

Structure and Layout

Marjorie Burghart

In this chapter you will learn:

- the general structure of a TEI document;
- how to represent the physical and logical structure of your document and/or text;
- a first hint at the resolution of concurrent hierarchies.

1. General structure of a TEI document

A TEI document always contains two mandatory parts: the header (**<teiHeader>**) which contains metadata about the document and its content, and the representation of the text itself (**<text>**). Those rules apply not only for editions, but for any kind of TEI document. In this chapter, we will only insist on the parts which are most essential for editing ancient texts and documents, without going over all the possibilities. For a general overview we recommend to read Burnard 2014, ch. [The structural organization of a TEI document](#) and [The TEI Header](#), and of course the TEI Guidelines. Note that, besides the mandatory header and text, the TEI root may also contain the **<facsimile>** and **<sourceDoc>** elements, which are mostly useful for documentary editions or a *dossier génétique*: they will be explained in chapter Digital facsimiles.

1.1. Header

The header (**<teiHeader>**) is particularly important as the ID card of your work, especially when it is distributed to others, or added to a corpus or digital library. It presents all the metadata in four sections.

The **<fileDesc>** (file description) “contains a full bibliographic description of an electronic file.”¹ This essential part of the header is subdivided into subsections, three of which are mandatory:

- **<titleStmt>** (title statement) “groups information about the title of a work and those responsible for its content”² It must contain at least one **<title>** element, used to indicate the title of the text edited here. It may be followed by an **<author>** element,

1. <http://www.tei-c.org/Vault/P5/3.0.0/doc/tei-p5-doc/en/html/ref-fileDesc.html>

2. <http://www.tei-c.org/Vault/P5/3.0.0/doc/tei-p5-doc/en/html/ref-titleStmt.html>

containing the name of the original author of the text you are editing, and an `<editor>` element to indicate the name of the scholarly editor of the text. A `<principal>` element is also available, to supply “the name of the principal researcher responsible for the creation” of this document, that the principal investigator who bears the scientific responsibility but may not be the actual editor.³ If you need to indicate more roles, for a collaborative effort for instance, you can use the `<respStmt>` element (statement of responsibility):⁴ each `<respStmt>` contains a `<resp>` and one or more `<name>` (or `<persName>`) elements, which supply respectively a description of the role played and the name of the people who played it. For instance, let us imagine that a fictional Dr. Anna Rossi created a digital scholarly edition of Cicero’s *De Amicitia*, a project for which she assigned tasks to her assistants John and Lenka. A basic `<titleStmt>` could be:

```
<titleStmt>
  <title>De Amicitia</title>
  <author>Marcus Tullius Cicero</author>
  <editor xml:id="arossi">Anna Rossi</editor>
  <respStmt xml:id="lnovakova">
    <resp>Identification of person names mentioned in the text,
and redaction of short biographies of these persons</resp>
    <persName>Lenka Nováková</persName>
  </respStmt>
  <respStmt xml:id="jsmith">
    <resp>Proofreading</resp>
    <persName>John Smith</persName>
  </respStmt>
</titleStmt>
```

Note that we have given a unique `@xml:id` value to each of these roles. This will allow us to point to these roles (and people) later, when we want to give them responsibility for an action, interpretation or change in the course of the creation or revision of the edition.

- `<publicationStmt>` (publication statement) “groups information concerning the publication or distribution” of the text.⁵ In its simplest form, it may consist in a simple prose paragraph (`<p>`). Alternatively it can contain elements dedicated to supply the name of the publisher (`<publisher>`), place (`<pubPlace>`) and date (`<date>`) of publication, and most importantly the conditions of availability of the text (`<availability>`), where you can indicate in a prose paragraph under which licence, if any, the text is distributed.
- `<sourceDesc>` (source description) “describes the source from which an electronic text was derived or generated”⁶ For scholarly editions, there can be two cases: the TEI

3. <http://www.tei-c.org/Vault/P5/3.0.0/doc/tei-p5-doc/en/html/ref-principal.html>

4. <http://www.tei-c.org/Vault/P5/3.0.0/doc/tei-p5-doc/en/html/ref-respStmt.html>

5. <http://www.tei-c.org/Vault/P5/3.0.0/doc/tei-p5-doc/en/html/ref-publicationStmt.html>

6. <http://www.tei-c.org/Vault/P5/3.0.0/doc/tei-p5-doc/en/html/ref-sourceDesc.html>

document may be a “born digital” edition, and the **<sourceDesc>** will contain the description of the original document (or documents) on which the edition is based.⁷ Or the TEI document may be the digital version of a pre-existing edition, and the **<sourceDesc>** will contain a bibliographic description of the book or article used as the source for this version. In either case, the simplest form of description is a simple prose paragraph (**<p>**), but it can be refined into a very elaborate manuscript or bibliographic description.

The other, non mandatory parts of **<fileDesc>** are **<editionStmt>**, **<extent>**, **<notesStmt>** and **<seriesStmt>**. We invite you to check their definitions in the Guidelines, but they are less crucial to scholarly editions than the elements described above.

The **<encodingDesc>** (encoding description) “documents the relationship between an electronic text and the source or sources from which it was derived.”⁸ In its simplest form, it can consist of a simple prose paragraph (**<p>**), but it can also be much more detailed, with several sections, one of which is the **<projectDesc>** (project description) which “describes in detail” (in simple prose paragraphs) “the aim or purpose for which an electronic file was encoded, together with any other relevant information concerning the process by which it was assembled or collected.”⁹ You can explain there the context of your edition (student assignment, part of a thesis, exercise, professional work, etc.), what you are willing to achieve, etc. Another optional section particularly relevant to critical edition is **<variantEncoding>** which “declares the method used to encode text-critical variants”¹⁰ (you will learn about the possible methods in chapter Textual Variants). For the other possible sections of **<encodingDesc>**, we refer you to the relevant sections of the Guidelines.

The **<profileDesc>** (text-profile description) “provides a detailed description of non-bibliographic aspects of a text, specifically the languages and sublanguages used, the situation in which it was produced, the participants and their setting.” When present, it contains a simple prose paragraph (**<p>**).¹¹

The **<revisionDesc>** (revision description) “summarizes the revision history for a file”¹² In this section, you have the possibility to record all the revisions of the current file. Each revision will be represented by a **<change>** element containing a prose description of the change made, and complete with a set of optional attributes: **@when** to indicate the date of the change, **@who** to indicate the person(s) responsible for this change typically by pointing to the element giving their full name, or **@target** which “points to one or more elements that belong to this change” (for instance when a change consisted in adding some elements).¹³ If desirable, related **<change>** elements can be grouped together in a **<listChange>** element.¹⁴ The recommended practice is

7. To learn more about the encoding of manuscript descriptions, see chapter Manuscript Description

8. <http://www.tei-c.org/Vault/P5/3.0.0/doc/tei-p5-doc/en/html/ref-encodingDesc.html>

9. <http://www.tei-c.org/Vault/P5/3.0.0/doc/tei-p5-doc/en/html/ref-projectDesc.html>

10. <http://www.tei-c.org/Vault/P5/3.0.0/doc/tei-p5-doc/en/html/ref-variantEncoding.html>

11. <http://www.tei-c.org/Vault/P5/3.0.0/doc/tei-p5-doc/en/html/ref-profileDesc.html>

12. <http://www.tei-c.org/Vault/P5/3.0.0/doc/tei-p5-doc/en/html/ref-revisionDesc.html>

13. <http://www.tei-c.org/Vault/P5/3.0.0/doc/tei-p5-doc/en/html/ref-change.html>

14. <http://www.tei-c.org/Vault/P5/3.0.0/doc/tei-p5-doc/en/html/ref-listChange.html>

to list the changes in reverse chronological order (most recent first). Going back to our example of the fictional edition of *De amicitia*, let us imagine that John Smith proofread the first version on June 5th 2017, then Lenka Novákova marked up the person names present in the text, without providing a full identification yet, on June 9th 2017. Here is how we could translate this information in the `<revisionDesc>`:

```
<revisionDesc>
  <change when="2017-06-09" who="#jnovakova">Added markup for each
person name mentioned in the text.</change>
  <change when="2017-06-05" who="#jsmith">Proofread the entire
text.</change>
</revisionDesc>
```

Note that, to point to an element from an attribute that allows this, like `@when` here, we use the value of the `@xml:id` of the attribute we are pointing to, but preceded with a `#` sign. This is a general rule when pointing to an attribute, whatever the target element or attribute used to point to it.

1.2. Text

Placed after the `<teiHeader>`, the `<text>` element is the other mandatory component of any TEI document. As the name suggests, it contains the text of the edition, which may be divided into three parts: `<front>`, `<body>`, and `<back>`. There can be different approaches regarding the distribution of the text among those parts:

- If you are preparing a digital facsimile of a book, you might choose to put all prefatory matter of the edition you are encoding (preface, dedication, preliminary letters, etc.) in the `<front>`. In this case, the text in the `<front>` and `<back>` as well as the `<body>` is the ancient text you are working on, and your own contribution (for instance to describe the project that led to the creation of this document) will be restricted to the `<teiHeader>`.
- If you are preparing a digital scholarly edition, you may consider that the TEI document represents not only the ancient text, but your edition as a whole, with its own front and back matter added by the critical editors. You could therefore choose to put your introduction, discussion of the choice of witnesses, etc. in the `<front>`, so that the `<body>` contains only the ancient text you have edited. At any rate, be careful not to mix together, without a clear distinction, the ancient text and your own prose as editor: they should not, for instance, be together in the `<body>` of the document. Keeping the edited text clearly separated from the editor's commentary and introduction will, for instance, allow for linguistic processing of the ancient text.

1.2.1. Front matter

The `<front>` (front matter) is an optional element which “contains any prefatory matter (headers, abstracts, title page, prefaces, dedications, etc.) found at the start of a document, before

the main body”¹⁵ The **<front>** may contain paragraphs of prose (**<p>**), possibly grouped in **<div>** elements. They may also contain other elements which are too specific to be presented here: we invite you to consult the Guidelines if you wish to know more about the possible contents of the **<front>**.

1.2.2. Body

The **<body>** is the only mandatory part of the **<text>**.

1.2.3. Back matter

The **<back>** (back matter) is, like **<front>**, an optional part of **<text>**. It “contains any appendixes, etc. following the main part of a text.”¹⁶ In the context of an edition, for instance, you could decide to put there your bibliography, or any type of index-like information.

To summarise, here is, on the following page, how we could structure the TEI edition, by a fictional Jane Doe, of The Saga of Haakon Haakonarson or Hákonar saga Hákonarsonar, an Old Norse saga written in the 1260s by the Icelandic historian Sturla Þórðarson. Note that we can use the **@xml:lang** on any relevant element to indicate in which language it is written. In this case, this allows us to indicate two versions of the title: one in the original language, Old Norse, which language code is “non,” and the English translation. Accordingly, most of the content of **<text>** is in Old Norse, but the front and back matter are added, in English, by the editor. In this case, we can indicate on **<text>** an **@xml:lang** with the value “non,” and this value will be inherited by all the descendants of **<text>**, unless they have their own, different value for **@xml:lang**.

Now that we are familiar with the basic structure of a TEI document, it is time to move to more important matters: how to represent, in a TEI document, the logical and physical structure of the text or document that we wish to edit.

15. <http://www.tei-c.org/Vault/P5/3.0.0/doc/tei-p5-doc/en/html/ref-front.html>

16. <http://www.tei-c.org/Vault/P5/3.0.0/doc/tei-p5-doc/en/html/ref-back.html>

```

<TEI @xmlns="http://www.tei-c.org/ns/1.0">
  <teiHeader>
    <fileDesc>
      <titleStmt>
        <title xml:lang="non">Hákonar saga Hákonarsonar</title>
        <title xml:lang="en">The Saga of Haakon Haakonarson</title>
        <author xml:lang="non">Sturla Þórðarson</author>
        <editor xml:id="jdoe">Jane Doe</editor>
      </titleStmt>
      <publicationStmt>
        <p>Publication Information would go here.</p>
      </publicationStmt>
      <sourceDesc>
        <p>Information about the source would go here.</p>
      </sourceDesc>
    </fileDesc>
    <encodingDesc>
      <projectDesc>
        <p>Optional description of the project's aim, scope, etc. </p>
      </projectDesc>
      <variantEncoding method="parallel-segmentation"
location="internal"/>
      <!-- The variantEncoding information is relevant only for
critical editions recording the variants of several witnesses. See
chapter "Textual Variants" -->
    </encodingDesc>
    <profileDesc>
      <p>Optional description of the language of the text, or the
circumstances of its production, </p>
    </profileDesc>
  </teiHeader>
  <text xml:lang="non">
    <front xml:lang="en">
      <p>Optional front matter, such as an introduction, discussion of
the witnesses, etc. could go here.</p>
    </front>
    <body>
      <p>The main text of the edition would go here.</p>
    </body>
    <back xml:lang="en">
      <p>Optional back matter, such as a bibliography, or index-like
information could go here.</p>
    </back>
  </text>
</TEI>

```

2. Representing the structure of a text or document

Texts have a logical structure: they have different parts, which may be arranged sequentially and / or hierarchically. For instance, a treatise may have a prologue, then several main sections called “books,” containing each several chapters, containing each in their turn several paragraphs. But, as philologists, the texts we edit are not born-digital: they are transmitted to us through objects like charters, codices or early printed books, which have their own physical structure. For instance, a manuscript codex comprises several folios, each of which may contain columns, and lines.

Of course, those logical and physical structures rarely coincide in ancient documents. Depending on your type of edition, you might be interested mostly in one or the other, but the TEI lets you encode both in a seamless way.

2.1. The logical structure

2.1.1. Main subdivisions of the text

The `<div>` (text division) “contains a subdivision of the front, body, or back of a text.”¹⁷ It is a powerful and versatile tool to represent the structure of a text: a `<div>` is a neutral subdivision, which can correspond to any relevant unit: book, section, chapter, subdivision of a chapter... In each case the same element is used, but the `@type` attribute can indicate the nature of the `<div>`. A `<div>` may contain other `<div>` elements, or paragraphs (`<p>`) and paragraph-like elements like `<ab>`, `<list>`, etc.,¹⁸ but never directly text.

For instance, a treatise containing a prologue and two books, comprising respectively two and three chapters could be represented like this:

```
<text>
  <body>
    <div type="prologue" xml:id="pro"/>
    <div type="book" xml:id="boo1">
      <div type="chapter" xml:id="cha1.1"/>
      <div type="chapter" xml:id="cha1.2"/>
    </div>
    <div type="book" xml:id="boo2">
      <div type="chapter" xml:id="cha2.1"/>
      <div type="chapter" xml:id="cha2.2"/>
      <div type="chapter" xml:id="cha2.3"/>
    </div>
  </body>
</text>
```

17. <http://www.tei-c.org/Vault/P5/3.0.0/doc/tei-p5-doc/en/html/ref-div.html>

18. For a full list of such elements, see <http://www.tei-c.org/Vault/P5/3.0.0/doc/tei-p5-doc/en/html/ref-model.pLike.html>

Note that we have added an `@xml:id` attribute to each `<div>`. Those attributes are optional, but may prove useful in certain situations. Each `@xml:id` must have a unique value (within the same document) and conform to certain rules (the value cannot start with a number, for instance). This attribute is used to identify without ambiguity the element. When an element has an `@xml:id`, other elements can point to it, for cross-references for instance. When processing and displaying the edition, the value of `@xml:id` can be used to link directly to an element, from the table of contents or from an index for instance.

Subdivisions of a text, like chapters, often have a title. In manuscripts, those titles are often represented as “rubrics,” which may be in a different colour from the main text. The title of a `<div>` may be encoded with a `<head>` (heading) element¹⁹ The `<head>` element, which is optional, must be the first to appear in the `<div>` (paragraphs or other `<div>` elements must come after the `<head>`). For instance, on this image, we can see the end of a section, a rubric in red, with slightly larger letters, giving the title of the new section (“Noua uectigalia institui non posse,” that is “New taxes cannot be established”), then the actual text of this new section begins with a drop capital:

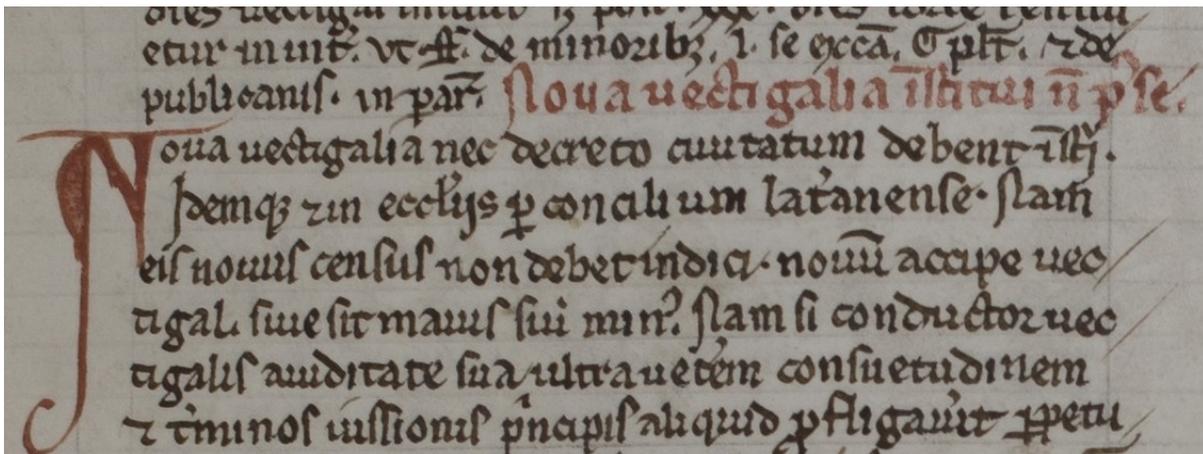


Figure 1: Klosterneuburg, Stiftsbibliothek, CCI 119, fol. 86r

We could represent this section, with its title, as follows:

```
<div>
  <head>Noua uectigalia institui non posse</head>
  <p>Noua uectigalia nec decreto ciuitatum debent institui...</p>
</div>
```

Sometimes, there may be more than a single heading, for instance “Distinctio XXIX. An in diuinis sit principium?” (that is, “Distinction XXIX. Is there a beginning to divine matters?”). In such cases, you may use more than one `<head>`, and differentiate them with a `@type` attribute if necessary:

19. <http://www.tei-c.org/Vault/P5/3.0.0/doc/tei-p5-doc/en/html/ref-head.html>

```

<div>
  <head type="section_number">Distinctio XXIX.</head>
  <head type="section_title">An in diuinis sit principium?</head>
  <p/>
</div>

```

The opening formula starting with “Incipit...” in Latin (or the equivalent in vernacular languages) sometimes found at the beginning of a text or chapter may also be encoded as a **<head>** element. The closing formula, starting with or simply consisting in the word “Explicit,” may be encoded as a **<trailer>**, which “contains a closing title or footer appearing at the end of a division of a text”²⁰ A section containing an “incipit” and an “explicit” formula could be encoded as follows:

```

<div>
  <head>Incipit liber tetrius</head>
  <p/>
  <trailer>Explicit liber tertius</trailer>
</div>

```

2.1.2. Paragraphs and blocks of text

Paragraphs and similar elements may appear within **<div>**s, but if the text is not significantly subdivided they may also appear directly in **<body>**. The **<p>** (paragraph) element “marks paragraphs in prose,”²¹ which may sound like a fairly simple definition. Yet, as an editor, you will need to decide what defines a paragraph in your edition. Typically, when preparing a diplomatic edition, you will follow the punctuation and paragraph marks (the “pieds-de-mouche”) of the manuscript. When preparing a critical edition of a text, based on several witnesses, it is fairly common for editors of Latin texts to apply a modern punctuation and division in paragraphs to the text. Traditions and scholarly practices do vary with the type of literature.²²

In some cases, you might need to wrap some text and phrase-level markup into a paragraph-like element, but without using the semantic of a paragraph. The **<ab>** element²³ may be used instead of **<p>** in this case.

2.1.3. Lists

Lists are another very common feature in all sorts of texts: library catalogues, recipes, accounts, etc. They are very straightforward to encode: the whole list is wrapped in a **<list>** element,²⁴ which contains as many **<item>** elements as necessary, and possibly an optional **<head>** element. Note that no text is allowed to be directly in the **<list>** element: all the text must be

20. <http://www.tei-c.org/Vault/P5/3.0.0/doc/tei-p5-doc/en/html/ref-trailer.html>

21. <http://www.tei-c.org/Vault/P5/3.0.0/doc/tei-p5-doc/en/html/ref-p.html>

22. For a practical overview of various editing practices and strategies, see: *The Arts of Editing Medieval Greek and Latin: A Casebook*, E. Göransson, G. Iversen, B. Crostini (eds.), Turnhout: Brepols, 2016.

23. <http://www.tei-c.org/Vault/P5/3.0.0/doc/tei-p5-doc/en/html/ref-ab.html>

24. <http://www.tei-c.org/Vault/P5/3.0.0/doc/tei-p5-doc/en/html/ref-ab.html>

contained in the `<item>` elements (or `<head>`). For instance, if there are punctuation marks separating the items in the list, they must be enclosed in the `<item>` elements. Lists may appear within a paragraph, or instead of a paragraph. They may be nested: a `<list>` may appear within the `<item>` of a parent `<list>`.

Let us consider, for instance, this short extract of a medieval catalogue, listing the books held at the Klosterneuburg monastic library in 1330:²⁵ the document starts with an indication that this is the list of works by saint Jerome held by the library, which we can consider a header. Each entry describes the contents of a codex kept in the library. Some volumes contain a single work, but others contain several works, also listed: these entries can be encoded as nested lists.

```
<list>
  <head>Item libri Ieronimi</head>
  <item>Primo Ieronimus super Ysaia prima pars</item>
  <item>Item Ieronimis super Ysaia secunda pars</item>
  <item>Item expositio Ieronimi super Ezechielem libri XIII</item>
  <item>
    <list>
      <item>Item epistule Ieronimi</item>
      <item>Ibidem Ieronimus super dicta quedam librorum et capitularum
        Pauli</item>
      <item>Ibidem epistula Augustini ad Ieronimum, in uno
        volumine</item>
      <!-- etc. -->
    </list>
  </item>
</list>
```

2.2. The physical structure: quires, pages, columns, lines

To represent the physical structure of the documents you encode, the TEI offers a series of “empty” or “milestone” elements: instead of wrapping a whole passage of text, they simply mark the beginning of a new quire, page, etc. The advantage of empty elements is that they will not interfere with the markup representing the logical structure of the document. That way, a paragraph can start on one page and finish on another one without causing any overlapping of the markup.

The `<gb>` (gathering beginning) element can be used to mark up the limits of different quires or gatherings constituting a manuscript. The `<pb>` (page beginning) element, marking the beginning of pages, is probably the most widely used of this series. Within the space of a page, `<cb>` (column beginning) and `<lb>` (line beginning) can be used to mark up the limits of columns and lines on the original document, respectively.

Note that all those milestones must be placed at the beginning of the phenomenon they are representing. Regarding lines, this goes against some traditional transcription practices, where editors are used to putting a marker at the end of lines. This may be the source of some errors

25. Klosterneuburg, 1251, f. 1r

when transitioning to digital editions: if you place `<lb>` milestones at the end of lines instead of the beginning, the count and display of the lines will be off.

Those milestones share similar attributes, especially; `@n`, `@ed` and `@edRef`, and `@break`.

The `@n` attribute may be used to record the number given to the element in the document. It is not mandatory, since it is possible to automatically count the relevant elements (quires, pages, columns or lines) with a script, allowing you for instance to automatically number the lines from 1 to x for each page of the document. But in some cases, you might want to record the precise number of one of those elements, as it is in the document. This is generally true of the page numbers which, in manuscripts, follow a particular naming pattern. For instance, the following paragraph is written across folio 4v and 5r in the manuscript we are transcribing. We could encode the beginning of f. 5r with a simple `<pb>` element in the text at the place where the page is changing, adding the `@n` attribute to indicate the folio number:

```
(...) <p> Comment messire Jehan le Boursier, gouverneur de Gand,
durant les treves avoit avitaillié la ville de Gand. Et comment ungue
maniere de gens qui s'appelloient <pb n="5r"/> les Comporselles
faisoient moult de maulx. LIIII chapitre. </p> (...)
```

When working on an edition involving more than one witness, it is possible to record separately the boundaries and numbers of the pages, lines etc. in each witness, if this is useful. The `@ed` (edition) and `@edRef` (edition reference) attributes let us attach a milestone to a particular “edition” or witness in general. The difference between those two attributes is that `@ed` contains a sigil or other identifier of a witness, while `@edRef` contains a pointer to the description of a witness. Let us imagine that we are editing the paragraph above using two manuscripts, one from London and the other from Cambrai, which naturally have different page boundaries. We could represent them for each manuscript using the `@ed` attribute:

```
(...) <p> Comment messire Jehan le Boursier, <pb ed="London"
n="3v"/>gouverneur de Gand, durant les treves avoit avitaillié la
ville de Gand. Et comment ungue maniere de gens qui s'appelloient <pb
ed="Cambrai" n="5r"/> les Comporselles faisoient moult de maulx. LIIII
chapitre. </p> (...)
```

To use the `@edRef` attribute, we would need to have described the relevant manuscripts elsewhere in the document (on the description of witnesses, see chapter Manuscript Description), for instance something like this:

```
<listWit>
  <witness xml:id="C">[Description of the Cambrai
manuscript]</witness>
  <witness xml:id="L">[Description of the London
manuscript]</witness>
</listWit>
```

Then in the edition, we would point to those descriptions, using **@edRef**:

```
(...) <p>Comment messire Jehan le Boursier, <pb edRef="#L"
n="3v"/>gouverneur de Gand, durant les treves avoit avitaillié la
ville de Gand. Et comment ungue maniere de gens qui s'appelloient <pb
ed="#C" n="5r"/> les Comporselles faisoient moult de maulx. LIIII
chapitre. </p> (...)
```

The advantage of **@edRef** over **@ed** is that the link between the milestone and the witness is established in a more formal way, and is easier to process; the coherence of the value of the attribute can also be controlled more easily.

A change of page, line, etc. can of course occur right in the middle of a word. When this happens, the **@break** attribute lets you indicate that the milestone bearing this attribute with the value **"no"** does not separate the word in two like a white space would.

```
<p><lb/>Comment messire Jehan le Boursier, gouver<lb
break="no"/>neur de Gand, durant les treves avoit avitaillié <lb/>la
ville de Gand.</p>
```

Additionally, since all those milestones correspond to identifiable elements in the witnesses, it is possible to link them with their digital reproduction thanks to the **@facts** (facsimile) attribute. The value of **@facts** can be the relative or absolute path to an image, or the URL of this image:

```
<pb edRef="#C" n="5r" facts="./images/f_005r.jpg"/>
```

This is a simple, basic way of linking an edited text to the facsimile of its witnesses. The TEI offers different and more refined ways of achieving this goal, which will be presented in Chapter Transcription.

3. Special structures

The particular structure of some texts, like poetry or drama, calls for special elements and attributes. We are just giving a quick overview here, and if you need to encode such texts we recommend that you read the relevant sections of the Guidelines for more details.²⁶

3.1. Verse

Each verse can be encoded with the **<1>** (verse line) element. This is the most basic form of encoding of verses, as illustrated with those opening lines of the Anglo-Irish poem entitled *Fall and Passion*:²⁷

26. <http://www.tei-c.org/Vault/P5/3.0.0/doc/tei-p5-doc/en/html/CO.html#COVE> and <http://www.tei-c.org/Vault/P5/3.0.0/doc/tei-p5-doc/en/html/VE.html> for verse, or <http://www.tei-c.org/Vault/P5/3.0.0/doc/tei-p5-doc/en/html/CO.html#CODR> and <http://www.tei-c.org/Vault/P5/3.0.0/doc/tei-p5-doc/en/html/DR.html> for drama.

27. See the full text online in *Anglo-Irish poems of the Middle Ages: The Kildare Poems*, <http://celt.ucc.ie/published/E300000-001/>

```

<1>The grace of God ful of might </1>
<1>That is a king and euer was, </1>
<1>Mote amang vs alight</1>
<1>And yiue vs alle is swet grace,</1>

```

Coherent groups of lines, like stanza, can be encoded with the **<lg>** (line group) element. The full poem is also a group of lines of verse, and can be encoded as an particular type of **<lg>**, containing other types of **<lg>** elements, optionally distinguished thanks to the **@type** attribute.

Some special attributes available on **<1>** and **<lg>** elements will allow you to encode the rhyme scheme and the metrical analysis of the verse. The **@rhyme** attribute lets you represent the rhyme scheme of a group of lines, typically using letters to symbolise the rhymes, or any other system defined by the encoder. The **@met** (metrical structure, conventional) and **@real** (metrical structure, realised) attributes can be used to represent, through a user-defined formula, the metrical structure of an element, be it a full poem, stanza, a line... The former represents the conventional metrical structure, while the later indicates that a line or group of lines follows a different structure from the conventional one.

Let us encode, for instance, the first two stanza of the poem above, while representing their rhyme scheme. Note that we here we have chosen to encode the whole poem as an **<lg>** element, but could just as well have encoded it as a **<div>** containing **<lg>** elements.

```

<lg type="poem">
  <lg type="stanza" rhyme="abab">
    <1>The grace of God ful of might </1>
    <1>That is a king and euer was, </1>
    <1>Mote amang vs alight</1>
    <1>And yiue vs alle is swet grace,</1>
  </lg>
  <lg type="stanza" rhyme="cdcd">
    <1>Me to spek and you to lere </1>
    <1>That hit be worsip, lord, to the; </1>
    <1>Me to teche and you to bere </1>
    <1>That helplich to ure sowles be,</1>
  </lg>
  <!-- etc. -->
</lg>

```

3.2. Drama

The main divisions of drama texts, like acts and scenes, can be encoded with the all-purpose **<div>** element discussed earlier. But besides those divisions, a number of specific elements are available.

Drama is characterised, most notably, by the alternance of utterances of speech by different characters, and by the presence of stage directions. Each utterance is encoded with the **<sp>**

(speech) element,²⁸ which typically contains a `<speaker>` element,²⁹ indicating which character is delivering this speech, and the speech itself, which can consist in a paragraph (`<p>`) or in lines of verse (`<l>`), possibly grouped in `<lg>` elements. Note that `<sp>` does not necessarily contain a `<speaker>` element. If the text does not explicitly give the speaker's name there, or if the editor want to create a more formal link between a speech and its performer, it is possible to use the `@who` attribute on `<sp>` to indicate who is speaking. This attribute may point to the `@xml:id` of a `<castItem>`, inside a `<castList>`, describing in more detail a character.³⁰ Stage directions, which may appear anywhere, are encoded with the `<stage>` element.³¹

Let us consider for instance a possible encoding for this passage from the 15th c. *Mystère des Actes des Apôtres* by Simon Gréban.³² We have here one example of stage direction, followed by two utterances in verse, for which the name of the speakers are given:

```
<stage>Il arrive à l'ostel d'Anne. </stage>
<sp>
  <speaker>Caiphaz</speaker>
  <l>Dy, Agripart, que fait ton maistre ?</l>
  <l>Est il pas bien embesoigné ?</l>
</sp>
<sp>
  <speaker>Agripart</speaker>
  <l>Il est lëans plus rechigné </l>
  <l>que n'est ung renart pourboully.</l>
</sp>
```

4. Concurrent hierarchies

Overlapping or non-nesting information is an issue with all XML-based languages, and the TEI is not an exception. An XML document must be a tree structure, with each element under the root nesting in a parent element. This poses a problem when users want to encode “concurrent hierarchies,” that is layers of information that overlap. A typical example of such concurrent hierarchies is the logical and physical structure of a document, i.e., the division into logical parts vs. the division into pages. In this particular case, the potential conflict between these layers of information is solved as we have seen by the use of milestones for the physical structure, to avoid any risk of overlapping elements, but other cases of concurrent hierarchy may occur, requiring the use of different encoding strategies to overcome this difficulty.

The TEI Guidelines support four XML-based methods for handling overlapping information, which are exposed in details.³³ Since, with scholarly editions, the encoding of the critical

28. <http://www.tei-c.org/Vault/P5/3.0.0/doc/tei-p5-doc/en/html/ref-sp.html>

29. <http://www.tei-c.org/Vault/P5/3.0.0/doc/tei-p5-doc/en/html/ref-speaker.html>

30. On the encoding of cast lists, see section [7.1.4 - Cast Lists](#) in the Guidelines

31. <http://www.tei-c.org/Vault/P5/3.0.0/doc/tei-p5-doc/en/html/ref-stage.html>

32. See the full text online: Simon Gréban. *Le Mystère des Actes des Apôtres*. CNRS-Lamop (UMR 8589). [online]. <http://eserve.org.uk/anr/>

33. See chapter 20 of the Guidelines, [Non-hierarchical Structures](#)

apparatus frequently overlaps with the one of citations, we have chosen to use this example to fully develop the explanation of the different strategies available in TEI, at the end of the Citations and references chapter. We direct the readers interested in this topic to the section Handling overlaps of this chapter.

Bibliography

Lou Burnard, *What is the Text Encoding Initiative?*, Marseille, OpenEdition, 2014, DOI: 10.4000/books.oep.426.