

¿QUÉ ES LA

INICIATIVA DE CODIFICACIÓN DE TEXTOS?

*Cómo añadir marcado inteligente a
los recursos digitales*

LOU BURNARD

¿Qué es la Iniciativa de Codificación de Textos?

Cómo añadir marcado inteligente a los recursos digitales

Lou Burnard

Traductor: Nicolás Vaughan

Editor: OpenEdition Press
Lugar de edición: Marseille
Año de edición: 2022
Publicación en OpenEdition Books: 11 mayo 2022
Colección: Encyclopédie numérique
EAN electrónico: 9791036592294



<https://books.openedition.org>

Referencia electrónica

BURNARD, Lou. *¿Qué es la Iniciativa de Codificación de Textos? Cómo añadir marcado inteligente a los recursos digitales*. Nueva edición [en línea]. Marseille: OpenEdition Press, 2022 (generado el 13 mai 2022). Disponible en Internet: <<http://books.openedition.org/oep/15662>>. ISBN: 9791036592294.

Este documento fue generado automáticamente el 13 mayo 2022.

© OpenEdition Press, 2022

Creative Commons - Atribución-NoComercial-SinDerivadas 4.0 Internacional - CC BY-NC-ND 4.0

RESÚMENES

Las Directrices de la Iniciativa de Codificación de Textos (TEI) se consideran desde hace tiempo el estándar de facto para la preparación de recursos digitales de texto en la comunidad de investigadores académicos.

Ofrecen un abanico de posibilidades abrumador para el principiante, que refleja la enorme gama de posibles aplicaciones de la codificación de textos: desde las ediciones académicas tradicionales hasta los corpus lingüísticos, los léxicos históricos, los archivos digitales y mucho más.

Apoyado en numerosos ejemplos de textos codificados con TEI procedentes de diversos campos de investigación, este libro sencillo y directo pretende ayudar al principiante a tomar sus propias decisiones entre toda la gama de posibilidades de TEI.

Explica la tecnología XML utilizada por TEI en un lenguaje accesible al lector lego, ofrece una visita guiada a las numerosas partes del universo TEI y explica cómo puede personalizarse para adaptarse a las necesidades de un proyecto individual.

LOU BURNARD

Antiguo subdirector de los servicios de computación de la Universidad de Oxford, Lou Burnard es uno de los editores originales de las Directrices TEI y ha estado estrechamente vinculado con TEI a lo largo de su evolución.

Es egresado de la Universidad de Oxford con una maestría en Lengua y Literatura Inglesas.

Lou ha trabajado en la frontera entre la computación y las humanidades desde la década de 1970, cuando fundó el Oxford Text Archive.

Durante los años noventa y el dos mil fue responsable de muchos otros desarrollos clave en lo que ahora llamamos las humanidades digitales, especialmente el British National Corpus y el Arts and Humanities Data Service del Reino Unido.

NICOLÁS VAUGHAN (TRADUCCIÓN)

Es profesor asociado del Departamento de Humanidades y Literatura de la Universidad de los Andes, Bogotá, donde también enseña en la Maestría en Humanidades Digitales.

Hizo su doctorado en filosofía medieval en la Universidad de Oxford. Actualmente es editor en el Programming Historian en Español, donde también ha publicado dos lecciones sobre codificación TEI. Es editor adjunto de la *Edición Crítica en línea de la 'Ordinatio' de Adam de Wodeham*, publicado por el Scholastic Commentaries and Texts Archive (SCTA).

ÍNDICE

Introducción

TEI y XML

Organización estructural de un documento TEI

Encabezado, texto y divisiones

Las variedades de la estructura textual

Un artículo de revista académica

Un texto poético

Un texto teatral

Una tarjeta postal

Un texto mínimamente estructurado

Textos complejos

Cornucopia TEI, primera parte

Hitos

Idiomas y sistemas de escritura

Idiomas

Caracteres no estándar

Representación

Nombres y fechas

Cuadros, figuras y bibliografías

Cuadros

Figuras

Referencias bibliográficas

Cornucopia TEI, segunda parte

Asuntos editoriales

Transcripción de un discurso oral

Diccionarios

Notas

Anotación lingüística

El encabezado TEI

Descripción del archivo

Variedad de descripciones de la fuente

Descripción de la codificación

Descripciones del perfil y de la revisión

Personalización de TEI

Construcción de una personalización

Añadir un nuevo elemento

Personalización y conformidad

Conclusión: ¿Qué es TEI?

Lecturas complementarias

Glosario

Introducción

- 1 La Iniciativa de Codificación de Textos (TEI, por sus siglas en inglés: *Text Encoding Initiative*) es uno de los proyectos más antiguos e influyentes en el campo que ahora se conoce como “humanidades digitales”. Su objetivo es proporcionar directrices para la creación y gestión en formato digital de todo tipo de datos creados y utilizados por investigadores en humanidades, como fuentes históricas, manuscritos, documentos de archivo, inscripciones antiguas y muchos otros. Como su nombre sugiere, su objeto principal es el texto, más que el sonido o el video, aunque puede ser aplicado exitosamente a toda forma de datos digitales. Es importante subrayar que ha sido creado y mantenido por la comunidad científica, para su propio uso. En este breve libro intentaremos explicar la naturaleza y los objetivos de TEI, y ofrecer algunos ejemplos concretos de cómo puede utilizarse en diversos proyectos de creación de recursos digitales. Antes de esto, sin embargo, intentaremos responder a una pregunta que textos introductorios como este tienden a dejar de lado: ¿por qué, en primer lugar, molestarse en aprender o usar TEI? Después de todo, casi todo el mundo sabe utilizar un procesador de textos para crear documentos digitales, o un programa de gestión de base de datos para crear y gestionar información, o un simple editor para crear páginas web.
- 2 Hay muchas respuestas a esto, aunque discutiremos tres en particular: la primera es que TEI se centra en el significado del texto más que en su apariencia; la segunda es que TEI es independiente de cualquier entorno de software particular; y la tercera es que TEI fue diseñado por y para la comunidad académica, que también es responsable de su desarrollo continuo.

- **El significado antes que el formato**

Si a usted le importan más las palabras de su texto y su significado que su disposición en la página, pronto encontrará algunas limitaciones frustrantes al usar un procesador de textos corriente. Casi cualquier programa de ese tipo encontrará fácilmente, entre cientos de páginas de texto, todas las veces que aparezca la cadena de caracteres “Londres”. Pero una búsqueda inteligente, capaz de distinguir “Londres” como un nombre de un lugar de Canadá del de un lugar de Inglaterra o del apellido de un autor, requiere algo más. Como señalaron Coombs et al. en 1987, hay una diferencia significativa entre la forma como un científico considera los recursos textuales y la forma como la mayoría de los sistemas automatizados de preparación de documentos están diseñados.

Retomando una expresión citada a menudo, si lo que ves es lo que obtienes (*What you see is what you get*, o WYSIWYG), entonces solo obtienes lo que ves; en otras palabras, los sistemas que se preocupan principalmente de reducir la brecha entre lo que se ve en la pantalla y lo que se escribe en el teclado tienden a conseguirlo reduciendo lo que se puede expresar y almacenar en formato digital a información sobre la apariencia de cadenas de caracteres, en lugar de sobre sus propiedades lingüísticas o su función. Usted podrá ajustar cuanto quiera el tipo de letra, tamaño, alineación o color de cada una de las apariciones de la palabra "Londres" en el documento; pero no podrá utilizar un programa de procesamiento de textos de forma fiable o sistemática para indicar cuáles hallazgos son nombres propios y cuáles son nombres de lugar.

- **Independencia del software**

Si uno quiere compartir con otras personas (o consigo mismo dentro de unas pocas décadas) los recursos textuales que ha creado, también debe preocuparse por la tendencia de muchos sistemas computacionales a aplicar su propia forma de almacenar la información. ¿Hasta qué punto es posible utilizar en un programa de computador diferente un documento o una base de datos creados con su actual programa favorito? ¿Puede transferirse desde su PC al Mac de ella o a mi Linux sin pérdida de información? Por supuesto, existen programas para realizar transformaciones entre las docenas de formatos de archivo diferentes utilizados actualmente. Esa es la buena noticia. La mala es que esas transformaciones son siempre necesarias y a menudo son complejas, susceptibles de pérdida de información o costosas de realizar. Incluso dentro del mismo sistema operativo sigue siendo problemática una verdadera interoperabilidad: un documento preparado en un procesador de textos (por ejemplo, Word) puede o no verse igual al abrirlo en otro (digamos, LibreOffice); una base de datos creada en Filemaker Pro puede o no ser fácil de convertir en otra que utilice Microsoft Access, y viceversa.

La información capturada en un documento TEI-XML tiene exactamente el mismo aspecto en todos los programas que lo utilizan. No hay una magia especial, accesible solo a un programa ejecutado en un entorno concreto. Un archivo TEI-XML, como otros archivos XML, es solo una secuencia de caracteres legibles, sin códigos secretos incrustados. De hecho, un documento TEI-XML representa un texto en gran medida *sin tener de ningún modo en cuenta* el software que podría utilizarse para visualizarlo, analizarlo o almacenarlo. Por el contrario, el contenido de un documento TEI representa directamente lo que su creador considera significativo en su estructura y contenido. Si hay una cadena de caracteres que quien codifica identifica como el nombre de una persona o de un lugar, o como una palabra extranjera, o como el título de una canción, o como un pasaje en discurso directo —por ejemplo—, TEI dispone de etiquetas para identificarlos como tales. Quizás sea deseable visualizar los títulos de las canciones en cursiva cuando se imprimen en la pantalla o en papel, pero también puede serlo añadir enlaces a una entrada de la base de datos cuando se muestran en un sitio web. Puede ser deseable incluir (o excluir) pasajes en discurso directo o palabras extranjeras al hacer un análisis de vocabulario de un texto.

Al fin y al cabo, los textos en formato digital pueden utilizarse de muchas maneras: por eso los creamos. Si lo único que se necesita es una nueva copia de un libro antiguo, bastará con una fotocopia digital. Pero si el objetivo es hacer que un texto antiguo sea accesible a todo tipo de nuevos lectores, entonces no bastará con reproducir su apariencia en la página. TEI nos da un marco conceptual para representar todo lo que se considere importante acerca del texto, no solo su apariencia, de modo que el software pueda actuar sobre las diferentes características identificadas, generando nuevas visualizaciones y nuevas perspectivas. En

consecuencia, decidir precisamente qué marcado TEI utilizar es siempre una actividad intelectual que involucra una elección consciente.

• **Impulsado por la comunidad**

TEI fue desarrollado por primera vez hacia finales de la década de 1980, como un proyecto de investigación en el campo entonces emergente de la “computación para las humanidades” (*humanities computing*). La idea original era proponer un conjunto de recomendaciones sobre la forma como los investigadores debían crear recursos textuales “legibles por computador”, que se adecuara a las necesidades de la investigación científica, siempre y cuando existiera un consenso sobre el asunto, pero que también fueran ampliables a medida que esas necesidades cambiaran y evolucionaran. Esta combinación entre flexibilidad y un enfoque en las necesidades científicas de una comunidad particular convierte a TEI en un caso inusual comparado con otros esfuerzos de estandarización, y es también uno de los factores que subyacen a su longevidad. En un sentido estrictamente darwiniano, TEI evoluciona respondiendo a los cambios que su comunidad de usuarios considera beneficiosos, mientras que ignora los que no lo son.

Esta estrecha vinculación de TEI con su comunidad de usuarios también tiene efectos beneficiosos para los principiantes: el conocimiento de TEI y la pericia pueden encontrarse ahora fácilmente en la mayoría de disciplinas de las humanidades: no está restringido a un área recóndita de las ciencias de la información. Instituciones como la lista de correo TEI-L ofrecen un testimonio vívido de la naturaleza heterogénea de la comunidad TEI, además de demostrar repetidamente su apertura, cordialidad y disposición a ayudar al principiante.

- 3 El objetivo de este libro es presentar una muestra de los numerosos tipos de marcado TEI, sin entrar en mucho detalle en ninguna de ellas. Esperamos que sea accesible a un amplio número de lectores, incluyendo tanto al humanista digital en activo como a aquellos que aún no han sucumbido del todo a la atracción del computador. Para entenderlo, basta con tener unas nociones muy generales sobre las tecnologías XML e interés por los problemas tradicionales de los estudios sobre textos. También hay que saber que utilizar TEI no implica utilizarlo en su totalidad: al contrario, la mayoría de los usuarios de TEI solo utilizan un pequeño subconjunto de lo que ofrece. Pero decidir cuál subconjunto es el mejor para sus propósitos es algo que solo usted puede hacer.
- 4 El primer capítulo, “TEI y XML”, ofrece una pequeña base técnica que introduce los conceptos del lenguaje XML en los que se basa el marcado TEI. El segundo capítulo, “La organización estructural de un documento TEI”, describe cómo se estructuran los documentos TEI, seguido de una serie de ejemplos típicos examinados en el tercer capítulo, “Variedades de la estructura textual”. Los dos capítulos siguientes profundizan en la “cornucopia TEI”, discutiendo primero algunas características textuales comunes a la mayoría de los tipos de documentos y luego algunas de las posibilidades más especializadas que ofrece el sistema TEI. El sexto capítulo describe el encabezado TEI, esto es, la *portada digital obligatoria* de un documento TEI que proporciona sus metadatos. El séptimo capítulo, “Personalización de TEI”, analiza cómo adaptar el sistema TEI a las necesidades de un proyecto concreto. Concluimos con una discusión sobre la organización del propio consorcio TEI, su presente y su futuro.

TEI y XML

- 1 TEI hace énfasis en aquello que es común a todo tipo de documento, ya sea representado físicamente en forma digital en disco o en tarjeta de memoria, en forma impresa como libro o periódico, en forma escrita como manuscrito o códice, o en forma inscrita en piedra o tablilla de cera. Esta continuidad facilita la migración del texto desde manifestaciones más antiguas, como la impresión o el manuscrito, a otras más recientes, como el disco o la pantalla. Por lo tanto, la visión de TEI de lo que el texto realmente *es* está condicionada en gran medida por lo que el texto *ha sido* en el pasado, sin comprometer demasiado lo que el texto puede llegar a ser en el futuro. Intenta tratar todos los tipos de documentos digitales de la misma manera, tanto si han “nacido digitales” como si no.
- 2 En consecuencia, el marco conceptual de TEI proporciona una forma útil de pensar acerca de la naturaleza del texto: constituye una especie de enciclopedia de nociones textuales generalmente admitidas. En esta breve guía intentaremos ejemplificar algunas de estas nociones, para lo cual utilizaremos el vocabulario definido por TEI en sus *Directrices*.
- 3 Actualmente, los documentos TEI en formato digital se expresan mediante un lenguaje de codificación formal muy extendido llamado XML, el *lenguaje de marcado extensible* (*extensible markup language*), publicado por primera vez por el World Wide Web Consortium (W3C) en 1998, cuyos orígenes se remontan a los sistemas de preparación de documentos de la década de 1980. XML ofrece una forma sencilla de representar datos estructurados como un flujo lineal de datos de caracteres, y de marcar partes particulares de ese flujo con *etiquetas* para indicar su función estructural o su semántica. Al haberse convertido XML en una tecnología tan generalizada, existen muchas guías introductorias excelentes, por lo que daremos por sentado que el lector tiene un conocimiento básico de conceptos clave como *q*, *atributo*, *esquema* y *espacio de nombres* (*namespace*), que a continuación se glosarán brevemente en aras de la inteligibilidad. Las *Directrices TEI* incluyen una *Introducción suave a XML*, que puede serle útil a los principiantes, aunque es posible encontrar muchos otros tutoriales sobre el tema en internet.
- 4 El siguiente es un ejemplo de un documento XML mínimo:


```
<?xml version="1.0"?>
<doc xmlns="http://example.org/namespace">
  <p n="1">Este es un párrafo.</p>
  <p n="2">
    Este párrafo menciona a <placeName>Bristol</placeName>.
  </p>
</doc>
```

- 5 La primera línea de un documento XML siempre adopta la forma mostrada anteriormente: un tipo especial de instrucción que indica que lo que sigue es un documento XML conforme a la versión del estándar XML indicada (en este caso, la versión 1.0). Un documento XML consiste en una secuencia de caracteres legibles por humanos, sin códigos especiales adicionales ni datos binarios. Los caracteres < y > se utilizan para marcar el inicio y el final de una *etiqueta (tag)* dentro de esta secuencia. Una etiqueta puede ser de inicio (como <p>) o de fin (como </p>). Una etiqueta siempre comienza con un nombre (<doc>, <p> y <placeName> en el ejemplo anterior) y también puede contener especificaciones de atributo (como n="1"). El propósito de una etiqueta de inicio es marcar el punto en la secuencia de caracteres donde comienza algún *elemento*, de un tipo indicado por el nombre de la etiqueta, y el propósito de una etiqueta de fin es marcar dónde termina ese elemento. El propósito de una especificación de atributo es añadir alguna información extra sobre la aparición de un elemento además de su nombre. En el ejemplo anterior tenemos un elemento llamado <doc> que contiene dos elementos <p>. Los elementos <p> tienen ambos un atributo @n que proporciona un número, y ambos contienen texto sin formato. El segundo elemento <p> también contiene un elemento llamado <placeName>.
- 6 Se dice que un documento XML como este está *bien formado* si respeta la sintaxis ejemplificada aquí, con las etiquetas de inicio y fin presentes y correctamente anidadas. Pero el estándar XML no dice nada en absoluto sobre cómo deben nombrarse los elementos o los atributos (a diferencia, por ejemplo, del HTML, que define un conjunto específico de etiquetas que deben utilizarse de una manera determinada en todos los documentos), y mucho menos sobre lo que significan sus nombres. Podemos suponer que los elementos <p> de arriba marcan párrafos numerados, pero no hay nada en la representación XML que justifique esa suposición: podrían marcar igualmente páginas, o entradas de un glosario, o versos. Por lo tanto, si encuentro otro documento que contenga elementos <p>, ¿cómo sabré si tienen la misma función? La función del atributo @xmlns de arriba es ayudar a resolver este problema, al proporcionar un valor por defecto para lo que se denomina el *espacio de nombres (namespace)* de todos los elementos contenidos por el elemento <doc>.
- 7 No es inusual encontrar elementos pertenecientes a muchos espacios de nombres en un mismo documento: por ejemplo, un documento que contenga notación musical, gráficos vectoriales y texto, todo ello representado en XML, podría utilizar etiquetas de tres espacios de nombres diferentes: uno para los elementos musicales, otro para los gráficos y otro para los textuales. Un espacio de nombres es una forma de etiquetar un grupo de elementos: en nuestro ejemplo, su uso deja claro que los elementos <p> aquí son diferentes de cualquier elemento <p> definido por otro espacio de nombres cualquiera.

- 8 El etiquetado en un documento se introduce a fin de marcarlo y organizarlo para su procesamiento automático. Si los párrafos están claramente marcados, un software de maquetación puede distribuirlos adecuadamente. Si los nombres de lugar están claramente marcados, un programa puede seleccionarlos automáticamente para elaborar un índice geográfico. Pero esto solo puede hacerse de forma fiable si tenemos algún control sobre cómo se introducen las etiquetas en el documento y dónde aparecen. La tecnología XML proporciona este nivel adicional de control mediante lo que se denomina un *esquema* (*schema*): un tipo de léxico y gramática combinados para los documentos XML válidos. Como mencionamos antes, se dice que un documento XML está *bien formado* si respeta las reglas sintácticas del estándar de XML. También se puede decir, alternativamente, que un documento es *válido* si las etiquetas que contiene se ajustan a un esquema.
- 9 Un *esquema* especifica un conjunto de nombres de elementos, los nombres y tipos de datos de los atributos asociados a ellos, y las reglas relativas a los contextos en los que pueden aparecer legalmente. Un esquema para nuestro sencillo ejemplo anterior dirá que existen elementos llamados <doc>, <p>, <placeName>, etc. También puede especificar que los elementos <p> pueden aparecer dentro de los elementos <doc>, que los <placeName> pueden aparecer dentro de los <p>, que el atributo @n debe tener un valor numérico, etc. Nótese, sin embargo, que un esquema XML aún no tiene modo de especificar que la etiqueta <placeName> indica el nombre de un lugar, ni tampoco qué entendemos por “un lugar”: esas restricciones semánticas adicionales deben especificarse en otra parte, por ejemplo mediante documentación como la que proporcionan las *Directrices TEI*.
- 10 TEI dispone de nombres y definiciones para cientos de etiquetas, junto con reglas sobre cómo pueden combinarse. Más exactamente, las *Directrices TEI* definen unos quinientos o seiscientos *conceptos* diferentes, junto con especificaciones detalladas para los elementos XML y las clases de elementos que pueden utilizarse para representarlos. La mayoría de los documentos TEI, si no todos, solo necesitan de una pequeña cantidad de lo que se proporciona. Por lo tanto, es un poco engañoso pensar en TEI como un único esquema monolítico. Para facilitar la interoperabilidad, todos los documentos TEI utilizan componentes extraídos del mismo esquema gigantesco, pero la mayoría de los proyectos TEI utilizan subconjuntos bastante pequeños de este, y un proyecto bien organizado generalmente tendrá su propia documentación personalizada que identifique ese subconjunto.
- 11 Uno puede utilizar software como la aplicación web Roma para elegir entre las especificaciones TEI y generar a partir de ellas un esquema adecuado a las necesidades de su propio proyecto. Alternativamente, uno puede simplemente diseñar un esquema a mano. Trataremos este asunto con más detalle en un capítulo posterior. Las *Directrices TEI*, a las que se puede acceder gratuitamente desde el sitio web <https://tei-c.org/Guidelines/>, constituyen un completo manual de referencia para estos conceptos, que combina las especificaciones técnicas con una discusión detallada de cómo deben utilizarse.
- 12 Hay una gran variedad de herramientas de software dedicadas a crear, transformar y procesar documentos XML en general, o TEI-XML en particular. Este es un tema muy amplio y en rápida evolución que no trataremos aquí, aunque es posible encontrar algunas indicaciones útiles en el sitio web de TEI y en la Wiki de TEI.

Organización estructural de un documento TEI

- 1 Todos los documentos TEI están organizados de una forma similar, independientemente del tipo de original que representen. Introduciremos algunas de las variaciones más comunes de esta estructura a través de algunos ejemplos sencillos.

Encabezado, texto y divisiones

- 2 Para empezar por el principio, todo documento TEI (representado mediante un elemento <TEI>) tiene al menos dos partes: un *encabezado* (representado mediante un elemento <teiHeader>) que contiene los metadatos que describen el documento, y el texto propiamente dicho (habitualmente representado por un elemento <text>). Por ejemplo:

```
<TEI xmlns='http://www.tei-c.org/ns/1.0'>
  <teiHeader>
    <!-- metadatos que describen un texto -->
  </teiHeader>
  <text>
    <!-- la representación del texto mismo -->
  </text>
</TEI>
```

- 3 Como cualquier otro documento TEI-XML, este documento declara explícitamente que los elementos que contiene deben entenderse por defecto como procedentes del espacio de nombres TEI, <http://www.tei-c.org/ns/1.0>.
- 4 Describiremos con más detalle el encabezado en un capítulo posterior; por el momento, señalaremos simplemente que un encabezado mínimo debe contener información que identifique el documento mismo (en el <titleStmt>), información sobre cómo se distribuye o publica (en el <publicationStmt>) y alguna indicación de sus orígenes (en el <sourceDesc>). El elemento <text> se utiliza para contener una versión codificada del texto mismo, en cuya estructura se representan elementos como <front> (para los

prólogos, etc.) <body> (para el cuerpo del texto propiamente dicho) y <back> (para cualquier apéndice, etc.). En el interior de estos componentes, también podemos representar subdivisiones adicionales como volúmenes, partes, capítulos, etc. utilizando el elemento <div>.

- 5 Por ejemplo, este es el comienzo de una versión TEI mínima de una novela famosa, tal como podría ser distribuida por una editorial digital imaginaria:

```
<TEI xmlns="http://www.tei-c.org/ns/1.0">
  <teiHeader>
    <fileDesc>
      <titleStmt>
        <title>La vida y las opiniones de Tristram Shandy,
          caballero: edición TEI</title>
      </titleStmt>
      <publicationStmt>
        <publisher>Web Head Press</publisher>
        <date>2013</date>
      </publicationStmt>
      <sourceDesc>
        <p>Transcrito de la primera edición, 1708</p>
      </sourceDesc>
    </fileDesc>
  </teiHeader>
  <text>
    <body>
      <div type="volumen" xml:id="TS01">
        <div type="capítulo" xml:id="TS0101">
          <head>Cap. I</head>
          <p>Desearía que mi padre o mi madre, o de hecho ambos,
            ya que ambos estaban igualmente obligados a ello,
            se hubieran preocupado de lo que hacían cuando
            me engendraron;...</p>
          <!-- resto del capítulo 1 aquí -->
        </div>
        <div type="capítulo" xml:id="TS0102">
          <head>Cap. II</head>
          <p>—Entonces, positivamente, no hay nada en la cuestión,
            que yo pueda ver, ni bueno ni malo. — Entonces déjeme
            decirle, señor, que fue una pregunta muy poco oportuna,
            al menos...</p>
          <!-- resto del capítulo 2 aquí -->
        </div>
        <!-- capítulos restantes del volumen 1 aquí -->
      </div>
      <!-- volúmenes restantes de la obra aquí -->
    </body>
  </text>
</TEI>
```

- 6 En este ejemplo, el cuerpo de un texto contiene divisiones más pequeñas, que a su vez contienen subdivisiones. Estas subdivisiones textuales tienen nombres diferentes en

distintas culturas y en distintos tipos de documentos; muchas culturas occidentales utilizan nombres como *sección*, *parte*, *libro* o *capítulo* (o sus equivalentes) para estas subpartes del cuerpo, pero a menudo de formas inconsistentes o incompatibles entre sí. Lo que se denomina *parte* dentro de un libro en un texto puede llamarse *sección* dentro de un capítulo en otro, o *libro* dentro de una parte en un tercero. En consecuencia, TEI propone un único elemento `<div>` para cualquier subdivisión estructural del cuerpo de un texto.

- 7 El elemento `<div>` puede acompañarse de una serie de atributos para indicar su función y sus propiedades con mayor exactitud. Aquí hemos utilizado el atributo `@type` para caracterizar o clasificar el contenido del elemento, distinguiendo así los elementos `<div>` que contienen “volúmenes” de los que contienen “capítulos”. También hemos utilizado el atributo `@xml:id` para proporcionar un identificador único a cada división de la novela.
- 8 Este mismo tipo de estructura jerárquica puede utilizarse para cualquier tipo de texto. Por ejemplo, en un poema épico dividido en libros podría etiquetarse cada libro como un `<div type="libro">`; en una obra de teatro dividida en actos y escenas podría etiquetárselos `<div type="acto">` y `<div type="escena">` respectivamente; y así sucesivamente. Los valores disponibles para su uso con los atributos `@type` y `@xml:id` no están definidos por TEI, sino que son elegidos por el codificador. Para `@xml:id` los valores utilizados son códigos arbitrarios que caracterizan de manera específica el elemento en el que aparecen: proporcionan una forma de etiquetar el elemento en cuestión para que otras partes de este documento, o de otros, puedan apuntar a él directamente. Para `@type` los valores son también códigos arbitrarios, elegidos por el codificador para indicar la función del `<div>` en cuestión. En las *Directrices TEI* se sugieren algunos valores habituales, como *chapter* (capítulo) o *volume* (volumen), pero corresponde a los proyectos individuales que utilizan TEI la definición de su propia taxonomía. Sin embargo, es fácil no solo documentar sino imponer cualquier taxonomía elegida mediante una personalización de TEI, como veremos en un próximo capítulo.
- 9 El elemento `<div>` se llama así porque es una *división*; por lo tanto, no debe utilizarse para algo que esté completo en sí mismo. Si el cuerpo de un texto no contiene divisiones, no es necesario proporcionar un único elemento `<div>` para contenerlo. En su lugar, el elemento `<body>` puede contener directamente uno o varios de los elementos que describiremos en la siguiente sección.
- 10 Un texto (es decir, aquello contenido en el `<text>`) es típicamente algo como un libro o un artículo, pero puede ser cualquier cosa que sea conveniente considerar como un objeto textual discreto pero completo, como un solo poema o un documento de archivo, o algo tan breve como una tarjeta postal. Como alternativa al elemento `<text>`, o además de este, TEI también proporciona un elemento `<facsimile>` que podría utilizarse para ofrecer una representación visual complementaria, por ejemplo en forma de una serie de imágenes de páginas digitalizadas. Y TEI también proporciona formas de representar colecciones de este tipo, como vamos a analizar en la sección final del siguiente capítulo.

Las variedades de la estructura textual

- 1 Desde la perspectiva de TEI, los textos y sus divisiones pueden contener una gama bastante limitada de componentes “estructurales”. Estos incluyen cosas como los encabezados o el asunto preliminar al comienzo o al final de una división, y una serie de componentes básicos que son característicos de la prosa, el verso o el drama. La prosa consiste principalmente en párrafos o listas, marcados con los elementos <p> (párrafo) o <list> (lista) respectivamente. El verso consta de líneas de verso, marcadas con el elemento <l> (línea), o de secuencias de líneas de verso marcadas con el elemento <lg> (grupo de líneas). En una obra dramática se proporciona un bloque de construcción adicional en forma del elemento <sp> (discurso), que combina un elemento <speaker> (hablante) con uno o más elementos <p> o <l>, dependiendo de si la obra está en verso o en prosa.

Un artículo de revista académica

- 2 El cuerpo de un artículo académico, por ejemplo, podría codificarse con una estructura como la siguiente:

```

<body xml:lang="es">
  <div type="sección">
    <head>Introducción</head>
    <p>Recomendamos el uso del marcado TEI como
    medio de representar la prosa científica. Presenta una
    serie de ventajas prácticas: </p>
    <list>
      <item>El marcado TEI es fácil de añadir;</item>
      <item>El marcado TEI es ampliamente comprendido;</item>
      <item>El marcado TEI es fácil de convertir a otros formatos.</item>
    </list>
    <p>El marcado TEI también tiene algunas ventajas, que discutimos en la
    sección <ptr target="#SEC3"/>.</p>
    <!-- más párrafos introductorios aquí -->
  </div>
  <div type="sección">
    <head>Orígenes de TEI</head>
    <p>La Iniciativa de Codificación de Textos nació en 1987...</p>
    <!-- muchos más párrafos aquí -->
  </div>
  <div xml:id="SEC3">
    <head>Propiedades científicas del marcado TEI</head>
    <p>El marcado TEI expresa una visión de la naturaleza del texto...</p>
    <!-- muchos más párrafos aquí -->
  </div>
</body>

```

- 3 Debemos tener en cuenta que este marcado solo indica la organización del documento. No dice nada sobre cómo debe visualizarse en pantalla o en papel. Indica que hay secciones que tienen títulos, y que contienen párrafos y listas. Distingue los elementos de la lista, pero no especifica si deben ir precedidos de una viñeta, un guion o un número. El elemento <ptr> (puntero) indica la existencia de una referencia cruzada desde una parte del documento (la ubicación del elemento <ptr>) a otra (el elemento <div> (división) titulado *Propiedades científicas...*). El valor del atributo @target introducido anteriormente especifica el destino de la referencia cruzada, pero no especifica si esta debe realizarse como (por ejemplo) un enlace HTML, o algún texto añadido (como el número de sección) o ambos. Por supuesto, no es difícil imaginar cómo podríamos mostrar estos componentes del documento utilizando un lenguaje de formato como Word, HTML o LaTeX, pero ese no es el objetivo principal de este marcado. Como todos estos aspectos se dejan sin especificar en la codificación, el mismo documento puede reutilizarse en diferentes contextos de procesamiento.

Un texto poético

- 4 Nuestro segundo ejemplo muestra un famoso soneto de Shakespeare, transcrito de una copia impresa específica

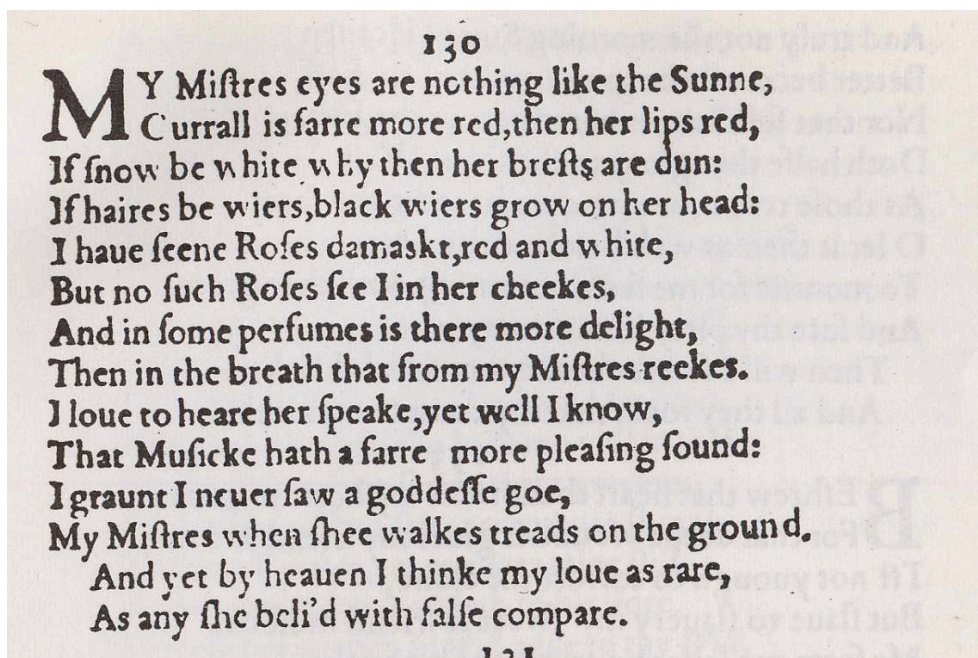
```

<lg type="soneto">
  <head>130</head>
  <l>My Mistres eyes are nothing like the Sunne,</l>
  <l>Currall is farre more red, then her lips red,</l>
  <l>If snow be white why then her brests are dun:</l>
  <l>If haire be wiers, black wiers grow on her head:</l>
  <l>I have seene Roses damaskt, red and white,</l>
  <l>But no such Roses see I in her cheekes,</l>
  <l>And in some perfumes is there more delight,</l>
  <l>Then in the breath that from my Mistres reekes.</l>
  <l>I loue to heare her speake, yet well I know.</l>
  <l>That Musicke hath a farre more pleasing sound:</l>
  <l>I graunt I never saw a goddesse goe,</l>
  <l>My Mistress when she walkes treads on the ground.</l>
  <lg type="estrofa">
    <l>And yet by heaven I thinke my love as rare,</l>
    <l>As any she beli'd with false compare.</l>
  </lg>
</lg>

```

- 5 Una vez más, este marcado captura solo la estructura del soneto: indicando los versos individuales de los que se compone, y marcando explícitamente el dístico al final. Dado que las líneas individuales se distinguen en la codificación, en lugar de ser tratadas como si fueran párrafos o accidentes del formato, se puede generar automáticamente un análisis métrico del poema.
- 6 Al compararlo con la siguiente imagen digital de la fuente original podemos ver que, aunque se ha conservado la ortografía original, esta codificación ha optado por ignorar rasgos de diseño como el uso de mayúsculas o la sangría asociada al dístico. Ha conservado la ortografía original, pero ha modernizado silenciosamente algunas de las variaciones tipográficas, como la S larga o las ligaduras.

Figura 1

Soneto 130, en *Shake-speares Sonnets. Never before Imprinted* (1609)

- 7 Veremos más adelante cómo se podría mejorar esta codificación para producir una versión de lectura moderna aceptable, así como la versión cuasi-diplomática mostrada aquí.

Un texto teatral

- 8 Nuestro tercer ejemplo muestra cómo podríamos codificar la estructura de un texto dramático: en este caso, el final de *Esperando a Godot* de Beckett.

```

<div type="scene">
  <!-- ... -->
  <sp>
    <speaker>Vladimir</speaker>
    <p>Ponte los pantalones.</p>
  </sp>
  <sp>
    <speaker>Estragón</speaker>
    <p>¿Quieres que me quite los pantalones?</p>
  </sp>
  <sp>
    <speaker>Vladimir</speaker>
    <p>Tira <emph>de</emph> tus pantalones.</p>
  </sp>
  <sp>
    <speaker>Estragón</speaker>
    <p>
      <stage>(Se da cuenta de que tiene los pantalones abajo)</stage>.
      Verdad
    </p>
  </sp>
  <stage>Se sube los pantalones</stage>
  <sp>
    <speaker>Vladimir</speaker>
    <p>¿Bien? ¿Nos vamos?</p>
  </sp>
  <sp>
    <speaker>Estragón</speaker>
    <p>Sí, vamos.</p>
  </sp>
  <stage>No se mueven.</stage>
</div>

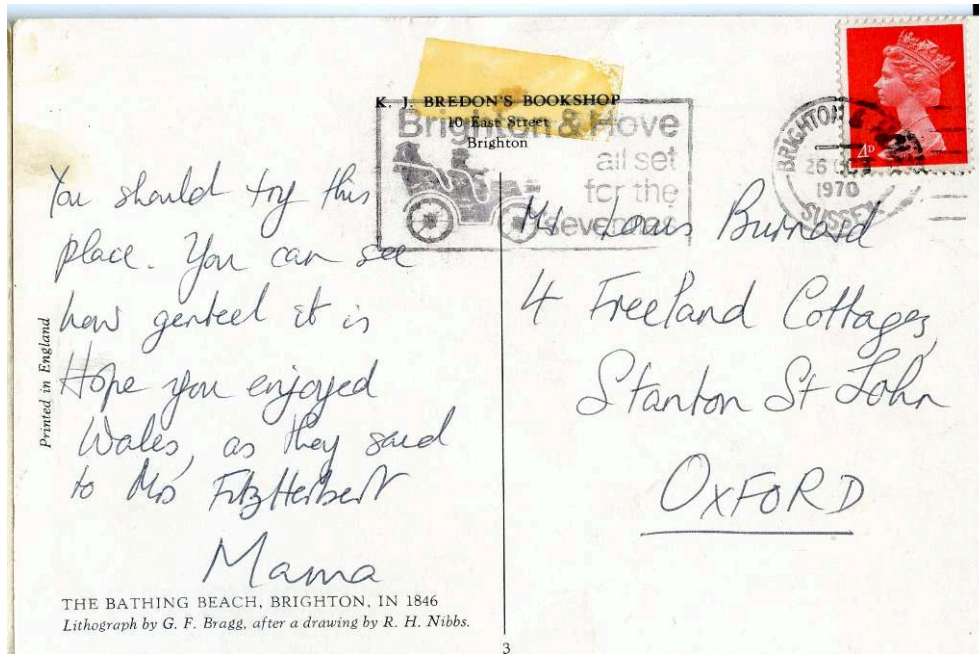
```

- 9 En esta codificación, hemos introducido algunos elementos TEI adicionales, para poder distinguir las acotaciones o indicaciones escénicas (<stage>, *stage directions*) de los discursos individuales de los personajes (<sp>, discurso). Nótese que una indicación escénica puede aparecer tanto dentro como entre los discursos; asimismo, un discurso o entrada de diálogo siempre contiene tanto una etiqueta de identificación para mostrar quién habla (<speaker>, hablante) como un párrafo (<p>) para contener lo que dice. En una obra teatral en verso, por supuesto, los discursos contendrían probablemente una secuencia de elementos <l>.

Una tarjeta postal

- 10 TEI se utiliza ampliamente para obras literarias existentes o publicadas formalmente. Sin embargo, también puede utilizarse para documentos de tipos totalmente diferentes, como manuscritos de autor, documentos de archivo o cualquier otro tipo de escritura informal. Nuestro cuarto ejemplo muestra cómo podríamos utilizar TEI para codificar la estructura de esta postal:

Figura 2



Una tarjeta postal (verso)

- 11 Todas las postales tienen dos caras: un anverso y un reverso; las representamos como elementos <div>. En la cara mostrada arriba, podemos ver que hay una clara distinción entre la parte que lleva el mensaje, a la izquierda, y la parte relativa al envío de la tarjeta, que contiene varios sellos y una dirección. A continuación se muestra una posible codificación, que utiliza los mismos elementos <div> y <p> que ya hemos visto, junto con algunos elementos y atributos adicionales más especializados:

```

<div type="verso">
  <div type="mensaje">
    <p>Deberías probar visitar este sitio. Ya ves
    lo elegante que es</p>
    <p>Espero que hayas disfrutado de Gales,
    como le dijeron a la señora Fitzherbert</p>
    <signed>Mamá</signed>
  </div>
  <div type="destino">
    <ab>
      <stamp type="publicidad">Silueta de automóvil antiguo y eslogan
      "Brighton & Hove todo listo para los setenta."
      </stamp>
      <stamp type="matasellos">
        Marca circular que especifica "Brighton & Hove-Sussex" y
        <date when="1970-10-26">26 oct. 1970</date>
      </stamp>
      <stamp type="estampilla">Diseño de Machin. 4d, bermellón.</stamp>
    </ab>
    <ab>
      <address>
        <addrLine>Mr Louis Burnard</addrLine>
        <addrLine>4 Freeland Cottages</addrLine>
        <addrLine>Stanton St John</addrLine>
        <addrLine>OXFORD</addrLine>
      </address>
    </ab>
  </div>
</div>

```

- 12 El elemento <signed> (firmado) utilizado para encerrar el bloque de firma es uno de los varios elementos especializados proporcionados por TEI para codificar firmas o encabezados que aparecen al principio o al final de un <div>. Parece útil diferenciarlo del resto del mensaje, ya que otro tipo de frases de firma o *fórmulas de cortesía* contienen un tipo de lenguaje diferente del resto, más formal.
- 13 En la otra división, hemos distinguido tres tipos de sellos diferentes: el sello postal propiamente dicho, el matasellos que indica cuándo y dónde se envió la tarjeta, y un sello publicitario adicional que promociona el Brighton and Hove Vintage Car Rally. El elemento TEI <stamp> (sello) está previsto para contener un texto descriptivo sobre cualquier tipo de sello; cuando nuestra descripción incluya palabras que aparezcan realmente como parte del sello, utilizaremos el elemento TEI <mentioned> (mencionado). En el caso del matasellos, también hemos distinguido la fecha indicada. Una gran cantidad de nuestras búsquedas efectuadas sobre un archivo de este tipo de documentos codificados querrá localizar y agrupar las tarjetas de forma fiable según la fecha en la que fueron enviadas. Las fechas indicadas en los matasellos suelen utilizar formatos diferentes y pueden estar incompletas: por ello, añadimos a la fecha codificada un valor normalizado, suministrado en el atributo @when (cuándo). Por último, hemos indicado la dirección a la que se envió la tarjeta, dividiéndola sencillamente en líneas.

- 14 En una codificación más detallada probablemente distinguiríamos los nombres de las personas y de los lugares, añadiendo quizá información de coordenadas geográficas para los nombres de lugares mencionados. Tal vez querríamos añadir alguna explicación de la bromita sobre Gales. Probablemente querríamos distinguir las partes formales del mensaje, como sus líneas de apertura y cierre, del contenido. Y también podríamos considerar la mejor manera de registrar la información textual impresa en el reverso de la tarjeta: el título y el nombre de su editor, por ejemplo. El alcance de estos metadatos, que proporcionan información útil sobre cómo se produjo y utilizó la tarjeta, es muy amplio y a veces es difícil saber dónde detenerse. Por ejemplo, la mancha amarilla visible en la imagen sugiere que esta tarjeta estuvo unida a algo mediante una cinta adhesiva barata. Si se tratara de un artefacto histórico muy raro, estas pruebas de procedencia podrían tener una importancia considerable y tendrían que codificarse utilizando (por ejemplo) el elemento <damage> (daño o desperfecto).

Un texto mínimamente estructurado

- 15 Finalmente, si todo esto parece demasiado, debemos tener en cuenta que TEI también puede admitir una visión estructural muy simple en la que todos estos componentes convencionales, con carga semántica, como los párrafos, los locutores, los versos, etc., se eluden o se ignoran en favor de una segmentación neutra basada en oraciones definidas por la ortografía. Desde este punto de vista, podríamos decidir mantener las divisiones principales del texto, para a continuación dividir las en cada signo de puntuación apropiado en segmentos lingüísticamente útiles, empleando los elementos genéricos <s> (segmento) y <ab> (bloque anónimo). Para repetir nuestro primer ejemplo:

```
<body xml:lang="es">
  <div type="sección">
    <ab>
      <s n="1">Introducción</s>
      <s n="2">Recomendamos el uso del marcado TEI como medio de
        representación de prosa científica.</s>
      <s n="3">Tiene una serie de ventajas prácticas:</s>
      <s n="4">el marcado TEI es fácil de añadir;</s>
      <s n="5">el marcado TEI es ampliamente comprendido;</s>
      <s n="6">el marcado TEI es fácil de convertir a otros formatos, y</s>
      <s n="7">el marcado TEI también tiene algunas ventajas que discutimos
        en la sección <ptr target="#SEC3"/>.</s>
    </ab>
  </div>
</body>
```

- 16 Aunque este tipo de marcado carece, obviamente, de la información presente en la versión más rica que hemos presentado antes y, por lo tanto, es menos útil en general, su simplicidad y regularidad significa que es mucho más fácil de procesar para tareas comparativamente mecánicas como el análisis del vocabulario o el enriquecimiento lingüístico. Como siempre, para sacar lo mejor de TEI es necesario definir cuidadosamente nuestras prioridades antes de tomar decisiones sobre la codificación.

Textos complejos

- 17 Un texto TEI puede ser *simple* (por ejemplo, un solo libro) o *complejo* (por ejemplo, una colección o antología). En cualquiera de los dos casos, el cuerpo principal del texto puede ir precedido o seguido de material adicional diferente (portadas, prólogos, dedicatorias, índices, etc.). TEI propone un elemento obligatorio <body> (cuerpo) para la parte principal de un texto, y elementos opcionales <front> (parte del adelante) y <back> (parte de atrás) para agrupar el material adicional que precede o sigue al cuerpo, respectivamente. En un texto complejo, el "cuerpo" del texto está representado por un elemento <group> (grupo) que puede contener múltiples elementos <text> (texto), o grupos de ellos, como en el siguiente esquema:

```

<TEI xmlns="http://www.tei-c.org/ns/1.0">
  <teiHeader>
    <!-- encabezado de un texto complejo ]-->
  </teiHeader>
  <text>
    <front>
      <!-- texto preliminar del texto complejo -->
    </front>
    <group>
      <text>
        <front>
          <!-- texto preliminar para el primer texto del complejo -->
        </front>
        <body>
          <!-- cuerpo del primer texto del complejo -->
        </body>
        <back>
          <!-- paratexto final para el primer texto del complejo -->
        </back>
      </text>
      <text>
        <front>
          <!-- texto preliminar para el segundo texto del complejo -->
        </front>
        <body>
          <!-- cuerpo del segundo texto del complejo -->
        </body>
        <back>
          <!-- paratexto final para el segundo texto del complejo -->
        </back>
      </text>
      <!-- más textos, simples o complejos -->
    </group>
    <back>
      <!-- paratexto final del texto complejo -->
    </back>
  </text>
</TEI>

```

- 18 Una estructura como esta sería apropiada para codificar una antología previa de poemas de diferentes autores reunidos por un editor.
- 19 En este modelo solo hay un encabezado TEI para todo el texto complejo. TEI también permite un texto complejo que representa una colección de documentos originalmente distintos pero reunidos por el codificador con algún propósito: los corpus lingüísticos son ejemplos típicos. Cada documento constitutivo de un corpus lingüístico tiene sus propios metadatos, representados por su propio encabezado <teiHeader>, pero hay una capa adicional de metadatos relativos al corpus en su conjunto. Esto se codifica en TEI utilizando una estructura como la siguiente:

```
<teiCorpus xmlns="http://www.tei-c.org/ns/1.0">
  <teiHeader>
    <!-- metadatos relacionados con todo el corpus -->
  </teiHeader>
  <TEI>
    <teiHeader>
      <!-- metadatos relacionados con el primer texto del corpus -->
    </teiHeader>
    <text>
      <!-- primer texto del corpus -->
    </text>
  </TEI>
  <TEI>
    <teiHeader>
      <!-- metadatos relacionados con el segundo texto del corpus -->
    </teiHeader>
    <text>
      <!-- segundo texto del corpus -->
    </text>
  </TEI>
</teiCorpus>
```

- 20 Una estructura como esta sería apropiada para codificar una colección de postales recién construida u otros objetos textuales independientes.

Cornucopia TEI, primera parte

- 1 En este capítulo y en el siguiente discutiremos y ejemplificaremos algunos de los elementos definidos por el estándar completo de TEI. Los elementos que se discuten en el presente capítulo pueden ser útiles en casi cualquier tipo de proyecto de codificación; los que se discuten en el siguiente son más especializados. Conviene destacar que todo lo que pretendemos aquí es dar una idea de la rica variedad de elementos que ofrece TEI; para obtener una información completa siempre se deben consultar las *Directrices*.

Hitos

- 2 Ya hemos señalado que TEI suele utilizarse la mayoría de veces para representar la organización de un libro en componentes como capítulos, párrafos, etc. o de un poema en versos y estrofas. Pero los libros y poemas impresos también existen como objetos físicos, que suelen estar formados por páginas con superficies de *verso* (esto es, la que se abre a la izquierda) y *recto* (la que se abre a la derecha). ¿Por qué entonces TEI no ofrece un elemento `<page>` (página) para representar cada página de un libro? Seguro que es al menos tan importante saber en qué página comienza una frase como saber que se encuentra en el quinto párrafo del segundo capítulo de la novena subsección de una obra, ¿cierto? Seguro que nos gustaría poder mostrar el texto paginado en pantalla de la misma manera que estaba paginado en la fuente original, ¿cierto?
- 3 La respuesta a estas preguntas perfectamente razonables es sorprendentemente compleja. De hecho, TEI proporciona un elemento para marcar los límites de las páginas (`<pb>` por *page beginning*, inicio de página), pero es un elemento vacío que debe colocarse al principio del texto transcrito de cada página:

```

<body xml:lang="es">
  <pb n="42"/>
  <p>
    Este párrafo empieza en la página 42...
    <!-- aquí hay más texto-->
    <pb n="43"/>
    <!-- aquí hay aun más texto -->
    ...y termina a mitad de la página 43.
  </p>
  <p>Este párrafo más corto empieza y termina en la página 43.</p>
</body>

```

- 4 Un estudio atento del ejemplo anterior puede ayudar a explicar por qué no es posible un elemento `<page>` que encierre realmente todo el texto de una página (en lugar de un `<pb>` vacío que simplemente señale el comienzo de una página). En XML y en lenguajes de marcado similares, es un requisito absoluto que los elementos definidos por las etiquetas de marcado estén correctamente anidados unos dentro de otros. Si iniciamos un elemento `<page>` y dentro de él iniciamos un elemento `<p>`, la sintaxis del lenguaje exige que cerremos ese `<p>` antes de poder cerrar el `<page>`, aunque es un hecho evidente que los párrafos se extienden con frecuencia a través de los límites de la página.
- 5 Los usuarios de XML (y antes del SGML) llevan décadas intentando encontrar mejores formas de representar la estructura textual sin caer en esta restricción, y no pretendemos debatir aquí el “problema del solapamiento” con más detalle. Nos limitaremos a señalar que una forma bien establecida en TEI de abordar este problema adopta la forma de las etiquetas de “hito”, de las que `<pb>` es probablemente el ejemplo más frecuente. Mientras que las etiquetas normales de un documento XML marcan explícitamente tanto el inicio como el final de un elemento y, por lo tanto, siempre encierran algo, una etiqueta de hito marca simplemente un punto en el que algo cambia, pero no encierra nada. El hito número 42 en la carretera de Londres a Bath marca el punto en el que comienza el tramo cuarenta y dos de esa carretera. Podemos suponer que si seguimos avanzando llegaremos al hito número 43, que marcará el punto en el que comienza el tramo cuarenta y tres, y por lo tanto (ya que un tramo de carretera no puede estar en dos secciones a la vez) el punto en el que termina el tramo 42, aunque no esté marcado explícitamente.
- 6 Los elementos `<lb>` (*line beginning*, inicio de línea), `<cb>` (*column beginning*, inicio de columna) y `<gb>` (*gathering beginning*, inicio de cuadernillo) son todos hitos en este sentido. Además, TEI proporciona el elemento genérico `<milestone>` (hito), que puede utilizarse para marcar cualquier tipo de unidad no prevista de otro modo. A modo de ejemplo, consideremos la práctica (común en el siglo XIX) de publicar novelas en forma de serie. Es posible que queramos marcar el límite de los números de las partes individuales, aunque estos no encajen necesariamente bien en la organización interna de la novela como una serie de capítulos. (Dickens, por ejemplo, no tenía reparos en cortar un capítulo en un momento de máximo suspenso.) Se puede utilizar un `<milestone unit="serialPart" n="12"/>` para representar el punto en el que comienza el episodio número 12, independientemente de la estructura de los elementos `<div>` y `<p>` utilizados para el texto mismo.

- 7 Este elemento genérico puede utilizarse para cualquier tipo de desplazamiento, por supuesto, no solo para unidades estructurales. Un análisis narratológico podría representarse utilizando elementos `<milestone>` para marcar puntos en los que cambia la voz narrativa; un análisis estilístico podría utilizarlos para marcar puntos de variación estilística. También hay otros elementos específicos que se comportan de forma parecida a los hitos, por ejemplo, para indicar cambios de calidad de voz en el discurso transcrito, o cambios de mano en el manuscrito transcrito.

Idiomas y sistemas de escritura

- 8 Salvo en el caso de los hitos vacíos, cada elemento de un documento TEI puede contener uno de los siguientes elementos: texto sin formato, “contenido del elemento” o (el caso más frecuente, con diferencia) “contenido mixto”. El texto sin formato consiste en caracteres codificados representados mediante el estándar Unicode; el contenido de elementos consiste en otros elementos XML; y el contenido mixto combina los dos. Los elementos y atributos TEI se utilizan para asociar propiedades semánticas y de otro tipo al contenido, de modo que un procesador pueda tratarlo correctamente en diversas situaciones. Una propiedad especialmente importante es el idioma humano en el que se expresa el contenido textual; otra es la forma en que se muestra o formatea el contenido textual. Estas propiedades son potencialmente importantes en casi todos los niveles: podemos querer decir que todo un documento utiliza un idioma específico, o que solo unas pocas palabras aquí y allá lo hacen; podemos querer decir que todo un documento fue tipografiado en un tipo de letra redonda de un tamaño determinado, y luego indicar aquellas partes suyas que utilizan algún otro tipo de letra o tamaño. Dado que esta información es en principio aplicable a todos los elementos, TEI propone *atributos globales* para especificarla.

Idiomas

- 9 El atributo global `@xml:lang` es el medio recomendado para especificar el idioma humano en el que se expresa el contenido de un elemento. (Como sugiere el prefijo `xml:`, este atributo es común a todos los documentos XML. Por lo tanto, no está definido por TEI sino por el W3C, como parte de la definición de XML.) Su valor se aplica jerárquicamente a todos los elementos contenidos por ese elemento, a menos que se anule:

```
<div xml:lang="la">
  <s>Pars haec lingua Latine composita est.</s>
  <s xml:lang="es">Salvo que esta oración que está en español.</s>
  <s>Vita brevis, ars longa.</s>
</div>
```

- 10 Aquí especificamos que todo el elemento `<div>` utiliza el idioma con el identificador codificado `la`, es decir, el latín (los códigos utilizados están tomados de la norma ISO 639 para los identificadores de idioma). Dado que está contenido por ese `<div>`, no es necesario volver a suministrar esta información para el primer elemento `<s>`. Sin embargo, el segundo elemento `<s>` anula este valor e indica que su contenido está en

español (la lengua con identificador es). El tercer elemento <s> está de nuevo en latín. Para encontrar el idioma del contenido de cualquier elemento, un procesador debe buscar primero si ese elemento, o cualquiera de sus ancestros, proporciona un valor para @xml:lang. Como es de suponer, el ancestro más cercano es el que se aplica, si hay más de uno. Supongamos, por ejemplo, que la segunda oración de arriba contiene también una frase en francés. Habrá que etiquetar dicha frase con algún elemento, por supuesto: el elemento <foreign> (extranjero) se proporciona para casos como este, en los que simplemente queremos mostrar que alguna secuencia de contenido utiliza un idioma diferente:

```
<div xml:lang="la">
  <s>Pars haec lingua Latine composita est.</s>
  <s xml:lang="es">
    Salvo <foreign xml:lang="fr">sacrebleu</foreign>
    esta oración que está en español.
  </s>
  <s>Vita brevis, ars longa.</s>
</div>
```

- 11 Los códigos utilizados para identificar el idioma pueden ser (como en este caso) códigos estándar de dos letras; se pueden utilizar códigos más complejos para distinguir cosas como variantes definidas geográfica o socialmente, o para distinguir el mismo idioma escrito con diferentes alfabetos. Estos códigos, y las formas de utilizarlos, están definidos por la ISO y no por TEI, por lo que no los discutiremos más.

Caracteres no estándar

- 12 Ya indicamos anteriormente que un documento TEI es un documento XML y, por lo tanto, se expresa con caracteres Unicode. El estándar Unicode, en el momento de escribir este artículo, cubre una gran proporción de los sistemas de escritura existentes y antiguos, pero, como es lógico, no intenta estandarizar los sistemas de escritura que aún están siendo estudiados o definidos, ni pretende proporcionar códigos para todas las posibles variantes de caracteres o glifos que puedan encontrarse en el estudio de los documentos antiguos. Por estas razones, TEI proporciona el elemento <g> (glifo o carácter), que puede utilizarse para marcar la presencia en un documento de algún carácter o glifo para el que no existe un punto de código Unicode estándar.
- 13 Supongamos, por ejemplo, que estamos estudiando una serie de manuscritos caracterizados por una forma variante particularmente notable de alguna letra (digamos una R); nuestras transcripciones TEI pueden distinguir este glifo variante de la forma habitual de la misma letra utilizando un elemento <g> para marcarlo. El contenido del <g> puede ser simplemente una forma normalizada, o el elemento puede estar vacío; en cualquier caso, llevará un atributo @ref que apunta a un elemento <glyph> (descripción de glifo) en otro lugar del documento, en el que definimos la naturaleza, la función, la apariencia, el nombre, etc. de esta variación. Así, suponiendo que hemos identificado y descrito tres formas variantes de la letra R, dándoles los códigos R1, R2, R3, podríamos representar los casos individuales de estas formas como <g ref="#R1"/>, <g ref="#R2"/> o <g ref="#R3"/>. Las definiciones asociadas pueden

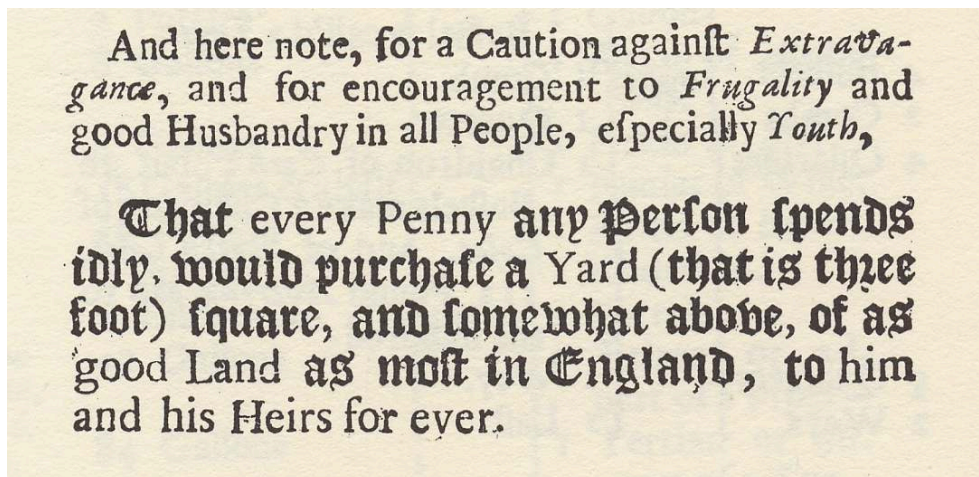
proporcionarse dentro del elemento <charsDecl> (declaración de caracteres) del encabezado TEI, del que hablaremos más adelante.

Representación

- 14 El término *representación* (o *renderización*, *rendition*) se utiliza en el contexto de TEI para describir la forma como se formatea o presenta un elemento en la pantalla o en la página. TEI **no** es un sistema diseñado para darle formato visual a los documentos; como indicamos anteriormente, en general pretende facilitar una codificación fiel al significado percibido de un texto, más que una que reproduzca su aspecto o su apariencia esperada. Este es uno de los muchos aspectos que distinguen a TEI de sistemas como los utilizados habitualmente para el procesamiento de textos, como ya hemos señalado. No obstante, si admitimos que puede ser importante conservar la información sobre la representación original de un documento con fines analíticos (entre otras cosas, porque puede ser que su apariencia sea todo lo que podemos describir de forma fiable en un documento antiguo), TEI proporciona una serie de mecanismos para codificar dichos datos. Volvemos a insistir en que la función de este tipo de marcado de representación es siempre describir el aspecto o la representación de uno o más documentos fuente originales; por supuesto, es probable que al mostrar un texto codificado de esta manera deseemos conseguir un aspecto similar utilizando una tecnología diferente, pero esto no es en absoluto siempre ni necesariamente el caso.
- 15 El método más sencillo es utilizar el atributo global @rend. Este atributo se comporta de forma similar al atributo @xml:lang comentado anteriormente: proporciona un valor codificado para la variante de representación asociada a un elemento y (a menos que se anule) a sus hijos. Sin embargo, a diferencia de @xml:lang, los valores disponibles para su uso con el atributo @rend no están formalmente definidos, ni por TEI ni por ningún otro organismo, aunque se utiliza convencionalmente un número comparativamente pequeño de códigos. El codificador es libre de establecer su propio conjunto de etiquetas y de aplicarlas con la consistencia (o inconsistencia) que desee.
- 16 El atributo @rendition, por el contrario, se comporta más bien como el atributo @ref del elemento <g>, en el sentido de que hace referencia a una definición (o conjunto de definiciones) proporcionada en otro lugar. El codificador se ve así obligado no solo a utilizar únicamente un conjunto especificado de etiquetas de representación, sino también a proporcionar (o remitir a) algún tipo de definición de los significados previstos para cada una de esas etiquetas.
- 17 Por supuesto, estas *definiciones de estilo* se expresan más fácilmente utilizando un lenguaje de codificación diseñado para el software de formateo, como un navegador web o un sistema de procesamiento de documentos. TEI permite al usuario expresar sus definiciones en cualquier lenguaje de este tipo que elija; en la actualidad, una recomendación bastante consolidada del W3C, llamada CSS (*Cascading Stylesheets*, hojas de estilo en cascada), es la más práctica, ya que es a la vez potente y está muy ampliamente implementada. La sintaxis de este lenguaje es muy sencilla e incluso puede utilizarse directamente dentro de un documento TEI, ya sea para proporcionar una interpretación por defecto para un elemento, o para anular ese valor por defecto utilizando otro atributo global: @style.

- 18 Mostraremos ahora cómo se podría utilizar cada uno de estos métodos para codificar el pasaje breve siguiente, tomado de un libro escolar de principios del siglo XVII.

Figura 3



Fragmento de *The Tutor to True English* (1687) de Henry Care, p. 70

- 19 En este pasaje podemos observar secuencias de palabras en un tipo de letra gótica, en un tipo de letra redonda normal, y en cursiva. Podemos especular sobre las razones de esta variación tipográfica, pero suponiendo que —en principio al menos— solo queremos registrarla, podríamos adoptar una codificación como la siguiente:

```
<p rend="redondas">
  <lb/>And here note, for a Caution against
  <hi rend="cursivas">Extrava-<lb break="no"/>gance</hi>, and
  for encouragement to <hi rend="cursivas">Frugality</hi> and
  <lb/>good Husbandry in all People, especially
  <hi rend="cursivas">Youth</hi>,
</p>
<p rend="góticas">
  <lb/>That <hi rend="redondas">every Penny</hi> any Person spends
  <lb/>idly, would purchase a <hi rend="redondas">Yard</hi>
  (that is three
  <lb/>foot) square, and somewhat above, of as
  <lb/><hi rend="redondas">good Land</hi> as most
  in England, to <hi rend="redondas">him
  <lb/>and his Heirs for ever.
  </hi>
</p>
```

- 20 Esta codificación utiliza el elemento hito `<lb>` para marcar explícitamente el inicio de las líneas tipográficas. También utiliza un nuevo elemento `<hi>` (resaltado o destacado). Este elemento no tiene ningún significado especial: simplemente indica que su contenido es visualmente distinto de lo que le rodea, del mismo modo que `<foreign>` simplemente indica que su contenido es lingüísticamente distinto de lo que le rodea. La naturaleza de esta distinción visual no está totalmente especificada, ya que esta codificación no explica qué significan los valores `redondas`, `cursivas` o `góticas`;

tampoco hay forma de proporcionar tal explicación salvo por medio de una nota documental. Por supuesto, un lector humano sabrá más o menos qué significa la cursiva, pero un programa formateador probablemente no. Supongamos, en cambio, que hemos proporcionado definiciones (expresadas en CSS) en el lugar del <teiHeader> asignado para ello:

```
<rendition xml:id="cu" scheme="css">
  font-family: roman; font-style: italic
</rendition>
<rendition xml:id="re" scheme="css">
  font-family: roman; font-style: normal
</rendition>
<rendition xml:id="go" scheme="css">
  font-family: unifaktur cursive
</rendition>
```

- 21 Establecidas dichas definiciones, podríamos ahora simplificar un poco nuestro ejemplo:

```
<p rendition="#go">
  That <hi rendition="#re">every Penny</hi> any Person spends idly...
</p>
```

- 22 Cualquier programa de renderización puede ahora procesar directamente el CSS y representar nuestro texto de un modo visualmente distintivo.
- 23 Además, dado que las expresiones CSS son de hecho muy simples, podríamos tan solo usar el atributo @style para presentarlas directamente en nuestro texto:

```
<p style="font-family: roman">
  And here note, for a Caution against
  <hi style="font-style: italic">Extravagance</hi>,
  and for encouragement to
  <hi style="font-style: italic">Frugality</hi>
  ...
</p>
```

- 24 Como indicamos anteriormente, el elemento <hi> no tiene otro significado que el de indicar que una cierta parte del documento que se está codificando tiene un aspecto significativamente diferente. En proyectos de codificación a gran escala, identificar tal variación puede ser lo único viable; en proyectos que trabajan con recursos muy antiguos, visualmente complejos o escasamente comprensibles, también puede ser muy difícil ir más allá de decir que *esta parte del texto se distingue visualmente de alguna manera*. Pero en la gran mayoría de los proyectos de codificación TEI se suele considerar deseable y viable añadir valor a una codificación deteniéndose a considerar **por qué** algo es visualmente distinto en la fuente. Cuando es posible llegar a responder esa pregunta con cierto grado de confianza, una codificación que nos muestre lo que motivó la decisión de representación original tiene más probabilidades de ser útil en general que otra que no lo haga.

- 25 No es difícil enumerar las principales razones por las que un impresor puede optar por resaltar algunas palabras en un texto impreso, según las convenciones elaboradas en Europa occidental durante los últimos tres siglos aproximadamente. Las palabras en un idioma extranjero casi siempre se resaltan, al igual que las palabras en las que el autor pone algún grado de énfasis lingüístico. Los títulos y los términos técnicos se identifican convencionalmente resaltándolos cuando aparecen en prosa continua. Hasta hace poco, las palabras que indican los nombres de personas, lugares o abstracciones también se destacaban convencionalmente.
- 26 Notemos que aquí utilizamos el término *resaltado* para incluir cualquier forma de resaltado visual, ya sea un pasaje en cursiva incrustado en uno predominantemente en letras redondas, o al revés, o (como en el caso anterior) un pasaje en redondas incrustado en uno predominantemente en góticas. Si ampliamos el significado del término para incluir pasajes delimitados por comillas de varios tipos, la lista de motivos para el resaltado se amplía hasta incluir pasajes de discurso directo, material citado o que se cita de otra parte, o que el escritor desea indicar que en algún sentido no es de su autoría, palabras que se mencionan en lugar de ser usadas..., y así sucesivamente.
- 27 TEI tiene elementos que le permiten al codificador realizar todas estas y otras distinciones semánticas dentro de las palabras simples de un texto, guiado por las pistas visuales implícitas en la interpretación original de un documento, y la comprensión que el codificador tiene de ella. Esta es a menudo una forma útil de distinguir un pasaje de texto que de otro modo sería ambiguo. En un artículo de una revista académica, por ejemplo, un pasaje en cursiva puede corresponder a una expresión extranjera, al título de otro artículo o de una película o canción, a un término técnico o a una frase que el autor desea destacar. Al utilizar la etiqueta adecuada — <foreign>, <title>, <term> o <emph> (énfasis) respectivamente—, el codificador enriquece el texto y facilita un procesamiento más inteligente de este, por ejemplo para enumerar los títulos de las obras mencionadas en un documento, excluyendo las palabras enfatizadas o extranjeras.
- 28 Por supuesto, estas etiquetas pueden utilizarse incluso cuando no se ha aplicado el resaltado. El atributo @rend también está disponible en todos estos elementos, para que el codificador pueda decir tanto lo que cree que indica un pasaje resaltado como en qué aspectos (si los hay) es visualmente destacable.

Nombres y fechas

- 29 Ya hemos mencionado la convención de resaltar los nombres de personas, lugares, etc. en el texto impreso. Esta práctica, que fue habitual hasta mediados del siglo XIX, es testimonio de un antiguo deseo de los lectores de distinguir claramente los nombres del resto de las palabras, y sigue vigente en los codificadores digitales de hoy. Identificar lo que ahora se suele denominar *entidades nombradas* del resto de un mar de palabras es una tarea tan importante para el lingüista computacional de hoy que construye un sistema de comprensión del lenguaje, o para el historiador digital de hoy que rastrea las redes sociales dentro de un corpus de correspondencia del siglo XVIII, como lo es para los organismos de inteligencia de hoy que tratan de rastrear las referencias a los sospechosos de terrorismo en vastos conjuntos de datos de vigilancia.

- 30 TEI proporciona varios elementos para codificar estas cosas una vez que se han identificado. Probablemente, el más general es <rs> (por *referring string*, cadena de referencia), que puede utilizarse para etiquetar cualquier secuencia de palabras que según el codificador se refiera a alguna entidad, como una persona o un lugar, aunque la referencia no adopte la forma de un nombre, sino que sea más bien una referencia pronominal (*ella, el rey*) o una frase deíctica (*esa chica, el tipo que vimos el jueves*), o un modismo (*el gran wen, la gran manzana*). Sin embargo, cuando una referencia de este tipo contiene solo uno o varios nombres propios, debe codificarse utilizando el elemento genérico <name> (nombre).
- 31 Tanto los elementos <rs> como <name> tienen un atributo @type que puede utilizarse para distinguir el tipo de entidad que se nombra, de modo que, por ejemplo, el lugar <name type="lugar">Londres</name> se puede distinguir del autor <name type="persona">Londres</name>. Sin embargo, TEI también proporciona elementos especializados que permiten hacer la misma distinción, y otras, con mayor precisión. Estos elementos más especializados distinguen los nombres relativos a diferentes tipos de entidad de forma explícita: así, proporciona <persName>, <placeName>, <orgName>, para apoyar la afirmación de que el nombre se refiere a una persona, lugar u organización. También es posible identificar los subcomponentes significativos de dichos nombres: TEI proporciona <surname> (para un apellido), <foreName> (para el primer nombre), <altName> (para un nombre alternativo como un apodo), <roleName> (para un papel o título socialmente definido como *Señora, Escudero, Pastor*, etc.), <genName> (para un nombre generacional como *Junior, el venerable*, etc.) y otros. Para los topónimos, TEI distingue componentes que sitúan el nombre de forma sociopolítica como <settlement> (asentamiento), <region> (región), <country> (país) y también componentes como <geogName> para cualquier nombre puramente geográfico como el *monte Sinaí* o el *río Támesis*, que también puede descomponerse en un nombre (*Támesis, Sinaí*) y el nombre de algún tipo de accidente geográfico (*monte, río*).
- 32 Estos elementos son especialmente útiles cuando se trata de documentos de archivo de cualquier tipo, en los que la indexación de los nombres de personas y lugares a los que se hace referencia en un documento es una de las principales preocupaciones, pero pueden ser útiles en cualquier tipo de edición digital o de estudio léxico.
- 33 Supongamos, por ejemplo, que estamos preparando una edición digital de un documento primario como el *Current Record of Events from 1792 to 1885, As recorded by The Shaker community of Canterbury, New Hampshire* (un manuscrito conservado en las Colecciones Especiales de la Biblioteca del Hamilton College). Este documento del siglo XIX comienza con la siguiente oración: *El primero de febrero de 1792 el padre Job Bishop y el venerable Edmond Lougee llegaron desde New Lebanon N.Y. para organizar y establecer una comunidad de creyentes en Canterbury, N.H.*
- 34 Podríamos contentarnos con registrar simplemente la presencia de los nombres propios (*el padre Job Bishop, Canterbury*, etc.), utilizando el elemento <name> de la siguiente manera:

```

<p>
  El primero de febrero de 1792 <name>el padre Job Bishop</name>
  y el <name>venerable Edmond Lougee</name> llegaron desde
  <name>New Lebanon N.Y.</name>
  para organizar y establecer una comunidad de creyentes en
  <name>Canterbury, N.H.</name>
</p>

```

- 35 Incluso este etiquetado mínimo permite distinguir fácilmente los nombres de las personas y los lugares mencionados y, por lo tanto, procesarlos por separado del resto del texto, por ejemplo, para construir un índice de nombres o para vincular estas referencias a más información sobre la persona o el lugar en cuestión.
- 36 Un etiquetado más cuidadoso distinguiría, sin embargo, los componentes de los nombres que (como ocurre a menudo) combinan apellidos (*Bishop, Lougee*) con nombres que indican un papel social (*padre, venerable*) y con nombres de pila (*Job, Edmond*). Estas distinciones solo pueden hacerse explícitas mediante un etiquetado adecuado: *Bishop* (obispo) puede ser un papel o un apellido, según el contexto. Del mismo modo, dentro de un <placeName> es conveniente distinguir el nombre de una región administrativa o de otro tipo, como un estado de EE.UU., por un lado, del nombre de un asentamiento, por el otro. (TEI prefiere utilizar este término neutral —*settlement*, “asentamiento”— en lugar de otros términos como *pueblo, aldea, ciudad*, etc., que son difíciles de definir fuera de un contexto muy preciso.)

```

<p>
  El primero de febrero de 1792
  <persName>
    <roleName>el padre</roleName>
    <forename>Job</forename>
    <surname>Bishop</surname>
  </persName>
  y
  <persName>
    <roleName>el venerable</roleName>
    <forename>Edmond</forename>
    <surname>Lougee</surname>
  </persName>
  llegaron desde <name>New Lebanon N.Y.</name>
  para organizar y establecer una comunidad de creyentes
  en <name>Canterbury N.H.</name>
  <placeName>
    <settlement>New Lebanon</settlement>
    <region>N.Y.</region>
  </placeName>
</p>

```

- 37 Por definición, un nombre se refiere a una entidad específica (una persona o un lugar), aunque la investigación histórica o genealógica no supondría un gran reto si lo hiciera de forma inequívoca o directa. ¿Es este *Edmond Lougee* la misma persona que el *Edmund*

Lougee registrado en otras fuentes? Un poco más adelante en la fuente, se menciona a *Elder Lougee*: ¿se refiere a la misma persona? Como siempre, TEI no puede ayudarle a responder a estas preguntas, pero sí proporciona formas de expresar la respuesta (si es que hay alguna) a la que ha llegado.

- 38 Proponer una forma estandarizada para un nombre ha sido, por supuesto, desde hace mucho tiempo, asunto de bibliotecarios y catalogadores. Para las obras literarias y las personas, y para las geografías modernas o clásicas, existen obras de referencia bien establecidas o "archivos de autoridad" que proporcionan un punto de referencia útil para muchas de las entidades que pueden ser nombradas en un documento TEI; para otras, podemos encontrar algún tipo de información canónica en otras referencias en línea como Wikipedia o ancestry.com. Una forma de dejar claro a cuál de las muchas personas posibles se refiere este *Edmond Lougee* sería, entonces, proporcionar una referencia a un registro definitivo en línea de la persona:

```
...
<persName
  ref="https://www.ancestry.co.uk/genealogy/records/edmund-
  lougee-24-12v0p19">
  el venerable Edmond Lougee
</persName>
...
```

- 39 Sin embargo, dado que es probable que solo algunas personas de un documento de archivo de este tipo dispongan de este tipo de registros, a menudo será necesario que un proyecto defina su propio archivo de autoridad, compuesto por una serie de registros, uno por cada individuo identificado. TEI proporciona elementos adicionales que pueden utilizarse para registrar dicha información con bastante detalle: la información biográfica o prosopográfica, por ejemplo, puede agruparse dentro de un elemento <person>, mientras que la información geográfica puede agruparse dentro de un elemento <place>; si se hace esto, puede utilizarse el atributo @ref para apuntar a dicha descripción:

```
<persName ref="#P1234">Elder Edmond Lougee</persName>
...
<person xml:id="P1234">
  <p>
    Edmund o Edmond Lougee nació en Exeter Newmarket, Rockingham,
    New Hampshire, Estados Unidos, en 1731,
    hijo de John Lougee y Anne Gilman.
    Se casó con Hannah Lord y tuvo 7 hijos.
    Falleció el 3 de junio de 1807 en
    Loudon, New Hampshire, EE.UU.
  </p>
</person>
```

- 40 Nótese que este ejemplo simplemente proporciona un resumen en prosa de la información disponible sobre el venerable Lougee dentro del elemento <person> referenciado; podría haberse construido alternativamente utilizando elementos

especializados como <birth> (nacimiento), <death> (muerte), <marriage> (matrimonio), <relation> (relación), etc.

- 41 Sin embargo, puede ser suficiente para las necesidades de un proyecto proporcionar con cada referencia un código arbitrario, o una forma normalizada del nombre, quizás incluyendo fechas o un marcador generacional para que esa forma sea inequívoca. El atributo @key está previsto para este fin:

```
<persName key="Lougee, Edmond (1731-1807)">
  el venerable Edmond Lougee
</persName>
```

- 42 La normalización mediante valores de atributos es un mecanismo que se encuentra en la totalidad de TEI. Es especialmente útil cuando algo que un motor de búsqueda u otro procesador preferiría encontrar en un formato concreto aparece en realidad en un documento de muchas formas diferentes, algunas de las cuales pueden tener algún interés o significado lingüístico. Si queremos recuperar de este documento todas las referencias a acontecimientos ocurridos antes o después de una fecha concreta, por ejemplo, la tarea será mucho más fácil si todas las fechas utilizan una forma estandarizada. Por tanto, en el ejemplo anterior, podríamos optar por codificar la primera fecha de la siguiente manera:

```
<p>
  En <date when="1792-02-01">el primero de febrero de 1792</date>
  <!-- ... -->
</p>
```

utilizando la recomendación del W3C para el formato de las fechas. El atributo @when utilizado aquí nos permite asociar una fecha con un punto específico en el tiempo. TEI también proporciona un conjunto de atributos adicionales que pueden combinarse para expresar una variedad de conceptos temporales:

```
<p>
  <date notAfter="1792-02-28" notBefore="1792-02-01">
    En febrero de 1792
  </date>
  <!-- ... -->
</p>
```

```
<p>
  <date from="1792-02-01" to="1792-02-07">
    Durante la primera semana de febrero de 1792
  </date>
  <!-- ... -->
</p>
```

Cuadros, figuras y bibliografías

- 43 Los documentos del mundo real contienen muchos componentes no textuales o solo semi-textuales, como figuras o cuadros, así como componentes textuales que tienen su propia estructura interna, como índices y bibliografías. Especialmente cuando TEI se utiliza para codificar un nuevo documento, aunque también cuando es necesario indicar la presencia de características de este tipo en uno antiguo, el codificador no puede prescindir de elementos como <table> (cuadro o tabla), <figure> (figura) o <bibl> (*bibliographic citation*, cita bibliográfica), a los que nos referiremos muy brevemente en las siguientes secciones.

Cuadros

- 44 Un cuadro es una forma de organizar varios fragmentos de información textual relacionados entre sí en filas (o columnas) compuestas por celdas. La mayoría de los sistemas de producción de documentos han desarrollado formas bastante complejas de indicar exactamente cómo deben mostrarse las celdas de un cuadro, aunque el modelo de cuadro TEI carece en gran medida de ellas. En TEI, un elemento <table> se compone de elementos <row> (fila), que a su vez se componen de elementos <cell> (celda); por lo tanto, un cuadro TEI no puede representarse fácilmente como una serie de columnas. Existe un atributo @role para indicar si una fila o celda particular contiene información de etiquetado o datos, y hay atributos @rows y @cols que pueden utilizarse para indicar cuándo una fila o una celda abarca más de una fila o columna respectivamente.
- 45 Supongamos que encontramos el siguiente cuadro (imaginario) en un documento que estamos codificando:

Figura 4

Frutas	manzana	plátano	cereza		dátil
Nueces	almendra	nuez del Brasil	nuez de coco	marañón	pistacho

Un cuadro imaginario

- 46 En nuestra versión TEI, podemos registrar que la primera columna de celdas contiene el título para las entradas del resto de cada fila, y también que la celda en la tercera columna de la primera fila abarca dos columnas:

```

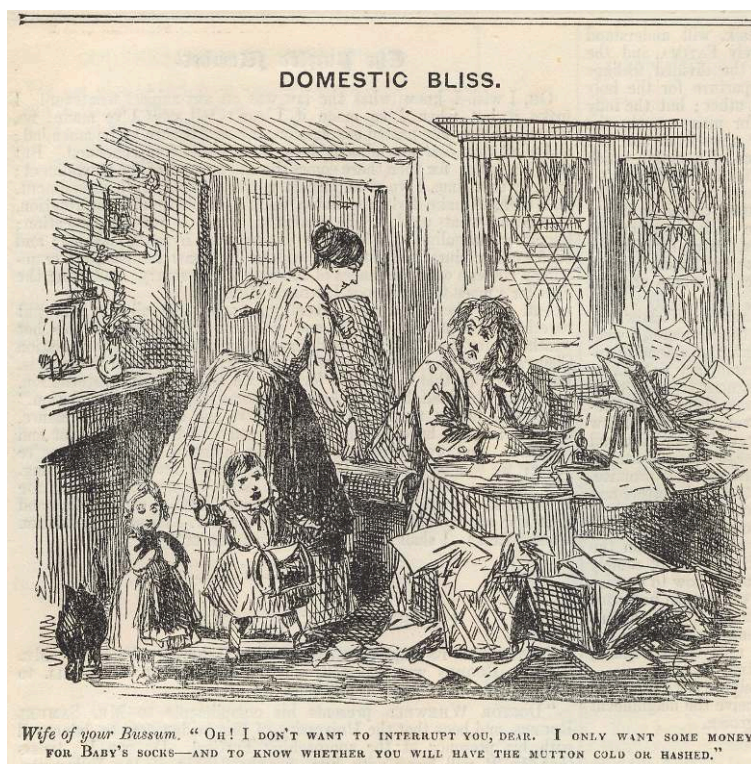
<table>
<row>
  <cell role="título">Frutas</cell>
  <cell>manzana</cell>
  <cell>plátano</cell>
  <cell cols="2">cereza</cell>
  <cell>dátil</cell>
</row>
<row>
  <cell role="título">Nueces</cell>
  <cell>almendra</cell>
  <cell>nuez del Brasil</cell>
  <cell>nuez de coco</cell>
  <cell>marañón</cell>
  <cell>pistacho</cell>
</row>
</table>

```

Figuras

- 47 Todo tipo de componente visual, por ejemplo una ilustración, un gráfico o un diagrama, puede estar incrustado en un documento, a veces como un componente independiente, otras veces como un frontispicio o una ilustración paginada por separado, y otras veces como parte de alguna división del texto. Tales **figuras** contienen a menudo encabezados o títulos, posiblemente asociados a algún texto de corrido, así como una imagen. Para codificarlas, TEI proporciona el elemento <figure> (figura), que suele contener al menos un elemento <graphic> (gráfica) y uno o más elementos <head> (encabezado), aunque también pueden utilizarse otros elementos textuales si la ilustración contiene texto.
- 48 El elemento <graphic> es un tipo especial de puntero. Su atributo @url apunta a una ubicación en la que puede encontrarse una representación digital de la imagen correspondiente. Los atributos @scale, @width y @height están disponibles para especificar respectivamente la escala, el ancho y/o el alto deseados de la imagen al mostrarla.
- 49 Consideremos el siguiente ejemplo tomado de *Punch*, una conocida publicación humorística británica del siglo XIX:

Figura 5



Domestic Bliss

50 Podemos codificarla así:

```
<figure type="caricatura" place="arriba-izquierda">
<head rend="mayúsculas">Domestic Bliss</head>
<graphic url="vol13p14.png"/>
<sp>
<speaker rend="italic">Wife of your bussum</speaker>
<p rend="versalitas">
<hi rendition="#cursivas">
  Oh! I don't want to interrupt you dear. I only want some money
  for Baby's socks – and to know whether you will have the mutton
  cold or hashed.
</hi>
</p>
</sp>
<figDesc>
  Un interior doméstico dibujado por Leech, que muestra a
  un hombre con el pelo desgreñado en bata sentado ante un escritorio
  cubierto de papeles, con su esposa, dos niños pequeños, uno de
  ellos golpeando un tambor, y un gato.
</figDesc>
</figure>
```

51 En nuestra codificación hemos transcrito el texto que aparece debajo del gráfico utilizando el elemento <sp> (discurso) mencionado anteriormente, para mostrar que el texto se presenta como una pieza de diálogo dramático. También hemos utilizado el

elemento <figDesc> (descripción de figura) para encerrar un texto adicional, que no aparece en la figura, pero que aporta metadatos descriptivos útiles que podrían mostrarse como alternativa al propio gráfico, o utilizarse para indexar su contenido.

Referencias bibliográficas

- 52 Las referencias bibliográficas, como las que suelen incluirse en forma de bibliografía o de lista de referencias al final de un artículo académico, o en una nota a pie de página, son una característica común de los textos académicos. Resulta útil distinguir explícitamente este tipo de elementos y particularmente sus componentes (autor, título, editorial, fecha de publicación, etc.), tanto para que se puedan buscar con mayor precisión como para facilitar de diferentes maneras su representación o formato. Cuando hay que manejar un gran número de referencias de este tipo, las herramientas especializadas como Zotero, que pueden importar o exportar en formato TEI, resultan eficaces para mantenerlas, pero pueden tratarse de la misma forma que cualquier otro componente de un documento TEI, ya sea que aparezcan en el cuerpo del texto, como una división separada de los paratextos finales (el *backmatter*) o en el encabezado.
- 53 En la escritura académica es habitual proporcionar referencias en un formato estándar, de modo que los componentes aparezcan en un orden especificado y utilicen un diseño consistente. Ya sea en una nota a pie de página, o en una lista de referencias independiente, un libro se describirá normalmente de esta manera:
- 54 Cameron, D. (1995) *Verbal Hygiene*. Londres y Nueva York: Routledge.
- 55 El marcado TEI puede utilizarse para mostrar qué parte es el título, cuál el lugar de publicación, etc. TEI proporciona dos elementos distintos para este fin: <bibl> (cita bibliográfica) y <biblStruct> (cita bibliográfica estructurada).
- 56 El elemento <biblStruct> proporciona una visión “estructurada” o *centrada en los datos* de una entrada bibliográfica, en la que cada parte de la descripción (autor, título, etc.) se etiqueta de forma diferenciada, tratándola como si formara parte de una base de datos. Tengamos en cuenta que no se permite la puntuación entre los componentes: esto se debe a que diferentes visualizaciones pueden ser necesarias para diferentes estilos bibliográficos. Un etiquetado detallado como el siguiente permite hacerlo de forma sencilla: el programa de formateo simplemente elige los elementos necesarios en el orden requerido, insertando la puntuación según el estilo requerido.


```
<bibl xml:id="Cameron1995">
  <monogr>
    <author>
      <surname>Cameron</surname>
      <forename>Deborah</forename>
    </author>
    <title>Verbal Hygiene</title>
    <imprint>
      <publisher>Routledge</publisher>
      <pubPlace>Londres</pubPlace>
      <pubPlace>Nueva York</pubPlace>
      <date>1995</date>
    </imprint>
  </monogr>
</bibl>
```

- 57 El elemento `<monogr>` (*monographic level*, nivel monográfico) indica aquí lo que los bibliógrafos llaman un ítem *monográfico*, es decir, una obra individual, a diferencia de un ítem *analítico* tal como un artículo de una revista.
- 58 El elemento `<bibl>` proporciona una visión menos estructurada —más *centrada en el texto*— de una entrada bibliográfica, donde se permite la puntuación entre los elementos etiquetados, así como cualquier otro contenido textual, y el orden de los componentes no es fijo. Resulta útil cuando el codificador desea seguir las convenciones del catálogo (normalmente impreso) del que se extrae la descripción o al cual se destina.
- 59 Con tal elemento podríamos etiquetar este artículo de forma que se parezca más a la versión original, a costa de dificultar un poco su tratamiento posterior:

```
<bibl>
  <author>Deborah Cameron</author>:
  <title level="m">Verbal Hygiene</title>.
  Londres, Nueva York:
  Routledge, <date>1995</date>.
</bibl>
```

- 60 La mayoría de los elementos disponibles dentro de `<biblStruct>` también están disponibles (pero son opcionales) dentro de `<bibl>`, por lo que es posible utilizar cualquiera de los dos elementos para proporcionar un etiquetado detallado de cualquier tipo de referencia bibliográfica. Abajo ofrecemos otros ejemplos.
- 61 Cuando se menciona un libro dentro de un texto de corrido, sobre todo si se menciona más de una vez, es habitual hacerlo mediante una referencia abreviada como “Cameron 1995”. Para vincular dicha referencia a una descripción bibliográfica más completa de la obra en cuestión podemos utilizar el atributo `@target` en un elemento `<ref>` (*reference*, referencia) o `<ptr>` (*pointer*, puntero):

<p>

Los humanos tenemos <q>una sana obsesión por el lenguaje</q>
(<ref target="#Cameron1995">Cameron 1995</ref>).

Sería sorprendente que no lo hiciéramos...

</p>

Cornucopia TEI, segunda parte

Asuntos editoriales

- 1 Cuando se emprende la digitalización de materiales de fuentes primarias, la primera prioridad suele ser producir imágenes digitales de buena calidad de los documentos originales de la fuente. La transcripción y codificación de esas imágenes es una operación costosa y difícil que no se puede automatizar fácilmente, si es que se hace, pero es un paso importante en la creación de una verdadera “edición digital”. Para el lector humano avezado, la imagen digital de sus páginas puede ser todo lo que se necesita para que la fuente sea útil. Sin embargo, para un público más amplio, y desde luego para cualquier tipo de análisis o búsqueda automatizada, una transcripción codificada será un complemento esencial a la imagen. TEI dispone de una variedad de mecanismos para facilitar la producción de ediciones digitales de todo tipo.

Imagen y transcripción

- 2 Anteriormente hemos mencionado el elemento `<facsimile>` (facsímil), que puede utilizarse para agrupar y organizar las imágenes digitales en función de las superficies físicas que representan —mediante el elemento `<surface>` (superficie)— y de las zonas lógicas de interés dentro de esas superficies —mediante el elemento `<zone>` (zona)—. Un `<facsimile>` puede ser todo lo que contenga un documento TEI, o puede estar complementado por una transcripción.
- 3 Cuando se dispone tanto de un facsímil como de una transcripción, se puede utilizar un atributo global `@facs` para vincularlos o vincular algunas de sus partes, según convenga. Supongamos, por ejemplo, que tenemos una sola hoja, con imágenes de sus dos superficies en dos archivos llamados `pag1r.png` y `pag1v.png` respectivamente. Transcribimos ambas superficies dentro de un elemento `<text>`, como es habitual, utilizando `<pb>` para marcar los puntos de la transcripción que se corresponden con el inicio de cada nueva página en la fuente. El documento resultante podría tener el siguiente aspecto:

```

<facsimile>
  <surface xml:id="s1">
    <graphic url="pag1r.png"/>
  </surface>
  <surface xml:id="s2">
    <graphic url="pag1v.png"/>
  </surface>
</facsimile>
<text>
  <body>
    <pb facs="#s1"/>
    <!-- texto transcrito de la página 1 recto aquí -->
    <pb facs="#s2"/>
    <!-- final texto transcrito de la página 1 verso aquí -->
  </body>
</text>

```

- 4 Nótese que asociar la transcripción con la superficie en lugar de con el archivo de imagen directamente nos ofrece la flexibilidad de añadir más archivos de imagen (de mayor o menor resolución, por ejemplo) sin alterar la estructura.
- 5 El atributo @facs utilizado anteriormente es global, por lo que puede utilizarse para asociar cualquier parte de una transcripción con cualquier parte de un elemento <facsimile>. Podríamos, por ejemplo, vincular algunas palabras de la transcripción con la parte de la imagen en la que aparecen.

Edición de una transcripción

- 6 Al transcribir una fuente, sobre todo una antigua, siempre existe una tensión entre el deseo de representar fielmente el estado real del documento y el deseo de conseguir que ese documento sea comprensible para un lector contemporáneo no especializado. El proceso de transcripción conlleva necesariamente una especie de normalización, en la que los diferentes símbolos escritos (los glifos) de un manuscrito o de un documento impreso antiguo se mapean en los códigos inequívocos de un conjunto de caracteres moderno. Por lo tanto, una codificación que intente representar las palabras de un documento producido según normas diferentes de las nuestras, o corregir errores “evidentes” en el texto de la copia resulta casi inevitable. TEI proporciona formas de marcar tramos de texto como normalizados o corregidos por el codificador, independientemente de que consideremos esto como “edición” en el sentido tradicional.
- 7 Ya hemos mencionado que TEI se inscribe conscientemente en la larga tradición filológica occidental que se ocupa de la identificación y recuperación de textos a partir de testimonios documentales dispersos. Por lo tanto, no es de extrañar que también ofrezca facilidades para representar fenómenos como la corrección o la alteración por parte de los escribas, cuando se manifiestan en un documento concreto, o para representar un cotejo de lecturas variantes. Mientras que los editores de textos anteriores a la era digital se ocupaban necesariamente de la producción de una única versión impresa de un texto, minuciosamente preparada y corroborada, la economía de

la producción digital ha facilitado enormemente la producción de ediciones digitales o documentales “diplomáticas”, que pretenden presentar objetivamente los hechos de una determinada colección de documentos de forma que el lector pueda determinar la plausibilidad de cualquier versión “editada” que la acompañe, o incluso derivar dicha edición por sí mismo.

- 8 Al transcribir un texto fuente, se puede utilizar el elemento <corr> (corrección) para encerrar una parte que ha sido corregida y el elemento <reg> (regularización) para una parte que ha sido normalizada. En el caso contrario, se puede utilizar el elemento <sic> (así) cuando algo debería haber sido corregido pero no lo ha sido; y el elemento <orig> (original) para los casos en que se ha conservado la ortografía original aunque parezca desconocida. Por ejemplo, una versión modernizada del soneto de Shakespeare mostrado anteriormente probablemente comenzaría:

```
<l>My Mistress' eyes are nothing like the sun,</l>
<l>Coral is far more red than her lips red,</l>
```

- 9 Incluso en una versión modernizada de este tipo, un codificador concienzudo que desee mostrar dónde se ha apartado de la fuente original puede simplemente encerrar las palabras *Mistress'*, *sun*, etc. en el elemento <reg>. Sin embargo, lo más habitual es que un codificador desee conservar la ortografía original junto con la versión modernizada, de modo que cualquiera de ellas esté disponible para su visualización. El elemento <choice> (elección) atiende a este requisito:

```
<l>
My
<choice>
  <reg>Mistress'</reg>
  <orig>Mistres</orig>
</choice>
eyes are nothing like the
<choice>
  <reg>sun</reg>
  <orig>Sunne</orig>
</choice>,
</l>
<l>
<choice>
  <reg>Coral</reg>
  <orig>Curral</orig>
</choice>
is far more red
<choice>
  <reg>than</reg>
  <orig>then</orig>
</choice>
her lips red,
</l>
```

- 10 El elemento `<choice>` puede utilizarse cuando sean posibles varias codificaciones diferentes, pero solo se pueda elegir una de ellas para una aplicación concreta. Es menos apropiado cuando haya muchos testimonios posibles diferentes, que proporcionen múltiples lecturas, todas las cuales deben tenerse en cuenta. TEI dispone de un conjunto más potente y complejo de etiquetas para marcar un aparato crítico tradicional de este tipo. El elemento `<app>` (aparato) se utiliza para marcar el punto donde se está considerando una variante; dentro de este elemento, el codificador puede suministrar una serie de elementos `<rdg>` (*reading*, lectura), uno por cada lectura variante identificada. Si la política editorial lo permite, una de estas lecturas puede considerarse como la “primaria” o “correcta”, en cuyo caso se etiquetará mediante el elemento `<lem>` en lugar de como otro `<rdg>`. El elemento `<rdg>` tiene un atributo `@wit` que puede utilizarse para identificar los testimonios en los que se atestigua la lectura en cuestión. Por ejemplo, la primera línea del *Cuento de la esposa de Bath* de Chaucer comienza: *Experience though noon Auctoritee ...*, aunque su primera palabra está escrita de forma muy diferente en tres de los manuscritos que se conservan. En conjunto, estas variaciones ortográficas aportan pruebas útiles para la reconstrucción de la historia de la tradición manuscrita en cuestión, por lo que nuestra codificación desea preservarlas. Podríamos hacerlo de la siguiente manera:

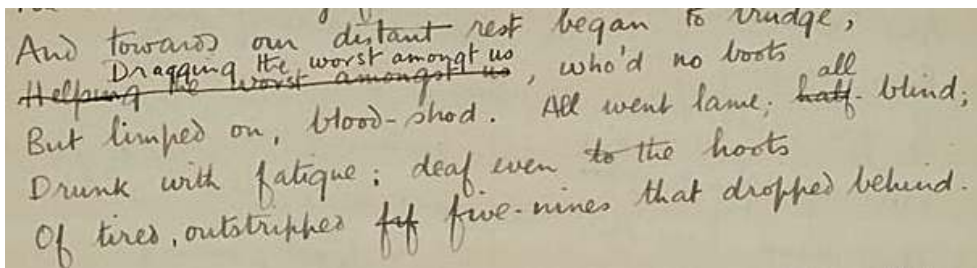
```
<l n="1">
  <app>
    <rdg wit="#El #Hg">Experience</rdg>
    <rdg wit="#La">Experiment</rdg>
    <rdg wit="#Ra2">Eryment</rdg>
  </app> though noon Auctoritee ...
</l>
```

- 11 Aquí los valores del atributo `@wit` identifican los manuscritos en cuestión: como indica el signo de numeral, son en realidad enlaces a elementos de una lista de los manuscritos que se cotejaron para producir este aparato.

Codificación del proceso de escritura

- 12 Consideremos ahora un último ejemplo de las posibilidades que TEI ofrece para trabajar en la transcripción detallada de materiales de fuentes primarias. Aquí se muestran algunos de los elementos necesarios para registrar el proceso de escritura real, por ejemplo para registrar los pasajes que han sido suprimidos, añadidos, corregidos, etc., ya sea por el autor de un texto literario o por un escriba que copia un manuscrito. El análisis de estas modificaciones documentales puede ser esencial antes de presentar un texto para lectura y resulta de evidente importancia en el proceso de redacción.
- 13 El ejemplo está tomado del manuscrito de autor que se conserva de un poema del escritor inglés Wilfred Owen, del que se muestra una parte:

Figura 6



Detalle de *Dulce et decorum est*, manuscrito autógrafo conservado en la English Faculty Library de la Universidad de Oxford

- 14 Owen escribió primero *Helping the worst among us*, pero luego lo borró, añadiendo *Dragging the worst amongst us* encima. Asimismo, revisó la frase *half-blind* borrando *half* y añadiendo *all* encima. En la última línea, Owen empezó a escribir una palabra que comenzaba por *fif*, antes de borrarla para escribir la palabra *five-nines*. Podemos codificar todo esto como sigue:

```
<l>And towards our distant rest began to trudge,</l>
<l>
<subst>
  <del>Helping the worst amongst us</del>
  <add>Dragging the worst amongst us</add>
</subst>, who'd no boots
</l>
<l>But limped on, blood-shod. All went lame;
<subst>
  <del>half-</del>
  <add>all</add>
</subst> blind;
</l>
<l>Drunk with fatigue; deaf even to the hoots</l>
<l>Of tired, outstripped
  <del>fif</del>
  five-nines that dropped behind.
</l>
```

- 15 Los elementos `<add>` (adición) y `` (*deletion*, eliminación) se utilizan para encerrar los pasajes añadidos o borrados, respectivamente. Existen atributos adicionales como `@resp` para indicar la responsabilidad de la modificación, o `@place` para indicar en qué parte del texto (por ejemplo, por encima o por debajo de la línea) se ha realizado la modificación. Cuando el codificador desee afirmar que la adición y la supresión constituyen un único acto editorial de sustitución, estos elementos pueden combinarse dentro de un elemento `<subst>`, como se muestra arriba.
- 16 Un examen muy cuidadoso de la segunda modificación muestra que Owen en realidad escribió *amongt* en lugar de *amongst*, presumiblemente por error. Un editor igualmente cuidadoso que desee restaurar la *s* que falta podría utilizar el elemento `<subst>` (sustitución) para indicar que lo ha hecho:

```
<add>
  Dragging the worst
  among<supplied resp="#ED">s</supplied>t us
</add>
```

- 17 Aquí se ha utilizado el atributo @resp para indicar que la s no la escribió Owen sino otra persona, en concreto la persona documentada en otro lugar por un elemento con el identificador ED.

Transcripción de un discurso oral

- 18 El soporte digital almacena y representa sonido o video con la misma facilidad y eficacia con la que lo hace con imágenes. Las transcripciones de las grabaciones de sonido o vídeo pueden utilizarse para complementarlas del mismo modo que se utilizan para complementar las imágenes de las páginas. Ya sean producidas de forma automática o manual, las transcripciones de discursos orales son de especial interés para numerosas ramas de la lingüística; por ejemplo, en los estudios de adquisición del lenguaje, en los estudios de las variaciones del lenguaje o de la relación entre el discurso oral y el escrito tanto en las sociedades alfabetizadas como en las prealfabetizadas. El estudio del habla transcrita es también un componente esencial de estudios sociológicos como la historia oral.
- 19 Sin embargo, no hay acuerdo universal sobre las unidades en las que se lleva a cabo dicho análisis —en términos de TEI, los elementos que componen un <texto> que contiene un discurso transcrito—, e incluso pueden ser controvertidas. Otro factor que complica las cosas es que el proceso de transcripción del habla es técnicamente difícil y a menudo se lleva a cabo con la ayuda de software para fines específicos; esto es una complicación porque la mayoría de estos programas tienden a utilizar sus propios formatos internos en lugar de un estándar definido externamente.
- 20 No obstante, TEI propone un conjunto de elementos que pueden ser útiles para la codificación detallada de las transcripciones de voz y que, según un artículo reciente de Thomas Schmidt, podrían utilizarse como una especie de *interlingua* para los métodos de codificación alternativos existentes.
- 21 El modelo de TEI se ocupa del habla como sistema comunicativo; por ello, propone el etiquetado explícito de cualquier parte de un sonido grabado que tenga una función o efecto comunicativo. Así, acontecimientos como el toque de una campana o el sobrevuelo de un avión se señalan explícitamente mediante el elemento <incident> (incidente). La comunicación oral se efectúa mediante la voz, pero no todos los efectos vocales se consideran léxicos y no toda la comunicación es vocal; por ello, TEI proporciona el elemento <vocal> (vocal) para los fenómenos no léxicos, como la tos o la risa, así como el elemento <kinesic> (cinético) para los fenómenos no léxicos y no vocales, como los gestos o la expresión facial. Estos elementos se complementan con otros relacionados con lo que podría llamarse la *representación (rendition)* del discurso, especialmente <pause> (pausa) y <shift> (elevación); este último se utiliza para indicar cambios en la calidad de la voz (por ejemplo, en el volumen, el tono o la velocidad). Estos funcionan de forma muy similar a los elementos de hito introducidos anteriormente.

- 22 Una de las nociones más controvertidas en la investigación del discurso hablado es la de *turno de palabra* (*turn*) o *proferencia* (*utterance*). TEI ofrece el elemento <u> (proferencia), que puede utilizarse para delimitar un fragmento de discurso de un solo hablante, sin hacer ninguna afirmación sobre el estatus o la función discursiva de dicho fragmento. El elemento lleva un atributo @who (quién) que puede utilizarse para vincularlo a otro elemento que proporcione información sobre la persona que habla, tal como hemos visto ya para el atributo @ref.
- 23 Para muchos tipos de análisis lingüístico, el elemento <s> mencionado anteriormente puede ser una forma más útil de organizar una transcripción, ya que permite al codificador categorizar cada segmento independientemente del enunciado que lo contiene, sobre todo porque los enunciados pueden ser mucho más largos (o cortos) que las unidades de análisis habituales.
- 24 A continuación presentamos un ejemplo simplificado tomado del British National Corpus (BNC) en el que se han utilizado tanto <s> como <u> para segmentar la señal del discurso transcrito:

<u who="#PS6P0">
 <s n="3">There ain't</s>
 </u>
 <u who="#PS6NY">
 <s n="4">
 <shift new="riendo"/>There is<shift/>
 </s>
 </u>
 <u who="#PS6P0">
 <s n="5">Oh shut up</s>
 </u>
 <u who="#PS6NY">
 <s n="6">Ach you do get corners on boats</s>
 </u>
 <u who="#PS6P0">
 <s n="7">No</s>
 <s n="8">Boats are shaped like a bloody rugby ball shape type</s>
 </u>
 <u who="#PS6NY">
 <s n="9">No they ain't</s>
 <s n="10">One end is and the other one ain't <pause/> and it was
 a yacht <pause/> and a yacht they got little rooms in or something cos
 <unclear/> innit</s>
 </u>
 <u who="#PS6P0">
 <s n="11">Oh that<pause/> th the rooms are shaped ni <pause/>
 like to the size of the boat<pause/> you nonce
 <vocal desc="chasqueo"/>
 </s>
 </u>
 <u who="#PS6NY">
 <s n="12">Let's ask your Mum if there's <pause/> if there's any
 corners on a boat</s>
 </u>
 <u who="#PS6P0">
 <s n="13">Of course there ain't</s>
 </u>
 <u who="#PS6NY">
 <s n="14">Just ask your Mum that</s>
 </u>
 <u who="#PS6P0">
 <s n="15">Yeah</s>
 </u>
 <u who="#PS6NY">
 <unclear/>
 </u>
 <u who="#PS6P0">
 <s n="16">I bet she'll probably side with you</s>
 </u>

- 25 Al transcribir el discurso oral, el transcriptor también debe tomar decisiones sobre cómo representar los sonidos no estándar o solo semi-lexicalizados que emiten los hablantes, como *hmmm*, o el *ach* o el *tsk* del ejemplo anterior (en este caso se ha optado por tratar el primero como palabra y el segundo como <vocal>). Incluso cuando está claro que el sonido es léxico, no siempre está claro cómo debe escribirse: por ejemplo, el sonido representado arriba como *yeah* podría escribirse igualmente *yeh* o *yuh* o incluso *yes*.
- 26 En los corpus lingüísticos de este tipo también es una práctica común marcar cada palabra explícitamente, a menudo utilizando el elemento <w> (*word*, palabra). Decidir cómo debe hacerse esta tokenización no siempre es evidente: en el BNC, por ejemplo, las formas elididas como *ain't* y *there's* se etiquetan como sendos elementos <w>, y se utilizan etiquetas adicionales para distinguir los inicios falsos o las formas truncadas como *th* y *ni* en el ejemplo anterior. Sin embargo, el uso principal de este etiquetado es proveer información adicional como el análisis de *categorías gramaticales* o el análisis *morfosintáctico*, tal como describiremos en la sección Anotación lingüística más adelante.
- 27 Hay otra función importante del modelo de TEI para la transcripción del habla: la alineación. Mientras que en una imagen la transcripción debe alinearse dentro de un espacio de coordenadas bidimensional, en un medio basado en el tiempo como el audio, las distintas partes de una transcripción deben situarse en el tiempo. Los hablantes suelen interrumpirse unos a otros, o hablar al mismo tiempo; la comprensión de un diálogo es imposible sin información sobre la secuencia de eventos transcritos. TEI propone una forma de registrar explícitamente una especie de escala temporal, utilizando el elemento <timeline> (línea de tiempo), que contiene definiciones de puntos en el tiempo, representados como elementos <when> (cuándo). Los eventos de una transcripción que tienen lugar en un punto determinado del tiempo pueden entonces alinearse con los elementos <when> apropiados utilizando atributos como @sync (sincrónico). La ventaja de este método es que permite que una transcripción represente el mínimo de información necesaria para alinear sus componentes, sin necesidad de una marca de tiempo precisa para cada uno de ellos; sin embargo, dicha marca de tiempo generalmente lo proporciona un sistema de transcripción automática y, por lo tanto, puede ser más sencillo utilizarlo.

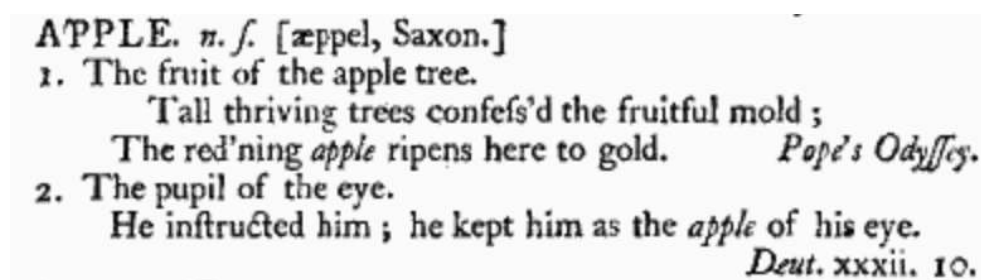
Diccionarios

- 28 TEI dispone de elementos para la codificación detallada de tan solo un reducido conjunto de tipos de documentos distintos, ya que en general prefiere definir elementos específicos para determinados tipos de análisis o enfoques. Sin embargo, sí proporciona un conjunto de etiquetas para el marcado de diccionarios y léxicos, quizás porque históricamente estos se encuentran entre el tipo de obras de referencia que primero demostraron poder manejarse manifiestamente mejor en forma digital que impresa.
- 29 Los diccionarios, sobre todo en forma impresa, se diferencian de la mayoría de documentos en que no consisten en un texto seguido, sino en entradas separadas dentro de las cuales se pueden reconocer fácilmente los subcomponentes individuales y semánticamente significativos (aunque no siempre se presenten tan sistemáticamente

como desearía un diseñador de bases de datos). Una entrada de diccionario siempre tiene una palabra principal o lema, seguida normalmente de alguna información lingüística, como su categoría gramatical, y de una serie de definiciones de sentido, a menudo ordenadas jerárquicamente. También puede haber información etimológica, ejemplos de uso, referencias a homónimos, etc.

- 30 En una codificación al estilo de TEI, todas estas y otras partes de una entrada se distinguirán tan claramente como sea posible, tanto si el propósito es producir un nuevo diccionario (donde todos estos elementos deben presentarse de modo coherente), como si se trata de representar uno ya existente (donde distinguirlos mejorará el rango de búsqueda inteligente).
- 31 En un primer ejemplo sencillo, consideremos la imagen siguiente:

Figura 7



Apple, en el *Dictionary of the English Language* (1755) de Samuel Johnson (1755)

- 32 Una codificación TEI mínima para esta entrada distinguiría la forma de la palabra utilizada como lema, su codificación ortográfica y gramatical, la etimología proporcionada y los dos sentidos:

```

<entry>
  <form type="lema">
    <orth rend="mayúsculas">A'PPLE</orth>.
    <gramGrp>
      <pos norm="nombre">n.s.</pos>
    </gramGrp>
  </form>
  <etym>
    [<foreign xml:lang="ang">æppel</foreign>,
    <lang>Saxon</lang>.]
  </etym>
  <sense n="1">
    <def>The fruit of the apple tree.</def>
    <cit>
      <quote type="verso">
        <l>Tall thriving trees confess'd the fruitful mold;</l>
        <l>The red'ning <hi>apple</hi> ripens here to gold.</l>
      </quote>
      <bibl>Pope's <title>Odissey</title>.</bibl>
    </cit>
  </sense>
  <sense n="2">
    <def>The pupil of the eye.</def>
    <cit>
      <quote>
        He instructed him; he kept him as the apple of his eye.
      </quote>
      <bibl>Deut. ii. 10.</bibl>
    </cit>
  </sense>
</entry>

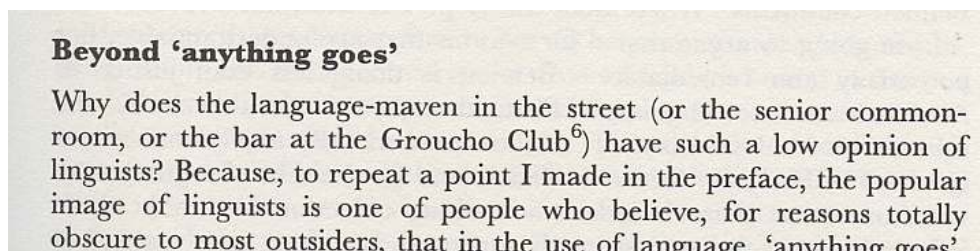
```

Notas

- 33 Los textos, sobre todo los antiguos, suelen ir acompañados de extensas anotaciones y comentarios, ya sea aportados por el escritor original o, más frecuentemente, por un editor o comentarista posterior. Todos los escritos académicos, del tipo que sean, se caracterizan también por las notas a pie de página, las notas al margen, las glosas, las referencias bibliográficas, etc. El modo más natural de etiquetar todas estas características es utilizar un elemento `<note>` (nota) para contener la anotación propiamente dicha, e incrustarla dentro del texto que se anota en el “punto de anclaje” (*point of attachment*), es decir, el lugar donde suele aparecer una llamada a nota a pie de página u otra marca en un texto impreso. Los atributos del elemento `<note>` pueden utilizarse para categorizarlo de alguna manera, para indicar su ubicación aproximada y para indicar quién es el responsable de ella.
- 34 Según la práctica moderna de impresión (especialmente en las publicaciones anglosajonas), las anotaciones explicativas suelen trasladarse al final del libro. Por

ejemplo, en una discusión sobre las actitudes contemporáneas hacia los lingüistas publicada en 1995, encontramos el siguiente comentario:

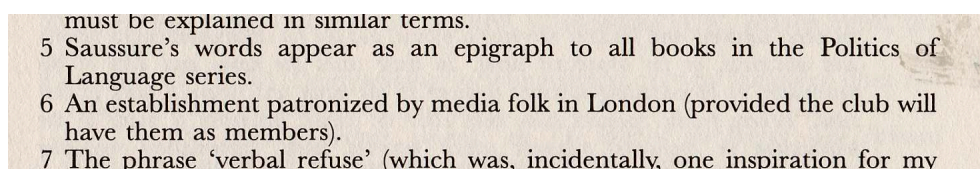
Figura 8



Fragmento de *Verbal Hygiene* de Deborah Cameron (p. 229)

- 35 El número 6, aquí en superíndice, significa aquí que la autora ha ofrecido un comentario adicional, que aparece debidamente recogido unas dieciocho páginas después:

Figura 9



Fragmento de *Verbal Hygiene* de Deborah Cameron (p. 247)

- 36 Un modo simple de codificar el texto principal podría ser el siguiente:

```
<div>
<head>Beyond "anything goes"</head>
<p>
  Why does the language-maven in the street
  (or the senior common-room, or the bar at the Groucho
  Club<note>An establishment patronized by media folk
  in London (provided the club will have them as
  members).</note>) have such a low opinion of
  linguists? Because ...
</p>
</div>
```

- 37 Aquí la nota se codifica incrustada dentro del texto, aunque en la fuente impresa se hayan separado las notas y el texto. Esto facilita la presentación de la nota de diversas maneras (por ejemplo, como una ventana emergente en una versión en línea) y también simplifica su codificación. Sin embargo, también es perfectamente posible tratar las notas como una sección separada del documento y colocar enlaces explícitos en el texto en el punto de anclaje, utilizando el elemento <ptr>:

```

<div>
  <head>Beyond "anything goes"</head>
  <p>
    Why does the language-maven in the street
    (or the senior common-room, or the bar at the Groucho
    Club <ptr target="#nota6"/>) have such a low opinion of
    linguists? Because ...
  </p>
</div>
<back>
  <head>Notes</head>
  <!-- otras notas aquí -->
  <note xml:id="nota6">
    An establishment patronized by media folk in London
    (provided the club will have them as members).
  </note>
  <!-- y aquí -->
</back>

```

- 38 Para representar el vínculo entre el texto y la nota, primero damos al elemento `<note>` un identificador único (`nota6`) y luego colocamos un elemento `<ptr>` en el punto del texto desde el que queremos referirnos a ella (el “punto de anclaje”). Este elemento puntero representa a la vez el 6 en superíndice en nuestra fuente original e indica que su función es apuntar a la nota cuyo identificador es suministrado por su atributo `@target` (destino). Alternativamente, si queremos registrar exactamente la forma como aparece el punto de anclaje en el texto, podríamos utilizar el elemento `<ref>`:

```

...
the Groucho Club
<ref target="#nota6">6</ref> have
...

```

Anotación lingüística

- 39 Durante la preparación de un texto codificado, es natural que el codificador tienda a querer añadir sus propias anotaciones. En cierto sentido, por supuesto, cada pieza de anotación añadida a un texto representa el resultado de un análisis, ya sea humano o automático, y por ello es natural pensar en representar dichas anotaciones de forma gradual mediante anotaciones a un texto digital. El elemento `<note>` puede utilizarse con este fin. Sin embargo, para muchas personas, los análisis con fines lingüísticos (como *esto es un verbo* o *esta secuencia de palabras tiene esta función sintáctica*, o incluso *esto es una metáfora*) son específicamente diferentes de las anotaciones discutidas en otras partes de este libro (como *esto es un título*, *esto es un nombre personal*, etc.). En particular, hay menos consenso sobre las categorías implicadas: la mayoría de la gente estará fácilmente de acuerdo en que “topónimo” es un concepto útil y definible, pero los conceptos que subyacen al análisis lingüístico son mucho más variados. Por ello, TEI **no** ofrece etiquetas como `<name>` (sustantivo) o `<adjunct>` (complemento circunstancial) o

<metaphor> (metáfora), sino que propone elementos de uso genérico para segmentar los textos en unidades más pequeñas, cualquiera de las cuales puede llevar etiquetas definidas por el proyecto.

- 40 Este método es particularmente útil cuando el texto digital se analiza automáticamente por medio de software (por ejemplo, por un programa analizador morfosintáctico o un etiquetador de categorías gramaticales), ya que el formato utilizado para las salidas de dichos sistemas automatizados difiere considerablemente. TEI ofrece una forma sencilla de segmentar explícitamente el texto en cuestión en unidades similares a palabras y de asociar códigos lingüísticos a cada uno de esos segmentos. Por ejemplo, al igual que muchos otros corpus lingüísticos, el BNC proporciona códigos que identifican la categoría gramatical (sustantivo, adjetivo, verbo, etc.) de cada palabra, así como su forma base o no flexionada. Por lo tanto, podríamos representar la oración citada anteriormente en TEI de la siguiente manera:

```
<u who="PS6NY">
  <s n="12">
    <w ana="#VM0" lemma="lets">Lets </w>
    <w ana="#VVI" lemma="ask">ask </w>
    <w ana="#DPS" lemma="you">your </w>
    <w ana="#NN1" lemma="mum">Mum </w>
    <w ana="#CJS" lemma="if">if </w>
    <w ana="#EX0" lemma="there">there</w>
    <w ana="#VBZ" lemma="be">'s</w>
    <w ana="#DT0" lemma="any">any </w>
    <w ana="#NN2" lemma="corner">corners </w>
    <w ana="#PRP" lemma="on">on </w>
    <w ana="#AT0" lemma="a">a </w>
    <w ana="#NN1" lemma="boat">boat</w>
  <pc>. </pc>
</s>
</u>
```

- 41 En esta codificación (producida en gran medida de manera automatizada, por supuesto) se ha utilizado el elemento <w> (word, palabra) para delimitar cada ítem léxico. En cada unidad de este tipo, se utiliza el atributo @ana (analysis, análisis) para apuntar a una definición de categoría gramatical, y el atributo @lemma (lema) para proporcionar una forma base.
- 42 Los análisis lingüísticos de este tipo se utilizan desde hace muchas décadas y son esenciales para la mayoría de los sistemas actuales utilizados en el Procesamiento del Lenguaje Natural (NLP, por sus siglas en inglés, *Natural Language Processing*). En consecuencia, tanto las técnicas utilizadas como las categorías representadas están en proceso de normalización, al menos para las principales lenguas de Europa occidental. El enfoque de TEI permite al codificador elegir cuán mucho o cuán poco desea explicitar el significado del análisis señalado por el atributo @ana: es decir, cómo definir lo que significa realmente VBZ o NN1.
- 43 Un método consiste en proporcionar una definición adicional para la categoría, normalmente como parte de la descripción de la codificación en el encabezado TEI, utilizando un elemento <taxonomy> (taxonomía). Se trata de una definición

exclusivamente significativa para un ser humano, pero aún así puede ser muy útil proporcionarla.

- 44 En el otro extremo, las categorías pueden definirse formalmente utilizando las recomendaciones conjuntas TEI/ISO para representar los análisis lingüísticos en términos de *estructuras de rasgos* (véase además el capítulo 18 de las *Directrices TEI*). Este tipo de análisis se utiliza actualmente de forma bastante generalizada en la comunidad del NLP, por ejemplo como medio para mapear la información entre diferentes léxicos computacionales.
- 45 Una solución intermedia consiste en realizar la categorización lingüística con referencia a una norma existente, como la que proporciona el Registro de Categorías de Datos de la ISO (véase Ide y Romary 2009 para mayor información). Este enfoque comparte muchos rasgos con otros intentos de armonización de datos en la web: simplifica la interoperabilidad de los sistemas, del mismo modo que los documentos TEI en general pueden ser interoperables.
- 46 También cabe señalar que los mismos mecanismos pueden utilizarse para cualquier tipo de análisis, no necesariamente lingüístico; los análisis estilísticos (que suelen caracterizar fragmentos de texto superiores a una sola palabra) también pueden codificarse de forma muy parecida, adjuntando el atributo @ana a cualquier elemento que se elija para encerrar dicho tramo de texto, por ejemplo un <p>, <div>, <s>, etc.
- 47 Por supuesto, las unidades de texto que se analizan no coinciden necesariamente con las unidades de texto directamente representadas dentro del texto. En la sección *Hitos* del capítulo anterior nos hemos referido a una forma de abordar el “problema de solapamiento”, bastante común. Otra es utilizar un “marcado separado” (*stand off*), en el cual las unidades analizadas se representan no mediante elementos XML sino mediante punteros. En el capítulo 20 de las *Directrices TEI* se analizan estas y otras técnicas con más detalle.

El encabezado TEI

- 1 Todo documento TEI debe tener un encabezado TEI, representado por un elemento <teiHeader>. Se trata de un contenedor para todos los metadatos asociados al documento digital mismo, análogo a la portada de un libro impreso. En las bibliotecas digitales y otros repositorios en línea, es habitual que los metadatos asociados a cada documento digital se almacenen o gestionen por separado, por ejemplo, en una estructura de base de datos, por razones de eficiencia; también es una práctica común exponer un subconjunto de dichos metadatos a los motores de búsqueda de la web, de acuerdo con uno o más estándares comunes como Dublin Core. El elemento de encabezado TEI se proporciona como una forma de almacenar toda esa información en un solo lugar, independientemente de cómo se utilice. Por supuesto, el alcance y la cantidad de datos recogidos en el encabezado pueden variar considerablemente en los documentos elaborados para diferentes fines o por diferentes proyectos. Además, al menos tal como se concibió originalmente, el contenido de un encabezado TEI no tiene por qué ajustarse a las normas de catalogación más estrictas o precisas. No obstante, cumple un propósito útil como lugar donde, por ejemplo, el creador y el conservador de datos pueden compartir información.
- 2 Originalmente, el encabezado TEI se diseñó para satisfacer dos requisitos bastante diferentes. Por un lado, buscaba ayudar a los bibliógrafos y bibliotecarios que se enfrentaban al (entonces) nuevo problema de documentar los “libros electrónicos”; por otro, buscaba satisfacer las necesidades de los investigadores que trabajaban con colecciones de textos digitales y necesitaban documentar las “prácticas de codificación” aplicadas a ellos. Para el investigador, lo importante del encabezado es que ofrece un lugar para todo, sin dejar de ser compatible con toda la gama de prácticas divergentes que existen en las diferentes comunidades de investigación, sin imponer barreras como un conocimiento detallado de las normas o prácticas de catalogación. Para el bibliotecario, lo importante del encabezado TEI es que se ajusta a los modelos bibliográficos estándar, pues utiliza una terminología similar, y que proporciona una única fuente de información para la descripción bibliográfica de un recurso digital, con algunas correspondencias establecidas con otros registros de este tipo (como MARC o EAD). Existe —naturalmente— cierta tensión entre estas dos perspectivas, así como una necesidad de elaborar perfiles detallados o “Guías de buenas prácticas” para las diferentes comunidades.

- 3 El encabezado TEI tiene cuatro componentes principales, que corresponden a las partes definidas en uno de los primeros intentos de descripción bibliográfica universal, la Descripción Bibliográfica Internacional Normalizada (ISBD, por sus siglas en inglés, *International Bibliographic Standard Description*):
- <fileDesc> (*file description*, descripción del archivo) contiene una descripción bibliográfica completa de un archivo electrónico;
 - <encodingDesc> (*encoding description*, descripción de la codificación) documenta la relación entre un texto electrónico y la fuente o fuentes de las que se ha derivado;
 - <profileDesc> (*profile description*, descripción del perfil) proporciona una descripción detallada de los aspectos no bibliográficos de un texto, concretamente los idiomas y subtipos de idiomas utilizados, el contexto en que se produjo, los participantes y su entorno (casi todo lo que no está contemplado en los demás elementos del encabezado);
 - <revisionDesc> (*revision description*, descripción de la revisión) resume el historial de revisiones de un archivo.

Descripción del archivo

- 4 Puesto que no suele tratarse de un documento preexistente, TEI puede dictar de forma bastante precisa la estructura de un encabezado TEI. La primera y única parte obligatoria, conocida como la descripción del archivo, está representada por un elemento <fileDesc>, que contiene tres partes obligatorias: declaración del título, declaración de la publicación y descripción de la fuente. Así, un encabezado TEI mínimo tiene un aspecto similar al siguiente:

```
<teiHeader>
<fileDesc>
  <titleStmt>
    <title>Título de la obra</title>
  </titleStmt>
  <publicationStmt>
    <p>Información sobre la publicación de la obra</p>
  </publicationStmt>
  <sourceDesc>
    <p>Información sobre la fuente de la que procede la obra</p>
  </sourceDesc>
</fileDesc>
</teiHeader>
```

- 5 Dentro de las cuatro secciones principales del encabezado TEI caben muchos elementos; aquí solo podemos ofrecer una visión superficial. Por ejemplo, el elemento <fileDesc> también puede contener un <editionStmt> que documente la edición particular del recurso que se describe: un <seriesStmt> si se publica como parte de una serie, un elemento <extent> para indicar su extensión, e incluso un <notesStmt> para contener notas o comentarios de todo tipo.
- 6 El enunciado obligatorio del título, como su nombre indica, proporciona un título para el recurso, junto con información sobre las personas o los organismos responsables de su contenido intelectual, de forma muy parecida al título de un libro impreso, o a un registro de catálogo convencional. Por ejemplo:

```

<titleStmt>
  <title xml:lang="sk">Yogadarśanam (arthāt yogasūtrapūphah)</title>
  <title>Los Yoga sūtras de Patañjali: una edición digital</title>
  <author>Patañjali</author>
  <funder>Wellcome Institute for the History of Medicine</funder>
  <principal>Dominik Wujastyk</principal>
  <respStmt>
    <name>Wieslaw Mical</name>
    <resp>ingreso de datos y corrección de pruebas</resp>
  </respStmt>
  <respStmt>
    <name>Jan Hajic</name>
    <resp>conversión a marcado TEI</resp>
  </respStmt>
</titleStmt>

```

- 7 En este ejemplo, el recurso tiene dos títulos, uno de ellos en sánscrito. También se indican los datos de su autor original, el financiador de la edición digital, el investigador principal y otras personas con responsabilidades más especializadas. Muchos proyectos de bibliotecas digitales tendrán sus propias normas sobre, por ejemplo, el formato de los nombres, o el uso de listas de autoridades para proporcionar los títulos, que pueden variar considerablemente: TEI permite, por lo tanto, un etiquetado más exacto de componentes como los nombres propios (por ejemplo, distinguiendo nombres y apellidos), aunque no lo impone.
- 8 El elemento <fileDesc> describe el *archivo* (es decir, el conjunto de un recurso digital; por supuesto, puede estar formado por varios archivos del sistema operativo). La *declaración de publicación* (*publication statement*) describe cómo se puede obtener este recurso, de la misma manera que el pie de imprenta de un libro impreso sugiere quién es el responsable de distribuirlo o publicarlo. Incluso si el documento TEI es un documento privado que en realidad no está disponible en ningún lugar fuera del disco duro privado del propietario, ese hecho debe mencionarse aquí; lo más habitual, por supuesto, es que se cree un encabezado solo para un recurso TEI que se está compartiendo. La declaración de publicación permite especificar quién pone a disposición el recurso, cualquier identificador que se utilice para él, como un número de catálogo o un identificador de recurso, e información sobre los términos bajo los que se distribuye, como una licencia, como se muestra aquí:

```

<publicationStmt>
  <publisher>Humanities Media and Computing center</publisher>
  <pubPlace>Universidad de Victoria</pubPlace>
  <date when="2011-08-04">4 de agosto de 2011</date>
  <idno type="nombre-de-archivo">maladies_des_femmes.xml</idno>
  <availability status="free">
    <p>
      Copyright 2011. Este texto es de libre acceso siempre y cuando se
      distribuya con la información del encabezado proporcionada.
    </p>
    <p xml:lang="fr">
      Les droits de reproduction des gravures ont été achetés de la
      Bibliothèque Nationale de France grâce à une subvention accordée par le
      <ref target="http://www.sshrc-crsh.gc.ca/">
        Conseil de recherches en sciences humaines du Canada
      </ref>.
      Les autres éléments du projet (les contributions des éditeurs, les
      transcriptions des textes, l'encodage et le code) sont distribués sous
      les termes de cette licence:
      <ref
        target="http://creativecommons.org/licenses/by-nc-nd/2.5/ca/deed.fr_CA">
        Creative Commons Paternité - Pas d'Utilisation Commerciale -
        Pas de Modification 2.5 Canada
      </ref>.
    </p>
  </availability>
</publicationStmt>

```

- 9 Este ejemplo muestra una mezcla típica de elementos estructurados y prosa informal. El nombre y la dirección de la institución responsable de la distribución del archivo se han distinguido como `<editor>` y `<pubPlace>` (lugar de publicación) respectivamente —claro que también podría utilizarse una dirección más detallada por medio del elemento `<address>` que vimos anteriormente—; además se ha proporcionado una fecha de publicación y un nombre de archivo para identificar el recurso. Dado que las condiciones de disponibilidad aquí son un poco complejas, se han presentado de manera informal como una secuencia de párrafos en prosa, que contienen referencias a otros documentos en línea según corresponda. En una situación más sencilla, se podría haber utilizado el elemento TEI `<licence>`.

Variedad de descripciones de la fuente

- 10 El siguiente y único otro componente obligatorio de `<fileDesc>` es la descripción de la fuente, representada por el elemento `<sourceDesc>`. Incluso en el caso de los documentos *nativos digitales* y que, en consecuencia, no tienen una fuente preexistente, debe proporcionarse este elemento. Su propósito es documentar formalmente el objeto u objetos de los que se ha derivado el documento TEI, utilizando términos bibliográficos tradicionales. El `<sourceDesc>` puede contener un simple párrafo de descripción, o

puede contener uno o más de los elementos especializados para la descripción bibliográfica que proporciona TEI.

- 11 En un primer ejemplo, consideremos un documento que no tiene otra forma de existencia fuera de su versión digital. Podemos representarlo utilizando un <bibl> reducido:

```
<sourceDesc>
  <bibl>
    <title>Manifeste des Digital humanities</title>.
    <author>Marin Dacos et al.</author>
    Disponible en
    <ref target="http://tcp.hypotheses.org/318">
      http://tcp.hypotheses.org/318
    </ref>
    <date when="2010-05-21"/>
  </bibl>
</sourceDesc>
```

o simplemente ofrecer un breve párrafo de explicación:

```
<sourceDesc>
  <p>Ninguna fuente: este es un documento nativo digital.</p>
</sourceDesc>
```

- 12 Suele suceder que una edición digital proceda de una fuente impresa específica. Esto también puede describirse por medio del elemento <biblStruct>:

```
<sourceDesc>
  <bibl xml:id="Sue1846">
    <author>
      <surname>Sue</surname>, <forename>Eugène</forename>
    </author>
    <title level="m">
      Martin, l'enfant trouvé :
      Mémoires d'un valet de chambre
    </title>
    <publisher>C. Muquardt</publisher>
    <pubPlace>Bruxelles</pubPlace>
    <pubPlace>Leipzig</pubPlace>
    <date when="1846">MDCCCXLVI</date>
  </bibl>
</sourceDesc>
```

- 13 Para las transcripciones digitales de grabaciones de audio o vídeo, debe utilizarse un elemento especializado <recordingStmt> (declaración de registro). Puede utilizarse para almacenar cualquier detalle técnico de la propia grabación y también para almacenar datos biográficos o sociológicos detallados sobre los participantes en un diálogo transcrito.

```

<sourceDesc>
  <recordingStmt>
    <recording type="audio" dur="P30M">
      <respStmt>
        <resp>Grabación del lugar por</resp>
        <name>Sound Services Ltd.</name>
      </respStmt>
      <equipment>
        <p>
          Múltiples micrófonos de proximidad mezclados en cinta de audio
digital
          estéreo, reproducción estándar, frecuencia de muestreo de 44,1 KHz
        </p>
      </equipment>
      <date>12 de enero de 1987</date>
    </recording>
  </recordingStmt>
</sourceDesc>

```

- 14 Para otras fuentes que no sean libros, como la tarjeta postal mencionada anteriormente, también se puede utilizar el elemento <bibl>:

```

<sourceDesc>
  <bibl>
    <title level="m">
      The Bathing Beach, Brighton, en 1845 [tarjeta postal]
    </title>
    <respStmt>
      <resp>Litografía por</resp>
      <name>G. F. Bragg</name>
    </respStmt>
    <respStmt>
      <resp>basada en un dibujo de</resp>
      <name>R. H. Nibbs</name>
    </respStmt>
    <publisher>Librería de K. J. Bredon</publisher>
    <pubPlace>10 East Street, Brighton</pubPlace>
  </bibl>
</sourceDesc>

```

- 15 En el caso de una edición digital de un manuscrito o de una fuente impresa antigua, donde es importante registrar las características del ejemplar concreto que estamos transcribiendo, los elementos detallados propuestos por TEI para la catalogación descriptiva de los manuscritos pueden ser más apropiados. Tales descripciones utilizan un elemento distinto, el <msDesc> (descripción de manuscrito), que tiene una estructura interna bastante elaborada, como muestra el siguiente ejemplo:

```

<sourceDesc>
  <msDesc>
    <msIdentifier>
      <country>Francia</country>
      <settlement>París</settlement>
      <repository>Archives nationales</repository>
      <collection>Commerce et Industrie</collection>
      <idno>F/12/5080</idno>
    </msIdentifier>
    <msContents>
      <p>
        Minuta de un informe de propuesta de concesión de la Legión de honor
        hecha, en 1850, por el ministro de Comercio y Agricultura y presidente
        de la Sociedad de Geografía, Jean-Baptiste Duma, al presidente
        de la República, a favor de los hermanos d'Abbadie, Antoine
        (1810-1897)
        y Arnadu (1815-1893), autores de un viaje a Abisinia.
      </p>
    </msContents>
    <physDesc>
      <p>Dos hojas de papel 24 x 12 cm; escrito con tinta negra.</p>
    <handDesc>
      <handNote xml:id="AA" scope="major">Antoine d'Abbadie</handNote>
      <handNote xml:id="DJB" scope="minor">Jean-Baptiste Dumas</handNote>
      <handNote xml:id="EPR" scope="minor">miembro desconocido del gabinete
        del ministro</handNote>
    </handDesc>
    </physDesc>
  </msDesc>
</sourceDesc>

```

Descripción de la codificación

- 16 La segunda gran división del encabezado TEI es la *descripción de la codificación*, que está representada por un elemento <encodingDesc>. Este elemento opcional puede utilizarse para suministrar información sobre casi cualquier aspecto del proceso de codificación mismo, ya sea simplemente resumido como texto de corrido, o suministrado dentro de elementos más específicos. Considerado en conjunto, el contenido de un <encodingDesc> completo se aproxima al tipo de información que suele encontrarse en un manual técnico asociado a un proyecto.
- 17 Algunos de sus componentes son puramente documentales, por ejemplo:
- <projectDesc>: notas sobre los objetivos generales del proyecto;
 - <samplingDecl>: notas sobre los principios de muestreo aplicados;
 - <editorialDecl>: notas sobre los principios editoriales aplicados, para los que se proporcionan otros elementos especializados como <correction> (corrección), <normalization> (normalización), <quotation> (cita), <hyphenation> (partición de palabras o silabación), <segmentation> (segmentación) o <interpretation> (interpretación).

- 18 En el siguiente ejemplo podemos ver el uso de algunos de estos elementos:

```

<encodingDesc>
  <projectDesc>
    <p>
      Textos recopilados para su uso en la Clínica Shakespeare de Claremont,
      junio de 1990
    </p>
  </projectDesc>
  <samplingDecl>
    <p>
      Cada texto contiene una muestra de hasta 2000 palabras, que va desde
      el inicio del documento hasta el final de la oración después de la
      marca de 2000 palabras. A efectos del recuento de palabras, los
      guiones y apóstrofes se trataron como espacios.
    </p>
  </samplingDecl>
  <editorialDecl>
    <normalization>
      <p>
        Las formas de las palabras partidas por el guion de final de línea
        se han reconstruido sin comentario. Se ha eliminado el guion, excepto
        en las
        formas con guion atestiguadas en otra parte del texto
      </p>
    </normalization>
    <quotation marks="all">
      <p>
        Se han eliminado todas las comillas. El discurso directo se
        representa mediante el uso de la etiqueta <gi>said</gi>;
        el resto del material citado se representa mediante
        la etiqueta <gi>q</gi>.
      </p>
    </quotation>
  </editorialDecl>
</encodingDesc>

```

- 19 Por supuesto, tal documentación solo es útil para un lector humano. Sin embargo, otros componentes de la descripción de la codificación están destinados a ser utilizados por procesos automatizados. Por lo general, proporcionan un conjunto de declaraciones para códigos particulares a los que posteriormente se hace referencia o se utilizan en el cuerpo del texto. Algunos ejemplos son:

- <charDecl>: XML requiere el uso de Unicode en todo el documento; sin embargo, los glifos o caracteres no estándar pueden indicarse en un documento TEI utilizando el elemento <g> para referenciar las definiciones de las cosas proporcionadas por este elemento;
- <classDecl>: cualquier tipo de sistema de clasificación o taxonomía puede declararse utilizando este elemento; las partes de un texto pueden entonces indicar el código o códigos de clasificación asociados a ellas utilizando (por ejemplo) el atributo @ana. En un corpus TEI, este elemento suele ser proporcionado por el encabezado TEI para que los encabezados de texto individuales puedan apuntar a él utilizando el elemento <catRef>.

- <refsDecl>, <geoDecl>, <metDecl>, <fsdDecl>, <variantEncoding>: proporcionan un lugar para la información relativa a la codificación de los sistemas de referencia, la información geográfica, los sistemas de análisis métrico, los sistemas de características para los análisis lingüísticos o la codificación de la variación textual.

- 20 A continuación daremos ejemplos más detallados de cada una de estas tres categorías.
- 21 En primer lugar, un elemento <charDecl> típico podría definir una forma variante del carácter Z que el codificador desea distinguir en una transcripción. Esta variante tiene dos trazos, pero puede ser sustituida por una Z normal.

```
<charDecl>
  <glyph xml:id="z103">
    <glyphName>LETRA LATINA Z CON DOS TRAZOS</glyphName>
    <mapping type="estandarizado">z</mapping>
    <mapping type="PUA">U+E304</mapping>
  </glyph>
</charDecl>
```

- 22 Las ocurrencias de esta forma de variante en la transcripción pueden distinguirse ahora utilizando un elemento <g> que haga referencia a la definición anterior mediante su valor @xml:id, como es habitual. Un procesador puede optar por representarlo utilizando la forma estandarizada correspondiente o el código no estandarizado dado por el mapeo PUA.¹

```
<p> ... mulct<g ref="#z103"/> ... </p>
```

- 23 A continuación consideramos el elemento <classDecl>. Este puede utilizarse para definir cualquier tipo de sistema de clasificación privado o taxonomía. Para una colección de artículos periodísticos podríamos utilizar una taxonomía como la siguiente:

```

<classDecl>
  <taxonomy xml:id="tamaño">
    <category xml:id="grande">
      <catDesc>la historia ocupa más de media página</catDesc>
    </category>
    <category xml:id="medio">
      <catDesc>la historia ocupa entre un cuarto y media página</catDesc>
    </category>
    <category xml:id="pequeño">
      <catDesc>la historia ocupa menos de un cuarto de página</catDesc>
    </category>
    <!-- etc. -->
  </taxonomy>
  <taxonomy xml:id="tópico">
    <category xml:id="pol-int">
      <catDesc>alude a eventos políticos nacionales</catDesc>
    </category>
    <category xml:id="pol-ext">
      <catDesc>alude a eventos políticos internacionales</catDesc>
    </category>
    <category xml:id="soc-mujer">
      <catDesc>alude al rol de la mujer en la sociedad</catDesc>
    </category>
    <category xml:id="soc-sirv">
      <catDesc>alude al rol de los sirvientes en la sociedad</catDesc>
    </category>
    <!-- etc. -->
  </taxonomy>
</classDecl>

```

- 24 Un texto individual que contenga (digamos) una historia de menos de un cuarto de página relativa al papel de la mujer en la sociedad haría entonces referencia a esta clasificación utilizando el elemento `<catRef>` de la siguiente manera:

```
<catRef target="#pequeño #soc-mujer"/>
```

- 25 El mismo mecanismo puede utilizarse para documentar los códigos utilizados en análisis lingüísticos sencillos como los que se han tratado en el capítulo anterior.
- 26 Por último, consideremos el elemento `<tagsDecl>`. Este puede utilizarse únicamente para enumerar los elementos realmente utilizados dentro de un documento, y también para definir los estilos de formato por defecto para ellos (normalmente utilizando CSS, el lenguaje de *Hojas de estilo en cascada* del W3C). En el siguiente ejemplo, primero definimos dos estilos de letra —cursiva y redonda—, utilizando CSS. A continuación, declaramos que dentro del espacio de nombres TEI cualquier cosa etiquetada como `<emph>` o `<hi>` está por defecto en cursivas (`italic`). También declaramos que el elemento `<text>`, y por defecto todo lo que contiene, utiliza la fuente en redondas (`normal`).

```

<tagsDecl>
  <rendition xml:id="IT" scheme="css">font-style: italic</rendition>
  <rendition xml:id="FontRoman" scheme="css">font-family: serif</rendition>
  <namespace name="http://www.tei-c.org/ns/1.0">
    <tagUsage gi="emph" rendition="#IT"/>
    <tagUsage gi="hi" rendition="#IT"/>
    <tagUsage gi="text" rendition="#FontRoman"/>
  </namespace>
</tagsDecl>

```

- 27 Dentro del cuerpo de un documento, el atributo @rendition puede sobrescribir el estilo por defecto de un elemento.

Descripciones del perfil y de la revisión

- 28 La tercera gran división del encabezado TEI es la extrañamente denominada *Descripción del perfil*, representada por un elemento <profileDesc>. Al igual que los demás, se trata de un grupo de notas opcionales o de elementos más especializados; tienen en común simplemente que son “no bibliográficos”. Los miembros por defecto de la clase `model.profileDescPart` incluyen:
- <creation>: información sobre el origen del contenido intelectual del texto, p. ej. tiempo y lugar. En una edición genética, esto puede incluir un elemento estructurado <listChange> (lista de cambios) que documente cada etapa significativa identificada en la evolución de un texto;
 - <langUsage>: información sobre los idiomas, registros, sistemas de escritura, etc., utilizados en el texto; cada idioma utilizado se identifica mediante un elemento <language> y un código de identificación tomado del estándar ISO;
 - <textDesc> y <textClass>: clasificaciones aplicadas al texto mediante una lista de criterios especificados o mediante una colección de punteros, respectivamente;
 - <particDesc> y <settingDesc>: información sobre los “participantes”, reales o representados, en el texto.
- 29 Ya hemos señalado la disponibilidad del elemento <catRef> como medio para clasificar un texto con respecto a una taxonomía predefinida. El elemento <profileDesc> ofrece varias formas complementarias de hacerlo:
- con <catRef>: para hacer referencia a una categoría definida localmente (por ejemplo, en la cabecera del corpus);
 - con <classcode>: para hacer referencia a alguna categoría comúnmente acordada y definida externamente, como el sistema de Clasificación Decimal Universal;
 - con <keywords>: mediante la asignación de palabras clave o términos descriptivos tomados de un vocabulario bibliográfico controlado o de una nube de etiquetas.
- 30 En el siguiente ejemplo, tomado del British National Corpus, un texto se clasifica utilizando los tres métodos posibles:

```

<profileDesc>
  <creation>
    <date when="1962"/>
  </creation>
  <textClass>
    <catRef target="#WRI #ALLTIM1 #ALLAVA2 #ALLTYP3 #WRIDOM5 #WRILEV2
      #WRIMED1 #WRIPP5 #WRISAM3 #WRISTA2 #WRITAS0"/>
    <classCode scheme="DLEE">W nonAc: humanities arts</classCode>
    <keywords scheme="COPAC">
      <hi rendition="#emph">Historia, Moderna - siglo XIX</hi>
      <hi rendition="#emph">Capitalismo, Historia - siglo XIX</hi>
      <hi rendition="#emph">Mundo, 1848-1875</hi>
    </keywords>
  </textClass>
</profileDesc>

```

- 31 Nótese que esta clasificación se aplica a todo el texto. Para una clasificación más fina, el atributo `@decls` puede utilizarse para seleccionar la clasificación aplicable a cualquier elemento declarable, por ejemplo a un `<div>` individual.
- 32 La cuarta y última parte del encabezado TEI es la *descripción de las revisiones*. Esta parte opcional y está representada por un elemento `<revisionDesc>` que contiene una lista de elementos `<change>` (cambio), cada uno con los atributos `@date` y `@who`, y cada uno de los cuales indica etapas significativas en la evolución de un documento; por convención, el elemento más reciente se da primero. El elemento `<listChange>`, mencionado anteriormente, también puede utilizarse aquí para referirse a etapas identificadas en la evolución del archivo electrónico, a diferencia del texto codificado. En un entorno de producción se utilizará un sistema automatizado de control de versiones, como Git, para hacer un seguimiento detallado de la evolución de un documento; la codificación TEI de las etapas significativas en el desarrollo de un documento puede realizarse de forma semiautomática por medio de dichas herramientas, o manualmente.

```

<revisionDesc>
  <listChange>
    <change when="2013-05-11">Primer borrador completo</change>
    <change when="2013-04-07">
      Encabezado y estructura del documento creados
    </change>
  </listChange>
</revisionDesc>

```

NOTAS

1. *Private Use Area*, “Área de Uso Privado”: un concepto de Unicode que permite la definición de códigos privados para caracteres no intercambiables.

Personalización de TEI

- 1 Como hemos visto, TEI está diseñado para soportar una gama muy amplia de opciones de codificación. Puede utilizarse para una simple transcripción orientada a la lectura de una fuente primaria, ya sea un manuscrito de autor, una obra literaria impresa, una emisión de audio o un diccionario. Puede utilizarse para codificaciones enriquecidas en las que se explicitan muchos aspectos de dichos textos, de modo que programas informáticos de todo tipo puedan operar sobre ellos, desde herramientas de visualización y sistemas de publicación digital hasta paquetes especializados de análisis estadístico. Puede utilizarse para proporcionar anotaciones complementarias y metadatos de todo tipo. Casi nadie necesita todo lo que define TEI, pero cada uno de sus elementos es de utilidad o interés para alguien. ¿Cómo deberíamos elegir solo las partes de TEI que necesitamos? Una de las principales motivaciones iniciales para crear documentos TEI es la posibilidad de compartirlos con otros e integrarlos con otros documentos TEI. ¿Cómo debemos comunicar a los demás las elecciones particulares de codificación TEI que hemos hecho, de manera que dicha integración siga siendo posible?
- 2 TEI ofrece una forma de abordar estas preocupaciones, así como de satisfacer la importante necesidad de una documentación detallada específica del proyecto, al proporcionar un conjunto de elementos que pueden utilizarse tanto para especificar un *esquema* en términos de nombres y propiedades formales de los elementos y atributos que contiene, como para documentar cómo se utilizan esos elementos y atributos en una aplicación determinada.
- 3 La noción de un esquema es fundamental para XML: proporciona una especie de gramática del documento, nombrando los posibles componentes y restringiendo la organización de un documento XML. Un esquema permite expresar restricciones como *los elementos <p> pueden aparecer dentro de elementos <div>, o cada elemento <list> debe contener al menos un <item>*. Un procesador automático (un validador) puede verificar fácilmente este tipo de reglas. Restricciones como *utilizar el elemento <p> para contener párrafos, no páginas, o utilizar <placeName> para topónimos, <persName> para nombres de personas y <name> para cualquier otro nombre* son mucho más difíciles de comprobar automáticamente, ya que se refieren a la semántica del contenido más que a su organización. Estas reglas se definen en documentos como las *Directrices TEI*, o en documentos que hacen referencia a ellas. Comprender estas restricciones semánticas es

extremadamente importante para quienes desarrollan software destinado a aprovechar al máximo el documento marcado, así como para quienes desean crear nuevos documentos codificados de la misma manera. Esto es especialmente cierto si el esquema en uso permite una gran variedad de elementos con significados muy similares, como será el caso de un esquema TEI no modificado y no personalizado.

- 4 Por esta razón, un uso serio de TEI requiere plantearse cuidadosamente qué elementos son los más apropiados para las necesidades de un proyecto, y también es probable que sea algo que el proyecto necesita especificar con más exactitud que TEI. Por ejemplo, TEI no establece requisitos obligatorios para los posibles valores del atributo `@type` utilizado en los elementos `<div>`, ya que es probable que estos varíen mucho entre los distintos proyectos. Sin embargo, en un proyecto determinado, es probable que la estandarización de un conjunto de valores acordados sea muy útil y, en consecuencia, será necesario tanto garantizar que todos los documentos utilicen solo el conjunto de valores acordado, como asegurar que la documentación sobre cuáles son esos valores y qué significan se mantenga junto con el resto del esquema, y que sea fácilmente accesible para los sistemas de preparación y edición de documentos con conocimiento de XML (como el programa ampliamente utilizado oXygen) para ayudar y guiar al editor humano de documentos.
- 5 TEI proporciona un conjunto especial de elementos que pueden utilizarse para crear una especificación de esquema de este tipo. Estos elementos (`<schemaSpec>`, `<moduleRef>`, `<elementSpec>`, `<classSpec>` y otros) combinan declaraciones formales XML para incluirlo en un DTD o un *esquema* con documentación detallada y ejemplos, para su inclusión en un manual técnico sobre el esquema de codificación que se está especificando. Por esta razón, un documento que utiliza estos elementos se conoce informalmente como un *ODD* (por *One Document Does it all*, *Un documento lo hace todo*); proporciona información para que la procese un computador junto con la documentación de dicha información para que la lea un ser humano en un único documento XML integrado. Como es lógico, el propio sistema TEI se expresa utilizando exactamente este mismo conjunto de elementos, aunque aquí nos centramos en su uso en la creación de una personalización TEI.

Construcción de una personalización

- 6 La personalización TEI más sencilla es la personalización nula, que simplemente dice: *permita todos los elementos definidos por TEI*. El esquema resultante, llamado `tei_all`, proporciona (en el momento de escribir este artículo) unos 450 elementos diferentes, y muchísimas formas diferentes de resolver problemas de codificación que a menudo se solapan. El tamaño y la permisividad de este esquema lo hacen más o menos inutilizable para fines prácticos, salvo uno muy importante relacionado con la *conformidad* con TEI, al que volveremos más adelante.
- 7 Probablemente la personalización TEI más utilizada y a la que más se hace referencia es `tei_lite`. TEI Lite es un subconjunto de unos cincuenta elementos que pretende satisfacer las necesidades del 90% de los usuarios de TEI, como evidencia su práctica real en la creación de textos digitales. Originalmente elaborado para un taller de formación en TEI en 1997, después de algunas actualizaciones continúa siendo útil en una amplia gama de contextos, hasta el punto de que algunas personas piensan que es

el esquema TEI. TEI Lite se incluye (como un Ejemplar) con la versión estándar de TEI, en formas de código fuente y derivadas.

- 8 Un rápido vistazo al código fuente XML del ODD de TEI Lite muestra que parece un documento TEI típico, con elementos <div> que contienen elementos <head>, <p> y <list>, con mucha prosa discursiva, así como elementos <ptr> para referencias cruzadas y algunos otros elementos especializados como <egXML> para ejemplos XML. El esquema real especificado por este documento está representado, hacia el final del documento, por un elemento <schemaSpec> que contiene declaraciones como las siguientes:

```
<schemaSpec ident="tei_lite" start="TEI teiCorpus">
  <moduleRef key="analysis" include="interp interpGrp pc s w"/>
  <moduleRef key="linking" include="anchor seg"/>
  <moduleRef key="tagdocs" include="att code eg gi ident val"/>
  <moduleRef key="tei"/>
  <moduleRef key="textstructure"
    include="TEI argument back body byline closer dateline div docAuthor
      docDate docEdition docImprint docTitle epigraph front group imprimatur
      opener postscript salute signed text titlePage titlePart trailer"/>
</schemaSpec>
```

- 9 El elemento <moduleRef> se utiliza para hacer referencia a un *módulo* TEI. Un módulo es simplemente un contenedor con nombre para una serie de declaraciones de elementos y clases TEI. TEI define actualmente 22 módulos, cada uno de los cuales corresponde a un capítulo de las *Directrices TEI* en el que se describe detalladamente su contenido. Por defecto, una referencia a un módulo implica que todas las declaraciones que este contiene deben ser transferidas al esquema que se especifica, aunque los atributos @include y @except pueden utilizarse para modificar este comportamiento por defecto. Así, en el ejemplo anterior, el esquema que se está creando incluirá todas las declaraciones encontradas en el módulo llamado `tei`, pero las declaraciones solo para los elementos <anchor> y <seg> del módulo llamado `linking`. Alternativamente, el atributo @except podría haberse utilizado para seleccionar todo de este módulo *con excepción de* los elementos nombrados como su valor.
- 10 Un módulo pone a disposición del <schemaSpec> una declaración para cada elemento seleccionado, en forma de un elemento <elementSpec>. Este tiene varios componentes, la mayoría de los cuales se visualizan en la documentación de referencia TEI. Incluye, por ejemplo:
- uno o varios elementos <desc> para proporcionar una breve descripción de la función del elemento, posiblemente en diferentes idiomas;
 - un elemento <classes> que contiene un elemento <memberOf> para cada clase de la que el elemento es miembro;
 - un elemento <attList>, que contiene una lista estructurada de todos los atributos definidos localmente para el elemento, junto con información sobre su tipo de datos y cualquier conjunto de valores predefinidos, expresados mediante los elementos <attDef>, <valList> y <valItem>, respectivamente;
 - un elemento <content> que proporciona una declaración formal del contenido legal del elemento, y opcionalmente elementos <constraint> utilizados para expresar cualquier restricción de uso adicional;

- uno o más elementos <exemplum>, cada uno de los cuales contiene un ejemplo de uso, en la medida de lo posible tomado de documentos reales, más un comentario.
- 11 El Apéndice C de las *Directrices TEI* contiene documentación en buen formato para cada elemento TEI definido sobre la base de dichas declaraciones.
 - 12 Además de seleccionar o excluir declaraciones, un <schemaSpec> puede modificar algunas partes de una declaración. Un poco más adelante, en las ODD para TEI Lite, encontramos lo siguiente:

```
<elementSpec ident="TEI" mode="change">
  <attList>
    <attDef ident="version" mode="delete"/>
  </attList>
</elementSpec>
```

cuyo efecto es proporcionar una segunda declaración para el elemento <TEI>, junto con la seleccionada del módulo `textstructure`. Un procesador ODD debe conciliar o unificar tales duplicaciones, bajo el control del atributo `@mode`. En este caso particular, el efecto es cambiar la <attList> declarada para el elemento <TEI>, borrando el atributo `@ident` de este.

- 13 Exactamente el mismo procedimiento se utiliza en el resto de la personalización de TEI Lite para eliminar de ella otros atributos no deseados. Por ejemplo, los atributos `@notBefore`, `@notAfter` y otros son proporcionados por la clase de atributos `att.dateable.w3c`. Por lo tanto, la forma más sencilla de eliminarlos es proporcionar una declaración adicional para la clase, como se indica a continuación:

```
<classSpec
  type="atts"
  ident="att.dateable.w3c"
  module="tei"
  mode="change">
  <attList>
    <attDef ident="notBefore" mode="delete"/>
    <attDef ident="notAfter" mode="delete"/>
    <attDef ident="from" mode="delete"/>
    <attDef ident="to" mode="delete"/>
  </attList>
</classSpec>
```

- 14 Una especificación ODD puede incluir o excluir elementos y atributos de estas y otras maneras. También puede modificar una declaración existente añadiéndole restricciones adicionales. Por ejemplo, consideremos la adaptación de TEI conocida como Epidoc, que es ampliamente utilizada por los epigrafistas y otras personas que trabajan con inscripciones del mundo antiguo. Como hemos señalado anteriormente, el atributo `@type` del elemento <div> no está limitado de ninguna manera por TEI. Sin embargo, la comunidad Epidoc ha decidido que desea imponer la presencia de este atributo y autorizar únicamente seis valores predefinidos para él. Veamos a continuación el fragmento de la ODD de Epidoc que permite obtener ese efecto:

```

<elementSpec ident="div" mode="change" module="textstructure">
  <attList>
    <attDef ident="type" mode="replace" usage="req">
      <valList type="closed">
        <valItem ident="apparatus">
          <desc>to contain apparatus criticus or textual notes</desc>
        </valItem>
        <valItem ident="bibliography">
          <desc>to contain bibliographical information, previous publications,
            etc.</desc>
        </valItem>
        <valItem ident="commentary">
          <desc>to contain all editorial commentary, historical/
prosopographical
          discussion, etc.
          </desc>
        </valItem>
        <valItem ident="edition">
          <desc>to contain the text of the edition itself; may include multiple
            text-parts</desc>
        </valItem>
        <valItem ident="textpart">
          <desc>used to divide a div[type=edition] into multiple parts
            (fragments, columns, faces, etc.)</desc>
        </valItem>
        <valItem ident="translation">
          <desc>to contain a translation of the text into one or more modern
            languages</desc>
        </valItem>
      </valList>
    </attDef>
  </attList>
</elementSpec>

```

- 15 Téngase en cuenta que se podrían añadir más restricciones para comprobar que estos valores se utilizan correctamente, por ejemplo para comprobar que un <div type="textpart"> siempre tiene un <div type="edition"> como ancestro.
- 16 El proyecto Epidoc muestra cómo una comunidad de investigadores particular puede adaptar TEI a sus propias necesidades, y su personalización es una buena demostración del tipo de cosas que el lenguaje ODD hace posible.

Añadir un nuevo elemento

- 17 El sistema ODD también permite añadir elementos completamente nuevos a un esquema. En su forma más sencilla, esto no requiere nada más que añadir a la especificación del esquema un <elementSpec> para el nuevo elemento, si bien para decidir correctamente cuál es el contenido más conveniente al nuevo elemento es necesario contar con ciertos conocimientos sobre cómo está diseñado el sistema TEI. Elegir qué membresías de clase deben especificarse para un nuevo elemento, por

ejemplo, es difícil sin saber qué clases existen y cómo interactúan. La elección de un modelo de contenido apropiado para el nuevo elemento requiere igualmente pensar en cómo se definen los demás elementos, suponiendo que deseemos que nuestro nuevo elemento actúe de forma coherente con el resto de TEI. Y por último, dado que se trata de un elemento que no pertenece a TEI, debemos tener cuidado de definirlo en un espacio de nombres distinto del espacio de TEI.

- 18 Por ejemplo, supongamos que deseamos añadir un nuevo elemento `<speciesName>` para marcar los nombres de las especies botánicas o de otro tipo que aparecen en un texto. Nuestro nuevo elemento es semánticamente similar a los elementos existentes `<persName>` o `<name>`, por lo que un buen punto de partida será buscar en la especificación ODD otros elementos similares a los nombres. Por ejemplo, en la `<elementSpec>` para el elemento `<persName>`, vemos que es miembro de una clase modelo `model.nameLike.agent`, que es una subclase de la clase más general `model.nameLike`. Añadir nuestro nuevo elemento a esta clase garantizará que aparezca en los modelos de contenido de otros elementos TEI en los mismos lugares que otros elementos de denominación, sin ningún trabajo adicional por nuestra parte. También podemos observar que `<persName>` es miembro de varias clases de atributos, especialmente `@att.global` (que tiene varias subclases propias) y `att.canonical`, que parecen útiles para nuestros fines. Si añadimos nuestro nuevo elemento a estas clases, nos aseguraremos de que utilice estos atributos del mismo modo que otros elementos similares a nombres.
- 19 A continuación debemos considerar el modelo de contenido de nuestro nuevo elemento. Lo más sencillo sería hacer lo mismo que con los demás elementos de denominación y decir que puede contener solamente texto, mezclado con otros elementos de nivel de frase, elementos tales como `<g>` y elementos globales. Este es un requisito tan habitual en el esquema TEI que existe un atajo (una “macro”) definido para ello llamado `macro.phraseSeq`. Sin embargo, si queremos representar el contenido interno de un `<speciesName>` (por ejemplo, para distinguir el “nombre de género” del componente “nombre de especie”), podríamos definir un modelo de contenido más específico, que posiblemente implique otros elementos nuevos u otras restricciones. Cualquiera que sea la opción que elijamos, se documentará explícitamente en nuestro ODD, para que otros usuarios de nuestros datos puedan ver cómo hemos personalizado el marco conceptual básico de TEI.

Personalización y conformidad

- 20 TEI debe reflejar la diversidad de prácticas dentro de su comunidad de usuarios si quiere seguir siendo útil. A menudo suele decirse que si se hace la misma pregunta a dos académicos, siempre se obtendrán al menos tres respuestas diferentes, por lo que no debería sorprendernos encontrar una diversidad considerable en las recomendaciones de TEI.
- 21 No obstante, uno de los propósitos de las *Directrices TEI* es *orientar* la práctica de la codificación. No se trata de una norma que le diga a uno lo que tiene que hacer (del mismo modo que las normas de ingeniería, por ejemplo, especifican exactamente la forma y las dimensiones de los accesorios eléctricos); por el contrario, le dice a uno cómo comunicar a otros lo que ha hecho. Por eso, en la sección de las *Directrices* que definen la conformidad con TEI, una de las características esenciales que se especifican

es la existencia de un ODD, que define un esquema que puede utilizarse para validar los documentos en cuestión, junto con otras posibles restricciones. Otros criterios esenciales para la conformidad con TEI son la validez formal con respecto al esquema `tei_all`, y el respeto de la semántica definida de los elementos TEI utilizados. Un documento conforme a TEI puede contener elementos de muchos espacios de nombres, pero cualquier elemento del espacio de nombres de TEI debe tener la semántica definida para ellos por TEI.

- 22 Consideremos, por ejemplo, al formidable novelista inglés del siglo XIX Bulwer-Lytton. En un corpus de documentos TEI podríamos esperar encontrar su nombre completo etiquetado, y asociado a una definición de la propia persona, de la siguiente manera:

```
<persName ref="https://en.wikipedia.org/wiki/Edward_Bulwer-Lytton">
  <forename>Edward</forename>
  <forename>George</forename>
  <surname type="linked">Bulwer-Lytton</surname>,
  <roleName>
    Baron Lytton of <placeName>Knebworth</placeName>
  </roleName>
</persName>
```

- 23 Pero en otros encontraremos versiones con marcado menos informativo o que utilicen elementos TEI diferentes, tales como:

```
<p>
  <rs type="nombre">Baron Lytton of Knebworth</rs>
</p>
```

o

```
<p>
  <name type="persona" key="BWLY">
    The Right Honourable the Lord Lytton PC
  </name>
</p>
```

o

```
<p>
  <persName ref="#BWLY">
    First Baron Lytton of Knebworth
  </persName>
</p>
```

- 24 Tales variaciones son perfectamente razonables: los proyectos varían considerablemente en sus prioridades. Para algunos, basta con mostrar que una cadena de palabras es un nombre; para otros, es importante distinguir los nombres de personas de los nombres de lugares; para otros, es esencial desambiguar las referencias de entidades con nombre, tal como se conocen. Está claro que los resultados de estos

proyectos no pueden combinarse de manera directamente interoperable sin un esfuerzo añadido. ¿Pueden considerarse todos ellos conformes con TEI?

- 25 Una personalización explícita, que utilice las características de ODD que hemos esbozado en esta sección, es la forma más fácil de que un proyecto comunique sus propias decisiones sobre cómo debe aplicarse TEI y, por lo tanto, la forma más fácil de garantizar que sus documentos sean intercambiables con otros. Un proyecto puede simplemente suprimir (por ejemplo) el elemento `<rs>` de su esquema, o especificar una lista cerrada de valores para su atributo `@type`. Puede optar por utilizar uno u otro (o ambos) de los atributos `@key` y `@ref` para asociar las diferentes versiones del nombre de una persona con los datos sobre dicha persona, y así desambiguarlos. Estas elecciones pueden hacerse explícitas en un esquema personalizado y, por lo tanto, nos indican cuál entre los muchos y muy diferentes enfoques existentes para etiquetar el nombre se ha adoptado dentro de un conjunto determinado de documentos. Y como también es posible generar un ODD automáticamente a partir de un conjunto de documentos TEI-XML, asimismo podemos distinguir una personalización prevista — expresando las prácticas de codificación deseadas por los creadores de un documento— de una personalización real, que expresa las prácticas de codificación realmente encontradas.
- 26 Si TEI está diseñado para ser personalizado, ¿cómo puede pretender ser además un formato de intercambio? En efecto, para el “intercambio ciego” (también conocido como “interoperabilidad”), el receptor de un documento debe estar seguro de que solo contendrá elementos TEI, y de que esos elementos se han utilizado según las reglas definidas por TEI. Sin modificar, esas reglas permiten tanta variación que a un desarrollador le resulta imposible prever todas las posibilidades que su software debe manejar. La disponibilidad de un ODD simplifica considerablemente esta tarea, tanto porque limita los elementos disponibles como por la posibilidad de añadir nuevas restricciones sobre el contenido.

Conclusión: ¿Qué es TEI?

- 1 Cuando hablamos de TEI, ¿a qué nos referimos? ¿Al organismo administrativo que es apoyado económicamente por los miembros del Consorcio de la Iniciativa de Codificación de Textos (*Text Encoding Initiative Consortium*)? Cuando decimos: *TEI me ha enviado este boletín* o *apoyamos a TEI*, probablemente estemos hablando del Consorcio como organización. Pero cuando decimos: *TEI ahora apoya el mercado de la notación musical* o *TEI ahora proporciona métodos para la codificación de ediciones genéticas*, estamos hablando de algo bastante diferente: el contenido técnico de las Directrices de TEI. Y cuando decimos: *ese es un punto de vista realmente de TEI* (por ejemplo, en referencia a las mejores prácticas en el archivo a largo plazo de recursos digitales) o *TEI ejemplifica nuestro enfoque del código abierto*, estamos hablando claramente de algo más que de un conjunto de recomendaciones técnicas o de las personas que las mantienen o que deciden utilizarlas. Algunas personas hablan de TEI como si fuera una especie de club o religión, con miembros y no creyentes, al que algunos aspiran y del que otros desean distinguirse. Dentro del ruidoso mercado de las *Humanidades Digitales*, TEI es una especie de miembro mayor, una figura paternal molesta para algunos, benévola para otros, algo demasiado anticuado incluso para ser considerado por otros. Sin embargo, a lo largo de la última década, ha quedado cada vez más claro que TEI forma parte de lo que hace posible las humanidades digitales: se ha convertido en una parte de la infraestructura con la que todo el mundo tiene que comprometerse, tanto técnica como socialmente, una vez que empieza a pensar en el texto o en otras formas de recursos culturales en formato digital. TEI proporciona un conjunto de herramientas con el que hacer esa reflexión y, lo que es más importante, también refleja la reflexión que se hace, tanto en sus preocupaciones como en sus rarezas ocasionales. Es este sentido de TEI, como arquitectura de la información, el que hemos utilizado a lo largo de la mayor parte de este texto. Sin embargo, parece apropiado concluir con alguna consideración sobre TEI como organización.
- 2 Una de las cosas más extrañas de TEI, al menos desde algunas perspectivas, es que no es una agencia gubernamental como la ANSI, ni una organización internacional bien financiada como la ISO, ni un consorcio industrial como el W3C, aunque organizacionalmente el Consorcio TEI tiene puntos en común con todos ellos. TEI es una organización sin ánimo de lucro, alimentada en su totalidad por una pequeña cantidad de fondos procedentes de instituciones y personas que se preocupan lo

suficiente por sus actividades como para poner dinero en su apoyo y mantenimiento, y también por una gran cantidad de esfuerzo voluntario de instituciones y personas que quieren contribuir a su desarrollo continuo. Todos los años, TEI organiza una conferencia en la que los miembros se reúnen para discutir sobre los asuntos oficiales necesarios para elegir a los directivos y, por supuesto, para debatir sobre los aspectos en los que TEI es intelectualmente fuerte o débil. Sin embargo, la comunidad de quienes utilizan TEI y contribuyen a su continuo desarrollo es mucho más amplia que la de los miembros abonados. Cualquier persona interesada es libre de proponer modificaciones o ampliaciones del sistema, o de informar de errores o inconsistencias— y de hecho mucha gente lo hace. TEI mantiene un Consejo Técnico elegido que se encarga de evaluar y actuar sobre todas las modificaciones propuestas y de gestionar la producción de nuevas versiones periódicas. Este proceso se lleva a cabo en público, en el espíritu del desarrollo de código abierto. Por lo tanto, nos parece natural referirnos a la "Comunidad TEI", aunque los límites y los componentes de esa comunidad de usuarios sigan siendo en gran medida desconocidos. Es esa comunidad la que posee TEI, la que habla en su nombre en la comunidad más amplia de las humanidades, y la que informa y determina su evolución.

- 3 La palabra *evolución* no se elige a la ligera. TEI de 2013 no es el mismo que el de 1998, ni siquiera el de 2009, aunque la mayoría de los nombres de los elementos siguen siendo los mismos. La organización y flexibilidad de TEI como arquitectura de sistema le permite adaptarse a las necesidades y prioridades cambiantes de su comunidad, de la misma manera que otras formas de vida han evolucionado en respuesta a un entorno cambiante. En Burnard 2013 he argumentado que este es el secreto de la longevidad de TEI: está organizado de tal manera que garantiza que pueda ser modificado de forma sencilla y eficaz, en función de las necesidades de sus usuarios. Por lo tanto, parece probable que TEI siga evolucionando, tanto en lo que se refiere a la adopción de nuevas áreas de codificación como a la modificación de lo que ya se ha propuesto para mantenerse al día con el cambiante panorama digital. A finales de 2010, por ejemplo, la versión 2.0 de TEI P5 introdujo una forma totalmente nueva de representar ediciones genéticas o documentales definida durante el año anterior por un grupo de trabajo financiado externamente. Aproximadamente al mismo tiempo se introdujo un método para hacer que los documentos TEI sean interoperables con un nuevo estándar para la representación de música anotada. Más recientemente, el trabajo en curso en el Grupo de Interés Especial en Ontologías de TEI garantiza que los textos codificados en TEI seguirán siendo aptos para las mejores prácticas emergentes en la representación de datos enlazados (*Linked Data*).
- 4 TEI como organización se compromete a mejorar la accesibilidad y la comprensibilidad de TEI como sistema técnico. Además del sitio web de TEI, que actúa como una completa referencia en línea para el esquema, existe una necesidad continua de proporcionar materiales de formación actualizados, y de promover el compromiso activo con los interesados en el uso, desarrollo y mantenimiento de TEI mediante wikis, grupos de discusión y cualquier otro medio apropiado que surja en el futuro. El desarrollo de estos recursos es una tarea continua para todos los miembros de la comunidad TEI y más allá. Este breve libro pretende ser una modesta contribución a esa agenda.

Lecturas complementarias

El sitio web de TEI mantiene una página de enlaces a tutoriales en línea. A continuación enumeramos algunos de ellos.

Las *Directrices TEI* incluyen una amplia lista de lectura de publicaciones sobre la teoría del marcado y TEI, que está disponible aquí. A continuación enumeramos algunos artículos de la misma que el lector general puede encontrar útiles o informativos.

El sitio web de Zotero alberga dos colecciones públicas de referencias bibliográficas relacionadas con TEI: una genérica y otra centrada en cuestiones lingüísticas.

Artículos tutoriales y manuales

BERTRAND Lauranne, DEMONET Marie Luce y FINS Jorge ed., *Manuel d'encodage XML-TEI. Renaissance et temps modernes (Imprimés - manuscrits)* [URL: <http://www.bvh.univ-tours.fr/XML-TEI/index.asp>] (Manual de codificación detallado para materiales impresos antiguos; en francés.)

BINGENHEIMER Marcus ed., 2009, *TEI使用指南—運用TEI處理中文文獻 Chinese TEI - A Guide to using TEI with Chinese Texts*, Taipei, Taiwan E-learning and Digital Archive Program 數位典藏與數位學習國家型科技計畫 [URL: <https://tei-c.org/Vault/Tutorials/TEI-ChinLoc-2ndPrintEd.pdf>] (Traducción al chino del tutorial TEI Lite (edición P5) y de los artículos que lo acompañan.)

BODARD Gabriel, ELLIOTT Tom y MYLONAS Elli ed., *EpiDoc Guidelines: Ancient documents in TEI XML* [URL: <https://www.stoa.org/epidoc/gl/latest/>] (Manual de codificación detallado para materiales epigráficos y otras fuentes primarias antiguas.)

BURNARD Lou y SPERBERG-MCQUEEN C.M., agosto de 2012, *TEI Lite: Encoding for Interchange: An Introduction to the TEI: Final Revised Edition for TEI P5* [URL: https://tei-c.org/release/doc/tei-p5-exemplars/html/tei_lite.doc.html] (El tutorial original de introducción genérica al uso de TEI. También disponible en una traducción al francés por Sophie David y Lou Burnard. [URL: https://www.tei-c.org/release/doc/tei-p5-exemplars/html/tei_lite_fr.doc.html])

CIOTTI Fabio ed., 2005, *Il Manuale TEI Lite: Introduzione Alla Codifica Elettronica Dei Testi Letterari*, Milán: Sylvestre Bonnard. (Traducción al italiano del tutorial TEI Lite (edición P4), acompañado de artículos introductorios sobre XML, sobre la transición de P4 a P5 y sobre las herramientas y técnicas TEI.)

DALMAU Michelle y HAWKINS Kevin ed., *Best Practices for TEI in Libraries* [URL: <https://tei-c.org/extra/teinlibraries/>] (Especificación detallada de un subconjunto de TEI apropiado para su uso en la captura de fuentes primarias en el trabajo en bibliotecas digitales.)

FLANDERS Julia y BAUMAN Syd ed., 2007, *Women Writers Project Guide to Scholarly Text Encoding* [URL: <https://www.wwp.northeastern.edu/research/publications/guide/>] (Sitio web que proporciona documentación y orientación sobre el uso de TEI para la codificación de incunables.)

HAWKINS Kevin, 2010, *Introduction to XML for Text* [URL: <http://www.ultraslavonic.info/intro-to-xml/>] (Breve introducción a la codificación con XML para principiantes absolutos.)

LOISEAU Sylvain, 2002, *Les Standards : autour d'XML et de la TEI* [URL: <http://www.revue-texto.net/Corpus/Manufacture/standards/sommaire.html>] (Visión detallada de TEI P4 y de los estándares XML relacionados; en francés.)

VAN DEN BRANDEN Ron y VANHOUTTE Edward ed., *TEI by Example*, 2010, [URL: <https://teibyexample.org/>] (Sitio de tutoriales en línea que ofrece ejercicios prácticos y ejemplos para toda la gama de módulos TEI.)

VAUGHAN, Nicolás, *Introducción a la codificación de textos en TEI (parte 1)* y *Introducción a la codificación de textos en TEI (parte 2)*, *Programming Historian en español* 5 (2021) [URL: <https://doi.org/10.46430/phes0053> y <https://doi.org/10.46430/phes0055> respectivamente]. (Tutorial introductorio para el uso de TEI, en español, dividido en dos partes.) DOI: 10.46430/phes0053 y 10.46430/phes0055

WITT Jeffrey C., STENSKJÆR C. Michael y VAUGHAN Nicolás, *LombardPress Critical Transcription Guidelines* [URL: <https://community.scta.info/pages/lombardpress-schema-critical.html>] (Manual de codificación detallado para ediciones críticas a partir de materiales manuscritos y otras fuentes primarias medievales.)

WITT Jeffrey C., STENSKJÆR C. Michael y VAUGHAN Nicolás, *LombardPress Diplomatic Transcription Guidelines* [URL: <https://community.scta.info/pages/lombardpress-schema-diplomatic.html>] (Manual de codificación detallado para ediciones diplomáticas a partir de materiales manuscritos y otras fuentes primarias medievales.)

Artículos de revistas y libros

BARNARD David T., BURNARD Lou y SPERBERG-MCQUEEN C. Michael, 1996, “Lessons Learned From Using SGML in the Text Encoding Initiative”, *Computer Standards & Interfaces*, vol. 18, nº 1, p. 3-10 [URL: [https://doi.org/10.1016/0920-5489\(95\)00035-6](https://doi.org/10.1016/0920-5489(95)00035-6)] (Artículo técnico que resume cómo se utilizó el SGML en TEI P3 para implementar la modularidad, la personalización y el hipervínculo. De interés para la historia de ODD.) DOI: 10.1016/0920-5489(95)00035-6

BURNARD Lou y RAHTZ Sebastian, 2004, “RelaxNG with Son of ODD”, *Proceedings of Extreme Markup Languages*, 2004, [URL: <https://ora.ox.ac.uk/objects/pubs:394056>] (Descripción técnica del nuevo sistema ODD desarrollado para TEI P5.)

BURNARD Lou, junio de 2000, “Text Encoding for Interchange: A New Consortium”, *Ariadne*, nº 24 [URL: <http://www.ariadne.ac.uk/issue24/tei/>] (Comunicado de prensa en el que se describe el relanzamiento de TEI como consorcio de miembros.)

BURNARD Lou, FIORMONTE Domenico y USHER Jonathan ed., 2001, “On the Hermeneutic Implications of Text Encoding”, *New Media and the Humanities: Research and Applications*, Oxford: Humanities Computing Unit, p. 31-38 [URL: <https://raw.githubusercontent.com/lb42/Works/master/Papers/2001-herman.xml>] (Breve artículo que subraya la naturaleza interpretativa del marcado y lo sitúa en una antigua tradición erudita.)

BURNARD Lou, 2013, “The Evolution of the Text Encoding Initiative: From Research Project to Research Infrastructure”, *Journal of the Text Encoding Initiative*, no. 5 [URL: <https://journals.openedition.org/jtei/811>] (Revisión histórica del desarrollo de TEI hasta la fecha.)

COOMBS James H, RENEAR Allen, DEROSE Steven J, 1987, “Markup Systems and the Future of Scholarly Text Processing”, *Communications of the ACM*, vol. 30, nº 11, p. 933-947 [URL: <http://>]

xml.coverpages.org/coombs-hallgren.html] [URL: <http://xml.coverpages.org/coombs.html>] [URL: <https://dl.acm.org/doi/10.1145/32206.32209>] (Polémica clásica e influyente en la que se defiende el marcado descriptivo en lugar del procedimental como actividad académica.) DOI: 10.1145/32206.32209

DEROSE Steven J, DURAND David G, MYLONAS Elli, RENEAR Allen H., 1997, “What is Text, Really?”, *Journal of Computer Documentation*, vol. 21, nº 3, p. 1-24 [URL: <https://doi.acm.org/10.1145/264842.264843>] (Declaración clásica de la hipótesis OHCO según la cual el texto es fundamentalmente una *jerarquía ordenada de contenidos*.) DOI: 10.1145/264842.264843

HOLMES Martin y ROMARY Laurent, 2010, “Encoding Models for Scholarly Literature: Does the TEI Have a Word to Say?”, *Publishing and Digital Libraries: Legal and Organizational Issues*, IGI Global, p. 88-110 [URL: https://hal.archives-ouvertes.fr/hal-00390966/PDF/romary_chap_iglezakis_book.pdf] (Discute el uso de TEI para la codificación y gestión de las publicaciones académicas.)

IDE Nancy y VERONIS Jean ed., 1995, *The Text Encoding Initiative: Background and Contexts*, Dordrecht, Boston: Kluwer (Colección esencial de 17 artículos publicados originalmente como número doble de *Computers and the Humanities* en 1994. Los principales colaboradores de TEI de la primera generación ofrecen una visión general del enfoque de TEI para la codificación de tipos de texto específicos, complementada con artículos clásicos sobre los objetivos y el diseño de TEI, y sobre SGML.)

IDE Nancy y ROMARY Laurent, 2004, “A Registry of Standard Data Categories for Linguistic Annotation”, *Proceedings of the 4th International Conference on Language Resources and Evaluation - LREC'04*, Lisboa, Portugal, p. 135-138 [URL: <https://hal.inria.fr/inria-00099858>] (Describe el desarrollo de un componente de Registro de Categorías de Datos (DCR) para el Marco de Anotación Lingüística; una norma ISO de gran importancia en la codificación del análisis lingüístico.)

MERTZ David, 2003, “XML Matters #30: The Text Encoding Initiative. An XML Dialect for Archival and Complex Documents” [URL: https://gnosis.cx/publish/programming/xml_matters_30.html] (Interesante perspectiva sobre TEI-XML vista desde fuera de la comunidad académica.)

MYLONAS Elli y RENEAR Allen, 1999, “The Text Encoding Initiative at 10: Not Just an Interchange Format Anymore — But a New Research Community”, *Computers and the Humanities*, vol. 33, nº 1-2 (Introducción a una colección retrospectiva de 12 ponencias seleccionadas entre las presentadas a la Conferencia del 10º Aniversario de TEI, celebrada en la Universidad de Brown en noviembre de 1997.)

SCHMIDT Thomas, 2011, “A TEI-based Approach to Standardising Spoken Language Transcription”, *Journal of the Text Encoding Initiative*, no. 1 [URL: <https://journals.openedition.org/jtei/142>] (Analiza los esquemas de transcripción existentes para el lenguaje hablado y la viabilidad de utilizar TEI como interlingua entre ellos.) DOI: 10.4000/jtei.142

SPERBERG-MCQUEEN C. Michael, 1991, “Text in the Electronic Age: Textual Study and Text Encoding, with Examples from Medieval Texts”, *Literary & Linguistic Computing*, vol. 6, nº 1, p. 34-46 [URL: <https://doi.org/10.1093/lc/6.1.34>] (Exposición clásica de la teoría del marcado aplicada a la edición de fuentes primarias.) DOI: 10.1093/lc/6.1.34]

UNSWORTH John, O'BRIEN O'KEEFE Katherine y BURNARD Lou ed., 2006, *Electronic Textual Editing*, New York: Modern Languages Association (Colección de artículos de 24 destacados profesionales que cubren todos los aspectos de la edición textual electrónica dentro del paradigma TEI. Una primera versión está disponible en el sitio web de TEI.)

WITTERN Christian, TUOHY Conal y CIULA Arianna, 2009, “The making of TEI P5”, *Literary and Linguistic Computing*, vol. 24, nº 3, p. 281-296 [URL: <https://doi.org/10.1093/lc/fqp017>] (Informe

detallado sobre el trabajo editorial que condujo a la primera edición de TEI P5.) DOI: 10.1093/lc/fqp017

Glosario

- **Analítico (*Analytic*)** En un contexto bibliográfico, un título analítico o de análisis se refiere a un solo componente de un elemento bibliográfico más amplio, por ejemplo, un solo artículo dentro de una revista.
- **Asentamiento (*Settlement*)** Cualquier lugar donde viven personas (un pueblo, una ciudad, etc.).
- **Atributos globales (*Global attributes*)** Atributos que están potencialmente disponibles en cada elemento definido por el TEI.
- **Bien formado (*Well-formed*)** Se dice que un documento está bien formado si se ajusta a las reglas sintácticas de XML.
- **Centrado en los datos (*Data-centred*)** Un enfoque de la estructuración de la información que da prioridad a su contenido informativo o estructura por encima de su forma o apariencia.
- **Centrado en texto (*Text-centred*)** Enfoque de la estructuración de la información que da prioridad a su presentación en forma textual sobre su contenido semántico.
- **Complejo (*Composite*)** Un documento complejo es aquel que tiene varias partes independientes, cada una de las cuales podría funcionar de manera autónoma.
- **Conformidad (*Conformance*)** Un documento conforme con TEI debe ser válido según el esquema `tei_all` y también debe respetar el modelo conceptual de TEI.
- **Declaración de la publicación (*Publication statement*)** La parte de un encabezado TEI que contiene la información sobre la publicación o la distribución de un recurso.
- **Descripción de la codificación (*Encoding description*)** La parte de un encabezado TEI que describe la codificación y las prácticas editoriales aplicadas en su producción.
- **Descripción de la revisión (*Revision description*)** La parte del encabezado TEI que se refiere a las revisiones y cambios introducidos durante la producción de un recurso.
- **Descripción del perfil (*Profile description*)** La parte de un encabezado TEI que se refiere a los metadatos descriptivos no bibliográficos.
- **Elemento (*Data-centred*)** Cualquier componente del documento XML identificado por medio del etiquetado.
- **Encabezado (*Header*)** En el contexto de TEI se refiere a todos los metadatos asociados a un corpus o documento.
- **Espacio de nombres (*Namespace*)** Un identificador asociado a un grupo particular de nombres de elementos XML o vocabulario concreto.

- **Esquema (Scheme)** Conjunto de reglas expresadas formalmente relativas al marcado de un documento.
- **Estructuras de rasgos (Feature structures)** Un formalismo abstracto ampliamente utilizado en lingüística para representar el significado en forma de estructuras de rasgos con nombre y valores de rasgos.
- **Etiquetas (Tags)** Códigos o etiquetas incrustados dentro de una cadena de caracteres para indicar el comienzo o el final de un elemento.
- **Grupo (Group)** Varios textos independientes que también pueden considerarse como uno solo, por ejemplo una antología.
- **Monográfico (Monographic)** En bibliografía, un único elemento bibliográfico identificable, como un libro.
- **Módulo (Module)** Una de las agrupaciones con nombre de las declaraciones TEI (para elementos, clases, macros, etc.) que componen el marco conceptual de TEI.
- **Representación (Rendition)** En el contexto de TEI se refiere a la forma en que algún aspecto de un documento fue presentado o formateado originalmente.
- **Resaltado (Highlighted)** En el contexto de TEI, se utiliza para describir cualquier forma de realce visual.
- **Simple (Simple)** Un texto o documento unitario, que puede ser independiente y está completo.
- **Validador (Validator)** Un software que comprueba que un documento XML es conforme a un esquema para determinar si es válido o no.
- **Válido (Valid)** Se dice que un documento XML es válido si se ajusta a un esquema.