

Estructura básica y elementos comunes de los documentos XML-TEI*

Susanna Allés Torrent

2019

I. Introducción

En esta lección nos centraremos en la estructura básica de los documentos XML-TEI, viendo solo aquellos elementos obligatorios e indispensables para empezar a trabajar con un documento TEI válido.

Recapitulando algunas nociones básicas, recordaremos que XML-TEI es:

- un marco de trabajo basado en el lenguaje de marcado XML que define una serie de guías directrices y recomendaciones para marcar electrónicamente textos de cualquier disciplina y de cualquier época.
- al ser XML, TEI es a su vez independiente de cualquier programa y plataforma.
- el objetivo final es el de organizar y estructurar los textos electrónicamente de manera que puedan ser procesados y explotados por las computadoras.
- el tipo de marcado utilizado puede ser personalizado según las necesidades de cada texto o cada proyecto.

En la [lección *El lenguaje XML \(eXtensible Markup Language\) y algunos conceptos generales*](#) explicamos los mecanismos que rigen el lenguaje XML, y ahora, debemos empezar a trabajar con las etiquetas que son propias del sistema TEI.

Un documento TEI es un documento XML de manera que sigue la misma sintaxis que cualquier otro documento XML. En esta imagen se reproduce la estructura mínima de un documento XML-TEI, que básicamente se compone de:

- el prólogo XML
- el encabezado
- el cuerpo del texto

*Estos materiales fueron creados en el marco del certificado de [Experto Universitario en Humanidades Digitales](#), ofrecido desde 2014 por el Laboratorio de Innovación en Humanidades Digitales de la Universidad Nacional de Educación a Distancia.

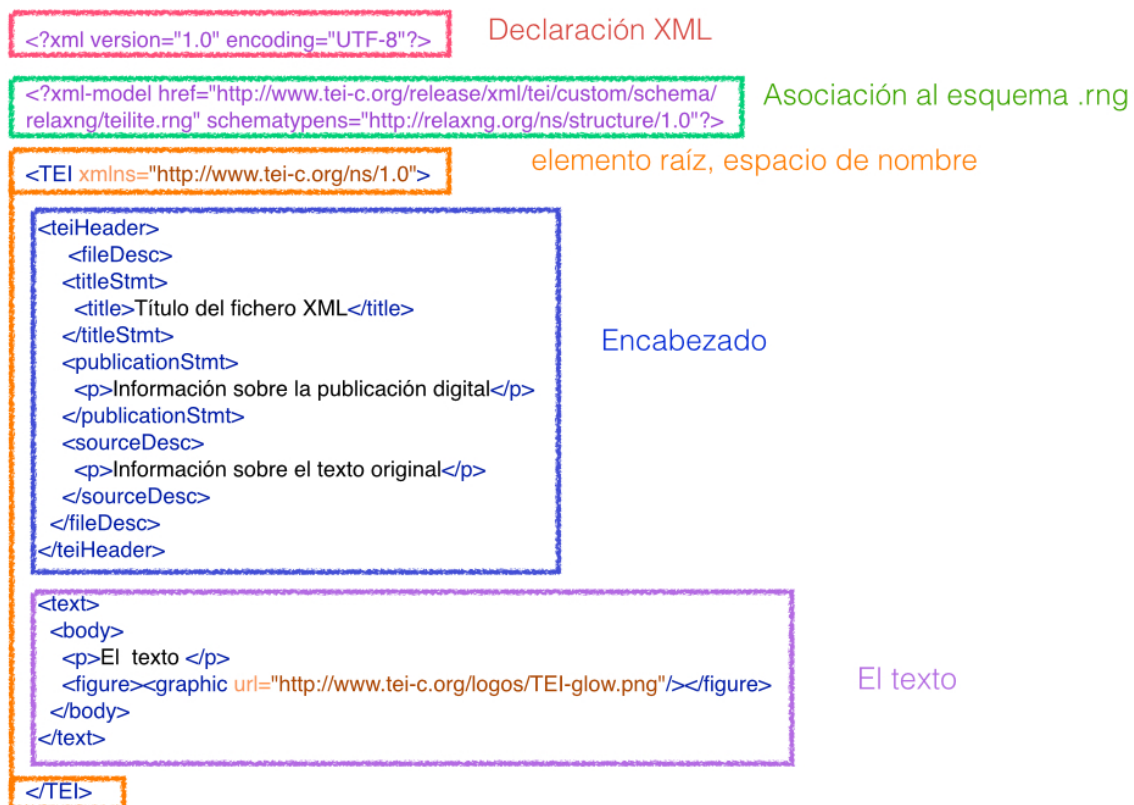


Figure 1: Estructura básica de un documento XML-TEI

En el prólogo del documento XML-TEI encontramos elementos que ya hemos visto anteriormente:

- La declaración XML nos indica que se trata de un documento XML y que el tipo de codificación utilizado es el UTF-8.
- La asociación del modelo de esquema: en este caso se trata de un esquema RelaxNG (.rng). Es la misma asociación que establecían las DTD a través del DOCTYPE (como vimos en el [Ejemplo 2](#)). Gracias a esta indicación sabemos que este documento XML-TEI depende de un esquema concreto, el [TEI-Lite](#), que es un modelo minimalista de TEI donde se contemplan los elementos obligatorios y algunos de los más utilizados.

Acabado el prólogo, aparece el elemento raíz <TEI> que engloba la totalidad del documento. Aquí irán todos los otros elementos XML-TEI: el encabezado y el texto, con la siguiente estructura:

```

1 <tei xmlns="http://www.tei-c.org/ns/1.0">
2   <teiheader>
3     ....
4   </teiheader>
5   <text>
6     ....
7   </text>
8 </tei>

```

II. Encabezado (teiHeader)

El documento XML-TEI consiste en dos grandes secciones: el encabezado que recibe el nombre de <teiHeader>, y el del texto propiamente dicho, llamado <text>.

En el encabezado encontramos los metadatos, es decir, las diferentes informaciones relativas al texto que estamos marcando.

Por ahora veamos cuáles son las partes obligatorias que debe tener cualquier documento XML-TEI:

```

1 <teiHeader>
2   <fileDesc>
3     <titleStmt>
4       <title>Titulo del fichero XML</title>
5     </titleStmt>
6     <publicationStmt>
7       <p>Informacion sobre la publicacion digital</p>
8     </publicationStmt>
9     <sourceDesc>
10      <p>Informacion sobre el texto original</p>
11    </sourceDesc>
12  </fileDesc>
13 </teiHeader>

```

En nuestro repositorio de ejemplos en GitHub, también encontraréis un ejemplo ([L3_documentoXML-TEI.xml](#)).

El `<teiHeader>` tiene un solo elemento obligatorio, llamado `<fileDesc>`. Este elemento es el responsable de contener tanto la información sobre el fichero XML-TEI con el que estamos trabajando, como los detalles de la fuente primaria:

- `<titleStmt>` contiene la información sobre el título del fichero propiamente dicho; éste a su vez debe ir dentro de un elemento más específico llamado `<title>`.
- `<publicationStmt>` recoge la información sobre la publicación digital; la información debe ir organizada en párrafos: `<p>`
- `<sourceDesc>` contiene la información de la fuente original.

Pongamos un ejemplo de la Biblioteca Virtual Miguel de Cervantes; imaginemos que estamos haciendo la edición digital de una edición antigua de la La Regenta (tal y como aparece en [este caso](#)). El encabezado podría marcarse de esta manera:

```

1  <teiheader>
2    <filedesc>
3      <titlestmt>
4        <title>La Regenta, por Leopoldo Alas (Clarín); prologo
5        de Benito Perez Galdos</title>
6      </titlestmt>
7      <publicationstmt>
8        <pubplace>Alicante</pubplace>
9        <publisher>Biblioteca Virtual Miguel de Cervantes</publisher>
10       <date>2000</date>
11      </publicationstmt>
12      <sourceDesc>La Regenta, por Leopoldo Alas (Clarín). Prologo
13      de Benito Perez Galdos, Madrid, Libreria de Fernando
14      Fe 1900</sourceDesc>
15    </filedesc>
16  </teiheader>

```

Podéis encontrar este documento, en nuestro Repositorio de Ejemplos en Github ([L3_Ejemplo_novela.xml](#)).

En el `<titleStmt>` hemos incluido el nombre del archivo que, en este caso, puede ser igual al de la obra original. Nada impediría que decidiéramos titularlo de una manera diferente, por ejemplo, “Edición digital de la Regenta (1900)”. En el elemento de `<publicationStmt>` aparece la información relativa a la publicación digital, en este caso el lugar (Alicante), el editor (Biblioteca Virtual Miguel de Cervantes) y el año de la publicación (2000).

Por otro lado, en el elemento `<sourceDesc>` debe aparecer la información de la obra original. Aquí se señala la información de la edición original de 1900. Podríamos, además, enriquecer las informaciones con la localización de la obra original, la división en tomos o cualquier otro dato que consideráramos pertinentes.

En las *Guías directrices* de TEI encontraréis toda la documentación relativa al `teiHeader`, así como múltiples ejemplos.

III. Cuerpo del documento

El cuerpo del documento corresponde al elemento `<text>` que puede contener, a su vez, tres (sub)elementos:

- `<front>`: se utiliza para marcar elementos paratextuales que preceden el texto, tales como prefacios, prólogos, cartas dedicatorias, una lista de personajes, etc. No es obligatorio.
- `<body>`: se consagra al texto propiamente dicho del documento y es obligatorio.
- `<back>`: puede utilizarse para añadir los apéndices, índices, cronologías, bibliografías, etc. No es obligatorio.

Siguiendo con el ejemplo anterior podríamos marcar el texto con la siguiente estructura:

```

1 <text>
2   <front>
3     <div type="prologo">
4   </div>
5 </front>
6 <body>
7   <div type="tomo" n="1">
8     <div type="capitulo" n="1"> ... </div>
9     <div type="capitulo" n="2"> ... </div>
10    <div type="capitulo" n="3"> ... </div>
11  </div>
12  <div type="tomo" n="2">
13    <div type="capitulo" n="1"> ... </div>
14    <div type="capitulo" n="2"> ... </div>
15    <div type="capitulo" n="3"> ... </div>
16  </div>
17 </body>
18 </text>

```

IV. Divisiones

Uno de los elementos más utilizados en el cuerpo del documento es el de “división”, `<div>`, que es utilizado para marcar cualquier tipo de sección estructural del texto, como tomos o volúmenes, libros, capítulos, partes, apartados, etc.

Del ejemplo anterior, se puede observar cómo hemos utilizado el elemento `<div>` para:

- caracterizar el “prólogo”.
- diferenciar los diferentes “tomos”.
- estructurar los diferentes “capítulos”.

Al afrontar cualquier estructuración de un texto, es pues un buen ejercicio empezar identificando las diferentes secciones del texto con `<div>` diferentes. Normalmente, cada una de las `<div>` tiene un atributo `@type` que ayuda a caracterizar esa sección en particular. El valor que puede conllevar `@type` es libre, con lo cual podemos utilizar el que mejor nos convenga (ej. libro, capítulo, escena, acto, poema, estrofa, etc.) y en la lengua que queramos.

Otro atributo muy utilizado es el `@xml:id` que asigna un identificador único a aquella sección particular del texto. Este mecanismo es imprescindible en el caso que queramos aislar determinadas porciones del texto. Por ejemplo, con este mecanismo al buscar una palabra podremos recuperar fácilmente en qué sección del libro estaba. El `@xml:id` se puede utilizar en una gran variedad de elementos, como bien podéis imaginaros es realmente útil en los casos de capítulos, líneas, versos, páginas, etc. Además, identificando las secciones de esta manera facilitamos el que otras partes del documento apunten a esa sección en concreto; de hecho, para la creación de “links”, este es un requisito indispensable: para crear un enlace necesitamos decirle al ordenador que vaya de un punto hacia otro, es decir, de un punto hacia un identificador concreto. Un elemento `<div>` con los atributos mencionados tendría esta forma:

```
1 <div type="tomo" xml:id="tomo_1" > ...</div>
```

El texto o la masa textual se incluye siempre al interior de elementos más concretos; no podemos añadir texto directamente dentro del elemento `<div>`. Por eso, en el caso que tengamos párrafos, utilizaremos el elemento `<p>`, en el caso que tuviéramos otros fragmentos de texto que no fueran párrafos, se podrían utilizar otros elementos más genéricos, como `<ab>` “anonymous block”, que indica cualquier unidad textual sin una semántica concreta (como es el caso, por ejemplo de “párrafo”, o “frase”).

V. Elementos comunes de todo documento TEI

Pasamos ahora a aquellos elementos generales y comunes en todos los documentos XML-TEI, como pueden ser párrafos, cuestiones de puntuación o de tipografía, nombres, fechas, abreviaturas, listas, notas a pie de página o de otra naturaleza, indexación, entre otros aspectos.

En primer lugar, conviene recordar que el sistema de codificación TEI se divide en módulos, cada uno destinado a una tipología textual: verso, drama, transcripciones de discurso, diccionarios, etc. Cada uno de estos módulos es descrito en las *Guías directrices*.

Los módulos pueden ser obligatorios o optativos. Así, los módulos obligatorios son cuatro: `tei`, `core`, `header`, `textstructure`; sin estos no tendríamos un documento válido. Los módulos optativos responden a tipologías textuales concretas.

Más arriba vimos aquellos elementos que constituían la estructura obligatoria y básica de un documento TEI. Vamos, aquí, a centrarnos en el módulo llamado «core», que contiene una serie de elementos o etiquetas que pueden utilizarse en cualquier documento, incluso en la versión más minimalista TEI Lite. La panorámica que ofreceremos no es completa, pero

corresponde a los más utilizados. Encontraréis la parte correspondiente a este tema de las Guidelines en la página web del consorcio.

Vamos a dividir los elementos en: Párrafos y cuestiones de puntuación, Elementos tipográficos, citas, Nombres, números, fechas, y Listas, notas, anotaciones, indexación, referencias, y un último apartado con Otros elementos frecuentemente utilizados.

Párrafos y puntuación

El párrafo es la unidad más pequeña en la que un texto en prosa puede dividirse. El elemento utilizado para marcar los párrafos es `<p>`.

```
1 <p>El único defecto que hallan en mi es el de que estoy muy delgadito, a fuerza
  ↳ de estudiar. Para que engorde se proponen no dejarme estudiar ni leer un
  ↳ papel mientras aquí permanezca, y además hacerme comer cuantos primores de
  ↳ cocina y de repostería se confeccionan en el lugar. Está visto: quieren
  ↳ cebarme. No hay familia conocida que no me haya enviado algun obsequio. Ya me
  ↳ envían una torta de bizcocho, ya un cuajado, ya una pirámide de piñonate, ya
  ↳ un tarro de almibar.</p>
2 <p>Los obsequios que me hacen no son sólo estos presentes enviados a casa, sino
  ↳ que también me han convidado a comer tres o cuatro personas de las más
  ↳ importantes del lugar.</p>
3 <p>Mañana como en casa de la famosa Pepita Jiménez, de quien Vd. habrá oído
  ↳ hablar sinduda alguna. Nadie ignora aquí que mi padre la pretende.</p>
```

(Juan Valera, *Pepita Jiménez*, Cap. I)

En los casos, por ejemplo, que quisiéramos dividir ulteriormente un párrafo, podríamos utilizar otros elementos como `<s>` (segment) para dividir otro tipo de segmentos, como por ejemplo, las frases:

```
1 <s type="frase">El unico defecto que hallan en mi es el de que estoy muy
  ↳ delgadito, a fuerza de estudiar.</s>
2 <s type="frase">Para que engorde se proponen no dejarme estudiar ni leer un papel
  ↳ mientras aquí permanezca, y además hacerme comer cuantos primores de cocina y
  ↳ de repostería se confeccionan en el lugar.</s>
3 <s type="frase">Está visto: quieren cebarme.</s>
4 <s>No hay familia conocida que no me haya enviado algún obsequio.</s>
5 <s type="frase">Ya me envían una torta de bizcocho, ya un cuajado, ya una
  ↳ pirámide de piñonate, ya un tarro de almibar.</s>
```

(Juan Valera, *Pepita Jiménez*, Cap. I)

El elemento `<p>` puede contener una gran variedad de elementos en su interior, prácticamente todos los que veremos a continuación.

Puntuación

De manera general, y especialmente en prosa, la puntuación del texto no suele codificarse; pero en algunos casos donde el marcado debe ser extremadamente sutil, se puede marcar. El elemento utilizado es:

- `<pc>` (punctuation character): utilizado para codificar el punto, el signo de interrogación y de exclamación, los puntos suspensivos, las comillas (aunque podrían eliminarse si se utiliza `<q>` o `<quote>`), apóstrofes, paréntesis, etc.

```
1 <s>
2 <seg>Está visto<pc>:</pc></seg>
3 <seg>quieren cebarme</seg><pc>.</pc>
4 </s>
```

La codificación de la puntuación puede ser interesante en aquellos casos en que se quiera dar fe de los diferentes cambios a lo largo de la historia. A veces, la ausencia o presencia de un punto, por ejemplo, puede cambiar el sentido.

Elementos tipográficos

Todos los textos impresos poseen elementos tipográficos diversos como pueden ser la cursiva, la negrita, las versalitas, incluso fuentes diversas, tamaño de la fuente, etc. La naturaleza de estos elementos tipográficos en las fuentes escritas pueden ser de diversa naturaleza:

- una palabra o frase en una lengua extranjera, un arcaísmo, un barbarismo, una palabra técnica, etc.
- una palabra o frase sobre la que se quiere poner un énfasis especial, que se resaltaría al hablar.
- un palabra o frase que en realidad no forma parte del discurso del texto, como las referencias cruzadas, los títulos o los encabezados, etc.
- cadenas narrativas diferentes, como un monólogo interno o un comentario.
- la atribución a una referencia externa al texto, como un discurso directo o una citación.

Ahora bien, como explicamos anteriormente, el sistema TEI no se concentra tanto en la apariencia del texto como en señalar qué son las cosas. De esta manera, al encontrarnos, por ejemplo con un extranjerismo, quizás no nos interesará tanto señalar que aparece en cursiva, como indicar que se trata de una palabra extranjera. Sea cual sea la opción que se elija, TEI ofrece la posibilidad de marcar estos elementos tipográficos de una manera general, sin implicaciones semánticas. El elemento utilizado a tal propósito es `<hi>` (highlight), que por lo general suele tener un atributo `@rend` (rendition) que indica de qué manera aparece en el texto:

```
1 <hi rend="italic">Palabra o frase en cursiva</hi>
2 <hi rend="bold">Palabra o frase en negrita</hi>
3 <hi rend="small-caps">Palabra o frase en versalitas</hi>
```

El valor de los elementos es libre, con lo cual nada nos impide señalarlo en otra lengua, por

ejemplo:

```
1 <hi rend="cursiva">Palabra o frase en cursiva</hi>
```

Podemos encontrar además otros elementos que responden a las tipografías diversas anteriormente mencionadas:

- `<foreign>` indica una palabra o frase perteneciente a otra lengua; normalmente posee el atributo `@xml:lang` para definir la lengua. Por ejemplo,

```
1 <p>El papa Francisco recuerda los conflictos activos en el mundo en la bendición
2 <foreign xml:lang="lat">urbi et orbi</foreign></p>
```

```
1 Te espero en el <foreign xml:lang="en">lobby</foreign>
```

- `<emph>` indica una palabra o frase a la que se atribuye un énfasis especial:

```
1 <p>Aquí, como en todas partes, la gente es muy aficionada al dinero. Y digo mal
  ↳ <emph>como en todas partes</emph>: en las ciudades populosas, en los grandes
  ↳ centros de civilización, hay otras distinciones que se ambicionan tanto o más
  ↳ que el dinero, porque abren camino y dan credito y consideracion en el
  ↳ mundo</p>
```

- `<distinct>` indica un término diferente lingüísticamente (arcaísmo, tecnicismo, dialectalismo, coloquialismo, etc.).
- `<mentioned>` es utilizado para señalar una palabra mencionada o citada.
- `<soCalled>` contiene una palabra o expresión respecto a la cual el autor manifiesta su rechazo de responsabilidad:

```
1 <p>El mencionado <soCalled>manuscrito</soCalled>, fielmente trasladado a la
  ↳ estampa, es como sigue.</p>
```

Citas

La marcas tipográficas también pueden utilizarse, como es sabido, para señalar citas de diferente naturaleza, ya sea un pensamiento, un discurso indirecto o directo, o incluso una cita literaria o bibliográfica. TEI ofrece una gran variedad de elementos para señalar estos fenómenos entre los que destacan:

- `<q>` (quoted) puede utilizarse para marcar el estilo directo, un pensamiento o algún término técnico, normalmente en las fuentes escritas aparece entre comillas:

```
1 <p>Al día siguiente, Bioy me llamo desde Buenos Aires. Me dijo que tenía a la
  ↳ vista el artículo sobre Uqbar, en el volumen XXVI de la Enciclopedia. No
  ↳ constaba el nombre del heresiarca, pero si la noticia de su doctrina,
  ↳ formulada en palabras casi idénticas a las repetidas por el, aunque -tal vez-
  ↳ literariamente inferiores. El había recordado: <q xml:lang="en">Copulation
  ↳ and mirrors are abominable</q>. El texto de la Enciclopedia decía: <q>Para
  ↳ uno de esos gnósticos, el visible universo era una ilusión o (más
  ↳ precisamente) un sofisma. Los espejos y la paternidad son abominables
  ↳ (<foreign xml:lang="en">mirrors and fatherhood are hateful</foreign>) porque
  ↳ lo multiplican y lo
```

```

2 divulgan.</q> Le dije, sin faltar a la verdad, que <q>me gustaria
3 ver ese articulo</q>. A los pocos dias lo trajo. Lo cual me sorprendió,
4 porque los escrupulosos indices cartográficos de la Erdkunde de Ritter
5 ignoraban con plenitud el nombre de Uqbar.</p>

```

(Borges, *Ficciones*)

- `<said>` (speech or thought) discurso hablado o pensamiento:

```

1 <p> Asi es que un día ambas se quedaron atonitas y pasmadas cuando, despues de
  ↳ varios requiebros, entre burlas y veras, D. Gumersindo soltó con la mayor
  ↳ formalidad y a boca de jarro la siguiente categórica pregunta:</p>
2 <said>Muchacha, ¿quieres casarte conmigo?</said>
3 <p> Pepita, aunque la pregunta venía después de mucha broma, y pudiera tomarse
  ↳ por broma, y aunque inexperta de las cosas del mundo, por cierto instinto
  ↳ adivinatorio que hay en las mujeres y sobre todo en las mozas, por cándidas
  ↳ que sean, conoció que aquello iba por lo serio, se puso colorada como una
  ↳ guinda, y no contestó nada. La madre contestó por ella:</p>
4 <said>Niña, no seas mal criada; contesta a tu tio lo que debes contestar: Tio,
  ↳ con mucho gusto; cuando Vd. quiera.</said>

```

(J. Valera, *Pepita Jiménez*)

- `<quote>` pasaje atribuido a una fuente externa:

```

1 <p> No conozco aún a Pepita Jimenez. Todos dicen que es muy linda. Yo sospecho
  ↳ que será una beldad lugareña y algo rústica. Por lo que de ella se cuenta, no
  ↳ acierto a decidir si es buena o mala moralmente; pero si que es de gran
  ↳ despejo natural. Pepita tendrá veinte años; es viuda; solo tres años estuvo
  ↳ casada. Era hija de doña Francisca Gálvez, viuda, como Vd. sabe, de un
  ↳ capitán retirado <quote>Que le dejo a su muerte, Solo su honrosa espada por
  ↳ herencia,</quote> según dice el poeta. Hasta la edad de diez y seis años
  ↳ vivió Pepita con su madre en la mayor estrechez, casi en la miseria.</p>

```

- `<cit>` se utiliza para citar de manera completa una referencia bibliográfica:

```

1 <p>El pasaje nos recuerda al
2 <cit>
3   <quote>"Arma virumque cano,..."</quote>
4   <bibl>(<author>Virgilio</author>,
5     <title>Eneida</title>, I)</bibl>
6 </cit></p>

```

Nombres, números y fechas

En múltiples ocasiones podemos encontrar en los textos nombres (ya sean topónimos, antropónimos, etc.), números y fechas. TEI propone para ello una serie de elementos que nos serán de gran utilidad.

Nombres

Un elemento general para codificar todo tipo de nombres o incluso frases es `<name>`; más adelante, veremos como existe un módulo específico para el marcado de nombres de lugar, de persona, etc. Otro elemento recurrente es `<rs>` (referring string) que indica un nombre con una denominación genérica o una cadena de referencia.

Los atributos más utilizados, a parte de `@xml:id`, son `@type` para especificarlos ulteriormente, y `@key` o `@ref` para referirse a un identificador concreto. Por ejemplo, supongamos esta frase: “Pepita Jiménez era hija de Francisca Gálvez quien la quería casar con D. Gumersindo”; la podríamos marcar de la siguiente manera:

```
1 <p><name xml:id="PJ">Pepita Jiménez</name> era hija de
2 <name xml:id="FG">Francisca Gálvez</name> que la hizo casar con
3 <name ref="#Gu">D. Gumersindo</name></p>
```

una vez establecidos los identificadores, podríamos referirnos a los personajes con facilidad en el fragmento que sigue:

```
1 <p>Era, con todo, tan inverosímil y tan desatinado el suponer que un hombre, que
  ↳ había pasado ochenta años sin querer casarse, pensase en tal locura cuando ya
  ↳ tenía un pie en el sepulcro, que ni la <rs ref="#FG">madre de Pepita</rs>, ni
  ↳ <name ref="#PJ">Pepita</name> mucho menos, sospecharon jamás los en verdad
  ↳ atrevidos pensamientos de <name ref="#Gu">D. Gumersindo</name>.</p>
```

Fechas

En lo que concierne a las fechas, se utiliza el elemento `<date>` y se regulariza su formato (de manera que la máquina la interprete correctamente) al interior del `@when`, cuyo formato es: año-mes-día. He aquí algunos ejemplos.

```
1 <date when="1980-03-23">23 de marzo de 1980</date>
2 <date when="1980"> el año 1980</date>
3 <date when="1999-10">octubre de 1999</date>
4 <date when="--02-14">cada 14 de febrero</date>
5 <date from="2014-02-10" to="2014-02-20"> del 10 de febrero al
6 20 de octubre de 2014</date>
```

Opcionalmente, se pueden especificar los diferentes sistemas de calendario (juliano, gregoriano, etc.) con `@calendar`:

```
1 <date when="1541" calendar="#juliano">1541</date>
```

En estos casos suele describirse el tipo de calendario utilizado en el elemento `<calendar>` en el TEI Header.

Podemos, además, especificar la hora con el elemento `<time>` que suele también incluir el atributo `@when` para regularizar la forma:

```
1 <time when="08:48:00">8:48</time>
```

Números y medidas

En lo que toca a los números y las medidas, se reducen al uso de los elementos `<num>`, para cualquier tipo de número (y a veces acompañados de los `@type`, `@value`), y de `<measure>` para marcar una cantidad (seguido opcionalmente de los atributos `@type`, `@quantity`, `@unit`, `@comodity`).

```
1 <num value="52">LII</num>
2 <num type="cardinal" value="15">quince</num>
3 <num type="percentage" value="20">veinte por ciento</num>
4 <num type="ordinal" value="6">6º</num>
5 <measure type="peso"><num>4</num> quilos de harina </measure>
```

Listas, notas, anotaciones, indexación, referencias

Listas

Las listas suelen ser un fenómeno habitual en muchos de los textos impresos. Los elementos más comunes que pueden constituir una lista son los siguientes:

- `<list>` secuencia de ítems que conforman una lista.
- `<item>` un componente de una lista.
- `<label>` etiqueta asociada a un ítem.

```
1 <list>
2   <head>Ingredientes:</head>
3   <item n="1">Harina</item>
4   <item n="2">Huevos</item>
5   <item n="3">Azucar</item>
6 </list>
7
8 <list type="gloss">
9   <head>Lista de medidas (traducción)</head>
10  <label>height</label>
11  <item>altura</item>
12  <label>width</label>
13  <item>ancho</item>
14  <label>depth</label>
15  <item>fondo</item>
16 </list>
```

Pueden también establecerse diferentes tipos de lista a través del atributo `@type`, que tiene estos valores por defecto:

- `bulleted`: ítems precedidos de puntos o marcas similares.

- **inline**: ítems presentados en prosa continua, sin saltos de página.
- **numbered**: ítems precedidos de un número.
- **simple**: ítems presentados como bloques, pero sin puntos ni números.

```

1 <list rend="numbered">
2   <head>Lista de temas a estudiar</head>
3   <item n="1">¿Qué es la TEI y qué aplicaciones prácticas tiene?</item>
4   <item n="2">Principios fundamentales de un documento XML</item>
5   <item n="3">La estructura básica de un documento TEI</item>
6   <item n="4">Elementos generales a todos los documentos TEI (core)</item>
7   <item n="5">...</item>
8 </list>

```

Notas

En muchos casos, podemos encontrarnos con notas o anotaciones de naturaleza diversa. El elemento para codificarlas es `<note>`, que sirve tanto para las que se encuentran en el texto impreso como las que son añadidas por el editor digital.

El elemento `<note>` tiene una serie de atributos que nos serán de gran utilidad, entre ellos destacamos:

- **@type**: permite caracterizar la nota en cuestión, si se trata de una nota al margen, de una nota a pie de página, etc.
- **@place**: es especialmente útil cuando se transcribe un manuscrito que posee notas en los márgenes; así por ejemplo podemos indicar la localización física de la nota.
- **@xml:id**: se utiliza, como hemos visto, para dar un identificador único a la nota en cuestión, de manera que podamos referirnos a esta con facilidad y crear, si lo deseamos, un enlace a la misma
- **@target**: lo utilizaríamos para conectar la referencia con la nota o viceversa.

Las notas pueden marcarse de dos maneras diferentes; una opción es situarla donde aparece la referencia a la misma en el cuerpo del texto (Opción 1); la otra consiste en crear una zona separada (una `<div type="notas">` por ejemplo) y apuntar a la referencia a través del **@target**:

Opción 1:

```

1 <head>Historia y organización<note n="1">Todas las discusiones e informaciones
  ↳ relativas entorno a TEI desde sus inicios se encuentran en el repositorio de
  ↳ la lista TEI</note></head>

```

Opción 2:

```

1 <head>Historia y organización<ref target="nota1"
  ↳ xml:id="referencia1">1</ref></head>
2 ....
3 [...]
4 <div type="notas">

```

```

5 <note xml:id="nota1" target="#referencia1">Todas las discusiones e informaciones
  ↳ relativas entorno a TEI desde sus inicios se encuentran en el repositorio de
  ↳ la lista TEI </note>
6 </div>

```

Indexación

En algunas ocasiones podemos también encontrarnos con índices de términos, para ello podemos utilizar estos dos elementos:

- `<index>` marca una entrada de índice, puede llevar `@indexName`; en su interior pueden anidarse diferentes índices.
- `<term>` marca un término en el interior del índice.

Puede ocurrir que la fuente original no tenga un índice, pero que queramos construir uno. Esto puede hacerse de diferentes maneras, una de ellas sería utilizando los dos elementos anteriores y marcar los términos que queramos introducir en un nuevo índice. Por ejemplo, imaginemos que queremos hacer un índice de personajes:

```

1 <index indexName="Personajes"><term xml:id="PJ">Pepita</term></index> tendrá
  ↳ veinte años; es viuda; sólo tres años estuvo casada. Era hija de doña <index
  ↳ indexName="Personajes"> <term xml:id="FG">Francisca Gálvez</term></index>,
  ↳ viuda, como Vd. sabe, de un capitán retirado.

```

De esta manera sería fácil autogenerar un índice con los nombres de los personajes.

Enlaces y referencias

Una de las ventajas de crear un hipertexto es justamente la posibilidad que tenemos de conectar las diferentes secciones del texto a otras y establecer todo tipo de enlaces. En TEI los enlaces pueden realizarse de dos maneras, con los siguientes elementos:

- `<ptr>` (pointer) un indicador hacia otra localización
- `<ref>` (reference) define una referencia hacia otra localización, normalmente contiene un texto que funciona como el enlace. Ambos suelen utilizar, entre otros, el atributo `@target` que toma la forma de una referencia URI.

Otros elementos

Existen mucho otros elementos entre los cuales podríamos destacar los cambios editoriales, tales como los errores, correcciones, adiciones, supresiones, etc. y otros elementos pertenecientes a la estructura física del texto, como los saltos de línea, de página o de columna.

Cambios editoriales

Como ocurre en la edición de fuentes escritas, a veces debemos señalar el proceso de constitución de un texto en sus diferentes fases (copistas, editor papel, editor digital, etc.). Para recoger las intervenciones editoriales disponemos de una serie de elementos básicos que podemos utilizar en todos los documentos TEI. Para entender la codificación de estos fenómenos debemos visualizar la intervención como una unidad en la que hay etapas o elementos diferentes. El elemento encargado de recoger estos marcados editoriales alternativos es `<choice>`, y puede utilizarse en los siguientes casos:

- Corrección de errores: ante un error, se utiliza el elemento `<sic>` para marcar el error aparente; con el elemento `<corr>`, en cambio, se recoge la corrección de dicho error:

```

1 <p>Hasta cinco mujeres han venido a verme que
2 <choice>
3   <sic>todos</sic><corr>todas</corr>
4 </choice> han sido mis amas y me han abrazado y besado.</p>
```

- Regularización de formas ortográficas: la forma original que aparece en el texto se marca con el elemento `<orig>`, mientras que su forma regularizada, con el elemento `<reg>`:

```

1 <choice>
2   <orig>Comiença</orig>
3   <reg>Comienza</reg>
4 </choice> la
5 <choice>
6   <orig>ystoria</orig>
7   <reg>historia</reg>
8 </choice>...
```

- Abreviatura: la forma abreviada aparece en el elemento `<abbr>`, mientras que la forma expandida bajo `<expan>`:

```

1 <p>...un largo
2 <choice>
3   <abbr>etc.</abbr>
4   <expan>etcétera</expan>
5 </choice>
6 </p>
```

Hay otro tipo de intervenciones que pueden codificarse de manera aislada, como las adiciones, las eliminaciones, las omisiones, e incluso las lecturas dudosas:

- `<add>`: adición al texto, como por ejemplo una glosa marginal.
- ``: una eliminación en el texto, como una tachadura.
- `<gap>`: indica un punto donde el material ha sido omitido; puede tener el atributo `@reason` para señalar el motivo.
- `<unclear>`: indica una lectura dudosa.

Milestones

Dentro de la categoría que las *Guías directrices* denominan “Milestone” encontramos diversos elementos que pueden indicar ciertas características físicas del documento. Entre estos elementos encontramos:

- `<milestone>`: marca un punto límite separando cualquier tipo de sección de un texto.
- `<gb>` (gathering begins) puede indicar el inicio o final de la foliación de un manuscrito, por ejemplo donde empieza y donde acaba un cuaderno.
- `<pb>` (page break): indica el cambio de página.
- `<lb>` (line break): indica el cambio de línea.
- `<cb>` (column break): indica el cambio de columna

VI. Bibliografía

- Burnard, L. (2014), What is the Text Encoding Initiative. How to add intelligent markup to digital resources, Marserville: OpenEdition Press <https://books.openedition.org/oep/681>

Tutoriales:

- “Module 1: Common Structure and Elements.” *TEI by Example*, <http://teibyexample.org/modules/TBED01v00.htm>
- Schreibman, Susan & Roman Bleier. “Text encoding and the TEI”. #dariahTeach. DH Teaching material, 2017.

Cita propuesta:

Allés Torrent, Susanna (2019). “Estructura básica y elementos comunes de los documentos XML-TEI”. *TTHUB. Text Technologies Hub: Recursos sobre tecnologías del texto y edición digital*. <https://tthub.io/> DOI: