

Jane Austen Project

Guidelines to Encoding

Table of content

Jane Austen Project Guidelines to Encoding.....	1
0. Manuscripts Survey	3
1. Structure overview	3
1.1. Outline of a manuscript containing just a text	3
1.2. Outline of a manuscript containing different texts	4
1.3. Overview of the <text> element internal structure	4
1.4. Manuscripts physical boundaries	4
1.4.1 Line breaks.....	4
1.4.2 Page breaks	7
1.4.3 Fascicule breaks	7
2. Structural features	7
2.1. Blocks	7
2.1.1 Blocks: special cases.....	8
2.1.2 Blocks titles.....	10
2.2. Front pages and covers	10
2.3. Epigraphs	10
2.4. Prose: paragraphs blocks	10
2.5. Drama.....	11
2.5.1 List of characters.....	11
2.5.2 Speeches.....	11
2.5.3 Stage directions.....	11
2.6. Poetry	12
2.6.1 Verses.....	12
2.6.2 Stanzas and other grouping of verses	12
2.7. Quotations	12
2.8. Drawings blocks	13
2.9 Trailers	13
3. Inline markup and transcriptional features	14
3.1. Special characters encoding	14
3.1.1. Long S (ſ)	14
3.1.2 Carets	14
3.1.3. Dashes	14
3.2. Revisions	15
3.2.1. Superlinear insertion	15
3.2.2. Sublinear insertion	16

3.2.3. Superiposition	16
3.2.4 Double stroked deletion	16
3.2.4 Silent deletion	17
3.2.5 Erasure (with overwritten text)	17
3.2.6 Other insertions and deletions.....	17
3.2.7 Rewriting	17
3.3 Underlined or variously emphasized text	18
3.4. Quotes for direct speeches	18
3.5 Filling rule.....	18
3.6 Hands	19
3.7 "Printer" rules	19
4. Editorial markup	19
4.1. Capitalization features	20
4.2. Spelling features (e.g. freind, veiw, etc.)	20
4.3. Abbreviations and relative expansions	20
4.4. Correction of text and diacritics.....	21
4.4.1. Error in the source.....	21
4.4.2. Text missing in the source	21
4.4.3. Uncertainty.....	21
4. Metadata.....	21
2.1. Title information	22
2.2. Publication statement	22
2.3. The manuscript description.....	22
2.3.1 The manuscript identifier.....	23
2.3.2. Manuscript content	23
2.3.3. Physical description	23
2.3.4. Provenance/history of the manuscript.....	24
2.3.5 Documentation on the digital reproduction	24
2.3.6. Bibliography	24
5. Second layer markup	25

0. Manuscripts Survey

- *Volume the First*: Bodleian Library, Oxford (193pp)
- *Volume the Second*: British Library, London (260pp)
- *Volume the Third*: British Library, London (146pp)
- *Lady Susan*: Pierpont Morgan Library, New York (170pp)
- *The Watsons*: Pierpont Morgan Library, New York (12pp)
- *The Watsons*: Queen Mary and Westfield College, London University (76pp +14)
- ‘Plan of a Novel’ + 2 holograph notes: Pierpont Morgan Library, New York (6pp)
- ‘Opinions of Mansfield Park’ and ‘Opinions of Emma’: British Library, London (4pp.)
- *Persuasion*: British Library, London (54pp + 8)
- *Sanditon*: King’s College, Cambridge (184pp + 16)

1. Structure overview

The present model assumes that each manuscript is transcribed within a single XML file.

Each manuscript and text within a manuscript should embed an unique identifier within an `xml:id=""` attribute.

A unique identifier system should be provided also for images (to be established).

1.1. Outline of a manuscript containing just a text

```
<TEI xmlns="http://www.tei-c.org/ns/1.0" xml:id="ms_identifier">
  <teiHeader>
    [metadata]
  </teiHeader>
  <text xml:id="text_identifier">
    <front></front>
    <body></body>
  </text>
</TEI>
```

1.2. Outline of a manuscript containing different texts

```
<TEI xmlns="http://www.tei-c.org/ns/1.0" xml:id="">
<teiHeader>
  [metadata]
</teiHeader>
<text>
  <group>
    <text xml:id="text1_identifier">
      <front></front>
      <body></body>
    </text>
    <text xml:id="text2_identifier">
      <front></front>
      <body></body>
    </text>
    [more texts here]
  </group>
</text>
</TEI>
```

1.3. Overview of the <text> element internal structure

The element <text> can have two different contents:

1. In case of a manuscript containing single text, we may have up to three elements:
 - a. <front>: contains all the front matters such as title pages, cover, dedications, table of contents, etc. The element is optional
 - b. <body>: contains the main content of the manuscript (the stories, chapters...). **The element is required**
 - c. <back>: contains any back matters such as indices, glossaries, bibliographies, etc. The element is optional.
2. In case of a manuscript containing more than one text, it will contain a <group> element, containing in turn several <text> elements, each of them structured as per item 2.

1.4. Manuscripts physical boundaries

1.4.1 Line breaks

Each line in the original source is followed by a line break element <lb/> unless it is at the end of the page where an element <lb/> is normally not used (nonetheless is used if some kind of hyphenation marks are present: see further down).

Notice that in case a line break occurs in the middle of a word and some sort of hyphenation marks is in the text, a `rend` attribute should be provided and the hyphenation marks deleted from transcription as follow:

Case 1: colon at the end and at the beginning of the lineText:

```

    for your pre:
:sence

```

Encoding:

```

for your pre<lb rend="double_colon"/>sence

```

Case 2: colon at the end and of the line onlyText:

```

    for your pre:
sence

```

Encoding:

```

for your pre<lb rend="double_colon1"/>sence

```

Case 3: hyphen at the end and at the beginning of the lineText:

```

    for your pre-
-sence

```

Encoding:

```

for your pre<lb rend="hy"/>sence

```

Case 4: hyphen at the end of the line onlyText:

```

    for your pre-
sence

```

Encoding:

```

for your pre<lb rend="hy1"/>sence

```

Case 5: colon at the beginning of the line onlyText:

```

    for your pre
:sence

```

Encoding:

```

for your pre<lb rend="double_colon2"/>sence

```

Case 6: colon at the end of the line and equal at the beginningText:

```

    for your pre:
=sence

```

Encoding:

```

for your pre<lb rend="colon_equal"/>sence

```

Case 7: hyphen at the end of the line and colon at the beginningText:

```

    for your pre-
:sence

```

Encoding:

```

for your pre<lb rend="hy_colon"/>sence

```

Case 8: equal at the end and at the beginning of the lineText:

for your pre
sence

Encoding:

for your pre<lb rend="hy_colon"/>sence

In case a line break with hyphenation mark occurs at the end of a page the <lb/> is required as well.

In case a word break occurs within an interlinear insertion, the encoding is pretty different as the hyphenation mark should be provided as a stand alone glyph element rather than with an attribute.

In the <teiHeader> the flowing glyphs are defined:

```
<char xml:id="dc">
  <charName>double colon</charName>
  <desc>double colon for word breaks</desc>
  <mapping type="standard">:</mapping>
</char>
<char xml:id="hy">
  <charName>hyphen</charName>
  <desc>hyphen for word breaks</desc>
  <mapping type="standard">-</mapping>
</char>
```

In a text like the following:

as if re:
at the end of one, ~~away form~~
:treating as far as he could
~~every body else, as if to enjoy~~

The encoding will be:

```
at the end of one, <choice>
  <del rend="overstrike">away form</del>
  <add rend="superlinear">as if re<g ref="#dc" xml:id="dc1"
next="#dc2"/></add>
</choice>
<lb/>
<choice><del>every body else, as if to enjoy</del>
  <add><g ref="#dc" xml:id="dc2" prev="#dc1"/>treating as far as
he could</add>
</choice>
```

The @ref attribute is due to display the kind of hyphenation mark needed (in case of an hyphen the value will then be "#hy"). The xml:id declares a milestone identifier for the specific point as the attributes @next and @prev are needed to connect the two halves of the word.

In case of a single hyphenation mark at the end of the line, the `<g/>` element is still required in the new line to guarantee the connection between the two halves of the word, but it will not carry the `@ref` attribute.

1.4.2 Page breaks

```
<pb n="" rend="" xml:id="" ed=""/>
```

The `xml:id` attribute contains file name of the image representing the page, without extension. **The attribute is mandatory**, while the `n` attribute is not, as it is possible that a page has not an original numeration.

The `rend` attribute can carry values like "pencil" or "left" as the default positioning of the page number would be on the top right corner of the page; it is an optional attribute.

In case a manuscript has different sequences of page numbers, they should be distinguished by the help of an `@ed` attribute which value should be established. The values used at present are:

`austen` : for the original page numbering

`hand1` : for another unidentified hand writing in pencil in blvolthird

1.4.3 Fascicule breaks

```
<au:fb n="" rend=""/>
```

2. Structural features

2.1. Blocks

Any structural block should be encoded within a `<div>` element provided with a `@type` attribute. A nomenclature for the `type` attribute should be collected (after a short time usage) and fixed.

Here are some examples of `type` values already in place:

```
<div type="dedication">
  for a dedication
</div>
```

```
<div type="story">
  for a short story without chapters
</div>
```

Notice that `<div>` elements can be nested at the occurrence:

```

<div type="book|novel">
  <div type="chapter">
    for each of the chapter of a book
  </div>
  <div type="chapter">
    for each of the chapter of a book
  </div>
</div>

```

2.1.1 Blocks: special cases

Case 1: Table of Content

```

<front>
  <div type="TOC">
    <ab>One line of toc <seg type="filling_dots" n="5">Page
      1</seg></ab>
    <ab>One line of toc <seg type="filling_dots" n="5">Page
      20</seg></ab>
  </div>
</front>

```

Notice that a `<seg type="filling_dashes">` is also available.

Case 2: Included correspondence

```

<floatingText="letter">
  <body>
    <div>
      <opener>
        <dateline>15 Oct </dateline>
        <salute>Dear Emma </salute>
      </opener>
      <p>Here the text of the letter</p>
      <closer>
        <signed>Here the signature</signed>
      </closer>
      <postscript><label>P.S.</label>
        <p>text</p>
      </postscript>
    </div>
  </body>
</floatingText>

```

Notice that `<floatingText>` applies only in case of a included letter that contains usual letter's features such as dates, salute, signature and it is formatted as such.

When the text includes just excerpt of a letter, a `<q>` element has to be used. In case of an epistolary novel it is not necessary to use the `<floatingText>` structure, but a sequence of `<div type="letter">` will be enough, i.e.:

```

<div type="letter">
  <opener>
    <dateline>15 Oct </dateline>

```



```

        <salute>Dear Emma </salute>
    </opener>
    <p>Here the text of the letter</p>
    <closer>
        <signatures>Here the signature</signatures>
    </closer>
    <postscript><label>P.S.</label>
        <p>text</p>
    </postscript>
</div>

```

Case 3. Patches leaves that substitute blocks of text

3.a. Patch that delete an existing portion of text

The text that was substituted by the one in a separate patch need to be encoded within a `<floatingText>` element.

The full patch content need to be encoded in a following `<floatingText>` element as well. Both `<floatingText>` elements need to carry an `xml:id` and a `corresp` attributes: the `xml:id` will provide a unique identifier for the block, while the `corresp` attribute will connect each to the other. In case the patch will contain a cue-mark, that need to be encoded within an optional `@rend` attribute.

```

<floatingText type="cancellandum" xml:id="canc1" corresp="#patch1">
    <body>
        <p>Text cancelled by a patch</p>
    </body>
</floatingText>

<floatingText type="patch" xml:id="patch1" corresp="#canc1"
rend="X">
    <body>
        <p>Text contained within a patch</p>
    </body>
</floatingText>

```

3.b Patch to be inserted somewhere within a paragraph level block:

In this case a patch is not supposed to delete anything, but just to add new text.

Somewhere in the text we need an element `<ptr/>` to mark the point of insertion, then the content of the patch need to be encoded in a `<div>` element located after the end of the `<div>` that contains the `<ptr/>` element, where the optional `@rend` attribute will contain any cue-mark eventually present in the text.

```

<div>
<p>In my text there is an insertion here <ptr type="insertion"
target="#xpatch1" xml:id="insertion_point1" rend="X"/></p>
[...]
```

```

</div>
<floatingText type="patch" xml:id="xpatch1"
corresp="#insertion_point1">
    <p>Text contained within a patch</p>

```

```
</floatingText>
```

2.1.2 Blocks titles

An `<head>` element might be provided as first child of a `<div>` or of a `<lg>` element.

```
<head>Chapter 1</head>
```

In case a title is followed by a line as for printer "rules", the `<head>` requires a `@rend` attribute:

```
<head rend="rule_after">CHAPTER 1</head>
```

In case of rules that appears on the top and bottom of a title, a `ruled2` value need to be used:

```
<head rend="center ruled2">CHAPTER 1</head>
```

In case the title should be centred, a convenient value to the `@rend` attribute should be provided:

```
<head rend="center">CHAPTER 1</head>
```

2.2. Front pages and covers

A front page or a cover needs to be encoded within the `<front>` element.

```
<titlePage type="cover">
  <docTitle>
    <titlePart>Volume the Third</titlePart>
  </docTitle>
</titlePage>
```

2.3. Epigraphs

Epigraphs may occur at the beginning or at the ending of a text, normally, but not necessarily, in the `<front>` element.

```
<epigraph>
  <p>Life is good<lb/>
  but sometimes it isn't</p>
</epigraph>
```

2.4. Prose: paragraphs blocks

```
<p>Here my paragraph</p>
```

In case of indented paragraphs one of the following value can be used, with dimension of the indentation increasing from `short` to `indent2`

```
<p rend="short-indent"> Here my paragraph </p>
<p rend="indent"> Here my paragraph </p>
<p rend="indent1"> Here my paragraph </p>
<p rend="indent2"> Here my paragraph </p>
```

The same range of value can be used for the `<space/>` empty element to be used in case of an arbitrary white space at the beginning of a line in a middle of a paragraph:

```
<p>here my first line <lb/>
<space scope="indent"/> here my second indented line
```

2.5. Drama

2.5.1 List of characters

To be encoded in the `<front>` element.

```
<front>
  <castList>
    <head>List of characters</head>
    <castItem>
      <role xml:id="char1">Character1</role>
      <roleDesc>Description</roleDesc>
    </castItem>
  </castList>
</front>
```

2.5.2 Speeches

```
<sp who="#char1">
  <speaker>Character1</speaker>
  <p>speech</p>
</sp>
```

The value of the `who` attribute must match one of the `xml:id` established in the list of characters.

2.5.3 Stage directions

Stage directions may occur

- a. between speeches

```

    <sp who="#char1">
      <speaker>Character1</speaker>
      <p>spoken text</p>
    </sp>
    <stage>stage direction</stage>
    <sp who="#char2">
      <speaker>Character2</speaker>
      <p>spoken text</p>
    </sp>

```

b. between the speaker's name and the speech

```

    <sp who="#char1">
      <stage>stage direction</stage>
      <speaker>Character1</speaker>
      <p>spoken text</p>
    </sp>

```

c. directly in the speech

```

    <sp who="#char1">
      <speaker>Character1</speaker>
      <p>spoken text <stage>stage direction</stage> other
        text. </p>
    </sp>

```

2.6. Poetry

2.6.1 Verses

```
<l>Here is a verse </l>
```

2.6.2 Stanzas and other grouping of verses

```

<lg>
  <head>Title of a poem</head>
  <l>a verse</l>
  <l>a verse</l>
  <l>a verse</l>
  <l>a verse</l>
</lg>

```

2.7. Quotations

A quotation, meaning by this a block of text quoted from someone else (from a book et similes) and therefore distinguished from direct speech quotes, has to be encoded within a `<quote>` element. Quotation marks can be variously present or absent and quotation marks can be single or double. This can be described by the combined usage of the attributes `type` and `rend` according to the following schema:

```

<quote> double quote and opening/closing quotes (default
value)</quote>
<quote type="o">text with opening quotes only</quote>
<quote type="c">text with closing quotes only</quote>
<quote type="l">text with line by line quotes</quote>

<quote rend="1">text with single quotes</quote>
<quote rend="2">text with double quotes</quote>
<quote rend="1-2">text with single opening quotes, double closing
quotes</quote>
<quote rend="2-1">text with double opening quotes, single closing
quotes</quote>

```

Notice that the two attributes can be combined:

```

<quote type="o" rend="2">Text with opening only double
quotes</quote>

```

Notice that `<quote>` is a block element. If the quotation is an inline one, please provide a `inline` value in the `rend` attribute. If the `rend` attribute is already used for marking the kind of quotation mark, please leave a space between the two values:

```

<quote rend="1-2 inline">

```

2.8. Drawings blocks

In case a drawing is found in the manuscript, it should be provided as an image as follow:

```

<figure>
  <graphic url="myfile.jpg"/>
</figure>

```

In case of two drawings on the same line, use two `<graphic>` elements within figure:

```

<figure>
  <graphic url="myfile1.jpg"/>
  <graphic url="myfile2.jpg"/>
</figure>

```

2.9 Trailers

Trailers and ending words can be encoded within a `<trailer>` element, after the end of the last `<div>` element containing the text.

```

<body>
<div type="novel"></div>
<trailer>Finis</trailer>
</body>

```

3. Inline markup and transcriptional features

3.1. Special characters encoding

Special characters need to be firstly defined in the metadata and the used in the text by referencing the used character.

In metadata, in section called `<profileDesc>`, within `<charDecl>` a list of `<char>` element can be provided, each of them being described in details and embedding an unique identifier through an `xml:id` attribute.

```
<charDecl>
  <char xml:id="ls">
    <charName>LATIN SMALL LETTER S WITH ESH</charName>
    <desc>long s</desc>
    <mapping type="standard">s</mapping>
    <mapping type="Unicode">U+0283</mapping>
  </char>
  <char xml:id="dash1">
    <charName>mdash</charName>
    <desc>long dash meaning...</desc>
    <mapping type="Unicode">U+2014</mapping>
    <mapping type="standard">-</mapping>
  </char>
  <char xml:id="dash2">
    <charName>ndash</charName>
    <desc>short dash meaning...</desc>
    <mapping type="Unicode">U+2013</mapping>
    <mapping type="standard">-</mapping>
  </char>
</charDecl>
```

In the text, a special character may be referenced by a `<g ref=""/>` element. the value of the `ref` attribute must match one of the defined character, preceded by #.

3.1.1. Long S (ſ)

```
<g ref="#ls"/>
```

3.1.2 Carets

Carets are to be supplied before a `<au:revision>` element.

```
<g ref="#c_o"/><au:revision>
  <del rend="overstrike">Instead of making any reply, Tom</del>
  <add place="superlinear">Tom made no reply. – "Upon my</add>
</au:revision>
```

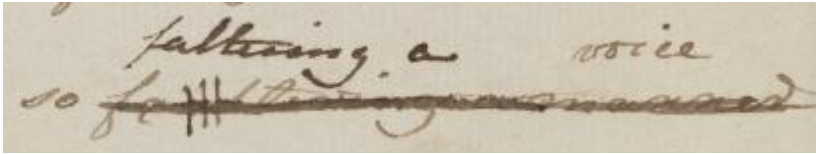
3.1.3. Dashes

```
<g ref="#dash1/> : long
<g ref="#dash2/> : short
<g ref="#dash3/> : very long
```

3.2. Revisions

The encoding of a **revisions** in the source is normally performed within a `<au:revision>` element. According to the place of the insertion and the kind of deletion we may have several possible encoding.

Notice that when an insertion portions need to be aligned to a specific word, the insertion need to be broken in several pieces as in the output an inserted word or segment will be aligned to next following word. So in the following example:



the applied mark up need to be:

```
<au:revision>
  <del rend="overstrike">faltering a</del>
  <add place="superlinear">faltering a</add>
</au:revision>
<au:revision>
  <del rend="overstrike">manner</del>
  <add place="superlinear">voice</add>
</au:revision>
```

3.2.1. Superlinear insertion

```
<au:revision>
  <del rend="overstrike">Instead of making any reply, Tom</del>
  <add place="superlinear">Tom made no reply. – "Upon my</add>
</au:revision>
```

Case a. double superlinear insertion

```
<au:revision>
  <add place="superlinear" n="2">Tom made no reply. – "Upon
my</add>
  <add place="superlinear" n="1">Tom sighed</add>
  <del rend="overstrike">Instead of making any reply, Tom</del>
</au:revision>
```

Case b. double superlinear insertion with deletion

```
<au:revision>
  <add place="superlinear" n="2">Tom made no reply. – "Upon
my</add>
  <add place="superlinear" n="1"><del>Tom sighed</del></add>
  <del rend="overstrike">Instead of making any reply, Tom</del>
</au:revision>
```

Case c. superlinear insertion split in two lines

```
<au:revision>
  <del rend="overstrike">on overhearing Margaret saying</del>
  <add place="superlinear" n="2">when she heard Margaret 5 minutes
<lb rend="interlinear"/>afterwards say</add>
</au:revision>
```

Case d. superlinear insertion as result of a rewriting

```
<au:revision>
  <del rend="overstrike">Instead of making any reply, Tom</del>
  <add place="superlinear" type="rewriting">Tom made no reply. –
  "Upon my</add>
</au:revision>
```

Case e. superlinear insertion displayed starting from a different position respect the insertion point

```
<au:revision>
  <del rend="overstrike">Instead of making any reply, Tom</del>
  <add place="superlinear" rend="-9.5 gl">Tom made no reply. –
  "Upon my</add>
</au:revision>
```

The value of the @rend attribute needs to contain the number of characters (including white spaces) before (or after) the insertion point. The – or + signs indicate whether the insertion have to stat before (-) or after (+). A punctuation mark counts as 0.5 character. In the example above "-9.5 gl" means that the insertion need to be displayed starting 9 character and a punctuation mark before the insertion point.

3.2.2. Sublinear insertion

```
<au:revision>
  <del rend="overstrike">Instead of making any reply, Tom</del>
  <add place="sublinear">Tom made no reply. – "Upon my</add>
</au:revision>
```

3.2.3. Superiposition

```
<au:revision>
  <del rend="overwritten">Instead of making any reply, Tom</del>
  <add place="superimposed">Tom made no reply. – "Upon my</add>
</au:revision>
```

3.2.4 Double stroked deletion

```
</au:revision>
  <del rend="double">Instead of making any reply, Tom</del>
</au:revision>
```


3.2.4 Silent deletion

```
<au:revision>
  <del rend="none">Instead of making any reply, Tom</del>
</au:revision>
```

3.2.5 Erasure (with overwritten text)

```
<au:revision>
  <del rend="erased">Instead of making any reply, Tom</del>
  <add place="superimposed">Text written on top</add>
</au:revision>
```

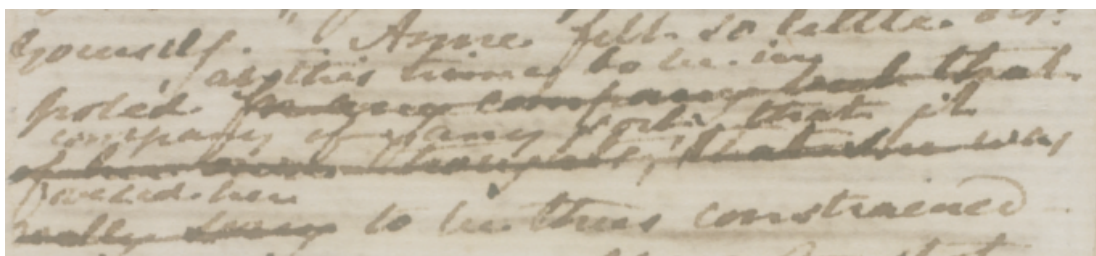
3.2.6 Other insertions and deletions

Case a. footer insertion with cue-mark

```
<au:revision>
  <add place="foot" rend="*">a post chaise </add>
</au:revision>
```

3.2.7 Rewriting

Some area of the text have not just been revised but completely rewritten, meaning that several lines have been deleted and substituted in the interlinear space with a completely different space. In this case again the deleted part need to be wrapped as always within a `<au:revision>` element, but each line break need to be added of an `@xml:id` attribute. The added text will follow the deleted one and the line breaks within that will carry a `@corresp` attribute which value will match the `xml:id` of the lines breaks in the deleted part. So for instance in *Persuasion*:



The encoding will be:

```
... Anne felt so little dis<lb rend="double_colon"/> posed
  <au:revision>
    <del rend="overstrike"> for any company but that <lb
xml:id="pers-del1"/>
      of her own thoughts, that she was <lb xml:id="pers-del2"/>
      really sorry </del>
```

```

    <add place="superlinear" type="rewriting">at this time to be
in<lb corresp="#pers-del1"/>
    company of any sort that it <lb corresp="#pers-del2"/>
    vexed her </add>
</au:revision> to be thus constrained<lb/>

```

Important! remember to put the `<lb corresp="">` in the interlinear insertion even is the insertion does not continue in the following line.

3.3 Underlined or variously emphasized text

Any emphasized text has to be encoded within an `<hi rend="">` element where the `rend` attribute describe the kind of emphasizing. At present, "underline", "dashed_underlined", "sup" (for superscript) and "bigger" values are supported.

```
<hi rend="underline">delightful</hi>
```

3.4. Quotes for direct speeches

Direct speeches are marked up using the **inline** `<q>` element. Quotation marks are not to be inserted in the transcription but are the output using a script interpreting markup.

Direct speeches are normally marked by simple or double quotes. Quotation marks can be variously present or absent and quotation marks can be single or double. The element This can be described by the combined usage of the attributes `type` and `rend` according to the following schema:

```

<q> double quotes and opening/closing quotes (default value)</q>
<q type="o">text with opening quotes only</q>
<q type="c">text with closing quotes only</q>
<q type="l">text with line by line quotes</q>

<q rend="1">text with single quotes</q>
<q rend="2">text with double quotes</q>
<q rend="1-2">text with single opening quotes, double closing
quotes</q>
<q rend="2-1">text with double opening quotes, single closing
quotes</q>

```

Notice that the two attributes can be combined:

```
<q type="o" rend="2">Text with opening only double quotes</q>
```

Notice that in case of extra quotation marks in addition to the "normal" ones a `<q ref="#quote"/>` can be used.

3.5 Filling rule

Sometimes a rule is drawn in the source to complete the line of text to the right margin. Such rule need to be marked as follow

```
<seg type="filling"/>
```

3.6 Hands

Several hand across text alternates in the manuscripts. All hands are defined in a single repository to be included in the metadata of each manuscript in order to enable the reference to them. The description of the different hands is embedded within the `<profileDesc>`, in a `<handNotes>` element.

Each hand carry a description of the kind of ink used and by a conventional identification of the scribe (notice that neither `@medium` or `@scribe` attribute may contain spaces). Each hand is also uniquely identified by an `xml:id` attribute.

```
<profileDesc>
  <handNotes>
    <handNote xml:id="a1" medium="pencil"
scribe="JaneAusten"/>
    <handNote xml:id="a2" medium="regular-black"
scribe="JaneAusten"/>
  </handNotes>
</profileDesc>
```

Once defined in the metadata section, the hand is referenced by a `<handShift/>` empty element. The element must embed a `new` attribute, which value need to match one of the value defined in the metadata, preceded by #.

- a. the first time: `<handShift new="#a1"/>`
- b. the other times: `<handShift new="#a2"/>`

Notice that a `<handShift/>` element **must be provided before any textual node** in any file, i.e. in the first element with a textual content within the `<text>` element (so not metadata), before typing any text, a `<handShift>` element need to be provided.

3.7 "Printer" rules

"Printer" rules coming after a certain block of text (a poetry, a header) need to be encoded as a `rend` attribute of the block element that is ruled after, so for instance a poetry that ha a rule after requires an `<lg rend="rule_after">`; for a paragraph use: `<p rend="rule_after">`.

In case of a rule before and after a given block the `"ruled2"` value should be provided.

Eg: `<head rend="ruled2">`

4. Editorial markup

The editorial markup encodes features of the text that are the result of an editorial evaluation.

4.1. Capitalization features

Capitalized words different from personal name and word occurring after a full stop are marked and their part of speech declared unless it is a noun. Capital and semi-capital are also distinguished.

```
<w rend="capital">Noun default value</w>
```

```
<w rend="semi-capital">Noun</w>
```

```
<w rend="capital" type="v">Verb</w>
```

```
<w rend="capital" type="a">Article</w>
```

Values – at present – for the type attribute:

- **a** : article
- **v** : verb
- **conj**: conjunction
- **adj**: adjective
- **pronoun** : pronoun
- **article**: article

4.2. Spelling features (e.g. freind, veiw, etc.)

To mark up peculiar Austen's spelling a `<distinct type="spelling">` element is required.

```
<distinct type="spelling">freinds</distinct>
```

4.3. Abbreviations and relative expansions

Abbreviated forms should be encoded within an `<abbr>` element, as the relative expanded form need to be embedded within an `<expan>`. Both elements are to be wrapped by a `<choice>` element.

In the eventuality that part of the abbreviated form carry some font features, an `<hi rend="">` attribute has to used as well.

```
<choice>
  <abbr>Marg<hi rend="sup">t</hi></abbr>
  <expan>Margaret</expan>
</choice>
```

4.4. Correction of text and diacritics

4.4.1. Error in the source

In case of a text that is obviously wrong in the source (typo, lapsus calami, missing word or letter), an editorial correction can be provided together with the original erroneous form, again wrapped within a `<choice>` element.

Case a. Text obviously wrong in the source

```
<choice>
  <sic>od</sic>
  <corr>of</corr>
</choice>
```

Case b. Word repetition in the source

```
the <choice>
  <sic>the</sic>
  <corr></corr>
</choice>
```

4.4.2. Text missing in the source

For any supplied text, a `<choice>` element is not necessary, as one can assume that the original form is the text without anything contained within the `<supplied>` element.

```
th<supplied>i</supplied>ng
```

In case missing text could not be inserted by the editor a `<gap/>` element is required. The `unit` attribute will be used to establish the entities to be measured and will include values as "words" or "letters"; the `n` attribute will encode the quantity of the unit encoded by scope:

```
<gap unit="words" n="2-3"/>
```

4.4.3. Uncertainty

When a certain passage is not entirely clear or readable un `<unclear>` element should be used together with the attribute `@reason` declaring the reasons of the uncertainty.

```
had indeed been <unclear reason="illeg">materially</unclear>
```

4. Metadata

Metadata are encoded within a `<teiHeader>` element. Here his ideal structural overview:

```
<teiHeader>
  <fileDesc>
```

```

        <titleStmt></titleStmt>
        <publicationStmt> </publicationStmt>
        <sourceDesc>
            <msDescription></msDescription>
        </sourceDesc>
    </fileDesc>
    <profileDesc>
        <handNotes></handNotes>
    </profileDesc>
    <encodingDesc>
        <charDecl></charDecl>
    </encodingDesc>
</teiHeader>

```

However, as mention before, both the `<profileDesc>` and the `<encodingDesc>` are stored in separate file and therefore included in each file thanks to an element `<xi:include/>`, so the real outline of the `<teiHeader>` will be:

```

<teiHeader>
    <fileDesc>
        <titleStmt></titleStmt>
        <publicationStmt> </publicationStmt>
        <sourceDesc>
            <msDesc></msDesc>
        </sourceDesc>
    </fileDesc>
    <xi:include href="hands.xml"/>
    <xi:include href="chars.xml"/>
</teiHeader>

```

ATTENTION! The following part is still under development and may sensibly change in future.

2.1. Title information

[TEI/teiHeader/fileDesc/titleStmt/title]

This must be completed with the full conventional name of the manuscript that will be probably displayed at the beginning of the page

e.g.: `<title>Volume the Third</title>`

2.2. Publication statement

[TEI.2/teiHeader/fileDesc/publicationStmt]

```

<publicationStmt>
    <publisher>The Austen Project</publisher>
</publicationStmt>

```

2.3. The manuscript description

The `<msDesc>` element is structured as follow:

```
<msDesc>
  <msIdentifier></msIdentifier>
  <msContents></msContents>
  <physDesc></physDesc>
  <history></history>
  <additional></additional>
</msDesc>
```

2.3.1 The manuscript identifier

A manuscript identifier is composed by several elements attesting the exact location of the source manuscript and a reference number or name for it. The element is mandatory

```
<msIdentifier>
  <country>United Kingdom</country>
  <settlement>Oxford</settlement>
  <repository>Bodleian Library</repository>
  <idno>[MS call number or other identifier]</idno>
</msIdentifier>
```

2.3.2. Manuscript content

The `<msContent>` element can be constituted by a simple prose description or by a much more structured description of the content of the manuscript. **This area should be revised later.**

```
<msContents>
  <p>Juvenilia</p>
</msContents>
```

2.3.3. Physical description

The `<pysDesc>` element contains an highly structured description of the physical support of the text. The element support different levels of detail and should be revised later.

```
<physDesc>
  <objectDesc>
    <supportDesc>
      <support>paper with watermark representing <watermark>a
      bird</watermark></support>
      <extent>34 folios <dimensions>
        <height>280</height>
        <width>190</width></dimensions>
      </extent>
    </supportDesc>
    <layoutDesc>
      <layout columns="1" ruledLines="23"> Single column of
      23 lines per page. </layout>
    </layoutDesc>
```

```

    </objectDesc>
</physDesc>

```

2.3.4. Provenance/history of the manuscript

The element `<history>` provide a mean to date the manuscript and to describe its provenance up to the present repository. What follow is a possible example of completion and should be revised.

```

<history>
  <origin>Created in <origDate notBefore="1814" notAfter="1816">the
summer of...</origDate> in <origPlace>Bath</origPlace>...</origin>

  <provenance>Bought by the Bodleian Library form a descendant of
the author</provenance>
</history>

```

2.3.5 Documentation on the digital reproduction

The element `<additional>` encodes information about any other existing reproduction of the source or of previous publication of the same material. In the `<surrogate>` element a documentation about the digital reproduction done for the present project can be encoded, together with the link from the present XML document to the METS document that encodes images metadata (to be revised).

```

<additional>
  <surrogates>
    <bibl>
      <title type="TIFF">digital pictures (master)</title>
      <!-- more details can go here -->
    </bibl>
    <bibl>
      <title type="JPG">digital pictures (web format)</title>
    </bibl>
  </surrogates>
[...]
```

2.3.6. Bibliography

After the `<surrogate>` element, the `<additional>` element allows to use a `<listBibl>` element, containing several `<bibl>` elements that may encode any previous edition of the present material.

```

<additional>
[...
  <listBibl>
    <bibl>here any relevant bibliographic item</bibl>
  </listBibl>
</additional>

```


5. Second layer markup

Will include possibly include:

- Punctuation semantic markup
- timed encoding for authorial corrections