Issues:

- are we using `<hi>` correctly?

- should we have used the part attribute for paragraphs?

Encoding Ticha Documents with TEI/XML
Guidelines by May Helena Plumb
last updated January 17, 2017

Contents

1   About this Documentation

2   Tags

2.1   TEI Prelude

2.1.1 `<TEI>`

Possible attributes: **xxx**, *... .*

http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-TEI.html

This tag...

2.1.2 `<teiHeader>`

Possible attributes: **xxx**, *... .*

http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-teiHeader.html

This tag... (marked as deprecated)

2.1.3 `<fileDesc>`

Possible attributes: **xxx**, *... .*

http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-fileDesc.html

This tag...

2.1.4 `<titleStmt>`

Possible attributes: **xxx**, *... .*

http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-titleStmt.html

This tag...

2.1.5 `<title>`

Possible attributes: **xxx**, *... .*

http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-title.html

This tag...

2.1.6 `<publicationStmt>`

Possible attributes: **xxx**, *... .*

http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-publicationStmt.html

This tag...

2.1.7 <profileDesc>

Possible attributes: ***xxx****, ...*
http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-profileDesc.html

This tag...

2.1.8 <encodingDesc>

Possible attributes: ***xxx****, ...*
http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-encodingDesc.html

This tag...

2.1.9 <sourceDesc>

Possible attributes: ***xxx****, ...*
http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-sourceDesc.html

This tag...

2.1.10 <langUsage>

Possible attributes: ***xxx****, ...*
http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-langUsage.html

This tag...

2.1.11 <language>

Possible attributes: ***xxx****, ...*
http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-language.html

This tag...

2.1.12 <projectDesc>

Possible attributes: ***xxx****, ...*
http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-projectDesc.html

This tag...

2.1.13 <editorialDecl>

Possible attributes: ***xxx****, ...*
http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-editorialDecl.
html

This tag...

2.1.14 <correction>

Possible attributes: ***xxx****, ...*
http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-correction.html

This tag...

2.1.15 <normalization>

Possible attributes: **xxx**, *... .*

http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-normalization.html

This tag...

2.1.16 <hyphenation>

Possible attributes: **xxx**, *... .*

http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-hyphenation.html

This tag...

2.1.17 <segmentation>

Possible attributes: **xxx**, *... .*

http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-segmentation.html

This tag...

2.1.18 <interpretation>

Possible attributes: **xxx**, *... .*

http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-interpretation.html

This tag...

2.1.19 <refsDecl>

Possible attributes: **xxx**, *... .*

http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-refsDecl.html

This tag...

2.1.20 <cRefPattern>

Possible attributes: **xxx**, *... .*

http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-cRefPattern.html

This tag...

2.1.21 <charDecl>

Possible attributes: **xxx**, *... .*

http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-charDecl.html

This tag...

2.1.22 <glyph>

Possible attributes: **xxx**, *... .*

http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-glyph.html

This tag...

2.1.23 <glyphName>

Possible attributes: ***xxx***, ...
http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-glyphName.html

This tag...

2.1.24 <figure>

Possible attributes: ***xxx***, ...
http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-figure.html

This tag...

2.1.25 <graphic>

Possible attributes: ***xxx***, ...
http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-graphic.html

This tag...

2.1.26 <add>

Possible attributes: ***xxx***, ...
http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-add.html

This tag...

2.1.27 <c>

Possible attributes: ***xxx***, ...
http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-c.html

This tag...

2.1.28 <del>

Possible attributes: ***xxx***, ...
http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-del.html

This tag...

2.1.29 <ptr>

Possible attributes: ***xxx***, ...
http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-ptr.html

This tag...

2.1.30 <ref>

Possible attributes: ***xxx***, ...
http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-ref.html

This tag...

2.1.31   <g>???

Possible attributes: ***xxx**, ...*
http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-g.html

This tag...

2.2   Structure Tags

2.2.1   <text>, <front>, <body>, and <back>

Possible attributes: none.
http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-text.html

The entire printed text of a Ticha document is enclosed in a <text> tag.

Printed documents in Colonial Mexico were divided into three parts: front matter, body, and end matter. The front matter included the cover, preliminary blank pages, a title page, and various letters of approval from local authorities. These sections are enclosed in a <front> tag. The body, the main portion of the text, is likewise enclosed in a <body> tag. The end matter, which included a list of printing errors, is enclosed in a <back> tag. The basic structure of a document, then, appears as below.

```
1  <text>
2    <front>
3      <!-- front matter -->
4    </front>
5    <body>
6      <!-- body text -->
7    </body>
8    <back>
9      <!-- end matter -->
10   </back>
11 </text>
```

None of these tags take attributes.

2.2.2   <div>

Possible attributes: *org, part, sample, xml:id, xml:lang .*
http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-div.html

The <div> tag indicates a text division. In Ticha we use this tag to indicate sections.

A <div> tag in Ticha always takes four attributes. The xml:id indicates the section number, using the convention "text#.#", for example "arte2.6.3". The part attribute indicates whether the parent element is fragmented, for example if a single poem was split across multiple <div>'s. The org (organization) attribute specifies how the the content of this div is organized. The sample attribute indicates whether any of the original source is missing.

Additionally, we can indicate the language used within a section with the xml:lang attribute.

2.2.3   <head>

Possible attributes: *type, xml:lang .*
http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-head.html

Ticha uses `<head>` tags to mark section headings. There are two kinds of section headings, those which are printed in the text and those which we have assigned to a particular section of text in our outline. As such, `<head>` can take the attribute `type`, with possible values of `"outline"` or `"[document]"`.

```
1 <head type="outline">2.2.2 Genitivo</head>
2 <head type="cordova">GENITIVO</head>
```

Additionally, we can indicate the language used within a heading with the `xml:lang` attribute.

2.2.4 `<p>`

Possible attributes: *`xml:lang`*, *`part`* .
http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-p.html

Each paragraph of text is surrounded by a `<p>` tag. This automatically assumes a new line. The `part` attribute indicates whether the parent element (in this case a paragraph) is fragmented. Additionally, we can indicate the language used within a paragraph with the `xml:lang` attribute.

2.2.5 `<pb/>`

Possible attributes: *`break`*, *`facs`*, *`n`*, *`type`* .
http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-pb.html

The `<pb/>` tag is a self-closing tag to indicate the begining of a new page. `<pb/>` tags come in pairs, one to indicate a new page of text and one to indicate a new page of the scan. The `type` attribute specifies which one. The `break` attribute indicates whether a word is broken across the page. The `n` attribute takes a value of the page number (e.g. `"iv"` or `"10r"`). The `facs` attribute supplies a link to the scan of that page.

2.2.6 `<lb/>`

Possible attributes: *`break`*, … .
http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-lb.html

The `<lb/>` tag is a self-closing tag to indicate the begining of a new line. The `break` attribute indicates whether a word is broken across the line.

2.2.7 `<fw>`

Possible attributes: *`type`, `place`, `rend`* .
http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-fw.html

The `<fw>` tag indicates forme work, that is running headers and footers on the page (e.g. page numbers and catchwords). The `type` attribute specifies what kind of forme work (e.g. `"pageNum"`). The `place` attribute describes where on the page it is located (e.g. `"bottom-right"`). The `rend` attribute describes how the forme work text is rendered.

2.2.8 `<pc>`

Possible attributes: *`type`* .
http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-pc.html

The tag `<pc>` stands for 'punctuation character'. Ticha uses it to surround hyphens when a word is broken across two lines. It takes a `type` attribute, in our case `"hyphen"`.

2.2.9   `<cb>`
   Possible attributes: **n** .
   http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-cb.html

The `<cb/>` tag is self-closing and marks the beginning of a new column (used only in passages with more than one column). The **n** attribute indicates the number of the column (moving in the direction of reading across the page).

2.3   Formatting Tags

2.3.1   `<hi>`
   Possible attributes: ***xxx****, ...*
   http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-hi.html

This tag...

2.4   Content Tags

2.4.1   `<choice>`
   Possible attributes: none.
   http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-choice.html

The `<choice>` allows for multiple encodings of a particular section of text; the reader has a "choice" of how to interpret that text. In Ticha this is used in two cases: when a word is spelled unconventionally, there is a choice between the original spelling (`<orig>`) and the regularized spelling (`<ref>`); when a word is abbreviated, there is choice between the abbreviation (`<abbr>`) and the expanded version (`<expan>`).

Note that `<choice>` tags may be nested and become quite complicated. For example, an abbreviation may have both an original and a regularized spelling. The example below shows the encoding for the Latin abbreviation *s*, short for *scilicet* 'that is to say', which is written in colonial texts using the long s.

```
1  <choice>
2    <abbr>
3      <choice>
4        <orig>f</orig>
5        <reg type="lat">s</reg>
6      </choice>
7    </abbr>
8    <expan>scilicet</expan>
9  </choice>
```

2.4.2   `<orig>`
   Possible attributes: none.
   http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-orig.html

The `<orig>` tag indicates the original version of the text, in a case where there is an alternate encoding. It always occurs within a `<choice>` tag. For example, the irregular

Spanish spelling *quatro* ('four') would be put within an `<orig>` tag, followed by a `<reg>` tag containing the regularized spelling, *cuatro*. Both tags would be contained with a `<choice>` tag.

2.4.3 `<reg>`

Possible attributes: type.
http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-reg.html

Where the original text has an unconventional or antiquated spelling of a word, the `<reg>` tag contains the regularized spelling of the word. It is contained with a `<choice>` tag and preceded by an `<orig>` tag containing the original spelling. For example, the irregular Spanish spelling *quatro* ('four') would be put within an `<orig>` tag, followed by a `<reg>` tag containing the regularized spelling, *cuatro*. Both tags would be contained with a `<choice>` tag.

In Ticha the type attribute indicates the language of the regularization (e.g. `"span"` for Spanish). This allows the functionality of enabling, for example, regularization of Spanish text but not Latin text.

2.4.4 `<abbr>`

Possible attributes: none.
http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-abbr.html

Where the original text contains an abbreviation, the `<abbr>` tag contains this abbreviation. It is always contained within a `<choice>` tag and followed by a `<expan>` tag containing the unabbreviated word or phrase.

2.4.5 `<expan>`

Possible attributes: none.
http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-expan.html

Where the original text contains an abbreviation, the `<expan>` tag contains the unabbreviated ('expanded') form o the word or phrase. It is always contained within a `<choice>` tag and preceded by a `<abbr>` tag containing the original abbreviation.

2.4.6 `<foreign>`

Possible attributes: xml:lang, xml:id, ana.
http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-foreign.html

The `<foreign>` tag surrounds a word or passage which is in a different language from the surrounding text. The `xml:lang` attribute indicates what language (e.g. `"cvz"`). The `xml:id` attribute has been used to match the foreign text with a linguistic analysis. still?? The `ana` attribute has been used irregularly. check with laurie

2.4.7 `<gap>`

Possible attributes: reason, agent.
http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-gap.html

The `<gap>` tag is used to encode text which is not physically represented on the original page, for example where the page has been torn or where a repeated phrase has been elided. The tag optionally takes the `reason` attribute indicating the reason the text is not on the page (e.g. `"illegible"`) and a `agent` attribute indicating the cause of the damage (e.g. `"mildew"`). In the *Arte* the `<gap>` tag is most commonly used with `reason="ellipsis"`, for example where Cordova has omitted a repeated verb stem and instead written just the verb endings.

3   Attributes

3.1   `xml:id`

del, div, foreign, glyph, refsDecl

3.2   `xml:lang`

foreign

3.3   `xml:space`

TEI

3.4   `part`

p, div

3.5   `org`

div

3.6   `sample`

div

3.7   `type`

pc, pb, reg, fw, add, del, ref, teiHeader, head

3.8   `break`

lb, pb

3.9   `n`

cb, pb

3.10   `facs`

pb

3.11 ref

g

3.12 rend

hi, fw, c

3.13 place

fw

3.14 ana

foreign

3.15 target

ptr, ref

3.16 default

sourceDesc, projectDesc, correction, normalization, hyphenation, segmentation, interpretation, editorialDecl, refsDecl

3.17 ident

language

3.18 usage

language

3.19 method

correction, normalization

3.20 status

correction

3.21 eol

hyphenation

3.22 matchPattern

cRefPattern

3.23 replacementPattern

cRefPattern

3.24 url

graphic

3.25 reason

The `reason` attribute is used with the `<gap>` tag to indicate the reason this text is not in the diplomatic transcription.

The value `ellipsis` is used when the author of the text has not included the gapped section at all. For example, in the *Arte* Cordova often elides repeated Zapotec morphemes (see 4.3).

The value `illegible` is used when the original text is unreadable due to damage. In this case, the `agent` attribute may also be used.

3.26 agent

The `agent` attribute is used with the `<gap>` tag to indicate the cause of illegible text. Possible values could include `mildew`, `bookworm`, or `printing error`.

4 Examples

4.1 Front Matter of Levanto's *Catecismo*

Printed books in Colonial Mexico commonly have three parts: the front matter, the body text, and the endmatter. The front matter includes the cover, preliminary blank pages, a title page, and various letters of approval from local authorities. We use the `<front>` tag to mark these pages. These tags take no attributes.

```
1 <front>
2   <!-- front matter -->
3 </front>
```

All text is contained within a section, which is marked by a `<div>` tag. `<div>` tags in Ticha can take four attributes. The `xml:id` indicates the section number, using the convention `"text#.#"`, in this case `"levanto1"`.

The `part` attribute indicates whether the parent element is fragmented, for example if a single poem was split across multiple `<div>`'s. Since Ticha uses the `<div>` tag only for sections, which are never fragmented, we use the value `"N"` (no).

The `org` (organization) attribute specifies how the the content of this `div` is organized; since this section is a logical unit and should be processed as a sequence, we use the value `"uniform"`. The `sample` attribute indicates whether any of the original source is missing. We use the value `"complete"` because we will transcribe everything in the section.

```
1 <front>
2     <div xml:id="levanto1" part="N" org="uniform" sample="complete">
3       <!-- front matter -->
4     </div>
```

```
5  </front>
```

The first element in each section is a `head` tag which prints the heading of the section. We differentiate between two types of section headings, headings which are printed in the text itself, and section titles which we assign to our outline of the text. For this we use the attribute `type`. In this case there is no printed text to mark the front matter of the book, so we use the value `"outline"`.

```
1  <front>
2    <div xml:id="levanto1" part="N" org="uniform" sample="complete">
3      <head type="outline">1 [Front matter]</head>
4      <!-- front matter -->
5    </div>
6  </front>
```

The first subsection of the front matter (Section 1.1) is the cover of the book. Again, while this is a section of the book, it is not marked with a section heading, and so it receives a `<head>` tag with `type="outline"`.

```
1  <front>
2    <div xml:id="levanto1" part="N" org="uniform" sample="complete">
3      <head type="outline">1 [Front matter]</head>
4      <div xml:id="levanto1.1" part="N" org="uniform" sample="complete">
5        <head type="outline">
6          1.1 [Cover of book and preliminary pages]
7        </head>
8        <!-- these preliminary pages have no text -->
9      </div> <!-- closing section levanto1.1 -->
10   </div> <!-- closing section levanto1 -->
11 </front>
```

The first page of front matter in Levanto's *Catecismo* that contains text is the title page in Figure 1. This is Section 1.2.

To begin we indicate the first relevant page break. The tag for this is `<pb/>`; notice that this is a self-closing tag. Page break tags can take several attributes. All `<pb/>`'s have the `break` attribute, which indicates whether it is a true break or whether a word is split between the two pages. This particular page break is indeed a break in the text, so here we use the value `"yes"`.

`<pb/>` tags also take the `type` attribute. We put two types of page beginnings at every page turn, one with `type` value `"levanto"`, indicating that the text itself changes pages, and one with value `"pdf"`, indicating tat the PDF scan of the text has a new page. The `"levanto"` type has a further attribute `n` which takes the value of the page number, in this case `"5"`. The `"pdf"` type has a further attribute `facs`, which has the value of the url to the scanned image of that page of the document.

```
1  <front>
2    <div xml:id="levanto1" part="N" org="uniform" sample="complete">
3      <head type="outline">1 [Front matter]</head>
4      <div xml:id="levanto1.1" part="N" org="uniform" sample="complete">
5        <head type="outline">
6          1.1 [Cover of book and preliminary pages]
7        </head>
8        <!-- these preliminary pages have no transcription -->
9        <pb break="yes" type="levanto" n="5"/>
```

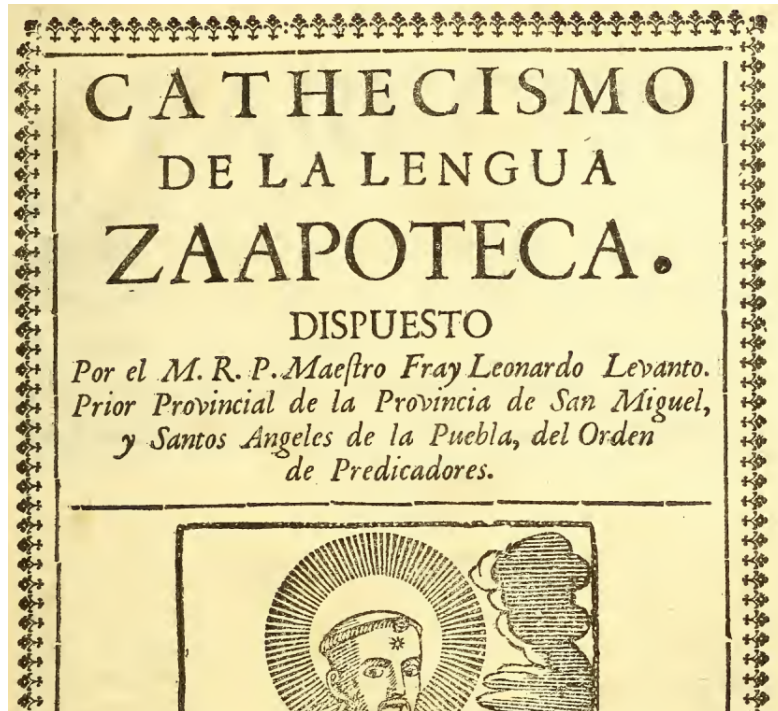Figure 1: Levanto's (1766) *Catecismo*, Title Page (Page 5)

```
10            <pb break="yes" type="pdf" facs="https://url/to/scan"/>
11         </div>
12      </div>
13   </front>
```

Now that we have begun the page, we want to begin Section 1.2. As before we use a `<div>` with the appropriate attributes and a `<head>` with `type="outline"`. (Here I will reduce the code to just the relevant `<div>`).

```
1 <div xml:id="arte1.2" part="N" org="uniform" sample="complete">
2    <head type="outline">1.2 [Title Page]</head>
3 </div>
```

This time we also have a heading which is printed in text, namely the title of the book, which requires a `<head>` tag with the attribute `type="levanto"`

```
1 <div xml:id="levanto1.2" part="N" org="uniform" sample="complete">
2    <head type="outline">1.2 [Title Page]</head>
3    <head type="levanto"></head>
4 </div>
```

Now we transcribe the text exactly as it appears in print. Line breaks are indicated with the self-closing `<lb/>` tag. Images, like the woodcut on this page, can be noted in a comment.

```
1 <div xml:id="levanto1.2" part="N" org="uniform" sample="complete">
2    <head type="outline">1.2 [Title Page]</head>
3    <head type="levanto">
4      C A T H E C I S M O
5      <lb/>D E L A L E N G U A
6      <lb/>Z A A P O T E C A.
7      <lb/>DISPUESTO
8      <lb/>Por el M. R. P.Maeftro Fray Leonardo Levanto.
```

```
 9      <lb/>Prior Provincial de la Provincia de San Miguel,
10      <lb/>y Santos Angeles de la Puebla, del Orden
11      <lb/> de Predicadores.
12    </head>
13    <!-- woodcut -->
14  </div>
```

Much of the Spanish in the title is non-standard, so we use `<choice>` tags to provide the regularized version. Notice that the original transcription is enclosed in the `<orig>` tag and the regularized version is enclosed in the `<reg>` tag. The `<reg>` tag also takes an attribute `type`, with the value of the language which is being regularized.

```
 1  <choice><orig>C A T H E C I S M O</orig><reg type="spanish">CATECISMO</
      reg></choice>
```

The final encoding the front matter so far is:

```
 1  <front>
 2    <div xml:id="levanto1" part="N" org="uniform" sample="complete">
 3      <head type="outline">1 [Front matter]</head>
 4      <div xml:id="levanto1.1" part="N" org="uniform" sample="complete">
 5        <head type="outline">
 6          1.1 [Cover of book and preliminary pages]
 7        </head>
 8        <!-- these preliminary pages have no transcription -->
 9        <pb break="yes" type="levanto" n="5"/>
10        <pb break="yes" type="pdf" facs="https://url/to/scan"/>
11      </div>
12      <div xml:id="levanto1.2" part="N" org="uniform" sample="complete">
13        <head type="outline">1.2 [Title Page]</head>
14        <head type="levanto">
15          <choice><orig>C A T H E C I S M O</orig><reg type="spanish">
      CATECISMO</reg></choice>
16          <lb/><choice><orig>D E</orig><reg type="spanish">DE</reg></
      choice>
17          <choice><orig>L A</orig><reg type="spanish">LA</reg></choice>
18          <choice><orig>L E N G U A</orig><reg type="spanish">LENGUA</reg
      ></choice>
19          <lb/><choice><orig>Z A A P O T E C A</orig><reg type="spanish">
      ZAPOTECA</reg></choice>.
20          <lb/>DISPUESTO
21          <lb/>Por el M. R. P.
22          <choice><orig>Maeftro</orig><reg type="spanish">Maestro</reg></
      choice>
23          Fray Leonardo Levanto.
24          <lb/>Prior Provincial de la Provincia de San Miguel,
25          <lb/>y Santos Angeles de la Puebla, del Orden
26          <lb/> de Predicadores.
27        </head>
28        <!-- woodcut -->
29      </div>
30    </div>
31  </front>
```

4.2   Body Text of Levanto's *Catecismo*
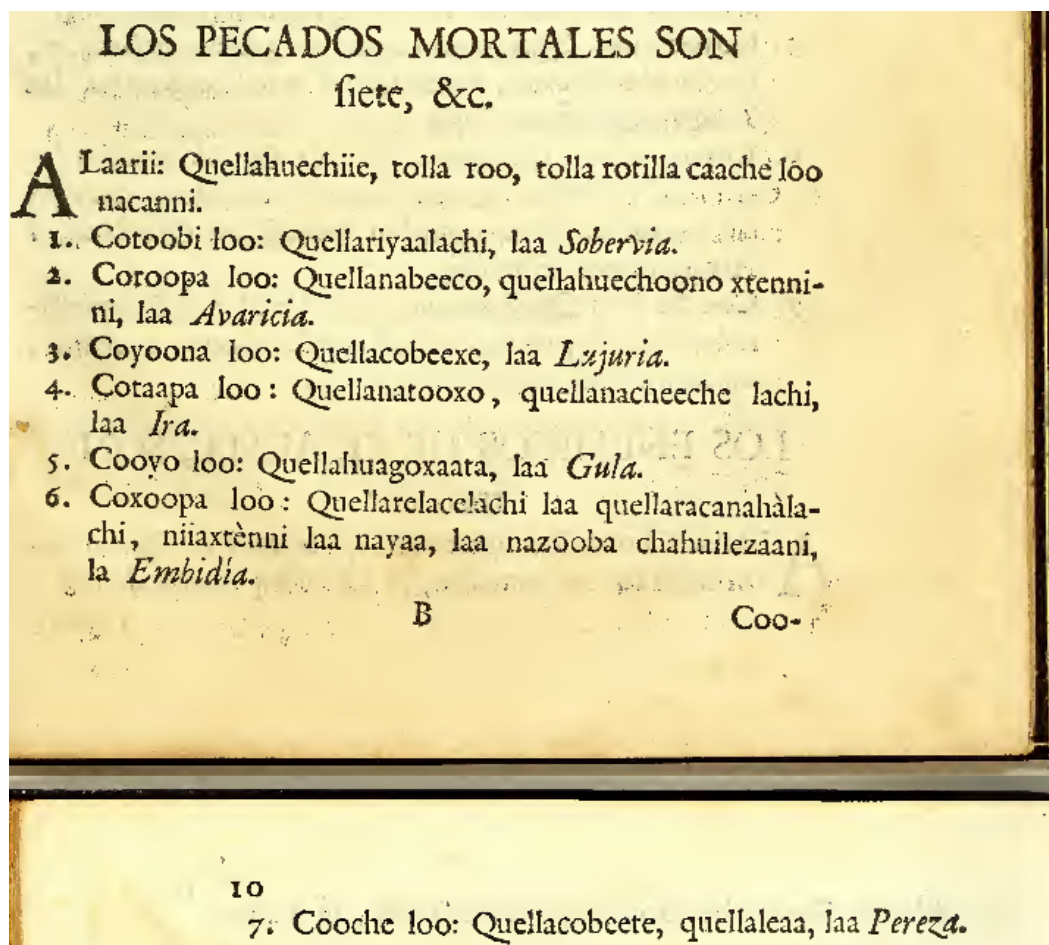
italics with ¡hi¿?

Figure 2: Levanto's (1766) *Catecismo*, Page 9–10

As before we mark the section with an appropriate marked `<div>`. Since this section is primarily in Zapotec, we using the `xml:lang` attribute with the value `"cvz"`.[1]

```
1  <div xml:id="levanto3.4" xml:lang="cvz">
2  </div>
```

We begin the section with two `<head>` marked with `type` attributes, one `"outline"` and the other `"levanto"`. Since we have marked the section as Zapotec, we use the `xml:lang` attribute with the value `"es"` to indicate that the headings are in Spanish. (Notice that ampersands are indicated using `&`, a predefined character reference.[2]

```
1  <div xml:id="levanto3.4" xml:lang="cvz">
2    <head type="outline" xml:lang="es" >
3      3.4 Los pecadores mortales
4    <head>
5    <head type="levanto" xml:lang="es" >
6      LOS PECADOS MORTALES SON
7      <lb/><choice><orig>fiete</orig><reg>siete</reg></choice>
8      , &amp;c.
9    </head>
10 </div>
```

---

[1]Depending on the structure of this part of the document, it may be appropriate to but the `xml:lang` tag at a higher level.

[2]See http://www.tei-c.org/release/doc/tei-p5-doc/en/html/SG.html#SG-er.

We then add the first paragraph of the body text, contained in a `<p>` tag.

```
1  <div xml:id="levanto3.4" xml:lang="cvz">
2    <head type="outline" xml:lang="es">
3      3.4 Los pecadores mortales
4    <head>
5    <head type="levanto" xml:lang="es">
6      LOS PECADOS MORTALES SON
7      <lb/><choice><orig>fiete</orig><reg>siete</reg></choice>
8      , &amp;c.
9    </head>
10   <p>
11     <choice><orig>ALaarii</orig><reg type="cvz">Alaarii</reg></choice>
12     :  Quellahuechiie, tolla roo, tolla rotilla caache loo
13     <lb/>nacanni.
14   </p>
15 </div>
```

The next paragraph has a word in Spanish. This is marked using the `<foreign>` tag with the `xml:lang` attribute.

```
1  <p>
2    1.   Cotoobi loo: Quellariyaalachi, laa
3    <foreign xml:lang="es">
4      <choice><orig>Sobervia</orig><reg type="spanish">Soberbia</reg></choice>
5    </foreign>.
6  </p>
```

In the last paragraph on the page, a word is broken across two lines. We indicate this by adding a `break` attribute with the value `"no"` to the `<lb/>` tag. The hyphen is contained in a `<pc>` tag with type `"hyphen"`,

```
1  <p>
2    6.   Coxoopa loo: Quellarelacelachi laa quellaracanahla
3    <pc type="hyphen">-</pc>
4    <lb break="no"/>chi, niiaxtnni laa nayaa, laa nazooba chahuilezaani,
5    <lb/>la <foreign xml:lang="es">Embidia</foreign>.
6  </p>
```

The bottom of page 9 and the top of page 10 both have forme work, which is headers and footers that are separate from the rest of the information on the page. This information is enclosed in `<fw>` tags described by a `type` attribute. The bottom of page nine has a signature marker "B", which is marked `type="sig"`, and a catchword "Coo-", which is marked `type="catchword"`. After the page break, at the top of page 10, we have a page number, which is marked `type="pageNum"`. Each `<fw>` tag also takes the `place` attribute, which describes where on the page it is located. The `<hi>` tag is used to indicate that this text is graphically distinct from the surrounding text.

```
1  <fw type="sig" place="bottom-center">
2    <hi>B</hi>
3  </fw>
4  <fw type="catchword" place="bottom-right">
5    <hi>Coo<pc type="hyphen">-</pc></hi>
6  </fw>
7  <pb break="yes" type="arte" n="10"/>
```

```
8  <pb break="yes" type="pdf" facs="https://url/to/scan"/>
9  <fw type="pageNum" place="top-left">10</fw>
```

The full encoding of this section is shown below:

```
1   <div xml:id="levanto3.4" xml:lang="cvz">
2     <head type="outline" xml:lang="es">
3       3.4 Los pecadores mortales
4     <head>
5     <head type="levanto" xml:lang="es">
6       LOS PECADOS MORTALES SON
7       <lb/><choice><orig>fiete</orig><reg>siete</reg></choice>
8       , &amp;c.
9     </head>
10    <p>
11      <choice><orig>ALaarii</orig><reg type="cvz">Alaarii</reg></choice>
12      : Quellahuechiie, tolla roo, tolla rotilla caache loo
13      <lb/>nacanni.
14    </p>
15    <p>
16      1.  Cotoobi loo: Quellariyaalachi, laa <foreign xml:lang="es"><
        choice><orig>Sobervia</orig><reg type="spanish">Soberbia</reg></
        choice></foreign>.
17    </p>
18  <!-- [snipped]  -->
19    <p>
20      6.  Coxoopa loo: Quellarelacelachi laa quellaracanahla<pc type="
        hyphen">-</pc>
21      <lb break="no"/>chi, niiaxtnni laa nayaa, laa nazooba
        chahuilezaani,
22      <lb/>la <foreign xml:lang="es">Embidia</foreign>.
23    </p>
24    <fw type="sig" place="bottom-center??">
25      <hi>B</hi>
26    </fw>
27    <fw type="catchword" place="bottom-right">
28      <hi>Coo<pc type="hyphen">-</pc></hi>
29    </fw>
30    <pb break="yes" type="arte" n="10"/>
31    <pb break="yes" type="pdf" facs="https://url/to/scan"/>
32    <fw type="pageNum" place="top-left">10</fw>
33    <p>
34      7.  Cooche loo: Quellacobcete, quellaleaa, laa <foreign xml:lang="
        es">Pereza</foreign>.
35    </p>
36  </div>
```

4.3   Ellipsis in Cordova's *Arte*

Throughout the *Arte*, Cordova frequently uses ellipsis to conserve space. This happens where there are multiple CVZ words that share a root. For example, he list two CVZ words for twenty-two, *callebitopa* and *callebicato*, but lists this as *callebitopa.l.cato*, eliding the second *callebi* and using *l*, the abbreviation for Latin *vel* 'or'.

To encode this we use the <gap> tag, which indicates letters which are supposed to be in the text but are not actually represented on the page. The <gap> tag takes a reason attribute and optionally an agent attribute. In this case we use the reason="ellipsis".
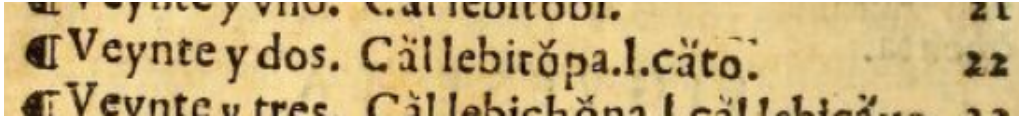
Figure 3: Cordova's (1578) *Arte*, Page 99v

(In another case we might say, for example, `reason="illegible"`). This line is encoded as below:

```
1  <p>
2    <pc>¶</pc>
3    <choice>
4      <orig>Veynte y dos</orig>
5      <reg type="spanish">Veintidós</reg>
6    </choice>.
7    <foreign xml:lang="cvz">Callebitopa</foreign>.
8    <foreign xml:lang="lat">
9      <choice><abbr>l</abbr><expan>vel</expan></choice>
10   </foreign>.
11   <foreign xml:lang="cvz">
12     <gap reason="ellipsis">Callebi</gap> cato
13   </foreign>. 22
14 </p>
```

4.4 Parallel Text in Levanto's *Catecismo*

Figure 4: Levanto's (1766) *Catecismo*, Page 17

IHS

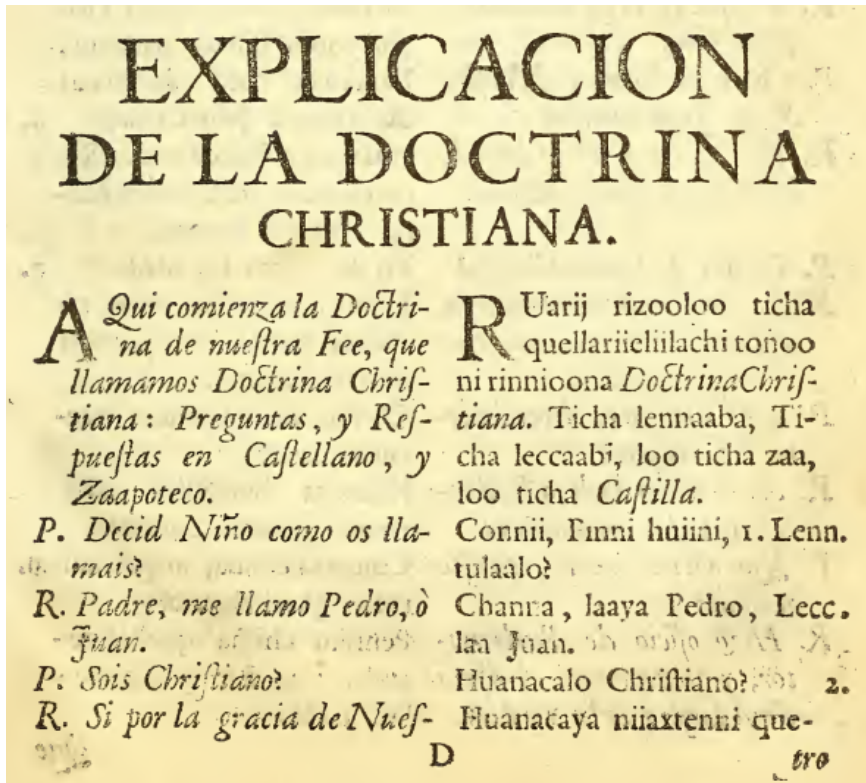VEIS aqui esta Doctrina, que ahora se sigue, es muy necessaria, que aprendais, y sepais de memoria: para que à la mañana, y â la tarde la digaîs, y rezeis siempre: esta Doctrina sirve para encomendaros â Dios, y agradecer sus beneficios, y alabar su grandeza: contiene tambien esta Doctrina las Leyes de nuestra Santa Madre Iglesia, que todos los Christianos han de saber, y aprender de memoria.

ALaarijricha, ni zeeta nallagaa annaá, nirij naquijñe teeté laató cozeetechahuito, cooquiqueto, nija gollato, quinnito zelli ricillalà, huachijla, ticha rii gonnizoniito, quinnixoniito, quinniicirooto, Bejuannana Dios: chela ticharii, ni naca xbea tichaa Xiñaana Sancta Iglesia, quiraalii benni Christianos gezeeteni, cooquiqueni.

***

CATHECISMO DE
LA DOCTRINA CHRISTIANA, EN LENGUA
Zaapoteca de el Valle.

Todo Fiel Christiano, &c.

TOobixeeneppe huazoopelao quiraali Benni Christiano: gapani quellarizinaachi loo Sancta Cruz xtenni Bejuannana Christo quellanayanii xtennina: quellani loo yoobi Sancta Cruz gocappelachi gatini: xitenni cociillaani tonoo loo quellatolla, ñeegaa loo xquellaritilie naxi-
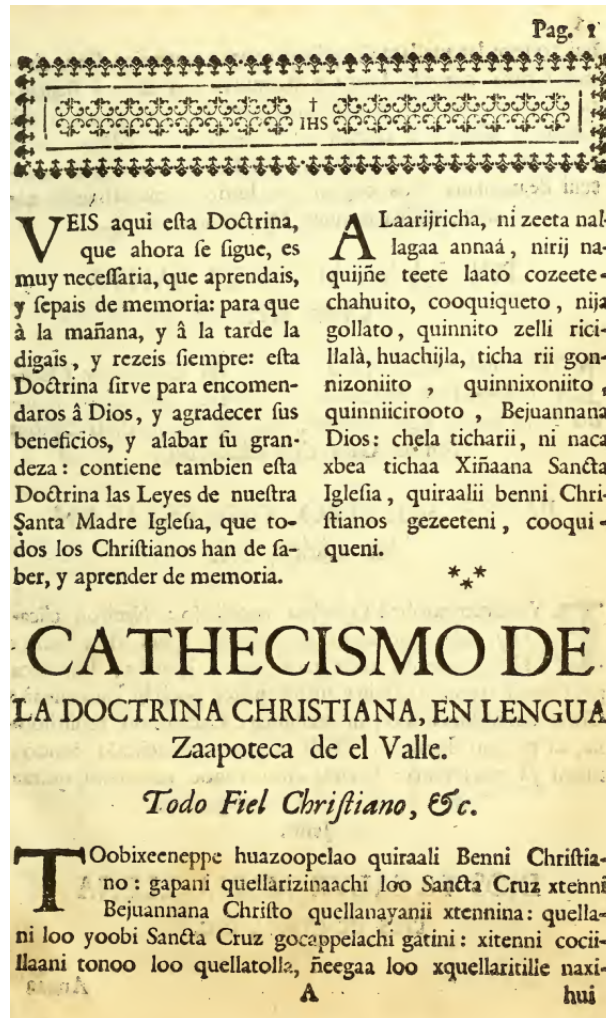
A   hui

Figure 5: Levanto's (1766) *Catecismo*, Page 1

21