



## La codificación XML/TEI de textos medievales

José Manuel Fradejas Rueda<sup>1</sup>  
*Universidad de Valladolid*

### Introducción

Hace años me propuse como objetivo investigador la edición de todas las obras de ceterería iberorrománicas. Con el advenimiento, difusión y popularización de internet comprendí que la red era el mejor medio para difundir los textos.

Comencé maquetando y transformando algunos de los textos que había editado al formato portable document file (pdf) de Acrobat y los colgué en la página web que inicié en la Universidad Nacional de Educación a Distancia y que posteriormente trasladé a la Universidad de Valladolid.<sup>2</sup> Pasé después a un proyecto más ambicioso y cuyo resultado es la web que he creado, con la colaboración de la Dra. Déborah Dietrick Smithbauer, durante el trienio 2006-09 dentro del proyecto de investigación subvencionado por el Ministerio de Educación y Ciencia (HUM00932-2006/FILO)<sup>3</sup> y

1. Este trabajo se ha realizado dentro de las labores del proyecto de investigación VA046A09 financiado por la Junta de Castilla y León.

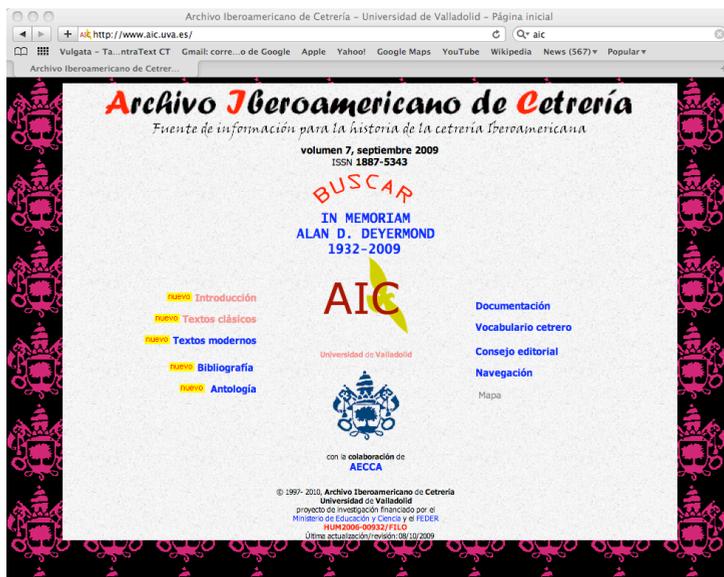
2. <http://gramola.fyl.uva.es/~ceterria/>. Puede accederse a algunos de los ficheros por medio de Internet Archive Wayback Machine ([http://web.archive.org/web/\\*/http://gramola.fyl.uva.es/~ceterria/](http://web.archive.org/web/*/http://gramola.fyl.uva.es/~ceterria/) [acceso 19.04.2010]) que tiene archivadas las páginas desde el 20 de mayo de 2001 a 4 de febrero de 2005: *Libro de los animales que cazan* (<http://web.archive.org/web/20010610225510/http://gramola.fyl.uva.es/~ceterria/textos/moamin.pdf>), el *Libro de la caza* de don Juan Manuel (<http://web.archive.org/web/20010610225510/http://gramola.fyl.uva.es/~ceterria/textos/juanmanuel.pdf>), el *Libro de la caza de las aves* de Pero López de Ayala (<http://web.archive.org/web/20010610225510/http://gramola.fyl.uva.es/~ceterria/textos/ayala.pdf>), el *Libro de aceterría y montería* de Juan Vallés (<http://web.archive.org/web/20010610225510/http://gramola.fyl.uva.es/~ceterria/textos/valles.pdf>) y el *Libro de ceterría* de Luis de Zapata (acceso 19.04.2010) que se ampliaron a en abril de 2004 a *Dancus Rex* (<http://web.archive.org/web/20040429174107/http://gramola.fyl.uva.es/~ceterria/textos/dancus.pdf>), *Guillelmus falconarius* (<http://web.archive.org/web/20040429174107/http://gramola.fyl.uva.es/~ceterria/textos/guillelmus.pdf>), *Libro de los azores* (<http://web.archive.org/web/20040429174107/http://gramola.fyl.uva.es/~ceterria/textos/azores.pdf>) y el *Tratado de ceterría* (<http://web.archive.org/web/20040429174107/http://gramola.fyl.uva.es/~ceterria/textos/tratado.pdf>), pero la Wayback Machine es incapaz de recuperar estos últimos como tampoco la obra de Juan Vallés.

3. <http://www.aic.uva.es> [fecha de acceso: 09.07.09]. En el momento de redactar estas líneas, se ha presentado una nueva solicitud de subvención al Ministerio de Ciencia e Innovación al amparo de la resolución de la Secretaría de Estado de Investigación de 30.12.2009 (BOE, 31.12.2009) a la que se le ha otorgado la referencia FFI2010-15128.

que se inició con una idea presentada durante el Simposio Internacional *Falconry: A World Heritage* celebrado en Abu Dhabi en septiembre de 2005 (Trimbell 2006: 69n1). El proyecto inicial se proponía:

1. Catalogar y digitalizar los manuscritos y los impresos, antiguos y modernos, que tratan de cetrería
2. Catalogar y digitalizar los artículos científicos y de divulgación publicados sobre cetrería
3. Localizar, catalogar y publicar transcripciones electrónicas de toda la documentación antigua que de una u otra manera se refiera a la cetrería
4. Localizar, catalogar, fotografiar y digitalizar todas las obras de arte cuyo motivo sea cetrero
5. Localizar todas las menciones que a la cetrería se haga en cualquier obra literaria y acumularla en una antología
6. Ofrecer ediciones electrónicas de los textos medievales y clásicos en cualquiera de las lenguas iberoamericanas.

Semestralmente (septiembre y marzo) se ha publicado una entrega con los avances y hallazgos del Archivo Iberoamericano de Cetrería (AIC). Desde el volumen 6 (marzo 2008) se decidió el cierre de varias secciones (Revistas, Recortes, Folletos, Imágenes y Miscelánea). Las dedicadas a Revistas y Recortes por meros problemas de copyright, a pesar de la gran cantidad de material digitalizado (3.514 ficheros jpg con un volumen de 4,08 GB), las otras dos secciones debido a la nula colaboración de los informantes externos. Por lo que en esa entrega se decidió centrar el AIC en la historia literaria (textos, manuscritos, menciones literarias) y documental de la cetrería Iberorrománicas.



Página principal del AIC

Durante la ejecución nos percatamos de un problema técnico a la hora de presentar los resultados y su reutilización. Las páginas web utilizan el *HyperText Markup Language*-

ge (HTML), un lenguaje de marcado o etiquetado presentacional. Este es un tipo de etiquetado útil para maquetar la presentación de un documento para su lectura, pero resulta insuficiente, cuando no directamente ineficaz, si se pretende procesar automáticamente la información. Por otra parte, aunque el marcado de presentación es fácil de elaborar, sin embargo, resulta complicado de mantener o modificar, por lo que su uso se ha ido sustituyendo por otros tipos de marcado más estructurados y que evitan las ambigüedades.

Un segmento marcado con cursiva (`<i>cursiva</i>`) puede emplearse para indicar énfasis o bien para señalar palabras en otro idioma, el título de una obra, el desarrollo de una abreviatura, etc. Esta ambigüedad, presente en el marcado de presentación que hemos venido utilizando, no puede soslayarse salvo con una tediosa revisión a mano. Sin embargo, si todas estas posibilidades que indicamos con las cursivas —énfasis, título, lengua, abreviatura,...— se hubieran diferenciado descriptivamente con etiquetas distintas (p. ej.: para los títulos `<título> ... </título>`, para marcar énfasis `<énfasis> ... </énfasis>`, para el desarrollo de abreviaturas `<abrevia> ... </abrevia>`, etc.), podrían representarse de manera diferente sin gran esfuerzo y, además serían fácilmente utilizadas para hacer otras labores de análisis. Los ordenadores ante ese tipo de codificación primitiva no saben qué hacer.

Puesto que el objetivo era (y es) crear un corpus que pueda ser reutilizado sin codificación ni mantenimiento ulterior, y que permita analizar los textos desde distintas perspectivas e incluso que se puedan convertir fácilmente en libros-electrónicos (*e-books*) decidimos codificarlos por medio de un *marcado descriptivo o semántico* que utilice etiquetas para describir los textos, pero sin especificar cómo se deben representar o en qué orden.

Una de las virtudes del marcado descriptivo es su flexibilidad: los textos se etiquetan *tal como son*, y no *tal como deben aparecer*, lo que implica que podrían utilizarse de formas no previstas inicialmente.

Por otra parte, el marcado descriptivo simplifica la tarea de reformatear un texto debido a que la información del formato está separada del propio contenido, es decir, es independiente. De todos los lenguajes informáticos que posibilitan el marcado descriptivo el mejor de todos es el XLM.

## El Sistema deHSMS

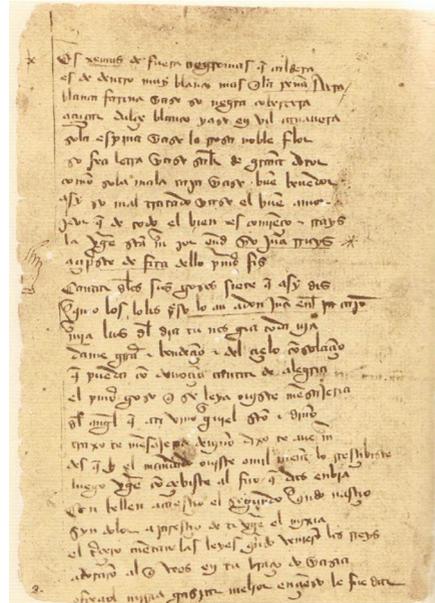
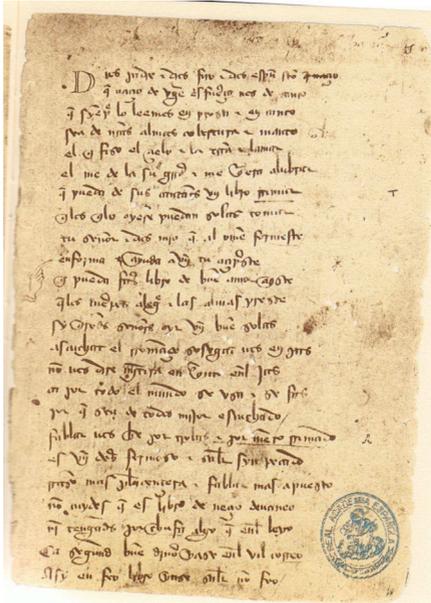
Durante años se ha utilizado el llamado sistema de Madison para codificar electrónicamente cualquier texto iberorrománico medieval. Fue formulado y explicitado por David Mackenzie y Kenneth Buelow en *A Manual of Manuscript transcription for the Dictionary of Old Spanish Language*, que se publicó por primera vez en 1977 y se ha reeditado a lo largo de veinte años, hasta 1997. Este manual tenía como finalidad unificar criterios con vistas a la creación de una base de datos cuyo objetivo era la redacción de un diccionario del Español Medieval.

El resultado más llamativo de este proyecto es el *Archivo Digital de Manuscritos y Textos Españoles* —ADMYTE—.<sup>4</sup> Ofrecía, en una serie de CD-ROM, transcripciones y facsímiles —estos solo en el volumen 0— de un buen número de manuscritos e incunables españoles. Desafortunadamente este material es actualmente inutilizable

4. Madrid: Micronet-V Centenario-Biblioteca Nacional, 1992 y 1993.

dado que se creó para una única plataforma, Windows 3.x, y no ha sido actualizada en ningún momento, aunque se menciona la existencia de una versión 32 bits<sup>5</sup> llamada ADMYTE-II<sup>6</sup> y existe una web de acceso restringido por suscripción que no funciona.<sup>7</sup>

Las normas madisonianas, según uno de sus artífices, «pretenden permitir una fiel representación del texto medieval, con un sistema de paréntesis de varios tipos para indicar la resolución de abreviaturas y modificaciones escritas [sic] o editoriales, ya sean rectificaciones, tachaduras, etc. ... [y] describir ciertos aspectos del texto y de la forma física del manuscrito mediante ciertos códigos, como por ejemplo rúbricas, glosas, texto en lengua extranjera, miniaturas, etc.» (Mackenzie 1992).



Sin embargo, el análisis de algunos de los textos editados según estas normas muestra que en ningún momento se intenta describir el texto más allá de una mínima disposición física y su interpretación lógica.<sup>8</sup> Veamos, a modo de ejemplo, la transcripción

5. <http://www.llf.uam.es/~fmarcos/informes/admyte/admytei.htm> [22.09.08].

6. <http://www.llf.uam.es/~fmarcos/informes/admyte/admyteii.htm> [22.09.08].

7. <http://admyte.com/home.htm> [19.04.2010]. Al cumplimentar la ficha de suscripción (<http://admyte.com/suscrip.htm>) y pulsar el botón de enviar arroja el resultado «404 Not Found | No encontrado | WeBMaSTer@admyte.com» y el webmaster al que manda el sistema responde con «<<mailbox unavailable>>, <<User unknown>>».

8. La descripción lógico-textual apenas interesaba al sistema del HSMS, pues se basaba en la transcripción secuencial de los folios de cada uno de los testimonios, sin advertir de posibles desórdenes internos. El *Libro de la caza de las aves* de Juan Manuel (ms. 6376) por un fallo de los copistas se halla un tanto enrevesado: el prólogo comienza en el folio 202, columna 1ª, línea 8, y se alarga hasta el folio 202, columna 1ª, línea 29, en ese punto hay que retroceder al folio 199; continúa la lectura hasta el folio 201, columna 1ª, línea 7; y desde ahí salta al folio 202, columna 1ª, línea 30 y sigue ya hasta el fin sin más desorden. La transcripción de Reinaldo Ayerbe-Chaux (O'Neill 1999) tan solo indica al comienzo del fichero, «[RMK: Libro de la cac'a; part of chapter 1; beginning found in folio 196r.]» y 359 líneas más adelante se indica «[RMK: five lines left blank; El libro de la cac'a really starts on next line.]».

del primer folio (recto y vuelto) del ms. G del *Libro de buen amor* (preparada por Erik W. Naylor 1999).

[fol. 1r]

{CB1.

Dios padre & djos fiyo & dios esp<iri>tu s<an>to (^q<ue> nac'io)  
 q<ue> nac'io de v<[ir]>ge<[n]> esfue<r>c'a nos de tanto  
 q<ue> sye<n>pre lo leemos en prosa & en canto  
 sea de n<uest>ras almas cobertura & manto  
 el q<ue> fizo el c'ielo & la t<ie>rra & la mar  
 el me de la su grr<ac'i>a & me q<uie>ra alubrar  
 q<ue> pueda de sus cantar<e>s vn ljbrr rrimar  
 q<ue> los q<ue> lo oyere<n> puedan solaz' tomar  
 tu sen-or dios mjo q<ue> al om<n>e formeste  
 enforma & ayuda a vn tu ac'ip<re>ste  
 q<ue> pueda faz'<er> libro de bue<n> amor aq<ue>ste  
 q<ue> los cue<r>pos aleg<r><<e>> & las almas preste  
 sy q<ue>red<e>s senor<e>s oyr vn bue<n> solaz'  
 ascuchat el rroma<n>c'e sosegatuos en paz'  
 no<n> vos djre mj<n>tira en q<ua>nto en <e>l jaz'  
 ca por todo el mundo se vsa & se faz'  
 por q<ue> sea de todos mjjor escuchado  
 fablaruos he por trobas & por cue<n>to rrimado  
 es vn dez'<ir> fermoso & sab<e>r syn pecado  
 rraz'o<n> mas plac'entera & fablar mas apuesto  
 no<n> cuydes q<ue> es ljbrr de nec'io deuaneo  
 nj<n> tengad<e>s por chufa algo q<ue> en <e>l leyo  
 Ca segund bue<n> djn<er>o yaze en <e>l vil correo  
 Asy en feo ljbrr yaz'e sab<e>r no<n> feo}

[fol. 1v]

{CB1.

Es xemuz' de fuera negro mas q<ue> caldera  
 es de dentro muy blanco mas q<ue> la pen~a vera  
 blanca farina yaz'e so negre cobertera  
 ac'ucar dul'e blanco yaz'e en vil canauera  
 so la espjna yaz'e la rrosa noble flor  
 so fea letra yaz'e sab<e>r de grant dotor  
 com<m>o so la mala capa yaz'e bue<n> beuedor  
 asy so mal tratado yaz'e el bue<n> amor  
 Por q<ue> de todo el bien es comje<n>c'o & rrayz'  
 la v<ir>ge<n> s<an>ta m<aria> por end<e> yo jua<n> rruyz'  
 ac'ip<re>ste de fita dello p<ri>mero fiz'  
 Cantar d<e> los sus goz'os siete q<ue> asy diz'  
 {GL. q<ua>n[d]o los lobos p<re>so lo an  
 a don jua<n> en <e>l ca<n>po}  
 marja luz' d<e>l dja tu nos gia todauja  
 dame grr<ac'i>a & bendec'io<n> & del c'ielo co<n>solac'io<n>  
 q<ue> pueda co<n> deuoc'io<n> cantar de alegria  
 el p<ri>m<er>o goz'o q<ue> se leya oujste me<n>sajeria  
 d<e>l ang<e>l q<ue> a ti vjno g<ra>ujel s<an>to & djno  
 traxote me<n>sajeria deujno dixote aue m<aria>

desq<ue> (^a) el mandato oujste omjlmte lo rresc'ibiste  
 luego v<ir>ge<n> co<n>c'ebiste al fiio q<ue> dios enbja  
 En bellen acaesc'io el segundo q<ua>ndo nasc'io  
 syn dolor aparesc'io de ti v<ir>ge<n> el mjxia  
 el t<er>c'ero cue<n>tan las leyes q<ua>ndo veniero<n> los rreys  
 adorar<n> al q<ue> vees en tu brac'o do yaz'ia  
 ofreciol mjrria gaspar meljor enc'ie<n>so le fue dar]

Lo primero que llama la atención es la gran cantidad de paréntesis angulares —esp<iri>tu s<an>to, cue<r>pos, aleg<r><<e>>—, llaves —{CB1., feo}, {GL., dar}— y virgulillas al lado de algunas letras (apóstrofes —nac'io, faz'<er>—, tildes de nasalidad tras la n —sen~or y pen~a—, o acentos circunflejos inmediatamente después de un paréntesis —(^q<ue> (^a)—.

El apóstrofo tras c' es un dígrafo para la ç, y lo mismo sucede con n~, que lo es para la ñ. La <s> sigmática –σ– se codifica como z' cuando representa el fonema africado dentoalveolar sonoro, pero el transcriptor, en este caso, no se ha molestado en indicar que esa misma grafía se emplea para el fonema fricativo dentoalveolar (ya sordo ya sonoro) a pesar de que las normas de transcripción lo indican («Likewise, s' or z' is transcribed if the sigma's function is either s or z» (Mackenzie y Buelow, 1977: 11, § 3.215a).

{CB. marca el inicio de una columna física del manuscrito y se cierra con }, que es un elemento común a muchas otras etiquetas —*mnemonics* en la terminología madisoniana— ({GL., {IN., {CW., {HD., {ADD., etc. etc.). Lo encerrado entre paréntesis o corchetes y cuyo primer elemento es un acento circunflejo denota obliteración (^) o adición [^] del escriba. El desarrollo de las abreviaturas se indica entre paréntesis angulares < > y se recurre al acento grave ` para marcar las letras sobreescritas («Superscript characters, [...], often accompany scribal abbreviation, and may be a letter or letters from the part of the suppressed» (Mackenzie y Buelow, 1977: 14, § 3.222b), aunque en el ejemplo aducido esa función se marca con dobles paréntesis angulares (línea 12: aleg<r><<e>>).

Este sistema, a pesar de la relativa sencillez de sus normas, no logró establecerse como *el sistema* para la transcripción electrónica (digital) de textos medievales iberorrománicos, así lo reconocía su creador, David Mackenzie (1992), al afirmar que «no han logrado una universal aceptación». No lo consiguió porque no ofreció un representación superficial fácil de leer (*friendly user interface*), aspecto que puso en evidencia Lucía Megías (1999: 233–34) al decir que «...una de sus grandes limitaciones se concreta en la forma de presentar tales transcripciones ... la lectura de la impresionante nómina de textos que se ofrecen en *Admyte II* ... se dificulta por la continua aparición de paréntesis y signos extraños». Esto llevó a Felipe Tenenbaum (2000–01: 160) a explicitar un sistema sencillo para «transformar desde un procesador de texto una transcripción de Madison en una libre de llaves, paréntesis, corchetes y otros añadidos». Por otra parte, la tecnología ha avanzado mucho y lo ha convertido en un sistema obsoleto.<sup>9</sup>

Recientemente ha recibido una furiosa crítica, ciertamente injustificada porque no ha tenido en cuenta la teoría editorial y los objetivos que se había propuesto el Hispanic

9. No hay que negarle el gran valor que ha tenido. Baste mirar la amplia colección de ediciones en microficha (texto, índices léxicos y concordancias) que han publicado y que posteriormente han publicado en CD: *Textos y Concordancia Electrónicos del Corpus Médico Español* (Madison, 1997); *Early Celestina Electronic Texts and Concordances* (Madison, 1997), *Electronic Texts and Concordances of the Madison Corpus of Early Spanish Manuscripts and Printings* (Madison y Nueva York, 1999).

Seminary of Medieval Studies, por parte de José Pérez Navarro quien, al describir la transcripción de la *General Estoria. Cuarta Parte* conservada en el ms. Urb. Lat. 539, dice que «Por ser paleográfica, esta transcripción conserva características gráficas sin trascendencia fonética» (Pérez Navarro 2008: 2070) ya que «concede caracter diacrítico a alógrafos que no tienen tal caracter...; mantiene todas las consonantes dobles que no tienen valor fonético distinto al de la sencilla, ... las grafías latinizantes y pseudolatinizantes» (Pérez Navarro 2008: 2070n19) por lo que «el respeto extremo del manuscrito es la causa de que vayan a parar al glosario voces inexistentes y que un mismo vocablo deba buscarse en varios lugares del glosario: emfermo, emffermo, enfermo, ffijo, fijo, fiio» (Pérez Navarro 2008: 2070n20).

### El sistema XML/TEI

Durante el Primer Congreso de la Lengua Española (Sevilla 1992), en la sección «La lengua española y las nuevas tecnologías», el profesor Charles B. Faulhaber (1992) daba a conocer la Text Encoding Initiative (TEI). Abundaría en ello siete años después Aurora Martín de Santa Olalla Sánchez (1999).

Sin embargo, no se ha desarrollado en España esta nueva tecnología cuya definición última, P5, se publicó en Noviembre de 2007,<sup>10</sup> aunque es la que subyace en la Biblioteca Virtual Cervantes<sup>11</sup> e incluso en los corpus de la Real Academia Española,<sup>12</sup> pero en su versión Lite de 1995.<sup>13</sup>

El sistema de codificación TEI, nació en un congreso patrocinado por la Association for Computers and the Humanities (ACH), financiado por la U.S. National Endowment for the Humanities (NEH), que tuvo lugar en el Vassar College en 1987. Al año siguiente obtuvo financiación de la Comisión Europea y otras instituciones y en junio de 1999 se constituyó en un consorcio internacional.

Se trata de un proyecto de investigación en humanidades que goza de una amplísima difusión y utilización en bibliotecas y colecciones de texto digitales y en la creación de corpora lingüísticos. Se basa en el lenguaje XML (*eXtensible Markup Language*),<sup>14</sup> una versión simplificada del SGML (*Standard Generalized Markup Language*), estándar creado por IBM en 1986 y reconocido internacionalmente bajo la norma ISO 8879.<sup>15</sup>

El SGML era un estándar para la creación y definición de documentos para la publicación electrónica a gran escala. Para ello, implementaba unos documentos de validación llamados DTD (*Document Type Definition* 'definición de tipo de documento'). Lo que hacía el SGML era definir la sintaxis de los documentos de datos además de

10. <http://www.tei-c.org/Guidelines/P5/>, accedido el 22.09.08.

11. <http://www.cervantesvirtual.com/>, accedido el 30.09.08.

12. <http://www.rae.es/>, accedido el 30.09.08.

13. <http://193.147.129.138/alex-bia/Research/tei/teilight-sp.htm>, accedido el 20.09.08.

14. De hecho, «Several TEI participants contributed heavily to the development of XML, most notably XML's senior co-editor C. M. Sperberg-McQueen, who served as the North American editor for the TEI Guidelines from their inception until 1999» (<http://www.tei-c.org/release/doc/tei-p5-doc/en/html/AB.html#Note3>; acceso 02.08.09).

15. En verdad las primeras formulaciones del TEI se hicieron en SGML, pero cuando el World Wide Web Consortium dictó una recomendación final para el XLM (febrero 1998), la TEI adoptó el nuevo metalenguaje «if they were to survive» (<http://www.tei-c.org/release/doc/tei-p5-doc/en/html/AB.html>; acceso 02.08.09).

la sintaxis y semántica de los documentos DTD. Sin embargo, el SGML tenía un gran problema de diseño: era muy complicado.

El XML ha heredado muchísimas referencias y similitudes con el SGML, tanto en la definición de entidades y elementos de los documentos DTD, como en la sintaxis del sistema de marcado. Las especificaciones del XML se crearon con un menor nivel de dificultad y, por tanto, ha propiciado su amplia difusión. Además, el XML es un lenguaje independiente de plataformas informáticas y fabricantes de software y, a la vez, es escalable y extensible, por lo que se ha convertido en la «*lingua franca* of the data exchange world» (St. Laurent y Fitzgerald 2005: 1). Sus características básicas son:

- Es posible definir etiquetas (marcas) propias.
- Se pueden asignar atributos a las etiquetas.
- Las etiquetas y atributos se definen de forma exacta mediante un esquema (DTD, XML Schema o Relax NG).<sup>16</sup>
- La estructura y el diseño están totalmente separados.
- Es un estándar abierto y no sujeto a licencias.
- Es multilingüe.<sup>17</sup>
- Está basado en texto plano.
- Es ideal para conservar datos a largo plazo.
- Es modular y ampliable.

Basándose en este metalenguaje, creado para documentos que contienen información estructurada,<sup>18</sup> la TEI proporciona un esquema útil porque:

1. Define y especifica varios cientos de distinciones textuales.
2. Proporciona un juego de módulos que se pueden emplear para definir esquemas que tengan en cuenta esas distinciones (véase el cuadro de la página siguiente).
3. Facilita un mecanismo de actualización y acomodación para modificar y combinar esas definiciones con las nuevas que se pudieran crear utilizando el mismo modelo conceptual.
4. Ofrece hojas de estilo para transformar los textos codificados a HTML o PDF.
5. Facilita el intercambio y la integración de los datos entre los investigadores.

16. Estos tres tipos de definición de documentos son sucesivas mejoras que se han ido creando e implementado. La DTD, heredada directamente del SGML, no está basada en el lenguaje XML, por lo que el World Wide Web Consortium (W3C) creó el XML Schema, que sí se basa en el lenguaje XML.

El XML Schema es un vocabulario XML que permite describir otros vocabularios XML de manera que los programas pueden comprobar si un documento dado cumple las reglas establecidas por el esquema (<http://www.xml.com/pub/a/2000/11/29/schemas/part1.html?page=1>; acceso 24.07.09).

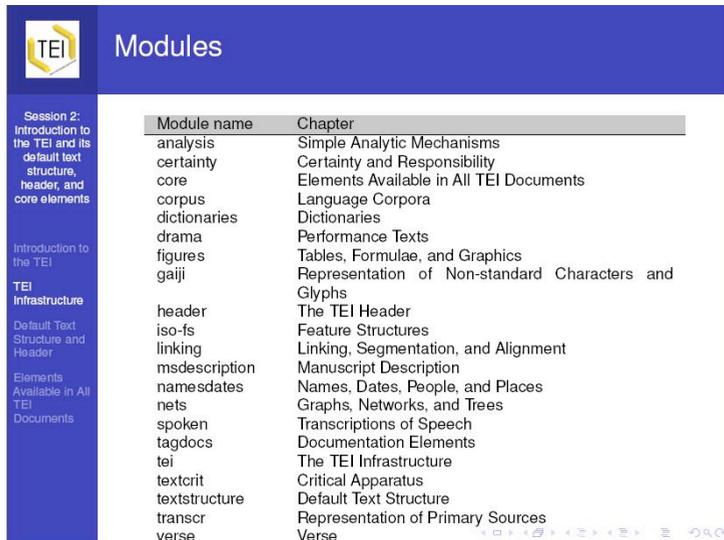
Relax NG (*Regular Expression Language for XML*, pronunciado como *relaxing*) es una mejora del XML Schema puesto que «[it] expresses rules about the possible structure of a document in terms of patterns; that is, it defines a number of named patterns, each of which acts as a kind of template against which an input document can be matched. The meaning of a pattern is expressed in a schema by reference to other patterns, or to a small number of built-in fundamental concepts» (*Guidelines for Electronic Text Encoding and Interchange. A Gentle Introduction to XML* <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/SG.html> [acceso 02.04.2009; comprobado 02.08.09]). La gran ventaja, además, es que es fácil de aprender y de utilizar y se puede expresar con la sintaxis XML o con una sintaxis no-XML mucho más compacta.

17. Por *multilingüe* se ha de entender que es válido para cualquier lengua escrita con independencia del sistema de escritura que emplee.

18. Toda información estructurada presenta un contenido (palabras, imágenes, etc.) e indicaciones de la función que cada contenido tiene dentro del documento. Por ejemplo, el contenido de un encabezamiento tiene un significado totalmente diferente del contenido de una nota a pie de página o del de un pie de foto. Prácticamente todos los documentos tienen alguna estructura.

6. Es válido para todo tipo de textos, de cualquier lengua y de cualquier época.

Las características anteriores convierten el XML/TEI en un entorno modular muy flexible para el tratamiento de los textos porque obliga a ver los textos como lo que realmente son. Se trata de una codificación pensada para las necesidades de los filólogos<sup>19</sup> y asume una serie de presupuestos y prioridades compatibles con la evolución del mundo digital puesto que se centra en el contenido y su función (no en la presentación) y ofrece soluciones genéricas (mucho más valiosas que las creadas *ad hoc*).



Module name	Chapter
analysis	Simple Analytic Mechanisms
certainty	Certainty and Responsibility
core	Elements Available in All TEI Documents
corpus	Language Corpora
dictionaries	Dictionaries
drama	Performance Texts
figures	Tables, Formulae, and Graphics
gaiji	Representation of Non-standard Characters and Glyphs
header	The TEI Header
iso-fs	Feature Structures
linking	Linking, Segmentation, and Alignment
msdescription	Manuscript Description
namesdates	Names, Dates, People, and Places
nets	Graphs, Networks, and Trees
spoken	Transcriptions of Speech
tagdocs	Documentation Elements
tei	The TEI Infrastructure
textcrit	Critical Apparatus
textstructure	Default Text Structure
transcr	Representation of Primary Sources
verse	Verse

Módulos disponibles en el TEI

Un excelente ejemplo de la aplicación de este sistema de codificación para textos medievales es el proyecto MENOTA (Medieval Nordic Texts). Otros ejemplos interesantes, aunque alejados de nuestro interés básico, son el Early America's Digital Archive<sup>20</sup> en el que se recogen transcripciones de los impresos en América realizados en América entre 1492 y 1820, o el Samyukta Agama,<sup>21</sup> edición digital comparativa del canon budista Samyukta Agama.

Por otra parte, el problema de la representación gráfica de las *letras* que constituyen cada uno de los textos, que encontraba graves limitaciones a causa de tecnología bajo la que se crearon las normas de Madison (código ASCII 7 bits), está resuelto con la codificación Unicode cuya última versión 5.1.0 tiene más de 100.000 caracteres diferentes (los importantes para nosotros son los del alfabeto latino y sus nueve extensiones que cubren un amplísimo espectro de representaciones).<sup>22</sup> Por este motivo es es

19. Así, Eito Brun (2008: 214) afirma que la «TEI propone un formato para codificar el texto completo de obras textuales en XML, y utiliza las marcas para destacar aspectos que tengan un valor para su estudio filológico y literario».

20. <http://www.mith2.umd.edu/eada/>; acceso 30.07.09.

21. <http://buddhistinformatics.chibs.edu.tw/BZA/>; acceso 30.07.09.

22. <http://www.unicode.org/charts/>; acceso 12.06.08.

de gran interés la Medieval Unicode Font Initiative<sup>23</sup> ya que cualquier procesador de textos actual puede codificar en UTF-8 (o superior -16 y -32).

## Estructura de un documento TEI

Veamos, como ejemplo, la transcripción, XML/TEI, del primer folio (recto y vuelto) del *Libro de buen amor* visto páginas atrás:

```
<?oxygen RNGSchema="http://www.tei-c.org/release/xml/tei/custom/schema/relaxng/tei_all.rng" type="xml"?>
<TEI xmlns:xi="http://www.w3.org/2001/XInclude"
  xmlns:svg="http://www.w3.org/2000/svg"
  xmlns:math="http://www.w3.org/1998/Math/MathML"
  xmlns="http://www.tei-c.org/ns/1.0">
  <teiHeader>
    <fileDesc>
      <titleStmt>
        <title>Libro de buen amor</title>
      </titleStmt>
      <publicationStmt>
        <p>Transcripción del ms. </p>
      </publicationStmt>
      <sourceDesc>
        <p>ms. T, BNE, ms.</p>
      </sourceDesc>
    </fileDesc>
  </teiHeader>
  <text>
    <body>
      <pb n="1r"/>
      <lg>
        <l>Dios padre ꝛ djos fijo ꝛ dios esp<ex>iri</ex>tu s<ex>an</ex>to<del>q
naçio</del></l>
        <l>q<ex>ue</ex> naçio de
          v<supplied><ex>ir</ex></supplied>ge<supplied><ex>n</ex></supplied>
          esfue<ex>r</ex>ça nos de tanto</l>
        <l>q<ex>ue</ex> sye<ex>n</ex>pre lo leemos en prosa ꝛ en canto</l>
        <l>sea de n<ex>uest</ex>ras almas cobertura ꝛ manto</l>
      </lg>
      <lg>
        <l>el q<ex>ue</ex> fizo el çielo ꝛ la t<ex>ie</ex>rre ꝛ la mar</l>
        <l>el me de la su grr<ex>açi</ex>a ꝛ me q<ex>uie</ex>ra alubrar</l>
        <l>q<ex>ue</ex> pueda de sus cantar<ex>e</ex>s vn ljbro rrimar</l>
        <l>q<ex>ue</ex> los q<ex>ue</ex> lo oyere<ex>n</ex> puedan solaz tomar</l>
      </lg>
      <lg>
        <l>tu señor dios mjo q<ex>ue</ex> al om<ex>n</ex>e formeste</l>
        <l>enforma ꝛ ayuda a vn tu açip<ex>re</ex>ste</l>
        <l>q<ex>ue</ex> pueda faz<ex>er</ex> libro de bue<ex>n</ex> amor
          aq<ex>ue</ex>ste</l>
        <l>q<ex>ue</ex> los cue<ex>r</ex>pos aleg<ex>r</ex><ex>e</ex> ꝛ las almas
          preste</l>
      </lg>
      <lg>
        <l>sy q<ex>ue</ex>red<ex>e</ex>s señor<ex>e</ex>s oyr vn bue<ex>n</ex>
          solaz</l>
        <l>ascuchat el rroma<ex>n</ex>çe sosegatuos en paz</l>
        <l>no<ex>n</ex> vos djre mj<ex>n</ex>tira en q<ex>ua</ex>nto en <ex>e</ex>l
          jaz</l>
        <l>ca por todo el mundo se vsa ꝛ se faz</l>
      </lg>
      <lg>
        <l>por q<ex>ue</ex> sea de todos mjjor escuchado</l>
        <l>fablaruos he por trobas ꝛ por cue<ex>n</ex>to rrimado</l>
        <l>es vn dez<ex>ir</ex> fermoso ꝛ sab<ex>e</ex>r syn pecado</l>
        <l>rrazo<ex>n</ex> mas plaçentera ꝛ fablar mas apuesto</l>
      </lg>
    </body>
  </text>
</TEI>
```

23. <http://gandalf.aksis.uib.no/mufi/>; acceso 12.06.08.

```

</lg>
<lg>
<l>no<ex>n</ex> cuydes q<ex>ue</ex> es ljbro de neçio deuaneo</l>
<l>nj<ex>n</ex> tengad<ex>e</ex>s por chufa algo q<ex>ue</ex> en
<ex>e</ex>l leyo</l>
<l>Ca segund bue<ex>n</ex> djn<ex>er</ex>o yaze en<ex>e</ex>l vil
correo</l>
<l>Asy en feo ljbro yaze sab<ex>e</ex>r no<ex>n</ex> feo</l>
</lg>
<pb n="1v"/>
<lg>
<l>Es xemuz de fuera negro mas q<ex>ue</ex> caldera</l>
<l>es de dentro muy blanco mas q<ex>ue</ex> la peña vera</l>
<l>blanca farina yaze so negra cobertera</l>
<l>açucar dulce blanco yaze en vil canauera</l>
</lg>
<lg>
<l>so la espjna yaze la rrosa noble flor</l>
<l>so fea letra yaze sab<ex>e</ex>r de grant dotor</l>
<l>com<ex>m</ex>o so la mala capa yaze bue<ex>n</ex> beuedor</l>
<l>asy so mal tratado yaze el bue<ex>n</ex> amor</l>
</lg>
<lg>
<l>Por q<ex>ue</ex> de todo el bien es comje<ex>n</ex>ço r rrayz</l>
<l>la v<ex>ir</ex>ge<ex>n</ex> s<ex>an</ex>ta m<ex>aria</ex>
porend<ex>e</ex> yo jua<ex>n</ex> rruyz</l>
<l>açip<ex>re</ex>ste de fita dello p<ex>ri</ex>me<ex>ro</ex> fiz</l>
<l>Cantar d<ex>e</ex> los sus gozos siete q<ex>ue</ex> asy diz</l>
</lg>
<p>
<add>q<ex>ua</ex>n<supplied>d</supplied>o los lobos p<ex>re</ex>so lo an a
don jua<ex>n</ex> en<ex>e</ex>l<del>pa</del> ca<ex>n</ex>po</add>
</p>
<lg>
<l>marja luz d<ex>e</ex>l dja tu nos gia todauja</l>
<l>dame grr<ex>açi</ex>a r benedeçio<ex>n</ex> r del çielo
co<ex>n</ex>solaçio<ex>n</ex></l>
<l>q<ex>ue</ex> pueda co<ex>n</ex> deuoçio<ex>n</ex> cantar de alegria</l>
<l>el p<ex>ri</ex>m<ex>er</ex>o gozo q<ex>ue</ex> se leya oujste
me<ex>n</ex>sajeria</l>
</lg>
<lg>
<l>d<ex>e</ex>l ang<ex>e</ex>l q<ex>ue</ex> a ti vjno g<ex>ra</ex>ujel
s<ex>an</ex>to r djno</l>
<l>traxote me<ex>n</ex>sajeria deujno dixote aue m<ex>aria</ex></l>
<l>desq<ex>ue</ex><del>y</del> el mandato oujste omjment<ex>e</ex> lo
rresçibiste</l>
<l>luego v<ex>ir</ex>ge<ex>n</ex> co<ex>n</ex>çebiste al fiio q<ex>ue</ex>
dios enbja</l>
</lg>
<lg>
<l>En bellén acaesçio el segundo q<ex>ua</ex>ndo nasçio</l>
<l>syn dolor aparesçio de ti v<ex>ir</ex>ge<ex>n</ex> el mjxia</l>
<l>el t<ex>er</ex>çero cue<ex>n</ex>tan las leyes q<ex>ua</ex>ndo
veniero<ex>n</ex> los rreys</l>
<l>adoraro<ex>n</ex> al q<ex>ue</ex> vees en tu braço do yazia</l>
</lg>
<lg>
<l>ofreciol mjria gaspar meljor ençie<ex>n</ex>so le fue dar</l>
</lg>
</body>
</text>
</TEI>

```

Evidentemente es mucho más complicado que el sistema de Madison, pero es mucho más rico. En el sistema de Madison todo se reducía a una secuencia de líneas físicas absolutamente mudas acerca de la estructura del texto, tan mudas que incluso carecen de información relativa a qué texto es, quién lo ha transcrito y otros datos identificativos que en el sistema XLM/TEI se tienen que declarar obligatoriamente en

la primera parte del fichero,<sup>24</sup> entre las etiquetas <teiHeader> ... </teiHeader> que es lo recogido en la caja amarilla, mientras que el texto es lo que se encuentra encerrado dentro de la caja verde. Las cajas marrones del inicio y del final constituyen el envoltorio último de un documento.

El contenido de la primera caja marrón es la declaración de que se trata de un documento XLM, la codificación y el namespace de la TEI. Pero dejemos este apartado, lo importante es el elemento raíz, en nuestro caso <TEI>, que se cierra en la última caja </TEI> del tipo de documento. Este elemento contiene en sí un único documento que cumpla las normas de codificación TEI y ha de estar constituido por un <teiHeader> y un <text> salvo que forme parte de un <teiCorpus>. En este caso el documento TEI estaría constituido por un único encabezamiento de corpus y la serie (colección) de elementos TEI conformado cada uno de ellos por un único encabezamiento de texto <teiHeader> y su texto <text>. Esto nos lleva a que un documento TEI puede presentar una de estas dos estructuras básicas: la de un documento TEI (que es la mínima):

```
<TEI>
  <teiHeader>
    <!-- obligatorio -->
  </teiHeader>
  <text>
    <!-- obligatorio -->
  </text>
</TEI>
```

y la de un corpus TEI:

```
<teiCorpus>
  <teiHeader>
    <!-- obligatorio -->
  </teiHeader>
  <TEI>
    <!-- obligatorio -->
  </TEI>
</teiCorpus>
```

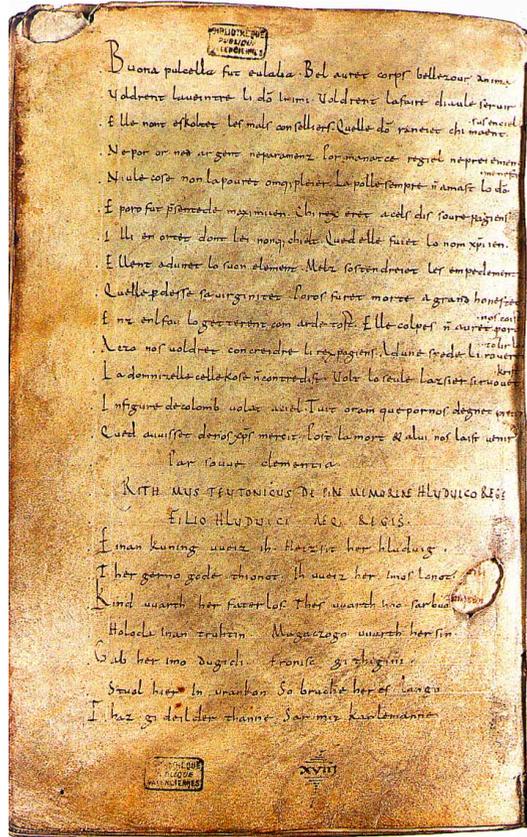
Volvamos sobre las primeras ocho líneas de la transcripción según el sistema del Hispanic Seminar of Medieval Studies de Madison:

```
Dios padre & djos fijo & dios esp<iri>tu s<an>to (^q<ue> nac'io)
q<ue> nac'io de v[<ir>]ge[<n>] esfue<r>c'a nos de tanto
q<ue> sye<n>pre lo leemos en prosa & en canto
sea de n<uest>ras almas cobertura & manto
el q<ue> fizo el c'ielo & la t<ie>rra & la mar
```

24. Los textos recogidos en el *Electronic Texts and Concordances...* (O'Neill 1999) se identifican en un fichero Word titulado intro.doc (también lo hay en texto plano intro.txt). Cada uno de los 201 textos recogidos está conformado por un conjunto de cinco ficheros cuyos nombres son text «the semi-paleographic text transcription», concord «complete reference concordances», alfindx «alphabetic frequency indexes», desindx «descending frequency indexes» y revindx «reverse alphabetic frequency indexes» (O'Neill 1999: díptico) y cada conjunto se diferencia de los demás en virtud de las extensiones de tres caracteres, las cuales se explican en los ficheros intro.doc (o intro.txt). En ese fichero la información bibliográfica de cada texto es código de tres letras, título, ciudad, repositorio (biblioteca, archivo), signatura y nombre del transcriptor.

el me de la su grr<ac'i>a & me q<uie>ra alubrar  
 q<ue> pueda de sus cantar<e>s vn ljbro rrimar  
 q<ue> los q<ue> lo oyere<n> puedan solaz' tomar

Sin el aviso previo de los editores, un lector no podría saber si el texto es prosa o verso salvo tras una demorada lectura y análisis.



Manuscrito de la Secuencia de Santa Eulalia

Recurramos a un caso *extremo*. La *Secuencia de Santa Eulalia* es una obra constituida por catorce períodos rítmicos de longitud desigual conformado cada uno de ellos por un pareado. En el manuscrito (ms. 150 de la Bibliothèque Municipale de Valennciennes) cada uno de los períodos rítmicos está copiado en una única línea en la que los dos constituyentes del pareado están separados por un punto. En el sistema de transcripción de Madison todo sería una secuencia, valga la redundancia, de catorce líneas que a veces riman (asonante) a veces no:<sup>25</sup>

Buona pulcella fut eulalia. Bel auret corps bellezour anima  
 Voldrent la ueintre li d<e>o inimi. Voldrent la faire diaule seruir.

25. Para no complicar más la codificación, no se tiene en cuenta el hecho de que algunos finales de línea se han copiado en la interlínea superior.

Elle nont eskoltet les mals conselliers. Quelle d<e>o raneiet chi maent sus en ciel.  
 Ne por or ned argent ne paramenz. Por manatce regiel ne preiemet  
 Niule cose non la pouret omq<ue> pleier. La polle sempre n<on> amast lo d<e>o menestier  
 E por o fut p<re>sentede maximiiien. Chi rex eret a cels dis soure pagin<e>s  
 Il li enortet dont lei nonq<ue> chielt. Qued elle fuiet lo nom xp<ist>iien.  
 Ellent adunet lo suon element. Melz sostendriet les empedementz.  
 Quelle p<er>desse sa uirginitet. Por os suret morte a grand honestet.  
 Enz enl fou la getterent com arde tost. Elle colpes n<on> auret por o nos coist.  
 A czo nos uoldret concreidre li rex pagiens. Ad une spede li roueret tolr lo chief.  
 La domnizelle celle kose n<on> contredist. Volt lo seule lazsier si ruouet krist.  
 In figure de colomb uolat a ciel. Tuit oram que por nos degnet preier.  
 Qued auisset de nos xp<istu>s mercit. Post la mort & a lui nos laist uenir  
 Par souue clementia.

Con la codificación XML/TEI esta peculiaridad gráfica del único testimonio se puede declarar y mantener. Cada línea física se marcaría con la etiqueta vacía <lb/> y cada verso se encerraría con el elemento <l>...</l> y cada pareado, si se estimara pertinente, se marcaría con <lg>...</lg>:

```
<lb n="1"/><lg><l>Buona pulcella fut eulalia.</l> <l>Bel auret corps belle-zour anima</l></lg>
<lb n="2"/><lg><l>Voldrent la ueintre li d<ex>e</ex>o inimi.</l> <l>Voldrent la faire diaule
seruir.</l></lg>
<lb n="3"/><lg><l>Elle nont eskoltet les mals conselliers.</l> <l>Quelle d<ex>e</ex>o ra-
neiet chi maent sus en ciel.</l></lg>
<lb n="4"/><lg><l>Ne por or ned argent ne paramenz.</l> <l>Por manatce re-giel ne preie-
ment</l></lg>
<lb n="5"/><lg><l>Niule cose non la pouret omq<ex>ue</ex> pleier.</l> <l>La polle sempre
n<ex>on</ex> amast lo d<ex>e</ex>o menestier</l></lg>
<lb n="6"/><lg><l>E por o fut p<ex>re</ex>sentede maximiiien.</l> <l>Chi rex eret a cels
dis soure pagin<ex>e</ex>s</l></lg>
<lb n="7"/><lg><l>Il li enortet dont lei nonq<ex>ue</ex> chielt.</l> <l>Qued elle fuiet lo
nom xp<ex>ist</ex>iien.</l></lg>
<lb n="8"/><lg><l>Ellent adunet lo suon element.</l> <l>Melz sostendriet les empedemen-
tz.</l></lg>
<lb n="9"/><lg><l>Quelle p<ex>er</ex>desse sa uirginitet.</l> <l>Por os su-ret morte a
grand honestet.</l></lg>
<lb n="10"/><lg><l>Enz enl fou la getterent com arde tost.</l> <l>Elle colpes n<ex>on</ex>
auret por o nos coist.</l></lg>
<lb n="11"/><lg><l>A czo nos uoldret concreidre li rex pagiens.</l> <l>Ad une spede li roue-
ret tolr lo chief.</l></lg>
<lb n="12"/><lg><l>La domnizelle celle kose n<ex>on</ex> contredist.</l> <l>Volt lo seule
lazsier si ruouet krist.</l></lg>
<lb n="13"/><lg><l>In figure de colomb uolat a ciel.</l> <l>Tuit oram que por nos degnet
preier.</l></lg>
<lb n="14"/><lg><l>Qued auisset de nos xp<ex>istu</ex>s mercit.</l> <l>Post la mort
& a lui nos laist uenir</l></lg>
<lb n="15"/><p><space/>Par souue clementia.</p>
```

Esto es lo que vería el codificador:

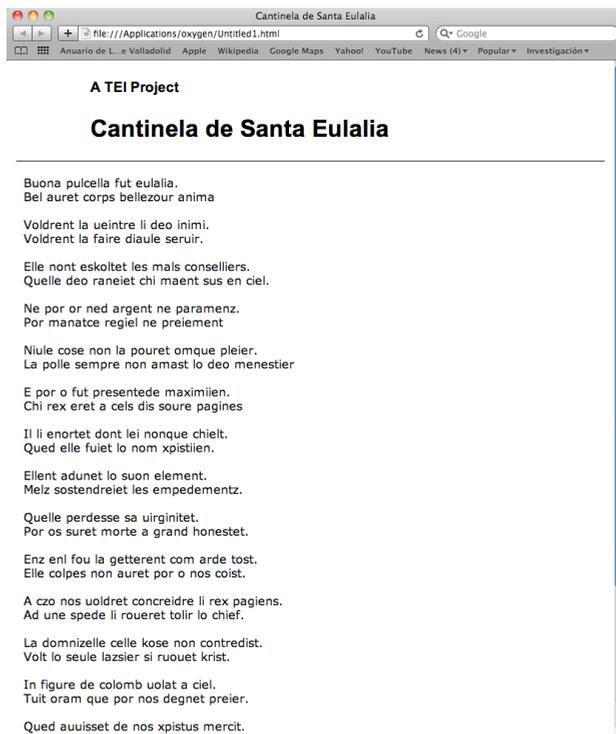
```

1 <?xml version="1.0" encoding="UTF-8" standalone="no" ?>
2 <xsl:stylesheet xmlns:xsl="http://www.w3.org/1999/XSL/Transform" version="1.0" ?>
3   <xsl:output method="text" encoding="UTF-8" ?>
4   <xsl:template match="/" ?>
5     <xsl:element name="text" ?>
6       <xsl:apply-templates ?>
7     </xsl:element>
8   </xsl:template>
9   <xsl:template match="text" ?>
10    <xsl:apply-templates ?>
11  </xsl:template>
12  <xsl:template match="text" ?>
13    <xsl:apply-templates ?>
14  </xsl:template>
15  <xsl:template match="text" ?>
16    <xsl:apply-templates ?>
17  </xsl:template>
18  <xsl:template match="text" ?>
19    <xsl:apply-templates ?>
20  </xsl:template>
21  <xsl:template match="text" ?>
22    <xsl:apply-templates ?>
23  </xsl:template>
24  <xsl:template match="text" ?>
25    <xsl:apply-templates ?>
26  </xsl:template>
27  <xsl:template match="text" ?>
28    <xsl:apply-templates ?>
29  </xsl:template>
30  <xsl:template match="text" ?>
31    <xsl:apply-templates ?>
32  </xsl:template>
33  <xsl:template match="text" ?>
34    <xsl:apply-templates ?>
35  </xsl:template>
36  <xsl:template match="text" ?>
37    <xsl:apply-templates ?>
38  </xsl:template>
39  <xsl:template match="text" ?>
40    <xsl:apply-templates ?>
41  </xsl:template>
42  <xsl:template match="text" ?>
43    <xsl:apply-templates ?>
44  </xsl:template>
45  <xsl:template match="text" ?>
46    <xsl:apply-templates ?>
47  </xsl:template>
48  <xsl:template match="text" ?>
49    <xsl:apply-templates ?>
50  </xsl:template>
51  <xsl:template match="text" ?>
52    <xsl:apply-templates ?>
53  </xsl:template>
54  <xsl:template match="text" ?>
55    <xsl:apply-templates ?>
56  </xsl:template>
57  <xsl:template match="text" ?>
58    <xsl:apply-templates ?>
59  </xsl:template>
60  <xsl:template match="text" ?>
61    <xsl:apply-templates ?>
62  </xsl:template>
63  <xsl:template match="text" ?>
64    <xsl:apply-templates ?>
65  </xsl:template>
66  <xsl:template match="text" ?>
67    <xsl:apply-templates ?>
68  </xsl:template>
69  <xsl:template match="text" ?>
70    <xsl:apply-templates ?>
71  </xsl:template>
72  <xsl:template match="text" ?>
73    <xsl:apply-templates ?>
74  </xsl:template>
75  <xsl:template match="text" ?>
76    <xsl:apply-templates ?>
77  </xsl:template>
78  <xsl:template match="text" ?>
79    <xsl:apply-templates ?>
80  </xsl:template>
81  <xsl:template match="text" ?>
82    <xsl:apply-templates ?>
83  </xsl:template>
84  <xsl:template match="text" ?>
85    <xsl:apply-templates ?>
86  </xsl:template>
87  <xsl:template match="text" ?>
88    <xsl:apply-templates ?>
89  </xsl:template>
90  <xsl:template match="text" ?>
91    <xsl:apply-templates ?>
92  </xsl:template>
93  <xsl:template match="text" ?>
94    <xsl:apply-templates ?>
95  </xsl:template>
96  <xsl:template match="text" ?>
97    <xsl:apply-templates ?>
98  </xsl:template>
99  <xsl:template match="text" ?>
100   <xsl:apply-templates ?>
101 </xsl:template>
102 </xsl:stylesheet>

```

## Codificación XML/TEI con &lt;Oxygen&gt;

no los usuarios (lectores) quienes, por medio de las hojas de estilo XLST (*Extensible Stylesheet Language–Transformation*),<sup>26</sup> tendrían una visión mucho más sencilla del texto, como se puede observar en la siguiente imagen:



La codificación XML transformada en página HTML (vista en Safari)

26. XLST es un lenguaje que sirve para transformar los documentos XML. La entrada de un programa XSLT es uno o más documentos XML y la salida es otro documento que puede estar en HTML, XML o texto. Con XSLT-FO se generan ficheros en formato PDF. Para una introducción *elemental* al XSLT véase <http://www.xml.com/pub/a/2000/08/holman/index.html> y para el XSL-FO <http://www.xml.com/pub/a/2000/08/holman/index.html>.



hasta el punto de que podrían diseñarse con distintos niveles de acceso y presentación, dependiendo de qué es lo que le interese al usuario (investigador).

La codificación según el modelo XLM/TEI puede ser mucho más rica. Por ejemplo, en el tachado del primer verso, marcado con `<del> ... </del>` puede indicarse quién y cómo se ha realizado la obliteración por medio de los atributos `@resp="..."` (quién borró) y `@type="..."` (cómo lo hizo). Las estrofas pueden ofrecer información más precisa con los atributos `@type="..."` (tipo de estrofa), `@subtype="..."` (subtipo), `@met="..."` (esquema métrico) y `@rhyme="..."` (esquema de rimas). Asimismo, el elemento `<l>` puede incorporar el atributo `@part="..."` con el que se especifica si el verso es métricamente completo o no (de especial interés en el teatro).

Tiene la posibilidad de agrupar entre las etiquetas `<choice> ... </choice>` codificaciones alternativas. Por ejemplo, en los errores se puede marcar lo que el transcriptor cree que es un error con `<sic> ... </sic>` y la enmienda con `<corr> ... </corr>`. También permite solventar, de una vez por todas, el viejo debate entre transcripciones regularizadas o no encerrando entre `<orig> ... </orig>` la forma original y entre `<reg> ... </reg>` la regularizada.<sup>27</sup> También puede poner fin a los problemas que supone el desarrollo de las abreviaturas al ofrecer entre `<abbr> ... </abbr>` la forma tal cual aparece en el original y entre `<expan> ... </expan>` el desarrollo de la abreviatura y entre `<ex> ... </ex>` la parte que el editor/transcriptor añade para desarrollar la abreviatura.<sup>28</sup>

Evidentemente, cuanto más detallado sea el nivel de etiquetado tanto mayor será el valor añadido de la transcripción y, por consiguiente, más compleja será la realización.

Este sistema de codificación es el que estamos utilizando en el Archivo Iberoamericano de Cetrería. Inicialmente se iba a codificar el ms. de la Biblioteca Bartolomé March, el cual contiene una versión textual de la traducción castellana del *De regimini principum* de Gil de Roma, sin embargo, hemos creído mucho más útil aplicarlo a los textos medievales del AIC. Para llevarlo a cabo, se están transcribiendo los manuscritos del *Libro de la caza de las aves* de Pero López de Ayala y los apéndices textuales que bajo el título genérico de *Recetas varias* aparecen al final de algunos de los manuscritos de la obra del Canciller.

Para la codificación utilizamos el programa `<oxygen/>`,<sup>29</sup> un editor XML que, sin duda alguna, es el mejor del mercado puesto que tiene implementado el sistema de codificación TEI, utiliza los esquemas de lenguaje XML más modernos como RELAX NG, incluye plantillas para la transformación de cualquier fichero XML/TEI a HTML y PDF,

27. En el Menota Project han desarrollado una ampliación del estándar TEI que han especificado en el *namespace* pertinente y que se declara en el elemento raíz de todos sus documentos (`<TEI xmlns="http://www.tei-c.org/ns/1.0" xmlns:me="http://www.menota.org/ns/1.0">`) por el que han establecido tres niveles diferentes de transcripción que se codifican con los elementos `<me:fac>`, `<me:dipl>` y `<me:norm>`. `<me:fac>` es una representación facsimilar del códice que editan. En este nivel se transcribe grafema a grafema y línea a línea y toda variación alográfica es reflejada con gran precisión y se copian las marcas abreviativas sin desarrollarlas. Este nivel para su correcta representación gráfica (en pantalla o sobre papel) necesita de una fuente tipográfica especial. En `<me:dipl>` no se codifican todas las variaciones alográficas y en la impresión (ya sea en pantalla o sobre otro soportes) no se mantiene la longitud de líneas y todas las abreviaturas se marcan encerradas con el elemento `<ex>...</ex>`. En `<me:norm>` se regulariza la ortografía «according to the norm found in grammars and dictionaries for the language in question» (Haugen 2008: § 3.2.3).

28. La palabra tierra podría codificarse `<expand>t<ex>ie</ex>rra</expand>`, en donde el elemento `<expand>...</expand>` indica que la palabra se encuentra abreviada en el original mientras que el elemento `<ex>...</ex>` marca las letras que el transcriptor ha incorporado para desarrollar la abreviatura, aunque no creemos pertinente tal minuciosidad, y para el desarrollo de las abreviaturas bastaría con el elemento `<ex>...</ex>`, con lo que el caso aducido se codificaría `t<ex>ie</ex>rra`.

29. <http://www.oxygenxml.com/> [acceso 27.04.2010].

además es multiplataforma, es decir, hay versiones para Windows, Mac OS X o Linux, las licencias académicas individuales son muy baratas (64 USD, aprox. 48 €) y cubren todas las máquinas que pueda tener el propietario de la licencia. Por otra parte, aunque es un programa muy complejo (y no se distribuye en español), es de uso relativamente sencillo que se preocupará de que el texto que se codifique esté bien formado (*well-formedness*), es decir, que cumple todas las reglas inherentes al lenguaje XML, por ejemplo:<sup>30</sup>

1. Toda etiqueta de inicio ha de tener su correspondiente etiqueta de cierre.
2. Los elementos se pueden anidar pero no solapar.
3. Que exista un único elemento raíz.
4. Que el valor de los atributos esté entrecomillado.
5. Que un mismo elemento no tenga dos atributos con el mismo nombre.
6. Que los comentarios e instrucciones de procesado no estén dentro de una etiqueta.
7. Que los signos < y &, y otros signos prohibidos, cuando han de aparecer como tales dentro del texto estén correctamente codificados como entidades.<sup>31</sup>

Y también controlará que no solo esté bien formado, sino que además sea un documento válido, es decir que se atenga a la sintaxis, en nuestro caso de la TEI, la cual especifica qué elementos y atributos son válidos, qué atributos pueden contener cada elemento y cómo se pueden anidar, es decir, qué elementos pueden encerrar en sí otros elementos ya que tiene incluida el esquema DTD (en realidad Relax NG). Se trata de un fichero que contiene la descripción de la estructura y sintaxis de la TEI. En ningún caso comprobará si el contenido es válido o no. Se puede tener un documento XML bien formado y válido según la TEI, pero que lo que contiene sea un absoluto sinsentido.

```

1 <?xslt xmlns:rs="http://www.tei-c.org/release/xml/tei/custom/schema/relaxng/tei_all.rng" type="xml"?>
2 <TEI xmlns:x1="http://www.w3.org/2001/XInclude" xmlns:svg="http://www.w3.org/2000/svg"
3   xmlns:math="http://www.w3.org/1998/Math/MathML" xmlns="http://www.tei-c.org/ns/1.0">
4   <teiHeader>
5     <fileDesc>
6       <titleStmt>
7         <title>Receta anónima</title>
8       </titleStmt>
9       <publicationStmt>
10        <p>Valladolid, AIC, mayo 2010</p>
11      </publicationStmt>
12      <sourceDesc>
13        <p>El texto se encuentra en el último folio (fol. 79v) del manuscrito
14          Krahe de Pero López de Ayala</p>
15      </sourceDesc>
16    </fileDesc>
17  </teiHeader>
18  <text>
19    <body>
20      <p>
21        <lb>>Tomad el cuero <ex></ex></lb>
22        <lb>>gato p<ex>ar</ex>a pieuelas &amp; p<ex>ar</ex>a
23        <lb>>lo fazer grueso, aca<supplied resp="editor"></supplied>
24        <lb>>bado de desollar lancete
25        <lb>>en pelambre &amp; q<ex>ue</ex> este
26        <lb>>ende fasta q<ex>ue</ex> sea pelado
27        <lb>>todo. E despu<ex></ex>s tomad
28        <lb>>agua caliente co<ex>n</ex> casca
29        <lb>>de cuma<ex>ue</ex> molida &amp;
30        <lb>>fasta veynte mojados
31        <lb>>ce<supplied rend="interlineado" resp="copista"></supplied>nida &amp; figos pasos
32        <lb>>pica<supplied resp="editor"></supplied> fierua el
33        <lb>>agua bu<ex>n</ex> rrato. E des<supplied resp="editor"></supplied>
34        <lb>>pues q<ex>ue</ex> estouiere tibia
35        <lb>>tomad el cuero &amp; echal<supplied resp="editor"></supplied>
36        <lb>>do dentro &amp; esfrie
37        <lb>>mucho el cuero en el
38        <lb>>agua con <ex></ex>l cuma<ex>ue</ex> &amp;
39        <lb>>con los figos &amp; este
40        <lb>>ende vn día &amp; vna
41        <lb>>noche
42      </p>
43    </body>
44  </text>
45 </TEI>
46

```

30. Una lista exhaustiva puede verse en Harold y Means 2004: 367-399.

31. Las entidades (ing. *entities*) comienzan por el símbolo & y finalizan con un punto y coma ;. Son una especie de comodines que evitan que el software pueda considerar un documento como no bien formado. Así los símbolos de menor que (<), mayor que (>), el ampersand (&) y las comillas, tanto simples (') como dobles (") cuando son ellos mismos deben aparecer como &lt; &gt; &amp; &quot; y &apos;.

La imagen anterior corresponde a un fichero (mínimo) de uno de los textos codificados para el AIC. Está bien formado de acuerdo con las reglas del lenguaje XML y, además, es válido según las especificaciones TEI (y el texto que codifica tiene sentido). Se pueden observar las dos grandes partes que nos conciernen: el <teiHeader> y el <text>.

### *El <teiHeader>*

Como se ha expuesto más atrás, todo documento TEI ha de contener un elemento <teiHeader> en el que se describe el fichero de manera que quede perfectamente documentado el texto mismo, la fuente (ya sea manuscrita o impresa), el proceso de codificación y un registro detallado de las revisiones que se han realizado. Este encabezado consta de cuatro partes:

<fileDesc> descripción del fichero; contiene la descripción bibliográfica del fichero electrónico.

<encodingDesc> descripción de la codificación; describe la relación que existe entre el fichero electrónico y su fuente (sea manuscrita, sea impresa). Debe dar cuenta de las normas y pautas de codificación seguidas durante el proceso.

<revisionDesc> crónica de las revisiones; da cuenta de todas las modificaciones, correcciones, adiciones, etc. que se hayan introducido en el fichero.

<profileDesc> Perfil del fichero. Es una especie de cajón de sastre dondese puede incluir toda aquella información que no es estrictamente bibliográfica.

Cuando se abre un fichero nuevo con <oXygen/> (File > New > From Template > TEI P5 – All > OK), la plantilla instalada en el programa ofrece un <teiHeader> mínimo cuyo aspecto es el siguiente:

```
<teiHeader>
  <fileDesc>
    <titleStmt>
      <title>Title</title>
    </titleStmt>
    <publicationStmt>
      <p>Publication Information</p>
    </publicationStmt>
    <sourceDesc>
      <p>Information about the source</p>
    </sourceDesc>
  </fileDesc>
</teiHeader>
```

Es, decir, de las cuatro partes que puede constar un <teiHeader> la única obligatoria es <fileDesc>, las otras tres son absolutamente opcionales. En la sección <fileDesc>, que puede contener en sí siete elementos más, tan solo <titleStmt>, <publicationStmt> y <sourceDesc> son obligatorios<sup>32</sup> y agrupan dentro de sí mismo otros elementos, generalmente especializados que recogen información estructurada. De nuevo hay unos elementos mínimos. En <titleStmt> se trata de <title> y encierra el título de la siguiente manera:

```
<title>Receta anónima</title>
```

32. Son opcionales <editionStmt>, <extent>, <seriesStmt> y <notesStmt>.

En <publicationStmt> se recoge la información referente a la publicación o distribución del fichero electrónico. Puede ser un sencillo párrafo:

```
<p>Valladolid, AIC, 2010</p>
```

u ofrecer una serie de elementos que presentan la información de forma más clara y estructura:

```
<publicationStmt>
  <publisher>Archivo Iberoamericano de Cetrería</publisher>
  <pubPlace>Valladolid</pubPlace>
  <date when="2010">2010</date>
</publicationStmt>
```

En <sourceDesc> se describe la fuente de la que se ha derivado el fichero electrónico. Puede ser un párrafo o serie de párrafos encerrado con los elementos <p> ... </p>:

```
<p>El texto se encuentra en el último folio (fol. 79v) del manuscrito Krahe de Pero López de Ayala</p>
```

o puede complicarse con referencias bibliográficas encerradas en el elemento <bibl> e incluso ser toda una descripción detallada de un manuscrito por medio del elemento <msDesc> (ofrecemos una mínima descripción):

```
<msDesc>
  <msIdentifier>
    <settlement>Madrid</settlement>
    <repository>Colección Krahe</repository>
    <idno>no disponible</idno>
  </msIdentifier>
  <msContents>
    <msItem>
      <author>Pero López de Ayala</author>
      <title>Libro de la caza de las aves</title>
    </msItem>
  </msContents>
  <physDesc>
    <objectDesc>
      <p>Papel. 208 x 146 mm. 79 folios (foliado 1-80, falta 48; último folio a dos columnas).</p>
    </objectDesc>
  </physDesc>
</msDesc>
```

Tras la <fileDesc> se incorporará la <encodingDesc> que puede ser un párrafo (o serie de párrafos) en texto libre encerrados entre <p>...</p> como, por ejemplo:

```
<p>Se ha mantenido la longitud de línea y se ha numerado cada una de ellas. Se han desarrollado las abreviaturas. Se han unido y separado las palabras con criterio lexicográfico. Las palabras cortadas al final de línea se han unido por medio de un guión suplido por el editor. La nota tironiana se ha transcrito con el ampersand (&);. </p>
```

El tercer elemento que constituye el <teiHeader> es la sección <revisionDesc>, en la que da se cuenta de cuantos cambios ha habido en el fichero, quién los ha realizado

y cuándo.<sup>33</sup> El elemento básico de esta parte es <change> y es de suma importancia el atributo @when ‘cuándo’:

```
<revisionDesc>
  <change when="2010-03-30">Se compulsó el texto con el
    original</change>
</revisionDesc>
```

Todas estas secciones, y la no examinada de <profileDesc>, pueden ser mucho más detalladas, es decir, contener un etiquetado mucho más rico y profundo que podría convertir el <teiHeader> en un auténtico estudio introductorio y, además, en un diario del proceso editorial del mismo.

### *El <text>*

El texto ofrecido como ejemplo es muy sencillo. Consta de un único párrafo (marcado con <p>) y, dado que hemos estimado que la división en líneas puede ser interesante, se ha marcado cada una de ellas con el elemento vacío <lb/>,<sup>34</sup> que se puede presentar como en el ejemplo, para mayor facilidad de lectura y comprobación, o se podría disponer como un bloque a línea tirada con las marcas de línea anidadas en el mismo:<sup>35</sup>

```
<p><lb/>Tomad el cuero d<ex>e</ex>l <lb/>gato p<ex>ar</ex>a pieuelas &amp;
p<ex>ar</ex>a <lb/>lo fazer grueso, acabado de desollar lançalo <lb/>en pelanbre &amp;
q<ex>ue</ex> este <lb/>ende fasta q<ex>ue</ex> sea pelado <lb/>todo. E despu<ex>e</
ex>s tomad <lb/>agua caliente co<ex>n</ex> casca <lb/>de çumaq<ex>ue</ex> molida
&amp; <lb/>çe<supplied rend="interlineado" resp="copista">r</supplied>nida &amp; figos
pasos <lb/>fasta veynte mojados <lb/>&amp; picados q<ex>ue</ex> fierua el <lb/>agua
bue<ex>n</ex> rrato. E des<lb/>pues q<ex>ue</ex> estouiere tibia <lb/>tomad el cuero
&amp; echal<lb/>do dentro &amp; esfrie <lb/>mucho el cuero en el <lb/>agua con <ex>e</
ex>l çumaq<ex>ue</ex> &amp; <lb/>con los figos &amp; este <lb/>ende vn día &amp; vna
<lb/>noche</p>
```

o sin ellas:

```
<p>Tomad el cuero d<ex>e</ex>l gato p<ex>ar</ex>a pieuelas &amp; p<ex>ar</ex>a
lo fazer grueso, acabado de desollar lançalo en pelanbre &amp; q<ex>ue</ex>
este ende fasta q<ex>ue</ex> sea pelado todo. E despu<ex>e</ex>s tomad
agua caliente co<ex>n</ex> casca de çumaq<ex>ue</ex> molida &amp;
çe<supplied rend="interlineado" resp="copista">r</supplied>nida &amp; figos
pasos fasta veynte mojados &amp; picados q<ex>ue</ex> fierua el
agua bue<ex>n</ex> rrato. E despues q<ex>ue</ex> estouiere tibia
tomad el cuero &amp; echaldo dentro &amp; esfrie mucho el cuero en el
agua con <ex>e</ex>l çumaq<ex>ue</ex> &amp; con los figos &amp; este
ende vn día &amp; vna noche</p>
```

33. Una recomendación que se suele hacer es que al describir los cambios se organicen de los más modernos a los más antiguos.

34. Este elemento puede llevar el atributo @n con el que se podrían numerar todas y cada una de las líneas, aunque contar es algo que los ordenadores saben hacer muy bien.

35. Estas dos visiones son muy complicadas a la hora de leerlas, por lo que en el AIC hemos decidido mantener la longitud de líneas y hacer uso del elemento vacío <lb/> junto con el atributo @n (<lb n="1"/>).

Hemos seleccionado un microtexto muy sencillo debido a la claridad estructural que ofrece. Sin embargo, dentro de <text> no solo aparece el <body> del texto codificado sino cualquier material que le pueda preceder <front> y seguir <back>. Tanto el elemento <front> como el <back>— no son obligatorios en ningún documento TEI.

Supongamos que queremos codificar una edición moderna de un texto medieval que consiste en un estudio introductorio, el texto editado, y una serie de apéndices. El estudio introductorio se transcribiría dentro del elemento <front>, mientras que los apéndices se codificarían dentro del elemento <back> y el texto medieval objeto de la edición se encontraría dentro del elemento <body>. El esquema estructural sería:

```
<TEI>
  <teiHeader>
    <!-- Aquí iría toda la información del encabezado -->
  </teiHeader>
  <text>
    <front>
      <!-- Estudio introductorio -->
    </front>
    <body>
      <!-- Edición del texto -->
    </body>
    <back>
      <-- Apéndices -->
    </back>
  </text>
</TEI>
```

Aparentemente, y de acuerdo con el esquema que se puede extraer de lo visto hasta ahora

```
<TEI>
  <teiHeader>
    <!-- obligatorio -->
  </teiHeader>
  <text>
    <!-- obligatorio -->
  </text>
</TEI>
```

solo podríamos incluir un único texto en cada documento TEI. Sin embargo, el sistema es lo suficientemente flexible como para permitir que dentro de un mismo documento TEI se editen varios textos. Dos casos diferentes, pero interesantes como ejemplos, son el ms. 6.376 de la Biblioteca Nacional de España y el ms. V.II.19 de El Escorial. El primero contiene todas las obras, salvo una, de don Juan Manuel, el segundo recoge una colección de cinco obras romances y una latina con unidad temática: todas ellas son tratados de cetrería. En este caso tras el elemento <text> se incorpora el elemento <group>, y después cada una de las obras que conforman la colección dentro del par <text><body>. La estructura básica del documento TEI que contendría la transcripción del ms, 6.376 de la BNE sería:

```

<TEI>
  <teiHeader>
    <!-- encabezamiento -->
  </teiHeader>
  <text>
    <group>
      <text>
        <body>
          <!-- Prólogo -->
        </body>
      </text>
      <text>
        <body>
          <!-- -->
        </body>
      </text>
    </group>
  </text>
  <!-- Fin de las obras de don Juan Manuel -->
</TEI>

```

la del ms. V.II.19 de El Escorial no sería muy diferente:

```

<TEI>
  <teiHeader>
    <!-- información del encabezamiento -->
  </teiHeader>
  <text>
    <group>
      <text>
        <body>
          <!-- Libro de los animales que cazan -->
        </body>
      </text>
      <text>
        <body>
          <!-- Tratado de las enfermedades de las aves -->
        </body>
      </text>
      <text>
        <body>
          <!-- Dancus rex -->
        </body>
      </text>
      <text>
        <body>
          <!-- Guillelmus falconarius -->
        </body>
      </text>
      <text>
        <body>
          <!-- Libro de los azores -->
        </body>
      </text>
      <text>
        <body>
          <!-- Practica avium -->
        </body>
      </text>
    </group>
  </text>
  <!-- Fin del ms. V.II.19 -->
</TEI>

```

Este modo de agrupar es el correcto cuando cada una de las obras constituye una unidad textual independiente, o lo que es lo mismo, cuando se desea que el documento TEI contenga más de una obra. No se debe emplear cuando una obra se subdivide en varios libros, pues en ese caso se debe utilizar el elemento <div>.

El cuerpo —<body>— de un texto se puede subdividir por medio del uso del elemento <div>, el cual se puede ir anidando recursivamente según se avanza en la estructura jerárquica del texto. Para hacer fácil el seguimiento de la estructura y del progresivo anidamiento de los diferentes niveles se utilizan los atributos globales @type y @n. Los

posibles valores que se fijen para el atributo @type se deben declarar en el <teiHeader>, en la sección dedicada a <encodingDesc>. Así, por ejemplo, la estructura del Libro de la caza de las aves de Pero López de Ayala se podría representar de la siguiente manera:

```
<TEI>
  <teiHeader>
    <!-- Aquí iría toda la información del encabezado -->
  </teiHeader>
  <text>
    <body>
      <div type="prologo">
        <!-- Texto del prólogo -->
      </div>
      <div type="tabla">
        <!-- Texto de la tabla -->
      </div>
      <div type="capitulo" n="1">
        <!-- Texto del capítulo 1 -->
      </div>
      <div type="capitulo" n="2">
        <!-- Texto del capítulo 2 -->
      </div>
      <!-- Texto de los capítulos 3 a 46 encerrados en <div>...</div> -->
      <div type="capitulo" n="47">
        <!-- Texto del capítulo 47 -->
      </div>
      <div type="colofon">
        <!-- Texto del colofón -->
      </div>
    </body>
  </text>
</TEI>
```

Cuando los textos son muy elementales en su estructura lógica no haría falta especificar ninguno de esos dos atributos. Sin embargo, cuando la obra presenta una estructura compleja dividida en libros, partes y capítulos o presenta otros elementos que complican la estructura lógica del texto, como pueden ser glosas, se puede anidar sucesivamente el elemento <div>.

El *Regimiento de príncipes* de Gil de Roma está constituido por tres libros que a su vez se subdividen en cuatro, tres y tres partes cada uno de ellos y cada parte tiene de un número variable de capítulos; su estructura sería la siguiente:

```
<TEI>
  <teiHeader>
    <!-- Aquí iría toda la información del encabezado -->
  </teiHeader>
  <text>
    <body>
      <div type="libro" n="1">
        <div type="parte" n="1.1">
```

```

    <div type="capitulo" n="1.1.1">
      <!-- Texto del capítulo 1 -->
    </div>
    <div type="capitulo" n="1.1.2">
      <!-- Texto del capítulo 2 -->
    </div>
    <!-- Los demás capítulos de la parte 1 -->
  </div><!-- Fin parte 1 -->
  <div type="parte" n="1.2">
    <div type="capitulo" n="1.2.1">
      <!-- Texto del capítulo 1 -->
    </div>
    <div type="capitulo" n="1.2.2">
      <!-- Texto del capítulo 2 -->
    </div>
    <!-- Los demás capítulos de la parte 2 -->
  </div><!-- Fin parte 2 -->
  <!-- Demás partes del libro 1 -->
</div><!-- Fin libro 1 -->
<div type="libro" n="2">
  <div type="parte" n="2.1">
    <div type="capitulo" n="2.1.1">
      <!-- Texto del capítulo 1 -->
    </div>
    <div type="capitulo" n="2.1.2">
      <!-- Texto del capítulo 2 -->
    </div>
    <!-- Los demás capítulos de la parte 1 -->
  </div><!-- Fin parte 1 -->
  <div type="parte" n="2.2">
    <div type="capitulo" n="2.2.1">
      <!-- Texto del capítulo 1 -->
    </div>
    <div type="capitulo" n="2.2.2">
      <!-- Texto del capítulo 2 -->
    </div>
    <!-- Los demás capítulos de la parte 2 -->
  </div><!-- Fin parte 2 -->
  <!-- Demás partes del libro 2 -->
</div><!-- Fin libro 2 -->
<div type="libro" n="3">
  <div type="parte" n="3.1">
    <div type="capitulo" n="3.1.1">
      <!-- Texto del capítulo 1 -->
    </div>
    <div type="capitulo" n="3.1.2">
      <!-- Texto del capítulo 2 -->
    </div>
    <!-- Los demás capítulos de la parte 1 -->
  </div><!-- Fin parte 1 -->
  <div type="parte" n="3.2">

```

```

<div type="capitulo" n="3.2.1">
  <!-- Texto del capítulo 1 -->
</div>
<div type="capitulo" n="3.2.2">
  <!-- Texto del capítulo 2 -->
</div>
  <!-- Los demás capítulos de la parte 2 -->
</div><!-- Fin parte 2 -->
  <!-- Demás partes del libro 2 -->
</div><!-- Fin libro 3 -->
</body>
</text>
</TEI>

```

Los elementos hasta aquí vistos se pueden utilizar tanto en obras en prosa como en verso. Pero ambos tipos de obras tienen unidades básicas totalmente diferentes.

### A modo de conclusión

Esta páginas han pretendido ser una introducción elemental y básica a la codificación TEI de textos medievales. El mundo que abre es inmenso debido a las múltiples aplicaciones y posibilidades que ofrece. Hasta ahora los hispanomedievalistas que se han dedicado a editar textos lo han hecho con procesadores de texto que codifican lo introducido de una manera peculiar. Los más veteranos empezaron con WordPerfect 4.0, o incluso antes con el WordStar<sup>36</sup> y, cuando han ido migrando hacia procesadores más modernos y más complejos, básicamente Word de Microsoft, se han encontrado con que mucha de la información se pierde en el camino. Esto no ocurrirá en las ediciones preparadas con el estándar XML/TEI, de ahí su gran futuro. Por otra parte, los textos introducidos en un procesador de textos, de cualquier índole, no están lo suficientemente estructurados y desambiguados como para poderlos procesar automáticamente puesto que su marcado es presentacional.

### Bibliografía

- FAULHABER, Charles B. (1992), «El *Text Encoding Initiative* y su aplicación a la codificación textual y explotación», *Actas del Congreso de la Lengua Española*. Sevilla.
- HAROLD, Elliotte Rusty y W. SCOTT MEANS (2004), *XML in a nutshell. A Desktop Quick Reference*. Sebastopol: O'Reilly.
- HAUGEN, Odd Einar (2008), *The Menota handbook. Guidelines for the electronic encoding of Medieval Nordic primary sources*. Bergen ([http://www.menota.org/guidelines-2/contents/contents\\_2-0.page](http://www.menota.org/guidelines-2/contents/contents_2-0.page))
- LENZ, Evan (2005), *XSLT 1.0. Pocket Reference*. Sebastopol: O'Reilly.

36. WordStar era un programa diseñado y fabricado por MicroPro. Originalmente se desarrolló para el sistema operativo CP/M (1976) y posteriormente se adaptó a MS/DOS (1982) e incluso a Windows 3.0 (1991). Fue el producto dominante durante la primera mitad de la década de los años ochenta del siglo XX. Su declive vino marcado por la aparición de WordPerfect, que se convirtió en el producto estándar durante la segunda mitad de los ochenta y primera parte de los noventa. Hubo versiones para MS/DOS (1982-1997), Apple II (1985-89) y Apple Macintosh (1988-1997), Microsoft Windows (desde 1991) e incluso para Linux (1996-2005).

- LUCÍA MEGÍAS, José Manuel. (1999), reseña a «ADMYTE II: Archivo Digital de Manuscritos y Textos españoles. 'La Celestina entre la escritura medieval y humanística'. Madrid, Micronet, 1999, 2 CD-Rom», en *Incipit*, 19: 231–237.
- MACKENZIE, David (1992), «Problemas de la transcripción textual electrónica: lenguas, dialectos, máquinas», *Actas del Congreso de la Lengua Española*. Sevilla.
- MACKENZIE, David y Kenneth BUELOW (1977), *A Manual of Manuscript transcription for the Dictionary of Old Spanish Language*. Madison: HSMS.
- MARTÍN DE SANTA OLALLA SÁNCHEZ, Aurora (1993), «Una propuesta de codificación morfosintáctica para corpus de referencia en lengua española», en *Estudios de lingüística española*, 3 (<http://elies.rediris.es/elies3/>)
- O'NEILL, John (1999), *Electronic Texts and Concordances of the Madison Corpus of Early Spanish Manuscripts and Printings*. Madison y Nueva York: HSMS–Hispanic Society of America.
- PÉREZ NAVARRO, José (2008) «Cuestiones metodológicas en la edición de las partes bíblicas de la *General Estoria*», *Actas del VII Congreso Internacional de Historia de la Lengua Española: Mérida (Yucatán), 4-8 septiembre de 2006*. Madrid: Arco Libros, 2008, pp. 2065-2078
- SAINT LAURENT, Simon y Michael FITZGERALD (2005), *XLM. Pocket Reference*. Sebastopol: O'Reilly.
- TENENBAUM, Felipe (2000–01), «El sistema de transcripción del Hispanic Seminary of Medieval Studies (Madison, Wisconsin), *Incipit*, 20-21: 153-168.
- TRIMBELL, Garry (2006), «A Brief History of World Falconry», *International Association for Falconry & Conservation of Birds of Prey*, 69-75 (<http://www.i-a-f.org/new/images/files/newsletter2006.pdf> [acceso 07.05.2010]).

FRADEJAS RUEDA, José Manuel, «La codificación XML/TEI de textos medievales», *Memorabilia* 12 (2009-2010), pp. 219-247.

#### RESUMEN

---

En este artículo, tras analizar el sistema de transcripción electrónico semipaleográfico del Hispanic Seminar of Medieval Studies (HSMS) de la Universidad de Madison, diseñado para un fin específico, la redacción de un diccionario del español medieval, se presenta el lenguaje XML y sus posibilidades para la codificación digital de textos medievales. Para ello se ofrece una introducción ampliamente ejemplificada de las posibilidades que puede tener el eXtended Markup Language y en la especificación de la Text Encoding Initiative (TEI). Y se da cuenta de aplicación que se está haciendo a los manuscritos del *Libro de la caza de las aves* de Pero López de Ayala.

PALABRAS CLAVE: codificación digital, XML, TEI, edad media, castellano, Pero López de Ayala, *Libro de la caza de las aves*.

#### ABSTRACT

---

This article begins with a detailed analysis of the semi-paleographic electronic-transcription system devised by the Hispanic Seminar of Medieval Studies (HMSM) for the construction of a database to be used for compiling a dictionary of Old Spanish. The reader is then introduced to the possibilities offered by the eXtended Markup Language (XML) and the Text Encoding Initiative (TEI) standard. These possibilities are exhaustively exemplified with digital transcriptions from the extant manuscripts of Pero López de Ayala's *Libro de la caza de las aves*.

KEYWORDS: digital encoding, XML, TEI, Middle Ages, Old Spanish, Pero López de Ayala, *Libro de la caza de las aves*.

