

# Encoding for ELTeC : a summary guide

2019-08-10

## 1 The checklist

Here's a list of the distinctions your encoding should make explicit:

- page breaks (`<pb>`) [optional]
- section and chapter divisions (`<div>`)
- headings, titles, footers (`<head>`,`<trailer>`)
- paragraphs (`<p>`)
- verse (`<l>`)
- typographic salience (`<hi>`) [optional]

A level 1 encoding permits the following additional distinctions, all optional but recommended:

- omitted material(`<gap>`)
- editorial correction (`<corr>`)
- chapter subdivision (`<milestone>`)
- linguistic emphasis (`<emph>`)
- code-shifting (`<foreign>`)
- floating headings and titles (`<label>`, `<title>`)
- quotation (`<quote>`)
- authorial notes (`<note>`, `<ref>`)

A level 2 encoding may additionally contain:

- `<s>` elements for segmentation within paragraphs
- `<rs>` for named entities (*@type* values to be determined)
- word level tokenization using `<w>` elements which bear attributes *@pos*, *@lem*, *@join*, *@msd* (values all to be determined)

## 2 A note on page breaks

- ELTeC transcription need not capture any aspect of a book's original appearance, including its pagination.
- However,
  - identifying a page number is useful when checking the encoding against an original
  - identifying all page breaks can also facilitate some kinds of analysis.
- Use `<pb/>` to mark the start of each new page or `<pb n="x"/>` for the page numbered 'x' in the source.
- If a word is hyphenated across a page break, re-assemble it on the page where it begins
- Discard any catchword, running page header or footer, etc.

### 3 The text itself

```
<text>
  <front>
    <div type="titlepage">...</div>
    <div type="liminal">...</div>
  </front>
  <body>
    <div type="chapter">....</div>
  </body>
  <back>
    <div type="notes">...</div>
  </back>
</text>
```

- Find the first and last chapters of the novel: everything before this is the `<front>`, everything after is the `<back>`, everything in between is the `<body>`.
- Discard any non-authorial sections from front or back
- Use `<div>`, with an appropriate *@type* to mark remaining sections
  - Use `<div type='titlepage'>` for the titlepage
  - Use `<div type='liminal'>` for any other subdivision of front or back
  - Use `<div type='chapter'>` for chapters
  - Use `<div type='group'>` for groupings of chapters
  - Use `<div type='letter'>` for epistolary novels
  - Use `<div type='notes'>` within the back for notes

### 4 Headings and footers

Le président fit un signe et la musique de la garde républicaine joua la Marseillaise.

## II

LE MARQUIS DE LA LANDE

Dans la coquette garçonnière des Champs-Élysées qu'habitait Pierre Lathon, l'on n'entendait ce matin-là aucun bruit.

- Use `<head>` for each heading or title provided at the beginning of a section, and `<trailer>` for each provided at the end
- Discard running titles or page headers

```
<div>...<p>Le président fit un signe et
  la musique de la garde républicaine joua la Marseillaise.</p>
</div>
<div type="chapter">
```

```
<head>II</head>
<head>LE MARQUIS DE LA LANDE</head>
<p>Dans la coquette garçonnière des Champs-Élysées...</p>
</div>
```

## 5 Prose and verse

At level 0, the element `<p>` is used for any stretch of prose text followed by a significant line break. It is usually, but not necessarily, a complete syntactic unit.

The element `<l>` should be used for each distinct line of verse, where this is clearly not prose, as in for example a song or an epigraph.

Erau ceaurile 10. Un bețiv trecând pe stradă  
cânta a lene:

Vin drăguță lângă mine:  
Drăguță, drăguță. . . .  
Căci mi-e dor, mi-e dor de tine,  
Drăguță, drăguță.

Eă la sinu-mă să te strâng,  
Drăguță, drăguță,  
Mijlocelul să ți-l frâng,  
Drăguță, drăguță.

Si românul bețiv, ca și fanariotul păgân, trecu

```
<p>Erau ceaurile 10. Un bețiv trecând  
pe stradă cânta a lene: </p>
<l>Vin drăguță lângă mine: </l>
<l>Drăguță, drăguță ...</l>
<l>Căci mi-e dor, mi-e dor de tine ...</l>
<!-- ... -->
<p>Si românul, bețiv, ca si fanaraiotul păgân...</p>
```

At level 1, quoted paragraphs or quoted verse may be wrapped in a `<quote>` element.

```
<p>Erau ceaurile 10. Un bețiv trecând  
pe stradă cânta a lene: </p>
<quote>
<l>Vin drăguță lângă mine: </l>
<l>Drăguță, drăguță ...</l>
<l>Căci mi-e dor, mi-e dor de tine ...</l>
</quote>
<!-- ... -->
<p>Si românul, bețiv, ca si fanaraiotul păgân...</p>
```

## 6 Typographic salience

- An ELTeC transcription need not capture any aspect of a book's original appearance, including its typography.
- However, it is often useful to distinguish (e.g.) italicized words or phrases because this usage may have a semantic implication.

- At level 0, we record simply that such words are *typographically salient*, using the `<hi>` element with no attributes
- At level 1, we may instead capture the semantic implication, using e.g. `<emph>` or `<foreign>`.

## 7 Omissions and corrections

At level 0, material which has been omitted from a transcription is simply omitted, and any editorial corrections to the text are silently applied.

At level 1, we record the fact that something has been omitted using the `<gap>` element, and the fact that some text has been corrected using the `<corr>` element.

```
<!-- a picture in the original has been omitted--><gap desc="graphic"/>
```

```
<!-- the word "indubitably" has been provided by an editor or transcriber --><p>...  
and so <corr>indubitably</corr> preferable .. </p>
```

The same element may be used to mark where something missing in the source (eg opening quotation mark) has been supplied by the encoder:

```
<div>  
<head>I. Dika Lombardije.</head>  
<p n="SL-WIKI001351">  
<corr>„</corr>Italija, kako lepa si, prostrana naša domovina! Kamorkoli se  
ozre oko, odpiraš mu nova čuda...dika Lombardije.“ – Tako je sanjaril sam s  
seboj pesnik Antonio Loschi...</p>  
</div>
```

## 8 Sub-divisions of chapters

‘  
And the lady arose, — again the stately,  
stern-looking woman.

\* \* \* \*

Months had passed by, spring was now  
bursting forth in all its verdant beauty,

Do not use `<div>` for subdivisions of chapters: instead mark the point at which the subdivision is indicated with a `<milestone>` element, using its `@unit` attribute to indicate the kind of subdivision concerned, and (optionally) its `@rend` attribute to indicate the appearance of the milestone.

At level 0, a milestone indication such as a row of stars may be retained as a `<p>`, but this should be changed at level 1.

At level 0:

---

```
<p>... stern-looking woman.</p>
<p>* * * *</p>
<p>Months had passed by .... </p>
```

At level 1:

```
<p>...stern-looking woman.</p>
<milestone unit="subchapter" rend="stars"/>
<p>Months had passed by... </p>
```

## 9 Emphasis and code switching

At level 1, emphasized words in speech or narrative may be marked using `<emph>`.

You said `<emph>what</emph>`?

At level 1, code switching (use of a language other than that of the main narrative) may be marked using `<foreign>`

— *Ite!... Un miş-maş, zise Kivu, care intrând auzise spusele Zamfirei.*

```
<p> -- Ite! Un <foreign>miş-maş</foreign>, zise Kiva, care intrand auzise spusele Zamfirei.</p>
```

The `@xml:lang` attribute may optionally be used to specify the language switched to, but is not obligatory

## 10 Floating headings and titles

At level 1, headings or labels not attached to a division should be marked using `<label>`

At level 1, titles of songs, literary works, etc. may optionally be marked using `<title>`.

## 11 Quotation

- At level 1, passages such as epigraphs or quoted songs, may be marked using `<quote>`.
- Do not however use `<quote>` for quoted phrases within a paragraph: simply retain the quotation marks (similarly, for direct speech)
- A citation e.g. giving the author of a quotation can be marked using `<label>` or as a free-standing `<p>`

### 12 Authorial notes

At level 0, all notes are omitted. At level 1 :

- Only authorial notes should be retained. Editorial or textual notes should be silently omitted
- The content of the authorial note should be marked using `<note>`
- The siglum indicating the point of attachment for the note within the text should be marked using `<ref>`
- All `<note>` elements should be given a unique identifier (e.g. use the text identifier followed by "N" and a number) as value for their `@xml:id`
- All `<ref>` elements should indicate the note they refer to by means of their `@target` attribute
- All `<note>` elements should be collected together in the `<back>` element, inside a `<div type="notes">`, no matter where they appear in the source

```
<body>
  <div type="chapter">
    <p> .... <ref target="#EN042N01">[1]</ref>
  </p>
  </div>
</body>
<back>
  <div type="notes">
    <note xml:id="EN042N01">Or so they say.</note>
  </div>
</back>
```