

7 Default Text Structure

This chapter describes the default high-level structure for all TEI documents. The majority of the different base tag sets described in part II simply embed the framework defined in this chapter while a few redefine it with some minor modifications. This chapter is therefore relevant to every kind of TEI document. For further details on the overall structure of the TEI document type definitions, in particular the use of base and additional tag sets, see chapter 3 *Structure of the TEI Document Type Definition*.

TEI texts may be regarded either as *unitary*, that is, forming an organic whole, or as *composite*, that is, consisting of several components which are in some important sense independent of each other. The distinction is not always entirely obvious: for example a collection of essays might be regarded as a single item in some circumstances, or as a number of distinct items in others. In such borderline cases, the encoder must choose whether to treat the text as unitary or composite; each may have advantages and disadvantages in a given situation.

Whether unitary or composite, the text is marked with the <text> tag and may contain front matter, a text body, and back matter. In unitary texts, the text body is tagged <body>; in composite texts, where the text body consists of a series of subordinate texts or groups, it is tagged <group>. The overall structure of any text, unitary or composite, is thus defined by the following elements:

- <text> contains a single text of any kind, whether unitary or composite, for example a poem or drama, a collection of essays, a novel, a dictionary, or a corpus sample.
- <front> contains any prefatory matter (headers, title page, prefaces, dedications, etc.) found at the start of a document, before the main body.
- <body> contains the whole body of a single unitary text, excluding any front or back matter.
- <group> contains the body of a composite text, grouping together a sequence of distinct texts (or groups of such texts) which are regarded as a unit for some purpose, for example the collected works of an author, a sequence of prose essays, etc.
- <back> contains any appendixes, etc. following the main part of a text.

The overall structure of a unitary text is:

```
<TEI.2>
  <teiHeader> <!-- ... --> </teiHeader>
  <text>
    <front>
      <!-- front matter of copy text goes here. -->
    </front>
    <body>
      <!-- body of text goes here. -->
    </body>
    <back>
      <!-- back matter of text, if any, here. -->
    </back>
  </text>
</TEI.2>
```

The overall structure of a composite text made up of two unitary texts is:

```
<TEI.2>
  <teiHeader> <!-- ... --> </teiHeader>
  <text>
    <front>
      <!-- front matter of composite text goes here. -->
    </front>
    <group>
      <text>
        <front>
          <!-- front matter of first unitary text, if any -->
        </front>
        <body>
          <!-- body of first unitary text -->
        </body>
        <back>
          <!-- back matter of first unitary text, if any -->
      </text>
    </group>
  </text>
</TEI.2>
```

```

        </back>
      </text>
    <text>
      <body>
        <!-- body of second unitary text -->
      </body>
    </text>
  </group>
</back>
<!-- back matter of composite text, if any -->
</back>
</text>
</TEI.2>

```

Each of these elements is further described in the following subsections. `<text>`, `<body>`, and `<group>` are formally declared as follows:

```

<!-- 7.: Top-level parts of default structure-->
<!ELEMENT text %om.RR; ((%m.Incl;)*, (front, (%m.Incl;)*)?,
    (body | group), (%m.Incl;)*, (back, (%m.Incl;)*)?)>
<!ATTLIST text
    %a.global;
    %a.declaring;
    TEIform CDATA 'text' >
<!ELEMENT body %om.R0; ( (%m.divtop; | %m.Incl;)*,
    ( ( (%component;), (%m.Incl;)*)+,
        ((divGen, (%m.Incl;)*)*,
            ( (div, (div|divGen|%m.Incl;)* ) |
                (div0, (div0|divGen|%m.Incl;)* ) |
                (div1, (div1|divGen|%m.Incl;)* )
            )? ))
        | ((divGen, (%m.Incl;)*)*,
            ((div, (div|divGen|%m.Incl;)* ) |
                (div0, (div0|divGen|%m.Incl;)* ) |
                (div1, (div1|divGen|%m.Incl;)* )
            )), (%m.divbot;), (%m.Incl;)* )*>
<!ATTLIST body
    %a.global;
    %a.declaring;
    TEIform CDATA 'body' >
<!ELEMENT group %om.R0; ((%m.divtop; | %m.Incl;)*, ((text | group),
    (text|group|%m.Incl;)*), (%m.divbot;), (%m.Incl;)*)*>
<!ATTLIST group
    %a.global;
    %a.declaring;
    TEIform CDATA 'group' >
<!-- end of 7.-->

```

Elements `<front>` and `<back>` are declared separately, and are further discussed in sections 7.4 *Front Matter* and 7.6 *Back Matter*. Textual elements, such as paragraphs, lists or phrases, which nest within these major structural elements, are discussed in chapter 6 *Elements Available in All TEI Documents* (for elements common to all kinds of document) and in part II (for elements specific to a particular base). The `<group>` element, used for composite texts, is further discussed in section 7.3 *Groups of Texts*.

7.1 Divisions of the Body

In some texts, the body consists simply of a sequence of low-level structural items, referred to here as *components* or *component-level elements* (see section 3.7 *Element Classes*). Examples in prose texts include paragraphs or lists; in dramatic texts, speeches and stage directions; in dictionaries, dictionary entries. In other cases sequences of such elements will be grouped together hierarchically into textual divisions and subdivisions, such as chapters or sections. The names used for these structural subdivisions of texts vary with the genre and period of the text, or even with the whim of the author, editor, or publisher. For example, a major subdivision of an epic or of the Bible is generally called a ‘book’, that of a report is usually called a ‘part’ or ‘section’, that of a novel a ‘chapter’ — unless it is an epistolary novel, in

which case it may be called a ‘letter’. Even texts which are not organized as linear prose narratives, or not as narratives at all, will frequently be subdivided in a similar way: a drama into ‘acts’ and ‘scenes’; a reference book into ‘sections’; a diary or day book into ‘entries’; a newspaper into ‘issues’ and ‘sections’, and so forth.

Because of this variety, these Guidelines propose that all such textual divisions be regarded as occurrences of the same neutrally named elements, with an attribute type used to categorize elements independently of their hierarchic level. Two alternative styles are provided for the marking of these neutral divisions: *numbered* and *un-numbered*. Numbered divisions are named <div0>, <div1>, <div2>, etc., where the number indicates the depth of this particular division within the hierarchy, the largest such division being “div0”, any subdivision within it being “div1”, any further sub-sub-division being “div2” and so on. Un-numbered divisions are simply named <div>, and allowed to nest recursively to indicate their hierarchic depth. The two styles may *not* be combined within a single <front>, <body> or <back> element.

7.1.1 Un-numbered Divisions

The following element is used to identify textual subdivisions in the un-numbered style:

<div> contains a subdivision of the front, body, or back of a text.

As a member of the class *divn*, this element has the following additional attribute:

type specifies a name conventionally used for this level of subdivision, e.g. “act”, “volume”, “book”, “section”, “canto”, etc.

Using this style, the body of a text containing two parts, each composed of two chapters, might be represented as follows:

```
<body>
  <div type="part" n="1">
    <div type="chapter" n="1">
      <!-- text of part 1, chapter 1 -->
    </div>
    <div type="chapter" n="2">
      <!-- text of part 1, chapter 2-->
    </div>
  </div>
  <div type="part" n="2">
    <div n="1" type="chapter">
      <!-- text of part 2, chapter 1 -->
    </div>
    <div n="2" type="chapter">
      <!-- text of part 2, chapter 2 -->
    </div>
  </div>
</body>
```

The <div> element has the following formal definition:

```
<!-- 7.1.1: Un-numbered divisions-->
<!ELEMENT div %om.R0; ( (%m.divtop; | %m.Incl; )*, (((div|divGen),
    (%m.Incl;)*)+ | ((%component; , (%m.Incl;)*)+,
    ((div|divGen), (%m.Incl;)*)) ),
    ((%m.divbot;), (%m.Incl;)*)* )>
<!ATTLIST div
  %a.global;
  %a.divn;
  %a.declaring;
  TEIform CDATA 'div' >
<!-- end of 7.1.1-->
```

7.1.2 Numbered Divisions

The following elements are used to identify textual subdivisions in the numbered style:

<div0> contains the largest possible subdivision of the body of a text.

<div1> contains a first-level subdivision of the front, body, or back of a text (the largest, if <div0> is not used, the second largest if it is).

<div2> contains a second-level subdivision of the front, body, or back of a text.

- <div3>** contains a third-level subdivision of the front, body, or back of a text.
- <div4>** contains a fourth-level subdivision of the front, body, or back of a text.
- <div5>** contains a fifth-level subdivision of the front, body, or back of a text.
- <div6>** contains a sixth-level subdivision of the front, body, or back of a text.
- <div7>** contains the smallest possible subdivision of the front, body or back of a text, larger than a paragraph.

As members of the class *divn* these elements all bear the following additional attribute:

type specifies a name conventionally used for this level of subdivision, e.g. “act”, “volume”, “book”, “section”, “canto”, etc.

The largest possible subdivision of the body may be regarded either as a **<div0>** or as a **<div1>** element,⁸³ and the smallest possible **<div7>**. If numbered divisions are in use, a division at any one level (say, **<div3>**), may contain only numbered divisions at the next lowest level (in this case, **<div4>**).

Using this style, the body of a text containing two parts, each composed of two chapters, might be represented as follows:

```
<body>

  <div0 type="Part" n="1">
    <div1 type="Chapter" n="1">
      <!-- text of part 1, chapter 1 -->
    </div1>
    <div1 type="Chapter" n="2">
      <!-- text of part 1, chapter 2-->
    </div1>
  </div0>

  <div0 type="Part" n="2">
    <div1 type="Chapter" n="1">
      <!-- text of part 2, chapter 1 -->
    </div1>
    <div1 type="Chapter" n="2">
      <!-- text of part 2, chapter 2 -->
    </div1>
  </div0>

</body>
```

Formal definitions for these elements are as follows:

```
<!-- 7.1.2: Numbered divisions-->
<!ELEMENT div0 %om.R0;
      ( (%m.divtop; | %m.Incl;)*, ( ((div1 | divGen), (%m.Incl;)*)+
      | ( (%component;; (%m.Incl;)*)+,
          ((div1 | divGen), (%m.Incl;)*)*)),
      ((%m.divbot;), (%m.Incl;)*)*>
<!ATTLIST div0
      %a.global;
      %a.divn;
      %a.declaring;
      TEIform CDATA 'div0' >
<!ELEMENT div1 %om.R0;
      ( (%m.divtop; | %m.Incl;)*, ( ((div2 | divGen), (%m.Incl;)*)+
      | ( (%component;; (%m.Incl;)*)+,
          ((div2 | divGen), (%m.Incl;)*)*)),
      ((%m.divbot;), (%m.Incl;)*)*>
<!ATTLIST div1
      %a.global;
      %a.divn;
      %a.declaring;
      TEIform CDATA 'div1' >
```

⁸³ This convention (corresponding with the idea that a type-set document may begin either with a “level 0” or a “level 1” heading) is provided for convenience and compatibility with some widely used formatting systems.

```

<!ELEMENT div2 %om.R0;
( (%m.divtop; | %m.Incl;)*, ( ((div3 | divGen), (%m.Incl;)*)+
  | ( (%component;;, (%m.Incl;)*)+,
    ((div3 | divGen), (%m.Incl;)*)*)),
  ((%m.divbot;), (%m.Incl;)*)*>
<!ATTLIST div2
  %a.global;
  %a.divn;
  %a.declaring;
  TEIform CDATA 'div2' >
<!ELEMENT div3 %om.R0;
( (%m.divtop; | %m.Incl;)*, ( ((div4 | divGen), (%m.Incl;)*)+
  | ( (%component;;, (%m.Incl;)*)+,
    ((div4 | divGen), (%m.Incl;)*)*)),
  ((%m.divbot;), (%m.Incl;)*)*>
<!ATTLIST div3
  %a.global;
  %a.divn;
  %a.declaring;
  TEIform CDATA 'div3' >
<!ELEMENT div4 %om.R0;
( (%m.divtop; | %m.Incl;)*, ( ((div5 | divGen), (%m.Incl;)*)+
  | ( (%component;;, (%m.Incl;)*)+,
    ((div5 | divGen), (%m.Incl;)*)*)),
  ((%m.divbot;), (%m.Incl;)*)*>
<!ATTLIST div4
  %a.global;
  %a.divn;
  %a.declaring;
  TEIform CDATA 'div4' >
<!ELEMENT div5 %om.R0; ( (%m.divtop; | %m.Incl;)*, ( ((div6 | divGen),
(%m.Incl;)*)+
  | ( (%component;;, (%m.Incl;)*)+,
    ((div6 | divGen), (%m.Incl;)*)*)),
  ((%m.divbot;), (%m.Incl;)*)*>
<!ATTLIST div5
  %a.global;
  %a.divn;
  %a.declaring;
  TEIform CDATA 'div5' >
<!ELEMENT div6 %om.R0;
( (%m.divtop; | %m.Incl;)*, ( ((div7 | divGen), (%m.Incl;)*)+
  | ( (%component;;, (%m.Incl;)*)+,
    ((div7 | divGen), (%m.Incl;)*)*)),
  ((%m.divbot;), (%m.Incl;)*)*>
<!ATTLIST div6
  %a.global;
  %a.divn;
  %a.declaring;
  TEIform CDATA 'div6' >
<!ELEMENT div7
%om.R0; ((%m.divtop; | %m.Incl;)*,
  (%component;;, (%m.Incl;)*)+, ((%m.divbot;),
  (%m.Incl;)*)*>
<!ATTLIST div7
  %a.global;
  %a.divn;
  %a.declaring;
  TEIform CDATA 'div7' >
<!-- end of 7.1.2-->

```

7.1.3 Numbered or Un-numbered?

Within the same <front>, <body>, or <back> element, all hierarchic subdivisions must be marked either using nested <div> elements, or using the <div0>, <div1>, <div2> tag appropriate at each level; the two styles may *not* be mixed.

The choice between numbered and un-numbered divisions will depend to some extent on the complexity of the material: un-numbered divisions allow for an arbitrary depth of nesting, while numbered divisions limit the depth of the tree which can be constructed. Where divisions at different levels should be processed differently (chapters, but not sections, for example, beginning on new pages), numbered divisions slightly simplify the task of defining the desired processing for each level. Some software may find numbered divisions easier to process, as there is no need to maintain knowledge of the whole document structure in order to know the level at which a division occurs; such software may however find it difficult to cope with some other aspects of the TEI scheme. On the other hand, in a collection of many works it may prove difficult or impossible to ensure that the same numbered division always corresponds with the same type of textual feature: a ‘chapter’ may be at level 1 in one work and level 3 in another.

Whichever style is used, the global *n* and *id* attributes (section 3.5 *Global Attributes*) may be used to provide reference strings or labels for each division of a text, where appropriate. Such labels should be provided for each section which is regarded as significant for referencing purposes (on reference systems, see further section 6.9 *Reference Systems*).

As indicated above, the *type* attribute is used to provide a name or description for the division. Typical values might be ‘book’, ‘chapter’, ‘section’, ‘part’, or (for verse texts) ‘book’, ‘canto’, ‘stanza’, or (for dramatic texts) ‘act’, ‘scene’.

In previous versions of these Guidelines this attribute had a declared value of #CURRENT, which in SGML implies that if defaulted, the value used will be that most recently specified on any element of the same kind, scanning the text left to right. Hence, if un-numbered divisions are used, the appropriate value must be specified each time a change of level occurs, both ‘down’ and ‘up’ the document hierarchy. This value is not available in XML, and so the present edition of these Guidelines does not use it.

The following extended example uses numbered divisions to indicate the structure of a novel, and illustrates the use of the attributes discussed above. It also uses some elements discussed in section 7.2 *Elements Common to All Divisions* and the <p> element discussed in section 6.1 *Paragraphs*.

```
<div0 type="book" n="I" id="JA0100">
  <head>Book I.</head>
  <div1 type="chapter" n="1" id="JA0101">
    <head>Of writing lives in general, and particularly of Pamela, with a word
      by the bye of Colley Cibber and others.</head>
    <p>It is a trite but true observation, that examples work more forcibly on
      the mind than precepts: ... </p>
    <!-- ... remainder of chapter 1 here ... -->
  </div1>
  <div1 type="chapter" n="2" id="JA0102">
    <head>Of Mr. Joseph Andrews, his birth, parentage, education, and great
      endowments; with a word or two concerning ancestors.</head>
    <p>Mr. Joseph Andrews, the hero of our ensuing history, was esteemed to
      be the only son of Gaffar and Gammar Andrews, and brother to the
      illustrious Pamela, whose virtue is at present so famous ... </p>
    <!-- ... remainder of chapter 2 here ... -->
  </div1>
  <!-- ... remaining chapters of Book 1 here ... -->
  <trailer>The end of the first Book</trailer>
</div0>
<div0 type="book" n="II" id="JA0200">
  <head>Book II</head>
  <div1 type="chapter" n="1" id="JA0201">
    <head>Of divisions in authors</head>
    <p>There are certain mysteries or secrets in all trades, from the highest
      to the lowest, from that of <term>prime-ministering</term>, to this of
      <term>authoring</term>, which are seldom discovered unless to members of
      the same calling ... </p>
    <p>I will dismiss this chapter with the following observation: that it
      becomes an author generally to divide a book, as it does a butcher to
      joint his meat, for such assistance is of great help to both the reader
      and the carver. And now having indulged myself a little I will endeavour
      to indulge the curiosity of my reader, who is no doubt impatient to know
```

```

        what he will find in the subsequent chapters of this book.</p>
    </div>
    <div1 type="chapter" n="2" id="JA0202">
        <head>A surprising instance of Mr. Adams's short memory, with the
            unfortunate consequences which it brought on Joseph.
        </head>
        <p>Mr. Adams and Joseph were now ready to depart different ways ... </p>
    </div>
</div0>

```

7.1.4 Partial and Composite Divisions

In most situations, the textual subdivisions marked by `<div>` elements will be both complete and identically organized with reference to the original source. For some purposes however, in particular where dealing with unusually large or unusually small texts, encoders may find it convenient to present as textual divisions sequences of text which are incomplete with reference to the original text, or which are in fact an ad hoc agglomeration of tiny texts. Moreover, in some kinds of texts it is difficult or impossible to determine the order in which individual subdivisions should be combined to form the next higher level of subdivision, as noted below.

To overcome these problems, the following additional attributes are defined for all elements in the *divn* class:

org specifies how the content of the division is organized. Legal values are:

- composite** composite content: i.e. no claim is made about the sequence in which the immediate contents of this division are to be processed, or their inter-relationships.
- uniform** uniform content: i.e. the immediate contents of this element are regarded as forming a logical unit, to be processed in sequence.

sample indicates whether this division is a sample of the original source and if so, from which part.

Legal values are:

- initial** division lacks material present at end in source.
- medial** division lacks material at start and end.
- final** division lacks material at start.
- unknown** position of sampled material within original unknown.
- complete** division is not a sample.

part specifies whether or not the division is fragmented by some other structural element, for example a speech which is divided between two or more verse stanzas. Legal values are:

- Y** the division is incomplete in some respect
- N** either the division is complete, or no claim is made as to its completeness.
- I** the initial part of an incomplete division
- M** a medial part of an incomplete division
- F** the final part of an incomplete division

For example, an encoder might choose to transcribe only the first two thousand words of each chapter from a novel. In such a case, each chapter might conveniently be regarded as a partial division, and tagged with a `<div>` element in the following form:

```

    <div n="xx" sample="initial" part="Y" type="chapter">
        <p> ... </p>
    </div>

```

where “xx” represents a number for the chapter. The `<sampling>` element in the TEI Header should also be used to record the principles underlying the selection of incomplete samples, as further described in section 5.3.2 *The Sampling Declaration*.

The following example demonstrates how a newspaper column composed of very short unrelated snippets may be encoded using these attributes:

```

    <div1 type="storylist" org="composite">
    <head>News in brief</head>
    <div2 type="story">
    <head>Police deny <soCalled>losing</soCalled> bomb</head>

```

```

<p>Scotland Yard yesterday denied claims in the Sunday
Express that anti-terrorist officers trailing an IRA van
loaded with explosives in north London had lost track of
it 10 days ago.</p>
</div2>
<div2 type="story">
<head>Hotel blaze</head>
<p>Nearly 200 guests were evacuated before dawn
yesterday after fire broke out at the Scandic
Crown hotel in the Royal Mile, Edinburgh.</p>
</div2>
<div2 type="story">
<head>Test match split</head>
<p>Test Match Special next summer will be split
between Radio 5 and Radio 3, after protests this
year that it disrupted Radio 3's music schedule.</p>
</div2>
<!-- other stories here -->
</div1>

```

The `org` attribute on the `<div1>` element is used here to indicate that individual stories in this group, marked here as `<div2>`, are really quite independent of each other, although they are all marked as subdivisions of the whole group. They can be read in any order without affecting the sense of the piece; indeed, in some cases, divisions of this nature are printed in such a way as to make it impossible to determine the order in which they are intended to be read. Individual stories can be added or removed without affecting the existing components.

This method of encoding composite texts as composite divisions has some limitations compared with the more general and powerful mechanisms discussed in section 7.3 *Groups of Texts*. However, it may be preferable in some circumstances, notably where the individual texts are very small.

7.2 Elements Common to All Divisions

The divisions of any kind of text may sometimes begin with a brief heading or descriptive title, with or without a byline, an epigraph or brief quotation, or a salutation such as one finds at the start of a letter. They may also conclude with a brief trailer, byline, or signature. Elements which may appear in this way, either at the start or at the end of a text division proper, are regarded as forming a class, known as *divtop* or *divbot* respectively.

The following special-purpose elements are provided to mark features which may appear only at the start of a division:

<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.

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

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

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

For further details of the `<head>` element, see section 7.2.1 *Headings and Trailers*; for `<epigraph>` and `<argument>`, see section 7.2.3 *Arguments and Epigraphs*; for `<opener>`, see section 7.2.2 *Openers and Closers*.

The following special-purpose elements are provided to mark features which may appear only at the end of a division:

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

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

For further details of the `<trailer>` element, see section 7.2.1 *Headings and Trailers*; for the `<closer>` element, section 7.2.2 *Openers and Closers*.

7.2.1 Headings and Trailers

The `<head>` element is used to identify a heading prefixed to the start of any textual division, at any level. A given division may of course contain more than one such element, as in the following example:

```
<div1 n="Etym">
  <head>Etymology</head>
  <head>(Supplied by a late consumptive usher to a
  grammar school)</head>
  <p>The pale Usher &#x2014; threadbare in coat, heart,
  body and brain; I see him now. He was ever
  dusting his old lexicons and grammars, ...</p></div1>
```

Unlike some other markup schemes, the TEI scheme does *not* require that headings attached to textual subdivisions at different hierarchic levels have different identifiers. All kinds of heading are marked identically using the `<head>` tag; the type or level of heading intended is implied by the immediate parent of the `<head>` element, which may for example be a `<div1>`, `<div2>`, etc., an un-numbered `<div>`, or a `<list>`.

In certain kinds of text (notably newspapers), there may be a need to categorize individual headings within the sequence at the start of a division, for example as ‘main’ headings, or ‘detail’ headings. Specific elements are provided for certain kinds of heading-like features, (notably `<byline>`, `<dateline>`, and `<salute>`; see further section 7.2.2 *Openers and Closers*), but the type attribute must be used to discriminate among other forms of heading.

In the following example, taken from a British newspaper, the lead story and its associated headlines have been encoded as a `<div>` element, with appropriate *divtop* elements attached:

```
<div type="story">
  <head rend="large underlined" type="sub">
  President pledges safeguards for 2,400 British
  troops in Bosnia</head>
  <head rend="very large bold" type="main">
  Major agrees to enforced no-fly zone</head>
  <byline>By George Jones, Political Editor, in Washington</byline>
  <p>Greater Western intervention in the conflict in
  former Yugoslavia was pledged by President Bush ...</p></div>
```

In older writings, the headings or *incipits* may be longer than in modern works. When heading-like material appears in the middle of a text, the encoder must decide whether or not to treat it as the start of a new division. If the phrase in question appears to be more closely connected with what follows than with what precedes it, then it may be regarded as a heading and tagged as the `<head>` of a new `<div>` element. If it appears to be simply inserted or superimposed — as for example the kind of ‘pull quotes’ often found in newspapers or magazines, then the `<quote>`, `<q>`, or `<cit>` element may be more appropriate.

The `<trailer>` element, which can appear at the end of a division only, is used to mark any heading-like feature appearing in this position, as in this example:

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

7.2.2 Openers and Closers

In addition to headings of various kinds, divisions sometimes include more or less formulaic opening or closing passages, typically conveying such information as the name and address of the person to whom the division is addressed, the place or time of its production, a salutation or exhortation to the reader, and so on. Divisions in epistolary form are particularly liable to include such features. Additional elements for the detailed encoding of personal names, dates, and places are provided in chapter 20 *Names and Dates*. For simple cases, the following elements should be adequate:

- <byline>** contains the primary statement of responsibility given for a work on its title page or at the head or end of the work.
- <dateline>** contains a brief description of the place, date, time, etc. of production of a letter, newspaper story, or other work, prefixed or suffixed to it as a kind of heading or trailer.
- <salute>** contains a salutation or greeting prefixed to a foreword, dedicatory epistle, or other division of a text, or the salutation in the closing of a letter, preface, etc.
- <signed>** contains the closing salutation, etc., appended to a foreword, dedicatory epistle, or other division of a text.

The **<byline>** and **<dateline>** elements are used to encode headings which identify the authorship and provenance of a division. Although the terminology derives from newspaper usage, there is no implication that **<dateline>** or **<byline>** elements apply only to newspaper texts. The following example illustrates use of the **<dateline>** and **<signed>** elements at the end of the preface to a novel:

```
<div type="preface">
  <head>To Henry Hope.</head>
  <p>It is not because this volume was conceived and partly
  executed amid the glades and galleries of the Deepdene,
  that I have inscribed it with your name. ... I shall find a
  reflex to their efforts in your own generous spirit and
  enlightened mind.
</p>
<closer>
  <signed lang="el">D.</signed>
  <dateline>Grosvenor Gate, May-Day, 1844</dateline>
</closer>
</div>
```

Where a sequence of such elements appear together, either at the beginning or end of an element, it may be convenient to group them together using one of the following elements:

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

The following examples demonstrate the use of the **<opener>** and **<closer>** grouping elements:

```
<div type="narrative" n="6">
  <head>Sixth Narrative</head>
  <head>contributed by Sergeant Cuff</head>
  <div type="fragment" n="6.1">
    <opener>
      <dateline>
        <name type="place">Dorking, Surrey,</name>
        <date>July 30th, 1849</date>
      </dateline>
      <salute>To <name>Franklin Blake, Esq.</name> Sir, &mdash;</salute>
    </opener>
    <p>I beg to apologize for the delay that has occurred in the
    production of the Report, with which I engaged to furnish you.
    I have waited to make it a complete Report ...</p>
    <!-- .... -->
    <closer>
      <salute>I have the honour to remain, dear sir, your
      obedient servant </salute>
```

```

        <signed> <name>RICHARD CUFF</name> (late sergeant in the
        Detective Force, Scotland Yard, London). </signed>
    </closer>
</div>
</div>

<div type="letter" n="14">
<head>Letter XIV: Miss Clarissa Harlowe to Miss Howe</head>
<opener> <dateline>Thursday evening, March 2.</dateline> </opener>
<p>On Hannah's depositing my long letter ...</p>
<p>An interruption obliges me to conclude myself
in some hurry, as well as fright, what I must ever be,</p>
<closer>
<salute>Yours more than my own,</salute>
<signed>Clarissa Harlowe</signed>
</closer>
</div>

```

For further discussion of the encoding of names of persons and places and of dates, see section 6.4.4 *Dates and Times* and chapter 20 *Names and Dates*.

7.2.3 Arguments and Epigraphs

The `<argument>` element may be used to encode the prefatory list of topics sometimes found at the start of a chapter or other division. It is most conveniently encoded as a list, since this allows each item to be distinguished, but may also simply be presented as a paragraph. The following are thus both equally valid ways of encoding the same argument:

```

<div type='chap' n='6'>
  <argument>
    <p>Kingston &mdash; Instructive remarks on early English history
    &mdash; Instructive observations on carved oak and life in general
    &mdash; Sad case of Stivvings, junior &mdash; Musings on antiquity
    &mdash; I forget that I am steering &mdash; Interesting result
    &mdash; Hampton Court Maze &mdash; Harris as a guide.</p>
  </argument>
  <p>It was a glorious morning, late spring or early summer, as you
  care to take it ...</p>
</div>

<div type='chap' n='6'>
  <argument>
    <list type='inline'>
      <item>Kingston</item>
      <item>Instructive remarks on early English history</item>
      <item>Instructive observations on carved oak and life in
      general</item>
      <item>Sad case of Stivvings, junior</item>
      <item>Musings on antiquity</item>
      <item>I forget that I am steering</item>
      <item>Interesting result</item>
      <item>Hampton Court Maze</item>
      <item>Harris as a guide.</item>
    </list>
  </argument>
  <p>It was a glorious morning, late spring or early summer, as you
  care to take it ...</p>
</div>

```

An *epigraph* is a quotation from some other work appearing on a title page, or at the start of a division. It may be encoded using the special-purpose `<epigraph>` element. Its content will generally be a `<q>` or `<quote>` element, often associated with a bibliographic reference, as in the following example:

```

<div n='19' type="chap"><head>Chapter 19</head>
  <epigraph>
    <cit><quote>I pity the man who can travel
    from Dan to Beersheba, and say <q>'Tis all
    barren;</q> and so is all the world to him
    who will not cultivate the fruits it offers.

```

```

    </quote>
    <bibl>Sterne: Sentimental Journey.</bibl>
  </cit></epigraph>

```

```

<p>To say that Deronda was romantic would be to
misrepresent him: but under his calm and somewhat
self-repressed exterior ...</p>

```

For discussion of quotations appearing other than as epigraphs refer to section 6.3.3 *Quotation*.

7.2.4 Content of Textual Divisions

Other than its initial sequence of *divtop* elements, and its closing sequence of *divbot* elements, every textual division (numbered or un-numbered) consists of a sequence of ungrouped *component* elements (see 3.7 *Element Classes*). The actual elements available will depend on the base tag set in use; in all cases, at least the component-level structural elements defined in the core will be available (paragraphs, lists, dramatic speeches, verse lines and line groups etc.). If the drama base has been selected, then additionally the low level dramatic structural elements (speeches or stage directions, as defined in chapter 10 *Base Tag Set for Drama*) will be available. If the dictionary base is in use, then dictionary entries, related entries, etc. (as defined in chapter 12 *Print Dictionaries*) will also be available; if the tag set for transcribed speech is in use, then utterances, pauses, vocals, kinesics, etc., as defined in chapter 11.2 *Elements Unique to Spoken Texts* will be available; and so on.

Where a text contains low level elements from more than one base, two options are available. The first option, selected by the ‘mixed’ base, allows for low level structural elements from any or all of the selected bases to appear at any point. The second option, selected by the ‘general’ base, allows for low level structural elements from different bases to appear in different textual divisions of the same text, but requires that any one division use elements from only one base. For further information, refer to section 3.4 *Combining TEI Base Tag Sets*.

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

```

<!-- 7.2.4: Tags for start and end of divisions-->
<!ELEMENT trailer %om.R0; %phrase.seq;>
<!ATTLIST trailer
    %a.global;
    TEIform CDATA 'trailer' >
<!ELEMENT byline %om.R0; (#PCDATA | %m.phrase; | docAuthor | %m.Incl;)*>
<!ATTLIST byline
    %a.global;
    TEIform CDATA 'byline' >
<!ELEMENT dateline %om.R0; ( #PCDATA | date | time
    | name | address | %m.Incl; ) * >
<!ATTLIST dateline
    %a.global;
    TEIform CDATA 'dateline' >
<!ELEMENT argument %om.RR; ((%m.Incl;)*, (head?, %component.seq;))>
<!ATTLIST argument
    %a.global;
    TEIform CDATA 'argument' >
<!ELEMENT epigraph %om.RR; (%component.seq;)>
<!ATTLIST epigraph
    %a.global;
    TEIform CDATA 'epigraph' >
<!ELEMENT opener %om.R0; (#PCDATA | %m.phrase; | argument | byline |
    dateline | epigraph | salute | signed | %m.Incl;)* >
<!ATTLIST opener
    %a.global;
    TEIform CDATA 'opener' >
<!ELEMENT closer %om.R0; (#PCDATA | signed | dateline | salute
    | %m.phrase; | %m.Incl;)* >
<!ATTLIST closer
    %a.global;
    TEIform CDATA 'closer' >
<!ELEMENT salute %om.R0; %phrase.seq;>

```

```

<!ATTLIST salute
    %a.global;
    TEIform CDATA 'salute' >
<!ELEMENT signed %om.RO; %phrase.seq;>
<!ATTLIST signed
    %a.global;
    TEIform CDATA 'signed' >
<!--The HEAD element is declared in the core tag set.-->
<!-- end of 7.2.4-->

```

7.3 Groups of Texts

The <group> element should be used to represent a collection of independent texts which is to be regarded as a single unit for processing or other purposes. Examples of such composite texts include anthologies and other collections. The presence of common front matter referring to the whole collection, possibly in addition to front matter relating to each individual text, is a good indication that a given text might usefully be encoded as a <group>, though encoders may choose to use this structure to represent other kinds of composite texts as well.

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

For example, the overall structure of a collection of short stories might be encoded as follows:

```

<TEI.2>
  <teiHeader>
    <!-- header information for the whole collection -->
  </teiHeader>
  <text>
    <front>
      <docTitle><titlePart>
        The Adventures of Sherlock Holmes
      </titlePart></docTitle>
      <docImprint>First published in <title>The Strand</title>
        between July 1891 and December 1892</docImprint>
      <!-- Any other front matter specific to the collection here ... -->
    </front>
    <group>
      <text>
        <front>
          <head rend="italic">Adventures of Sherlock
            Holmes</head>
          <docTitle><titlePart>Adventure I. &mdash;</titlePart>
            <titlePart>A Scandal in Bohemia</titlePart></docTitle>
          <byline>By A. Conan Doyle.</byline>
        </front>
        <body>
          <p>To Sherlock Holmes she is always
            <emph>the</emph> woman. ... </p>
        </body>
      </text>
      <text>
        <front>
          <head rend="italic">Adventures of Sherlock Holmes</head>
          <docTitle><titlePart>Adventure II. &mdash;</titlePart>
            <titlePart>The Red-Headed League</titlePart></docTitle>
          <byline>By A. Conan Doyle.</byline>
        </front>
        <body>
          <p><!-- text of The Red-Headed League here --> </p>
        </body>
      </text>

    <!-- more texts here -->

```

```

<text>
  <front>
    <head rend="italic">Adventures of Sherlock Holmes</head>
    <docTitle><titlePart>Adventure XII. &mdash;</titlePart>
      <titlePart>The Adventure of the Copper Beeches</titlePart>
    </docTitle>
    <byline>By A. Conan Doyle.</byline>
  </front>
  <body>
    <p><q>&odq;To the man who loves art for its
      own sake,&cdq;</q> remarked Sherlock Holmes ...
    <!-- rest of the the Copper Beeches here -->
    ... she is now the head of a private school
    at Walsall, where I believe that she has
    met with considerable success.</p>
  </body>
</text>
<!-- end of the Copper Beeches -->
</group>
</text>
<!-- end of the Adventures of Sherlock Holmes -->
</TEI.2>

```

A text which is a member of a group may itself contain groups. This is quite common in collections of verse, but may happen in any kind of text. As an example, consider the overall structure of a typical collection, such as the *Muses Library* edition of Crashaw's poetry (ed. J.R. Tutin, [ca. 1900]). Following a critical introduction and table of contents, this work contains the following major sections:

- *Steps to the Temple* (a collection of verse first published in 1648)
- *Carmen deo Nostro* (a second collection, published in 1652)
- *The Delights of the Muses* (a third collection, published in 1648)
- *Posthumous Poems*, I (a collection of fragments all taken from a single manuscript)
- *Posthumous Poems*, II (a further collection of fragments, taken from a different manuscript)

Each of the three collections published in Crashaw's lifetime has a reasonable claim to be considered as a text in its own right, and may therefore be encoded as such. It is rather more arbitrary as to whether the two posthumous collections should be treated as two groups, following the practice of the *Muses Library* edition. An encoder might elect to combine the two into a single group, or simply to treat each fragment as an ungrouped unitary text.

The *Muses Library* edition reprints the whole of each of the three original collections, including their original front matter (title pages, dedications etc.). These should be encoded using the <front> element and its constituents (on which see further section 7.4 *Front Matter*), while the body of each collection should be encoded as a single <group> element. Each individual poem within the collections should be encoded as a distinct <text> element. The beginning of the whole collection would thus appear as follows (for further discussion of the use of the elements <div> and <lg> for textual subdivision of verse, see section 6.11.1 *Core Tags for Verse* and chapter 9 *Base Tag Set for Verse*):

```

<text>
  <front>
    <titlePage>
      <docTitle><titlePart>The poems of Richard Crashaw</titlePart></docTitle>
      <byline>Edited by J.R. Tutin</byline>
    <!-- ... -->
  </titlePage>
  <div type="preface"><head>Editor's Note</head>
    <p>A few words are necessary ... </p>
    <!-- ... -->
  </div>
</front>
<group>

```

```

<text>
  <front>
    <titlePage>
      <docTitle>
        <titlePart>Steps to the Temple, Sacred Poems</titlePart>
      </docTitle>
      <!-- ... -->
    </titlePage>
    <div type="address"><head>The Preface to the Reader</head>
      <p>Learned Reader, The Author's friend will not usurp much
        upon thy eye ... </p>
      <!-- ... -->
    </div>
  </front>
  <group>
    <text>
      <front>
        <docTitle><titlePart>Sospetto D'Herode</titlePart></docTitle>
      </front>
      <body>
        <div1 type="book" n="Herod I">
          <head>Libro Primo</head>
          <epigraph>
            <l>Casting the times with their strong signs</l>
            <!-- ... -->
          </epigraph>
          <lg n="I.1" type="stanza">
            <l>Muse! now the servant of soft loves no more</l>
            <l>Hate is thy theme and Herod whose unblest</l>
            <l>Hand (O, what dares not jealous greatness?) tore</l>
            <l>A thousand sweet babes from their mothers' breast,</l>
            <l>The blooms of martyrdom ...</l>
            <!-- ... -->
          </lg> </div1> </body> </text> <!-- end of Sospetto D'Herode -->
        <text>
          <front><docTitle><titlePart>The Tear</titlePart></docTitle></front>
          <body>
            <lg n="I">
              <l>What bright soft thing is this</l>
              <l>Sweet Mary, thy fair eyes' expense?</l>
              <!-- ... -->
            </lg> </body> </text> <!-- end of The Tear -->

            <!-- the remaining poems of the Steps to the Temple appear -->
            <!-- here, each within its own <text> element -->

          </group>
          <back> <!-- back matter for Steps to the Temple here --> </back>
        </text>
        <text> <!-- Carmen deo Nostro -->
          <front> <!-- ... --> </front>
          <group>
            <text> <!-- ... --> </text>
            <text> <!-- ... --> </text>
            <!-- more texts here -->
          </group>
        </text>
        <text> <!-- The delights of the Muses -->
          <group>
            <text> <!-- ... --> </text>
            <text> <!-- ... --> </text>
            <!-- more texts here -->
          </group>
        </text>
      <!-- ... -->
    </group>
  </text>

```

```

    </group>
    <back> <!-- back matter for the whole collection --> </back>
</text>

```

The <group> element may be used in this way to encode any kind of collection of which the constituents are regarded by the encoder as texts in their own right. Examples include anthologies of verse or prose by multiple authors, collections, florilegia or commonplace books, journals, day books, etc. As a fairly typical example, we consider *The Norton Book of Travel*, an anthology edited by Paul Fussell and published in 1987 by W. W. Norton. This work comprises the following major sections:

1. Front matter (title page, acknowledgments, introductory essay)
2. The Beginnings
3. The Eighteenth Century and the Grand Tour
4. The Heyday
5. Touristic Tendencies
6. Post Tourism
7. Back matter (permissions list, index)

Each titled section listed above comprises a group of extracts or complete texts from writers of a given historical period, preceded by an introductory essay. For example, the second group listed above contains, inter alia, the following:

1. Prefatory essay
2. Five letters by Lady Mary Wortley Montagu
3. An extract from Swift's *Gullivers Travels*
4. Two poems by Alexander Pope
5. Two extracts from Boswell's *Journal*
6. A poem by William Blake

Each group of writings by a single author is preceded by a brief biographical notice. Some of the extracts are quite lengthy, containing several chapters or other divisions; others are quite short. As the above list indicates, the texts included range across all kinds of material: verse, prose, journals and letters.

The easiest way of encoding such an anthology is to treat each individual extract as a text in its own right. A sequence of texts by a single author, together with the biographical note preceding it, can then be treated as a single <group> element within the larger <group> formed by the section. The sequence of single or composite texts making up a single section of the work is likewise treated, together with its prefatory essay, as a single <group> within the work. Schematically:

```

<text> <!-- the whole anthology -->
  <front>
    <!-- title page, acknowledgments, introductory essay for anthology -->
  </front>
  <group> <!-- 'body' of the anthology -->
    <group><head>The Beginnings</head>
      <!-- sequence of texts or groups -->
    </group>
    <group> <!-- The Eighteenth Century and the Grand Tour -->
      <text> <!-- prefatory essay by editor --> </text>
      <group> <!-- Lady Mary Wortley Montagu -->
        <text> <!-- biographical notice, by editor --> </text>
        <text> <!-- first letter --> </text>
        <text> <!-- second letter --> </text>
        <!-- ... -->
      </group> <!-- end of Montagu section -->
      <text> <!-- single text by Jonathan Swift -->
        <front> <!-- biographical notice, by editor --> </front>
        <body> <!-- ... --> </body>
      </text> <!-- end of Swift section -->
    </group> <!-- Alexander Pope -->

```



```

        <text> <!-- biographical notice, by editor --> </text>
        <text> <!-- first poem --> </text>
        <text> <!-- second poem --> </text>
    </group> <!-- end of Pope section -->
    <!-- ... -->
</group> <!-- end of 18th Century Section -->
<group><head>The Heyday</head>
    <!-- texts and subgroups ... -->
</group>
<!-- ... -->
</group> <!-- end of 'body' of anthology -->
<back> <!-- back matter for whole anthology --> </back>
</text> <!-- end of the anthology -->

```

Note that the editor's introductory essays on each author may be treated as texts in their own right (as the essays on Lady Mary Wortley Montagu and Alexander Pope have been treated above), or as front matter to the embedded text, as the essay on Swift has been. The treatment in the example is intentionally inconsistent, to allow comparison of the two approaches. Consistency can be imposed either by treating the Swift section as a <group> containing one text by Swift and one by the editor, or by treating the Montagu and Pope sections as <text> elements containing the editor's essays as front matter. Marked in the second way, the Pope section of the book would look like this:

```

<text>      <!-- Alexander Pope -->
    <front> <!-- biographical notice --> </front>
    <group>
        <text> <!-- first poem --> </text>
        <text> <!-- second poem --> </text>
        <!-- ... -->
    </group>
</text>    <!-- end of Pope section -->

```

The essays on "The Eighteenth Century and the Grand Tour" and other larger sections could also be tagged as 'front' matter in the same way, by treating the larger sections as <text> elements rather than <group> elements.

Where, as in this case, an anthology contains different kinds of text (for example, mixtures of prose and drama, or transcribed speech and dictionary entries, or letters and verse), the elements to be encoded may well need to be drawn from more than one of the base tag sets described in part II. In such a situation, either the mixed or the general base should be specified, as further described in section 3.4 *Combining TEI Base Tag Sets*. The elements provided by the core tag set described in chapter 6 *Elements Available in All TEI Documents* should however prove adequate for most simple purposes, where prose, drama, and verse are combined in a single collection.

For anthologies of short extracts such as commonplace books, it may often be preferable to regard each extract not as a text in its own right but simply as a quotation or <cit> element. The following component-level elements may be used to encode quotations of this kind:

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

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

For example, the chapter of 'extracts' which appears in the front matter of Melville's *Moby Dick* might be encoded as follows:

```

<div n="2" type="chap">
    <head>Extracts</head>
    <head>(Supplied by a sub-sub-Librarian)</head>
    <p>It will be seen that this mere painstaking burrower and
    grubworm of a poor devil of a Sub-Sub appears to have gone
    through the long Vaticans and street-stalls of the earth,
    picking up whatever random allusions to whales he could
    anyways find ...
    Here ye strike but splintered hearts together &dash; there,
    ye shall strike unsplinterable glasses!</p>
    <p>

```

```

<cit>
  <quote>And God created great whales.</quote>
  <bibl>Genesis</bibl>
</cit>
<cit>
  <quote>
    <l>Leviathan maketh a path to shine after him;</l>
    <l>One would think the deep to be hoary.</l>
  </quote>
  <bibl>Job</bibl>
</cit>
<!-- ... -->
<cit>
  <quote>By art is created that great Leviathan,
  called a Commonwealth or State &dash; (in Latin,
  <mentioned lang="lat">civitas</mentioned>), which
  is but an artificial man.</quote>
  <bibl>Opening sentence of Hobbes's Leviathan</bibl>
</cit>
</p>
</div>

```

For more information on the use of the `<quote>` and `<bibl>` elements, see sections 6.3.3 *Quotation* and 6.10 *Bibliographic Citations and References* respectively.

Where one or more whole texts are embedded within other texts, without necessarily forming a composite, the encoder may also choose to represent the nested structure directly. The `<text>` element is itself a *component* element, and thus can appear within any division level element in the same way as a paragraph. For example, texts such as the *Decameron* or the *Arabian Nights* might be regarded as sequences of discrete texts embedded within another single text, the framing narrative, rather than as groups of discrete texts in which the fragments of framing narrative are regarded as front matter.

7.4 Front Matter

By *front matter* we mean distinct sections of a text (usually, but not exclusively, a printed one), prefixed to it by way of introduction or identification as a part of its production. Features such as title pages or prefaces are clear examples; a less definite case might be the prologue attached to a play. The front matter of an encoded text should not be confused with the TEI header described in chapter 5 *The TEI Header*, which serves as a kind of front matter for the computer file itself, not the text it encodes.

An encoder may choose simply to ignore the front matter in a text, if the original presentation of the work is of no interest, or for other reasons; alternatively some or all components of the front matter may be thought worth including with the text as components of the `<front>` element.⁸⁴ With the exception of the title page, (on which see section 7.5 *Title Pages*), front matter should be encoded using the same elements as the rest of a text. As with the divisions of the text body, no other specific tags are proposed here for the various kinds of subdivision which may appear within front matter: instead either numbered or un-numbered `<div>` elements may be used. The following suggested values⁸⁵ for the type attribute may be used to distinguish various kinds of division characteristic of front matter:

- preface*** A foreword or preface addressed to the reader in which the author or publisher explains the content, purpose, or origin of the text.
- ack*** A formal declaration of acknowledgment by the author in which persons and institutions are thanked for their part in the creation of a text.
- dedication*** A formal offering or dedication of a text to one or more persons or institutions by the author.
- abstract*** A summary of the content of a text as continuous prose.

⁸⁴ This decision should be recorded in the `<sampling>` element of the header.

⁸⁵ As with all lists of 'suggested values' for attributes, it is recommended that software written to handle TEI-conformant texts be prepared to recognize and handle these values when they occur, without limiting the user to the values in this list.

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

frontispiece A pictorial frontispiece, possibly including some text.

The following extended example demonstrates how various parts of the front matter of a text may be encoded. The front part begins with a title page, which is presented in section 7.5 *Title Pages* below. This is followed by a dedication and a preface, each of which is encoded as a distinct `<div>`:

```
<div type='dedication'>
  <p>To my parents, Ida and Max Fish</p>
</div>
<div type='preface'><head>Preface</head>
  <p>The answer this book gives to its title question is <q>there is
    and there isn't</q>.</p>
  <!-- ... -->
  <p>Chapters 1&#x2013;12 have been previously published in the
    following journals and collections:
    <list>
      <item>chapters 1 and 3 in <title>New literary History</title></item>
      <!-- ... -->
      <item>chapter 10 in <title>Boundary II</title> (1980)</item>
    </list>.
    I am grateful for permission to reprint.</p>
  <signed>S.F.</signed>
</div>
```

The front matter concludes with another `<div>` element, shown in the next example, this time containing a table of contents, which contains a `<list>` element (as described in section 6.7 *Lists*). Note the use of the `<ptr>` element to provide page-references: the implication here is that the target identifiers supplied (P1, P68 etc.) may correspond with identifiers used either for `<div>` elements representing chapters of the text, or for `<pb>` elements marking page divisions of the text. (For the `<ptr>` element, see 6.6 *Simple Links and Cross References*.) Alternatively, the literal page numbers present in the source text might be transcribed, but they are likely to be of little direct use in work with the electronic text.

```
<div type='contents'>
  <head>Contents</head>
  <list>
    <item>Introduction, or How I stopped Worrying and Learned to Love
      Interpretation <ptr target='P1'></item>
    <item>
      <list>
        <head>Part One: Literature in the Reader</head>
        <item n='1'>Literature in the Reader: Affective Stylistics
          <ptr target='P21'></item>
        <item n='2'>What is Stylistics and Why Are They Saying Such
          Terrible Things About It? <ptr target='P68'></item>
      </list></item></list>
</div>
```

The following example uses numbered divisions to mark up the front matter of a medieval text. (Entity references are used to represent the characters thorn, yogh, and ampersand, as discussed in section 4.2 *Entry and display of characters*.) Note that in this case no title page in the modern sense occurs; the title is simply given as a heading at the start of the front matter. Note also the use of the type attribute on the `<div>` elements to indicate document elements comparatively unusual in modern books such as the initial prayer:

```
<front>

  <div1 type='incipit'>
    <p>Here bygynni&th; a book of contemplacyon, &th;e whiche
      is clepyd <title>&Th;E CLOWDE OF VNKNOWYNG</title>,
      in &th;e whiche a soule is onyd wi&th; GOD.</p>
  </div1>
```

```

<div1 type='prayer'>
  <head>Here biginne&th; &th;e preyer on &th;e prologe.</head>
  <p>God, unto whom alle hertes ben open, &amp; unto whome alle wille
    speki&th;; &amp; unto whom no priue &th;ing is hid: I beseche
    &th;ee so for to clyense &th;e entent of myn hert wi&th; &th;e
    unspekable &yog;ift of &th;i grace, &th;at I may parfiteliche
    loue &th;ee &amp; wor&th;ilich preise &th;ee. Amen.</p>
</div1>

<div1 type='preface'>
  <head>Here biginne&th; &th;e prolog.</head>
  <p>In &th;e name of &th;e Fader &amp; of &th;e Sone &amp;
    of &th;e Holy Goost.</p>
  <p>I charge &th;ee &amp; I beseche &th;ee, wi&th; as moche
    power &amp; vertewe as &th;e bonde of charite is sufficient
    to suffre, what-so-euer &th;ou be &th;at &th;is book schalt
    haue in possession ...</p>
</div1>

<div1 type='contents'>
  <head>Here biginne&th; a table of &th;e chapitres.</head>
  <list>
    <label>&th;e first chapitre </label>
    <item>Of foure degrees of Cristen mens leuing; &amp; of &th;e
      cours of his cleping &th;at &th;is book was maad vnto.</item>
    <label>&th;e secound chapitre</label>
    <item>A schort stering to meeknes &amp; to &th;e werk of &th;this
      book</item>
    <!-- ... -->
    <label>&th;e fiue and seuenti chapitre</label>
    <item>Of somme certain tokenes bi &th;e whiche a man may proue
      whe&th;er he be clepid of God to worche in &th;is werk.</item>
  </list>
  <trailer>&amp; here eende&th; &th;e table of &th;e chapitres.</trailer>
</div1>

</front>

```

7.5 Title Pages

Detailed analysis of the title page and other *preliminaries* of older printed books and manuscripts is of major importance in descriptive bibliography and the cataloguing of printed books; such analysis may require a rather more detailed tag set than that proposed here.⁸⁶ The following elements are therefore proposed as an interim measure; they constitute a useful descriptive tag set for the major features of most title pages:

<titlePage> contains the title page of a text, appearing within the front or back matter. Attributes include:

type classifies the title page according to any convenient typology.

Values Any string, e.g. 'full', 'half', 'Series', etc.

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

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

Attributes include:

type specifies the role of this subdivision of the title.

Suggested values include:

main main title of the work

sub subtitle of the work

alt alternative title of the work

⁸⁶ Definition of such a tag set remains a work item for the TEI; such tag sets for contemporary printed matter already exist or are being created within the publishing industry, for example the Majour (Modular Application for Journals) Project of the European Workgroup on SGML. See for example *MAJOUR: Modular Application for Journals: DTD for Article Headers* ([n.p.]: EWS, 1991).

- desc descriptive paraphrase of the work included in title
- <argument>** A formal list or prose description of the topics addressed by a subdivision of a text.
- <byline>** contains the primary statement of responsibility given for a work on its title page or at the head or end of the work.
- <docAuthor>** contains the name of the author of the document, as given on the title page (often but not always contained in a **<byline>**).
- <epigraph>** contains a quotation, anonymous or attributed, appearing at the start of a section or chapter, or on a title page.
- <imprimatur>** contains a formal statement authorizing the publication of a work, sometimes required to appear on a title page or its verso.
- <docEdition>** contains an edition statement as presented on a title page of a document.
- <docImprint>** contains the imprint statement (place and date of publication, publisher name), as given (usually) at the foot of a title page.
- <docDate>** contains the date of a document, as given (usually) on a title page. Attributes include:
- value** gives the value of the date in a standard form.
- Values* a date in ISO standard form, generally ISO 8601:2000 5.2.1.1 Complete representation, extended format (yyyy-mm-dd). If the date in the source was not in the Gregorian calendar it should be converted to Gregorian or proleptic Gregorian here; alternatively, the fact that it was not converted should be documented in the **<stdVals>** element in the TEI header.

Together with the **<figure>** element described in chapter 22 *Tables, Formulae, and Graphics*, these elements constitute the element class **tpParts**, which is defined by the parameter entity **m.tpParts**. Any number of elements from this class can appear grouped together within a **<titlePage>** element. (The **<figure>** element is included so as to enable encoders to record the presence of printers' ornaments or other illustrative material found within a title page; its use implies that the TEI tagset for figures and tables has been selected, as discussed in chapter 22 *Tables, Formulae, and Graphics*).

The elements listed above, together with the **<head>** element, also constitute the element class **fmchunk**, which is defined by the parameter entity **m.fmchunk**. The elements in this class can appear within a 'minimal' **<front>** element without any need to group them together and encode a complete title page.

Encoders wishing to add new elements to either class may do so by modifying or redefining this parameter entity, as further described in chapter 29 *Modifying and Customizing the TEI DTD*. Two examples of the use of these elements follow. First, the title page of the work discussed earlier in this section:

```
<front>
  <titlePage>
    <docTitle>
      <titlePart type="main">Is There a Text in This Class?</titlePart>
      <titlePart type="sub">The Authority of Interpretive Communities</titlePart>
    </docTitle>
    <docAuthor>Stanley Fish</docAuthor>
    <docImprint>
      <publisher>Harvard University Press</publisher>
      <pubPlace>Cambridge, Massachusetts</pubPlace>
      <pubPlace>London, England</pubPlace>
    </docImprint>
  </titlePage></front>
```

Second, a characteristically verbose 17th century example. Note the use of the **<lb>** tag to mark the line breaks of the original where necessary:

```
<titlePage>
  <docTitle>
    <titlePart type="main">THE
      <lb/>Pilgrim's Progress
      <lb/>FROM
      <lb/>THIS WORLD,
      <lb/>TO
      <lb/>That which is to come:</titlePart>
```

```

<titlePart type="sub">Delivered under the Similitude of a
  <lb/>DREAM</titlePart>
<titlePart type="desc">Wherein is Discovered,
  <lb/>The manner of his setting out,
  <lb/>His Dangerous Journey; And safe
  <lb/>Arrival at the Desired Countrey.</titlePart>
</docTitle>
<epigraph>
  <cit><q>I have used Similitudes,</q><bibl>Hos. 12.10</bibl></cit>
</epigraph>
<byline>By <docAuthor>John Bunyan</docAuthor>.</byline>
<imprimatur>Licensed and Entred according to Order.</imprimatur>
<docImprint>
  <pubPlace>LONDON,</pubPlace>
  Printed for <name>Nath. Ponder</name>
  <lb/>at the <name>Peacock</name> in the <name>Poultrey</name>
  <lb/>near <name>Cornhil</name>, <docDate>1678</docDate>.
</docImprint></titlePage>

```

Those elements in the above list which are not defined elsewhere have the following formal declarations:

```

<!-- 7.5: Tags for title pages-->
<!ELEMENT titlePage %om.R0; ((%m.Incl;)*, (%m.tpParts;),
  (%m.tpParts; | %m.Incl;)* ) >
<!ATTLIST titlePage
  %a.global;
  type CDATA #IMPLIED
  TEIform CDATA 'titlePage' >
<!ELEMENT docTitle %om.R0; ((%m.Incl;)*, (titlePart, (%m.Incl;)*)+)>
<!ATTLIST docTitle
  %a.global;
  TEIform CDATA 'docTitle' >
<!ELEMENT titlePart %om.R0; %paraContent;>
<!ATTLIST titlePart
  %a.global;
  type CDATA "main"
  TEIform CDATA 'titlePart' >
<!ELEMENT docAuthor %om.R0; %phrase.seq;>
<!ATTLIST docAuthor
  %a.global;
  TEIform CDATA 'docAuthor' >
<!ELEMENT imprimatur %om.R0; %paraContent;>
<!ATTLIST imprimatur
  %a.global;
  TEIform CDATA 'imprimatur' >
<!ELEMENT docEdition %om.R0; %paraContent;>
<!ATTLIST docEdition
  %a.global;
  TEIform CDATA 'docEdition' >
<!ELEMENT docImprint %om.R0; (#PCDATA | %m.phrase; | pubPlace |
  docDate | publisher | %m.Incl;)* >
<!ATTLIST docImprint
  %a.global;
  TEIform CDATA 'docImprint' >
<!ELEMENT docDate %om.R0; %phrase.seq;>
<!ATTLIST docDate
  %a.global;
  value %ISO-date; #IMPLIED
  TEIform CDATA 'docDate' >
<!-- end of 7.5-->

```

Where title pages are encoded, their physical rendition is often of considerable importance. One approach to this requirement would be to use the <seg> tag, described in chapter 14 *Linking, Segmentation, and Alignment*, to segment the typographic content of each part of the title page, and then use the global *rend* attribute to specify its rendition. Another would be to use a tag set specialized for the description of typographic entities such as pages, lines, rules, etc., bearing special-purpose attributes to describe line

height, leading, degree of kerning, font, etc. Further discussion of these problems is provided in chapter 18 *Transcription of Primary Sources*.

Front matter elements are defined in a distinct DTD file called *TEIfron2.dtd*.

```
<!-- 7.5: Additional Tag Set for Front Matter-->
<!ELEMENT front %om.R0; ( (%m.front; | %m.Incl;)*,
                           (( (%m.fmchunk;), (%m.fmchunk; | titlePage | %m.Incl;)* )
                            | ( div, (div | %m.front; | %m.Incl;)* )
                            | ( div1, (div1 | %m.front; | %m.Incl; )*) )? ) >
<ATTLIST front
    %a.global;
    %a.declaring;
    TEIform CDATA 'front' >
<!--declarations from 7.5: Tags for title pages inserted here -->
<!-- end of 7.5-->
```

7.6 Back Matter

Conventions vary as to which elements are grouped as back matter and which as front. For example, some books place the table of contents at the front, and others at the back. Even title pages may appear at the back of a book as well as at the front. The content model for <back> and <front> elements are therefore identical.

The following suggested values may be used for the type attribute on all division elements, in order to distinguish various kinds of division characteristic of back matter:

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

glossary A list of terms associated with definition texts (“glosses”): this should be encoded as a <list type=“gloss”> (see section 6.7 *Lists*).

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

bibliogr A list of bibliographic citations: this should be encoded as a <listBibl> (see section 6.10 *Bibliographic Citations and References*).

index Any form of index to the work.

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

No additional elements are proposed for the encoding of back matter at present. Some characteristic examples follow; first, an index (for the case in which a printed index is of sufficient interest to merit transcription):

```
<back>
<div type="index">
  <head>Index</head>
  <list type="index">
    <item>Actors, public, paid for the contempt attending
      their profession, <ptr target="P209"/></item>
    <item>Africa, cause assigned for the barbarous state of
      the interior parts of that continent, <ptr target="P125"/></item>
    <item>Agriculture
      <list type="indexentry">
        <item>ancient policy of Europe unfavourable to, <ptr target="P371"/></item>
        <item>artificers necessary to carry it on, <ptr target="P481"/></item>
        <item>cattle and tillage mutually improve each other, <ptr target="P325"/></item>
      <!-- ... -->
        <item>wealth arising from more solid than that which proceeds
          from commerce <ptr target="P520"/></item>
      </list></item>
    <item>Alehouses, not the efficient cause of drunkenness, <ptr target="P461"/></item>
  <!-- ... -->
  </list>
</div>
```

```
<!-- ... -->
</back>
```

Next, a back-matter division in epistolary form:

```
<back>
  <div type="letter">
    <head>A letter written to his wife, founde with this booke
      after his death.</head>
    <p>The remembrance of the many wrongs offred thee, and thy
      unreproued vertues, adde greater sorrow to my miserable state,
      than I can utter or thou conceiue. ...
      ... yet trust I in the world to come to find mercie, by the
      merites of my Saiuour to whom I commend thee, and commit
      my soule.</p>
    <signed>Thy repentant husband for his disloyaltie,
      <name>Robert Greene.</name></signed>
    <epigraph lang="LA"><p>Faelicem fuisse infaustum</p></epigraph>
    <trailer>FINIS</trailer>
  </div>
  <!-- ... -->
</back>
```

And finally, a list of corrigenda and addenda with pseudo-epistolary features:

```
<back>
  <div type="corrigenda">
    <head>Addenda</head>
    <salute lang="LA">M. Scriblerus Lectori</salute>
    <p>Once more, gentle reader I appeal unto thee, from the shameful
      ignorance of the Editor, by whom Our own Specimen of
      <name>Virgil</name> hath been mangled in such miserable manner,
      that scarce without tears can we behold it. At the very
      entrance, Instead of <q lang='GR'>prolego/mena</q>, lo! <q
        lang='GR'>prolegw/mena</q> with an Omega! and in the same line
      <q lang='LA'>consul&acirc;s</q> with a circumflex! In the next
      page thou findest <q lang="LA">leviter perlabere</q>, which his
      ignorance took to be the infinitive mood of <q
        lang="LA">perlabor</q> but ought to be <q
        lang="LA">perlabi</q> ... Wipe away all these monsters,
      Reader, with thy quill.</p>
  </div>
</back>
```

The `<back>` element is defined in file *TEIback2.dtd*; since there are no other specialized back-matter tags, nothing else is defined there.

```
<!-- 7.6: Tags for Back Matter-->
<!ELEMENT back %om.R0; ( (%m.front; | %m.Incl;)*,
  ( ( (%m.divtop;), (%m.divtop; | titlePage | %m.Incl;)* )
  | ( div, (div | %m.front; | %m.Incl;)* )
  | ( div1, (div1 | %m.front; | %m.Incl;)* ) )? ,
  (trailer | closer)* ) >
<!ATTLIST back
  %a.global;
  %a.declaring;
  TEIform CDATA 'back' >
<!-- end of 7.6-->
```

7.7 DTD Fragment for Default Text Structure

The DTD fragment described by the present chapter is found in file *teistr2.dtd*; it has the following overall structure:

```
<!-- 7.7: Default text structure-->
<!--This definition of the basic text structure is used by most
TEI base tag sets; some bases, however, use slight variations
upon it.-->
<!--declarations from 7.: Top-level parts of default structure inserted here -->
```



```
<!--declarations from 7.1.1: Un-numbered divisions inserted here -->
<!--declarations from 7.1.2: Numbered divisions inserted here -->
<!--declarations from 7.2.4: Tags for start and end of divisions inserted here -->
<!--Front matter is defined in TEI.front file.-->
<!ENTITY % TEI.front.dtd PUBLIC "-//TEI P4//ELEMENTS Front Matter//EN"
'teifron2.dtd' >%TEI.front.dtd;
<!--Back matter is defined in TEI.back file.-->
<!ENTITY % TEI.back.dtd PUBLIC "-//TEI P4//ELEMENTS Back Matter//EN"
'teiback2.dtd' >%TEI.back.dtd;
<!-- end of 7.7-->
```


III: Base Tag Sets

