter, harsh</item>
  <label lang="la">ager, agr&imacr;, M. </label> <item>field</item>
  <label lang="la">audi&omacr;, &imacr;re,
    &imacr;v&imacr;, &imacr;tus         </label> <item>hear, listen (to)</item>
  <label lang="la">bellum, -&imacr;, N. </label> <item>war</item>
  <label lang="la">bonus, -a, -um       </label> <item>good</item>
  <!-- etc. -->
</list>
```

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

```
<list type="gloss">
  <head>Unit Three &mdash; Vocabulary</head>
  <label><term lang="la">acerbus, -a, -um</term> </label>
  <item><gloss>bitter, harsh</gloss> </item>
  <label><term lang="la">ager, agr&imacr;, M. </term> </label>
  <item><gloss>field</gloss> </item>
  <label>
 <term lang="la">audi&omacr;, -&imacr;re, -&imacr;v&imacr;, -&imacr;tus</term>
  </label>
  <item><gloss>hear, listen (to)</gloss> </item>
  <label><term lang="la">bellum, -&imacr;, N. </term> </label>
  <item><gloss>war</gloss> </item>
  <label><term lang="la">bonus, -a, -um</term> </label>
  <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 3.5 *Global Attributes*, and section 4.3 *Code shifting*. 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 *Terminological Databases* or 12 *Print Dictionaries*.

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 <headLabel> and <headItem>:

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

The elements <label>, <head>, <headLabel>, and <headItem> 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 &mdash; 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 %om.RR; ((%m.Incl;)*, (head, (%m.Incl;)*)?, ( ( (item,
(%m.Incl;)*)*)
      | ((headLabel, (%m.Incl;)*)?, (headItem, (%m.Incl;)*)?, (label,
      (%m.Incl;)*, item, (%m.Incl;)*)+)))>
<!ATTLIST list
      %a.global;
      type CDATA "simple"
      TEIform CDATA 'list'  >
<!ELEMENT item %om.RO; %specialPara;>
<!ATTLIST item
      %a.global;
      TEIform CDATA 'item'  >
<!ELEMENT label %om.RO;  %phrase.seq;>
<!ATTLIST label
      %a.global;
      TEIform CDATA 'label'  >
<!ELEMENT head %om.RO; %paraContent;>
<!ATTLIST head
      %a.global;
      type CDATA #IMPLIED
      TEIform CDATA 'head'  >
<!ELEMENT headLabel %om.RO;  %phrase.seq;>
```

```
<!ATTLIST headLabel
      %a.global;
      TEIform CDATA 'headLabel'  >
<!ELEMENT headItem %om.RO;  %phrase.seq;>
<!ATTLIST headItem
      %a.global;
      TEIform CDATA 'headItem'  >
<!-- end of 6.7-->
```

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:

type describes the type of note.

Values Values can be taken from any convenient typology of annotation suitable to the work in hand; e.g. annotation, gloss, citation, digression, preliminary, temporary

resp (responsible) 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.

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.

target indicates the point of attachment of a note, or the beginning of the span to which the note is attached.

Values reference to the ids of element(s) which begin at the location in question (e.g. the id of an <anchor> element).

targetEnd 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.

Values reference to the id(s) of element(s) which *end* at the location(s) in question, or to an empty element at the point in question.

A *note* is any additional comment found in a text, marked in some way as being out of the main textual stream. 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 targetEnd 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 section 6.6 *Simple Links and Cross References*.

Examples:

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

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

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 element is provided for such applications; it is available only when the additional tag set for simple analysis is selected (see section 15.3 *Spans and Interpretations*).

The formal declarations for the <note> element is this:

```
<!-- 6.8.1: Annotation-->
<!ELEMENT note %om.RO;  %specialPara;>
<!ATTLIST note
      %a.global;
      type CDATA #IMPLIED
      resp CDATA #IMPLIED
      place CDATA 'unspecified'
      anchored (yes | no) "yes"
      target IDREFS #IMPLIED
      targetEnd IDREFS #IMPLIED
      TEIform CDATA 'note'  >
<!--declarations from 6.8.2: Index Entries inserted here -->
<!-- end of 6.8.1-->
```

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 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 <divGen> 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 (first-level index entry) gives the form under which the index entry is to be made.

Values any string of characters.

level2 (second-level index entry) gives the second-level form, if any.
Values any string of characters.
level3 (third-level index entry) gives the third-level form, if any.
Values any string of characters.
level4 (fourth-level index entry) gives the fourth-level form, if any.
Values any string of characters.
index (index number) indicates which index (of several) the index entry belongs to.
Values any string of characters; valid values are application-dependent.

<divGen> indicates the location at which a textual division 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 <divGen> 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 type="appendix">
      <head>Examples</head>
      <p> ... </p>
   </div>
   <div type="appendix">
      <head>Bibliography</head>
      <listBibl>
         <bibl> ... </bibl>
      </listBibl>
   </div>
   <divGen n="Index Nominum" type="index 1"/>
   <divGen n="Index Rerum" type="index 2"/>
</back>
```

The formal declaration for these elements is as follows. The <index> element is a member of the element class metadata and may thus be used anywhere within the <text> element.

```
<!-- 6.8.2: Index Entries-->
<!ELEMENT index %om.RO;  EMPTY>
<!ATTLIST index
      %a.global;
      index CDATA #IMPLIED
      level1 CDATA #REQUIRED
      level2 CDATA #IMPLIED
      level3 CDATA #IMPLIED
      level4 CDATA #IMPLIED
      TEIform CDATA 'index'  >
```

```
<!ELEMENT divGen %om.RO;  EMPTY>
<!ATTLIST divGen
      %a.global;
      type CDATA #IMPLIED
      TEIform CDATA 'divGen'  >
<!-- end of 6.8.2-->
```

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 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*.
- 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*.
- where a reference system exists which is not based on the same logical structure as that of the text's 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), any of a variety of methods for encoding the logical structure representing the reference system may be employed, as described in chapter 31 *Multiple Hierarchies*.
- 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 6.9.3 *Milestone Tags*.

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 6.9.4 *Declaring Reference Systems* and in chapter 32 *Algorithm for Recognizing Canonical References.*

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.[76]

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 marked-up 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

[76] 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.

chapter 9 *Base Tag Set for Verse* 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" type="volume">
  <div1 n="1" type="book"> <!-- ... --> </div1>
  <div1 n="2" type="book">
      <div2 n="1" type="poem"> <!-- ... --> </div2>
      <div2 n="2" type="poem"> <!-- ... --> </div2>
      <!-- ... -->
      <div2 n="10" type="poem">
          <l n="1"> ... </l>
          <l n="2"> ... </l>
          <!-- ... -->
          <l n="7"> ... </l>
          <!-- ... -->
      </div2>
      <!-- ... -->
  </div1>
  <!-- ... -->
</div0>
```

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

If the names used by the traditional reference system can be formulated as 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.[77] 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" type="volume">
   <div1 id="am.2" type="book">
      <div2 id="am.2.10" type="poem">
         <l id="am.2.10.7"> ... </l>
      </div2>
   </div1>
</div0>
```

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. For examples of declarations for the reference systems just shown, see section 6.9.4 *Declaring Reference Systems*.

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 could employ any of a variety of methods discussed in chapter 31 *Multiple Hierarchies*.

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 markup of the text. As with any reference system

[77] XML allows many more international characters in identifiers; the legal form of identifiers in SGML 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 SGML documents.

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. ID attributes must be unique 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 structure of the document is to construct, for each element, a *domain-style address* comprising a series of components separated by full stops, with one component for each level of the document hierarchy. Two methods may be used. In the *typed path* form of identifier, each component in the identifier takes the form element-type '-' number. The element name specifies what type of element to be sought, and the number specifies which occurrence of that element type is to be selected. (The hyphen and number may be omitted if there is only one element of the given type.) In the *untyped path* form of identifier, each component consists of a number, indicating which element in the sequence of nodes at each level is to be selected. A fixed prefix beginning with a letter may be used to make the untyped path legal as an ID value.

Identifiers generated with these methods should use the <text> element as their starting point, rather than the <TEI.2> or <body> elements. The <TEI.2> element may be taken as a starting point only if identifiers need to be generated for the <teiHeader>, which is not usually the case; using the <body> element as a root would prevent assignment of identifiers for the front and back matter. The component corresponding to the root element can be omitted from identifiers, if no confusion will result. In collections and corpora, the component corresponding to the root may be replaced by the unique identifier assigned to the text or sample.

In the following example, each element within the <text> element has been given a typed-path identifier as its id value, and an untyped-path identifier as its n value; the latter are prefixed with the string 'AB', which may be imagined to be the general identifier for this text.

```
<text id='TEXT-1' n='AB'>
  <front id='FRONT'       n='AB.1'>
    <div id='FRONT.div-1' n='AB.1.1'>
      <p> ... </p>
    </div>
    <titlePage id='FRONT.titlePage' n='AB.1.2'>
      <titlePart> ... </titlePart>
    </titlePage>
    <div id='FRONT.div-2' n='AB.1.3'>
      <p> ... </p>
    </div>
  </front>
  <body id='BODY' n='AB.2'>
    <p id='BODY.p-1' n='AB.2.1'> ... </p>
    <p id='BODY.p-2' n='AB.2.2'> ... </p>
    <div id='BODY.div-1' n='AB.2.3'>
      <head id='BODY.div-1.head' n='AB.2.3.1'> ... </head>
      <p    id='BODY.div-1.p-1'  n='AB.2.3.2'> ... </p>
      <p    id='BODY.div-1.p-2'  n='AB.2.3.3'> ... </p>
    </div>
    <div id='BODY.div-2' n='AB.2.4'>
      <head id='BODY.div-2.head' n='AB.2.4.1'> ... </head>
      <p    id='BODY.div-2.p-1'  n='AB.2.4.2'> ... </p>
      <p    id='BODY.div-2.p-2'  n='AB.2.4.3'> ... </p>
    </div>
  </body>
</text>
```

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 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 section 5.3.5 *The Reference System Declaration*.

When the hierarchy of the 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 of an SGML (but not XML) TEI text may choose to represent the alternative reference system(s) as elements in one or more concurrent document hierarchies. For an introduction to the concept of concurrent hierarchies, see the discussion of the CONCUR feature in section 2.5 *Complicating the issue*. For further discussion of this and other mechanisms, see chapter 31 *Multiple Hierarchies*.

6.9.3 Milestone Tags

Often concurrent markup is not a viable method because the document is XML, the available SGML parser does not support the CONCUR feature, or the desired reference system does not correspond to any particular structural hierarchy. In these cases it is often desirable to mark up changes in the reference system by using one or more of the following *milestone* elements:

<milestone> marks the boundary between sections of a text, as indicated by changes in a standard reference system. Attributes include:

ed (edition) indicates which edition or version the milestone applies to.
Values Any string of characters; usually a siglum conventionally used for the edition.

unit indicates what kind of section is changing at this milestone.
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.

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

ed (edition) indicates the edition or version in which the page break is located at this point
Values Any string of characters; usually a siglum conventionally used for the edition.

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

ed (edition) indicates the edition or version in which the line break is located at this point
Values Any string of characters; usually a siglum conventionally used for the edition.

<cb> marks the boundary between one column of a text and the next in a standard reference system. Attributes include:

ed (edition) indicates the edition or version in which the column break is located at this point
Values Any string of characters; usually a siglum conventionally used for the edition.

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, rather in the same way as milestones divide a road into segments. The elements <pb>, <cb>, and <lb> are provided to mark specific types of milestone, namely page, column, and line boundaries, as further described in chapter 18 *Transcription of Primary Sources*. The global n attribute is used in each case to provide a value for the value (for example, the page or line number). Validation of a reference system based on <milestone> tags is not directly provided

by SGML or XML parsers, so it will be the responsibility of the encoder or the application software to ensure that milestone tags occur in a correct order.

Milestone tags may be useful where a text has two competing structures. For example, many English novels were first published as serial works, individual parts of which do not always contain a whole number of chapters. An encoder may decide to represent the chapter-based structure using <div1> elements, with <milestone> elements to mark the points at which individual parts end; or the reverse. Thus, an encoding in which chapters are regarded as more important than parts might encode some work in which chapter three begins in part one and is concluded in part two as follows:

```
<text><body>
    <milestone unit='part'/>
    <div1 n='1' type='chapter'>
      <!-- text of chapter 1 here -->
    </div1>
    <div1 n='2' type='chapter'>
      <!-- text of chapter 2 here -->
    </div1>
    <div1 n='3'>
      <!-- part of text of chapter 3 here -->
      <milestone unit='part'/>
      <!-- remainder of text of chapter 3 here -->.
    </div1>
  </body></text>
```

An encoding of the same work in which parts are regarded as more important than chapters might begin as follows:

```
<text>
  <body>
    <div1 n="1" type="part">
      <milestone unit="chapter"/>
      <p><!-- text of chapter 1 here -->
        <milestone unit="chapter"/>
      </p>
      <p><!-- text of chapter 2 here -->
        <milestone unit="chapter"/>
      </p>
      <p><!-- part of text of chapter 3 here -->
      </p>
    </div1>
    <div1 n="2" type="part">
      <p><!-- remainder of text of chapter 3 here -->
        <milestone unit="chapter"/>
        <!-- ... -->
      </p>
    </div1>
  </body>
</text>
```

Milestone tags also make it possible to record the reference systems used in a number of different editions of the same work. The reference system of any one edition can be recreated from a text in which all are marked by simply ignoring all elements that do not specify that edition on their ed attribute.

As a simple example, assuming that edition E1 of some collection of poems regards the first two poems as constituting the first book, while edition E2 regards the first poem as prefatory, a markup scheme like the following might be adopted:

```
<milestone ed="E1" unit="work"/>
<milestone ed="E2" unit="work"/>
<milestone ed="E1" unit="book"/>
<milestone ed="E1" unit="poem"/>
<milestone ed="E2" unit="poem"/>
  <!-- text of first poem here -->
<milestone ed="E2" unit="book"/>
<milestone ed="E1" unit="poem"/>
```

```
<milestone ed="E2" unit="poem"/>
  <!-- text of second poem here -->
```

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 n="Amores" ed="E1" unit="work"/>
<milestone n="1" ed="E1" unit="book"/>
<milestone n="1" ed="E1" unit="poem"/>
  <!-- text of Amores 1.1 -->
<milestone n="2" ed="E1" unit="poem"/>
  <!-- text of Amores 1.2 -->
<milestone n="3" ed="E1" unit="book"/>
```

or

```
<milestone n="Amores" ed="E1" unit="work"/>
<milestone n="1" ed="E1" unit="book"/>
<milestone n="1.1" ed="E1" unit="poem"/>
 <!-- text of Amores 1.1 -->
<milestone n="1.2" ed="E1" unit="poem"/>
  <!-- text of Amores 1.2 -->
<milestone n="1.3" ed="E1" unit="book"/>
```

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).

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 29 *Modifying and Customizing the TEI DTD*.)

See below, section 6.9.4 *Declaring Reference Systems*, for examples of declarations for the reference systems just shown.

The milestone elements are formally defined as follows:

```
<!-- 6.9.3: Milestone tags-->
<!ELEMENT milestone %om.RO; EMPTY>
<!ATTLIST milestone
      %a.global;
      ed CDATA #IMPLIED
      unit CDATA #REQUIRED
      TEIform CDATA 'milestone'  >
<!ELEMENT pb %om.RO; EMPTY>
<!ATTLIST pb
      %a.global;
      ed CDATA #IMPLIED
      TEIform CDATA 'pb'  >
<!ELEMENT lb %om.RO; EMPTY>
<!ATTLIST lb
      %a.global;
      ed CDATA #IMPLIED
      TEIform CDATA 'lb'  >
<!ELEMENT cb %om.RO; EMPTY>
<!ATTLIST cb
      %a.global;
      ed CDATA #IMPLIED
      TEIform CDATA 'cb'  >
<!-- end of 6.9.3-->
```

6.9.4 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 5.3.5 *The Reference System Declaration*. 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* would 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:

```
<teiHeader>
    <fileDesc> <!-- ... --> </fileDesc>
    <encodingDesc>
         <!-- ... -->
         <refsDecl>
              <step refunit='work' delim=' '
                    from='DESCENDANT (1 DIV0 N %1)'/>
              <step refunit='book' delim='.'
                    from='CHILD (1 DIV1 N %1)'/>
              <step refunit='poem' delim='.'
                    from='CHILD (1 DIV2 N %1)'/>
              <step refunit='line'
                    from='CHILD (1 L N %1)'/>
         </refsDecl>
         <!-- ... -->
    </encodingDesc>
</teiHeader>
```

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. If canonical references are made only to lines, the reference system could be declared as follows:

```
<refsDecl>
   <step refunit="line" from="DESCENDANT (1 L N %1)"/>
</refsDecl>
```

Since no delimiter is specified, the entire canonical reference string is sought as the value of the n attribute on an <l> element.

In order to handle references to works, books, and poems as well as to individual lines, the declaration for the reference system must be more complicated:

```
<refsDecl>
   <step from="DESCENDANT (1 (DIV[012]|L) N %1)"/>
</refsDecl>
```

This declaration indicates that the entire reference string must be sought as the value of the n attribute on a <div0>, <div1>, <div2>, or <l> element.

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 would be:

```
<refsDecl>
   <step from="ID (%1)"/>
</refsDecl>
```

As in the previous example, no single value can be given for the refunit attribute in this declaration, as the single step handles references to works, books, and poems, as well as to lines. The type attribute on the <div0>, <div1>, and <div2> elements may be used, however, to indicate the type of the result returned from a match.

Reference systems recorded by means of milestone tags can also be declared; the following prose description could be used to declare the example given in section 6.9.3 *Milestone Tags*.

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

Or in this way, using a formal declaration for this reference scheme derived from edition E1.

```
<refsDecl>
   <state ed="E1" unit="work" delim=" "/>
   <state ed="E1" unit="book" delim="."/>
   <state ed="E1" unit="poem" delim=":"/>
   <state ed="E1" unit="line"/>
</refsDecl>
```

This is synonymous with the following declaration using the <step> element:

```
<refsDecl>
   <step refunit="work" delim=" "
      from="DESCENDANT (1 MILESTONE EDITION E1 UNIT work N %1)"
      to="FOLLOWING (1 MILESTONE EDITION E1 UNIT work)"/>
   <step refunit="book" delim="."
      from="DESCENDANT (1 MILESTONE EDITION E1 UNIT book N %2)"
      to="FOLLLOWING (1 MILESTONE EDITION E1 UNIT book)"/>
   <step refunit="poem" delim="."
      from="DESCENDANT (1 MILESTONE EDITION E1 UNIT poem N %3)"
      to="FOLLLOWING (1 MILESTONE EDITION E1 UNIT poem)"/>
   <step refunit="line"
      from="DESCENDANT (1 MILESTONE EDITION E1 UNIT line N %4)"
      to="FOLLLOWING (1 MILESTONE EDITION E1 UNIT line)"/>
</refsDecl>
```

6.10 Bibliographic Citations and References

Bibliographic references (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 5.2.7 *The Source Description*; 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 bibliographic reference 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,[78] quite apart from the need to distinguish the reference itself as a textual object with particular linguistic properties.

It should be emphasized that for references 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.[79] The features distinguished and described below (in section 6.10.2 *Components of Bibliographic References*) constitute 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*; for a brief mention of related library standards see section 5.7 *Note for Library Cataloguers*.

6.10.1 Elements of Bibliographic References

The following elements are used to mark individual bibliographic references as wholes, or in groups:

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

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

[79] Among the bibliographic software systems and subsystems consulted in the design of the <biblStruct> structure were BibTeX, Scribe, and ProCite. The distinctions made by all three may be preserved in <biblStruct> 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*.

<biblStruct> contains a structured bibliographic citation, in which only bibliographic subelements appear and in a specified order.

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

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

These elements all share a number of possible component sub-elements. For the <bibl> and <biblStruct> elements, exactly the same sub-elements are concerned, and they are described together in section 6.10.2 *Components of Bibliographic References*; for the <biblFull> element, the sub-elements concerned are fully described in section 5.2 *The File Description*.

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 reference itself, with perhaps a few distinctions being made within it (for example, between the part of the reference which identifies a title or author and the rest). Such references, 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.</p>
```

Indeed, some encoders may find it unnecessary to mark the bibliographic reference 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.</p>
```

Some bibliographic references 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 reference 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 *reference*, which is a self-sufficient description of a bibliographic item, and a bibliographic *pointer*, which is a short-form citation (e.g. ‘Baxter, 1983’) which serves usually as a place-holder or pointer to a full long-form reference found elsewhere in the text. The usual encoding of short-form references such as ‘Baxter, 1983’ is not as <bibl> elements but as cross-references to such elements; see section 6.10.3 *Bibliographic Pointers* 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 style-sheet or retrieval processor, the <biblStruct> element should be used. Note that several of the features in this and later examples are explained later in the current section.

```
<biblStruct>
   <monogr>
      <author>Edward R. Tufte</author>
      <title>Envisioning Information</title>
      <imprint>
         <pubPlace>Cheshire, Conn.</pubPlace>
         <publisher>Graphics Press</publisher>
         <date>1990</date>
      </imprint>
   </monogr>
</biblStruct>
```

The highest level of detail and the most complex structure supported by the current proposals is provided by the <biblFull> element, which closely resembles the <fileDesc> element of the TEI Header (section 5.2 *The File Description*).

```
<biblFull>
   <titleStmt>
      <title>Envisioning Information</title>
```

```
      <author>Tufte, Edward R[olf]</author>
   </titleStmt>
   <extent>126 pp.</extent>
   <publicationStmt>
      <publisher>Graphics Press</publisher>
      <pubPlace>Cheshire, Conn. USA</pubPlace>
      <date>1990</date>
   </publicationStmt>
</biblFull>
```

A list of bibliographic items, of whatever kind, may be treated in the same way as any other list (see section 6.7 *Lists*). Alternatively, the specialized <listBibl> element may be used. The difference between the two is that a <list> contains <item> elements, within which bibliographic elements (<bibl>, <biblStruct> or <biblFull>) may appear, as well as other phrase- and paragraph-level elements, whereas the <listBibl> may contain only bibliographic 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 <listBibl>:

```
<listBibl>
   <head>Bibliography</head>
   <biblStruct id="NEL80">
      <analytic>
         <author>Nelson, T. H.</author>
         <title>Replacing the printed word:
             a complete literary system.</title>
      </analytic>
      <monogr>
         <title>Information Processing '80:  Proceedings of the IFIPS
             Congress, October 1980</title>
         <editor>Simon H. Lavington</editor>
         <imprint>
            <publisher>North-Holland</publisher>
            <pubPlace>Amsterdam</pubPlace>
            <date>1980</date>
         </imprint>
         <biblScope>pp 1013&ndash;23 </biblScope>
      </monogr>
      <note>Apparently a draft of section 4 of
          <title>Literary Machines</title>.</note>
   </biblStruct>

   <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>
</listBibl>
```

or a simple <list>:

```
<list>
   <head>Bibliography</head>
   <item>
      <bibl id="NEL80">
         <author>Nelson, T. H.</author>
         <title level="a">Replacing the printed word:
          a complete literary system.</title>
         <title level="m">Information Processing '80:
          Proceedings of the IFIPS Congress, October 1980</title>
         <editor>Simon H. Lavington</editor>
         <publisher>North-Holland</publisher>
         <pubPlace>Amsterdam</pubPlace>
```

```
         <date>1980</date>
         <biblScope>pp 1013&ndash;23
       </biblScope>
         <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 References-->
<!ELEMENT bibl %om.RO; (#PCDATA | %m.phrase; |
    %m.biblPart; | %m.Incl;)*>
<!ATTLIST bibl
      %a.global;
      %a.declarable;
      TEIform CDATA 'bibl'  >
<!ELEMENT biblStruct %om.RO; ((%m.Incl;)*, (analytic, (%m.Incl;)*)?,
                  ((monogr, (%m.Incl;)*), (series, (%m.Incl;)*)*)+,
                  ((note, (%m.Incl;)*) | (idno, (%m.Incl;)*))*)>
<!ATTLIST biblStruct
      %a.global;
      %a.declarable;
      TEIform CDATA 'biblStruct'  >
<!ELEMENT biblFull %om.RO; ((%m.Incl;)*, (titleStmt, (%m.Incl;)*),
               (editionStmt, (%m.Incl;)*)?, (extent, (%m.Incl;)*)?,
               (publicationStmt, (%m.Incl;)*), (seriesStmt, (%m.Incl;)*)?,
               (notesStmt, (%m.Incl;)*)?, (sourceDesc, (%m.Incl;)*)*)>
<!ATTLIST biblFull
      %a.global;
      %a.declarable;
      TEIform CDATA 'biblFull'  >
<!ELEMENT listBibl %om.RR; ((%m.Incl;)*, (head, (%m.Incl;)*)?,
                   (bibl | biblStruct | biblFull),
                   (bibl | biblStruct | biblFull | %m.Incl;)*,
                   (trailer, (%m.Incl;)*)?)>
<!ATTLIST listBibl
      %a.global;
      %a.declarable;
      TEIform CDATA 'listBibl'  >
<!--continued in 6.10.1: Levels of bibliographic information-->
<!--continued in 6.10.1: Author, title, etc.-->
<!--continued in 6.10.1: Bibliographic subelements-->
<!-- end of 6.10.1-->
```

6.10.2 Components of Bibliographic References

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

- elements for grouping components of the *analytic*, *monographic*, and *series* levels in a structured bibliographic reference
- 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 <biblStruct> 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 reference must be distinguished, and must also be presented in a specific order, discussed further below (section 6.10.2.6 *Order of Components within References*).

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 <biblStruct> 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"> ... </title> 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 reference 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 reference 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 German language series:

```
<biblStruct>
   <analytic>
      <author>Thaller, Manfred</author>
      <title level="a">A Draft Proposal for a Standard for the
                       Coding of Machine Readable Sources</title>
   </analytic>
   <monogr>
      <!-- In -->
      <title level="j">Historical Social Research</title>
      <imprint>
         <biblScope type="vol">40</biblScope>
         <date>October 1986</date>
         <biblScope type="pages">3-46</biblScope>
      </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>
         <pubPlace>St. Katharinen</pubPlace>
         <publisher>Max-Planck-Institut f&uuml;r Geschichte
                    In Kommission bei
                    Scripta Mercaturae Verlag</publisher>
         <date>1991</date>
      </imprint>
   </monogr>
   <series lang="DEU">
      <title level="s">Halbgraue Reihe
                      zur Historischen Fachinformatik</title>
      <respStmt>
         <resp>Herausgegeben von</resp>
         <name type="person">Manfred Thaller</name>
         <name type="org">Max-Planck-Institut f&uuml;r Geschichte</name>
      </respStmt>
      <title level="s">Serie A: Historische Quellenkunden</title>
      <biblScope>Band 11</biblScope>
   </series>
</biblStruct>
```

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 <biblStruct> element in systems which can render bibliographic references in any of several styles.

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

```
<!-- 6.10.2.1: Levels of bibliographic information-->
<!ELEMENT analytic %om.RO; (author | editor | respStmt
                  | title | %m.Incl;)*>
<!ATTLIST analytic
      %a.global;
      TEIform CDATA 'analytic'  >
<!ELEMENT monogr %om.RO;  (( (%m.Incl;)*,
               ((
                 (author | editor | respStmt),
                 (author | editor | respStmt | %m.Incl;)*,
                 (title, (%m.Incl;)*)+,
                 ( (editor | respStmt), (%m.Incl;)* )* )
               |(
                 (title, (%m.Incl;)*)+,
                 ((author | editor | respStmt), (%m.Incl;)*)*
               )))?,
               ((note | meeting), (%m.Incl;)*)*,
               (edition, (editor | respStmt | %m.Incl;)*)*, imprint,
               (imprint | extent | biblScope | %m.Incl;)*
              )>
<!ATTLIST monogr
      %a.global;
      TEIform CDATA 'monogr'  >
<!ELEMENT series %om.RO;  (#PCDATA | title | editor | respStmt | biblScope |
%m.Incl;)*>
<!ATTLIST series
      %a.global;
      TEIform CDATA 'series'  >
<!-- end of 6.10.2.1-->
```

6.10.2.2 Authors, Titles, and Editors

Bibliographic references 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 (bibliographic level (or class) of title) indicates whether this is the title of an article, book, journal, series, or unpublished material.

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)

type (type of title) 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> in a bibliographic reference, contains the name of the author(s), personal or corporate, of a work; the primary *statement of responsibility* for any bibliographic item.

<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. Attributes include:

role specifies the nature of the intellectual responsibility

Values semi-open list (examples might include: translator, editor, compiler, illustrator, etc.)

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

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

<name> contains a proper noun or noun phrase. Attributes include:

type indicates the type of the object which is being named by the phrase.

Values Values such as person, place, institution, product, acronym.

<meeting> in bibliographic references, contains a description of the meeting or conference from which the bibliographic item derives.

In bibliographic references, 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:

```
<biblStruct>
   <analytic>
      <author>Lucy Allen Paton</author>
      <title>Notes on Manuscripts of the
         <title level="m" lang="FRA">Proph&eacute;cies de Merlin</title>
      </title>
   </analytic>
   <monogr>
      <title level="j">PMLA</title>
      <imprint>
         <biblScope type="vol">8</biblScope>
         <date>1913</date>
         <biblScope type="pages">122</biblScope>
      </imprint>
   </monogr>
</biblStruct>
```

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 reference, from a national standard for bibliographic references,[80] illustrates this type of analysis with its distinction between main and subordinate titles. Note that this uses the more flexible <bibl>, rather than the structured <biblStruct> element: consequently, there is no requirement to tag all the components of the reference (notably the authors).

```
<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>
   <title level="j">Ann. Acad. Sci. Fenn., Ser. A4</title>
   <date>1972</date>
   <biblScope type="vol: pp">187: 1-4</biblScope>
</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 Bibliographic Description (ISBD).[81] Again, it is only because this example uses <bibl> rather than <biblStruct>, that specific punctuation may be included between the component elements of the reference.

```
<bibl>Tchaikovsky, Peter Ilich.
<title level="m" type="main">The swan lake ballet</title>
= <title level="m" type="parallel" lang="FRA">Le lac des cygnes</title>
: <title level="m" type="subordinate" lang="FRA">grand ballet en 4 actes</title>
: <title level="m" type="subordinate">op. 20</title>
[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

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

[81] 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 in the example or in the reference structure offered by the standard.

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 <respStmt> element should be used. The nature of the responsibility is indicated by means of a <resp> element, and the person, organization etc. responsible by a <name> element. At least one of each of these should be given within the <respStmt> element, followed optionally by any number of either. (This constraint is required for TEI conformance, but is not enforced by the current SGML or XML DTD). Examples of secondary responsibility of this kind include the roles of illustrator, translator, editor, annotator. The <respStmt> element 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 *Elements of Bibliographic References*, and 6.10.2.1 *Analytic, Monographic, and Series Levels*; wherever <author> and <editor> may occur, the <respStmt> 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 <respStmt> elements apply to the work as a whole, not merely to the first edition:

```
<bibl>
   <author>Lominadze, D. G.</author>
   <title level="m">Cyclotron waves in plasma.</title>
   <respStmt>
      <resp>translated by</resp>
      <name>A. N. Dellis;</name>
      <resp>edited by</resp>
      <name>S. M. Hamberger.</name>
   </respStmt>
   <edition>1st ed.</edition>
   <imprint>
      <pubPlace>Oxford:</pubPlace>
      <publisher>Pergamon Press,</publisher>
      <date>1981.</date>
   </imprint>
   <extent>206 p.</extent>
   <title level="s">International series in natural philosophy.</title>
   <note place="inline">Translation of:
   <title lang="ru" level="m">Ciklotronnye volny v plazme.</title>
   </note>
</bibl>
```

In this example, by contrast, the <respStmt> 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 reference 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.

```
<biblStruct>
   <monogr lang="DEU">
      <title>Des Minnesangs Fr&uuml;hling</title>
      <note place="inline">Mit 1 Faksimile</note>
      <edition>36., neugestaltete und erweiterte Auflage</edition>
      <respStmt>
         <resp>Unter Benutzung der Ausgaben von <name>Karl
            Lachmann</name> und <name>Moriz Haupt</name>, <name>Friedrich
            Vogt</name> und <name>Carl von Kraus</name> bearbeitet von</resp>
         <name>Hugo Moser</name>
         <name>Helmut Tervooren</name>
      </respStmt>
      <imprint>
         <biblScope type="volume">I</biblScope>
         <biblScope type="volume title">Texte</biblScope>
         <pubPlace>Stuttgart</pubPlace>
         <publisher>S. Hirzel Verlag</publisher>
         <date>1977</date>
      </imprint>
```

```
   </monogr>
</biblStruct>
```

With the exception of the <name> element (for which see section 6.4 *Names, Numbers, Dates, Abbreviations, and Addresses*), the elements described in this section are defined as follows:

```
<!-- 6.10.2.2: Author, title, etc.-->
<!ELEMENT author %om.RO; %phrase.seq;>
<!ATTLIST author
      %a.global;
      TEIform CDATA 'author'  >
<!ELEMENT editor %om.RO; %phrase.seq;>
<!ATTLIST editor
      %a.global;
      role CDATA "editor"
      TEIform CDATA 'editor'  >
<!ELEMENT respStmt %om.RO; (resp | name | %m.Incl;)+ >
<!ATTLIST respStmt
      %a.global;
      TEIform CDATA 'respStmt'  >
<!ELEMENT resp %om.RO; %phrase.seq;>
<!ATTLIST resp
      %a.global;
      TEIform CDATA 'resp'  >
<!ELEMENT title %om.RO; %paraContent;>
<!ATTLIST title
      %a.global;
      level (a | m | j | s | u) #IMPLIED
      type CDATA #IMPLIED
      TEIform CDATA 'title'  >
<!ELEMENT meeting %om.RR; %paraContent;>
<!ATTLIST meeting
      %a.global;
      TEIform CDATA 'meeting'  >
<!-- end of 6.10.2.2-->
```

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 references 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.

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

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

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

<date> contains a date in any format. Attributes include:

calendar indicates the system or calendar to which the date belongs.

Values Recommended values include: *Gregorian*, *Julian*, *Roman*, *Mosaic*, *Revolutionary*, *Islamic*.

value gives the value of the date in some standard form, usually yyyy-mm-dd.

Values Any string representing a date in standard format; recommended form is ISO 8601:2000 5.2.1.1 Complete representation, extended format (yyyy-mm-dd)

certainty indicates the degree of precision to be attributed to the date.

Values Any appropriate value, e.g. *ca.*, *approx*, *after*, *before*.

<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.
Values A name or abbreviation indicating what type of identifying number is given (e.g. ISBN, LCCN).

<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”.
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.

For bibliographic purposes, usually only the place (or places) of publication are required, possibly including the name of the country, rather than a full address; the element <pubPlace> is provided for this purpose. Where however the full postal address is likely to be of importance in identifying or locating the bibliographic item concerned, it may be supplied and tagged using the <address> element described in section 6.4.2 *Addresses*. Alternatively, if desired, the <rs> or <name> elements described in section 6.4.1 *Referring Strings* may be used; 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 should be marked using the <publisher> tag even if the item is made public (‘published’) by an organization other than a conventional publisher, as is frequently the case with technical reports:

```
<biblStruct>
   <monogr>
      <author>Nicholas, Charles K.</author>
      <author>Welsch, Lawrence A.</author>
      <title>On the interchangeability of SGML and ODA</title>
      <imprint>
         <pubPlace>Gaithersburg, MD</pubPlace>
         <publisher>National Institute of Standards and Technology
      </publisher>
         <date value="1992-01">January 1992</date>
      </imprint>
      <extent>19 pp.</extent>
   </monogr>
   <idno type="NIST">NISTIR 4681</idno>
</biblStruct>
```

and with dissertations:

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

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.

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

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 classes 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:

```
<biblStruct>
   <analytic>
      <author>Wrigley, E. A.</author>
      <title>Parish registers and the historian</title>
   </analytic>
   <monogr>
      <editor>Steel, D. J.</editor>
      <title>National index of parish registers</title>
      <imprint>
         <pubPlace>London</pubPlace>
         <publisher>Society of Genealogists</publisher>
         <date value="1968">1968</date>
      </imprint>
      <biblScope type="volume">vol. 1</biblScope>
      <biblScope type="pages">pp. 155&ndash;167.</biblScope>
   </monogr>
</biblStruct>
```

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

```
<biblStruct>
   <analytic>
      <author>Boguraev, Branimir</author>
      <author>Neff, Mary</author>
      <title>Text Representation, Dictionary Structure,
             and Lexical Knowledge</title>
   </analytic>
```

```
   <monogr>
      <title level="j">Literary &amp; Linguistic Computing</title>
      <imprint>
         <biblScope type="volume">7</biblScope>
         <biblScope type="issue">2</biblScope>
         <date>1992</date>
         <biblScope type="pages">110-112</biblScope>
      </imprint>
   </monogr>
</biblStruct>

<biblStruct>
   <analytic>
      <author>Chesnutt, David</author>
      <title>Historical Editions in the States</title>
   </analytic>
   <monogr>
      <title level="j">Computers and the Humanities</title>
      <imprint>
         <biblScope>25.6</biblScope>
         <date value="1991-12">(December, 1991):</date>
         <biblScope>377&ndash;380</biblScope>
      </imprint>
   </monogr>
</biblStruct>
```

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

```
<!-- 6.10.2.3: Bibliographic subelements-->
<!ELEMENT imprint %om.RO;
        (pubPlace | publisher | date | biblScope | %m.Incl;)*>
<!ATTLIST imprint
      %a.global;
      TEIform CDATA 'imprint'  >
<!ELEMENT publisher %om.RO; %phrase.seq;>
<!ATTLIST publisher
      %a.global;
      TEIform CDATA 'publisher'  >
<!ELEMENT biblScope %om.RO; %phrase.seq;>
<!ATTLIST biblScope
      %a.global;
      type CDATA #IMPLIED
      TEIform CDATA 'biblScope'  >
<!ELEMENT pubPlace %om.RR;  %phrase.seq;>
<!ATTLIST pubPlace
      %a.global;
      %a.names;
      TEIform CDATA 'pubPlace'  >
<!--Note and date are defined elsewhere, as are extent, address,
and idno.-->
<!-- end of 6.10.2.3-->
```

6.10.2.4 Series Information

Series information may (in <bibl> elements) or must (in <biblStruct> elements) be enclosed in a <series> element or (in a <biblFull> 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*) may be tagged <editor> or <respStmt>.

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 any descriptive annotation of a bibliographic entry in a database.

<note> contains a note or annotation. Attributes include:

type describes the type of note.

Values Values can be taken from any convenient typology of annotation suitable to the work in hand; e.g. annotation, gloss, citation, digression, preliminary, temporary

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.

For example:

```
<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&ndash;947.</biblScope>
   <note>Classic polemic supporting descriptive over procedural
         markup in scholarly work.</note>
</bibl>
```

6.10.2.6 Order of Components within References

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

In <biblStruct> 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 <respStmt> elements
- <imprint>
- <biblScope>

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

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

If more detailed structuring of a bibliographic description is required, the <biblFull> 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 *The File Description*.

6.10.3 Bibliographic Pointers

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*). 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 bibliographic reference 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> ...
```

It may be important to distinguish between the short form of a bibliographic reference and some qualifying or additional information. The latter should not appear within the scope of the <ref> element when this is the case, as for example in an application concerned to normalize bibliographic references:

```
Nelson claims (<ref target="NEL80">Nelson [1980]</ref>,
pages 13&ndash;37) ...
```

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.

Listed below are the equivalences between the various bibliographic fields defined for use in the Scribe and BibTeX systems of bibliographic databases and the elements defined in this tag set.[82] 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 <respStmt>
editors tag as multiple <editor> or <respStmt> elements
fullauthor use the reg attribute on <author> or <name>
fullorganization use the reg attribute on <name type="org">
howpublished tag as <note>, possibly using the form <note place="inline">
institution used only for issuer of technical reports; tag as <publisher>
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 a normalized equivalent in ISO 8601 format
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 <name type="org"> within <respStmt> 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 <publisher>
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

[82] 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.

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.

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

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

<sp> An individual speech in a performance text, or a passage presented as such in a prose or verse text. Attributes include:

who identifies the speaker of the part by supplying an `IDREF` value.

Values The values used are derived from the `id` attribute on the `<role>` elements in the cast list or from a list of the participants.

<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.

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

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 9 *Base Tag Set for Verse*; for performance texts, see the drama base described in chapter 10 *Base Tag Set for Drama*). 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 Core Tags for 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 8 *Base Tag Set for Prose* and 9 *Base Tag Set for Verse*. 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*. 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>
<l>From 500 to 600 poor fellows &mdash; the groans
and screams &mdash;</l>
<l>The odor of blood, mixed with the fresh scent
```

```
of the night, <lb/>the grass, the trees &mdash;
that Slaughter-house!</l>
```

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 31 *Multiple Hierarchies*, or to use the part attribute to indicate that the verse line is incomplete, as in the following example:

```
<l>On a tree by a river a little tomtit</l>
<l>Sang <q>Willow, titwillow, titwillow!</q></l>
<l part="I">And I said to him,</l> <q><l part="F">Dicky-bird, why do you sit</l>
<l>Singing <q>Willow, titwillow, titwillow!</q></l></q>
```

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 rend attribute to indicate whether or not a line is indented.

```
<lg type="stanza">
   <l>Come fill up the Glass,</l>
   <l rend="indent">Round, round let it pass,</l>
   <l>'Till our Reason be lost in our Wine:</l>
   <l rend="indent">Leave Conscience's Rules</l>
   <l rend="indent">To Women and Fools,</l>
   <l>This only can make us divine.</l>
</lg>
<lg n="Chorus" type="refrain">
   <l>Then a Mohock, a Mohock I'll be,</l>
   <l>No Laws shall restrain</l>
   <l>Our Libertine Reign,</l>
   <l>We'll riot, drink on, and be free.</l>
</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 element described in chapter 8 *Base Tag Set for Prose*. For example:

```
<lg type="poem.sonnet">
  <lg type="octet">
    <l>Thus speaks the Muse, and bends her brow severe:&mdash;</l>
    <l>&ldquo;Did I, <name>L&aelig;titia</name>, lend my choicest lays,</l>
    <l>And crown thy youthful head with freshest bays,</l>
    <l>That all the' expectance of thy full-grown year</l>
    <l>Should lie inert and fruitless? O revere</l>
    <l>Those sacred gifts whose meed is deathless praise,</l>
    <l>Whose potent charms the' enraptured soul can raise</l>
    <l>Far from the vapours of this earthly sphere!</l>
  </lg>
  <lg type="sestet">
    <l>Seize, seize the lyre! resume the lofty strain!</l>
    <l>'T is time, 't is time! hark how the nations round</l>
    <l>With jocund notes of liberty resound,&mdash;</l>
    <l>And thy own <name>Corsica</name> has burst her chain!</l>
    <l>O let the song to <name>Britain's</name> shores rebound,</l>
    <l rend="indent(-1)">Where Freedom's once-loved voice is heard,
      alas! in vain.&rdquo;</l>
  </lg>
</lg>
```

The part attribute may also be attached to an <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</speaker>
```

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

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 14 *Linking, Segmentation, and Alignment*. For discussion of other elements and attributes specific to the encoding of verse, see chapter 9 *Base Tag Set for Verse*.

These elements are defined as follows:

```
<!-- 6.11.1: Verse-->
<!ELEMENT l %om.RO; %paraContent;>
<!ATTLIST l
      %a.global;
      %a.metrical;
      %a.enjamb;
      part (Y | N | I | M | F) "N"
      TEIform CDATA 'l'  >
<!ELEMENT lg %om.RO; ((%m.divtop; | %m.Incl;)*,
                    (l | lg), (l | lg | %m.Incl;)*,
                    ((%m.divbot;), (%m.Incl;)*)*)>
<!ATTLIST lg
      %a.global;
      %a.divn;
      TEIform CDATA 'lg'  >
<!-- end of 6.11.1-->
```

6.11.2 Core Tags for 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 or <div1> (etc.) elements described below in chapters 8 *Base Tag Set for Prose* and 10 *Base Tag Set for Drama*. 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 n="I.2" type="scene">
   <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.</p>
   </sp>
   <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.</p>
   </sp>
   <sp>
      <speaker>FILCH.</speaker>
      <p>Tom Gagg, sir, is found guilty.</p>
```

```
   </sp>
</div2>
```

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

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

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="unknown">Now call'd Bernardo &amp;
Francesco.</add></stage>
<sp who="francisco"> <speaker>1.</speaker>
   <l part="Y">STand: who is that?</l>
</sp>
<sp who="barnardo"> <speaker>2.</speaker>
   <l part="Y">Tis I.</l>
</sp>
<sp who="francisco"> <speaker>1.</speaker>
   <l>O you come most carefully vpon your watch,</l>
</sp>
<sp who="barnardo"> <speaker>2.</speaker>
   <l>And if you meete Marcellus and Horatio,</l>
   <l>The partners of my watch, bid them make haste.</l>
</sp>
<sp who="francisco"> <speaker>1.</speaker>
   <l part="Y">I will: See who goes there.</l>
</sp>
<stage>Enter Horatio and Marcellus.</stage>
<sp who="horatio"> <speaker>Hor.</speaker>
   <l part="I">Friends to this ground.</l>
</sp>
```

```
<sp who="marcellus"> <speaker>Mar.</speaker>
   <l part="F">And leegemen to the Dane,</l>
   <l>O farewell honest souldier, who hath releeued you?</l>
</sp>
<sp who="francisco"> <speaker>1.</speaker>
   <l>Barnardo hath my place, giue you good night.</l>
</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 n="I" type="act">
   <div2 n="1" type="scene">
      <head rend="italic">Actus primus, Scena prima.</head>
      <stage rend="italic" type="setting">A tempestuous
        noise of Thunder and Lightning heard: Enter
        a Ship-master, and a Boteswaine.</stage>
      <sp>
         <speaker>Master.</speaker> <p>Bote-swaine.</p>
      </sp>
      <sp>
         <speaker>Botes.</speaker> <p>Heere Master: What cheere?</p>
      </sp>
      <sp>
         <speaker>Mast.</speaker>
         <p>Good: Speake to th' Mariners: fall
           too't, yarely, or we run our selues a ground,
           bestirre, bestirre.  <stage type="move">Exit.</stage>
         </p>
      </sp>
      <stage type="move">Enter Mariners.</stage>
      <sp>
         <speaker>Botes.</speaker>
         <p>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.</p>
      </sp>
   </div2>
</div1>
```

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. The following example is taken from a 19th century novel in which passages of narrative and passages of dialogue are mixed within the same chapter:

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

These elements are defined as follows:

```
<!-- 6.11.2: Drama-->
<!ELEMENT sp %om.RO; ((%m.Incl;)*, (speaker, (%m.Incl;)*)?,
```

```
                    ((p | l | lg | ab | seg | stage), (%m.Incl;)*)+)>
<!ATTLIST sp
      %a.global;
      who IDREFS #IMPLIED
      TEIform CDATA 'sp'  >
<!ELEMENT speaker %om.RO; %phrase.seq;>
<!ATTLIST speaker
      %a.global;
      TEIform CDATA 'speaker'  >
<!ELEMENT stage %om.RR; %specialPara;>
<!ATTLIST stage
      %a.global;
      type CDATA #IMPLIED
      TEIform CDATA 'stage'  >
<!-- end of 6.11.2-->
```

6.12 Overview of the Core Tag Set

All the elements described in this chapter (except for those tags designed to be used in concurrent markup streams, which are availble in SGML only) occur in the *core* of TEI tags, defined by the following DTD fragment.

```
<!-- 6.12: Elements available in all forms of the TEI main DTD-->
<!--Definition of elements, sub-group by sub-group.-->
<!--declarations from 6.1: Paragraph inserted here -->
<!--declarations from 6.3.2.1: Highlighted phrases inserted here -->
<!--declarations from 6.4.1: Proper Nouns inserted here -->
<!--declarations from 6.4.3: Numbers and measures inserted here -->
<!--declarations from 6.4.4: Dates and times inserted here -->
<!--declarations from 6.4.5: Abbreviations inserted here -->
<!--declarations from 6.5.1: Editorial tags for correction inserted here -->
<!--declarations from 6.5.2: Editorial tags for regularization inserted here -->
<!--declarations from 6.5.3: Other editorial tags inserted here -->
<!--declarations from 6.4.2: Addresses and their components inserted here -->
<!--declarations from 6.6: Simple cross references inserted here -->
<!--declarations from 6.7: Lists and List Items inserted here -->
<!--declarations from 6.8.1: Annotation inserted here -->
<!--declarations from 6.9.3: Milestone tags inserted here -->
<!--declarations from 6.10.1: Tags for Bibliographic References inserted here -->
<!--declarations from 6.11.1: Verse inserted here -->
<!--declarations from 6.11.2: Drama inserted here -->
<!-- end of 6.12-->
```