

**RESEÑA A ISASI MARTÍNEZ, CARMEN; MARTÍN AIZPURU, LEYRE; PÉREZ ISASI, SANTIAGO; PIERAZZO, ELENA Y SPENCE, PAUL (2020):
EDICIÓN DIGITAL DE DOCUMENTOS ANTIGUOS: MARCACIÓN XML-TEI
BASADA EN LOS CRITERIOS CHARTA. SEVILLA: UNIVERSIDAD DE SEVILLA.
COLECCIÓN LINGÜÍSTICA (ISBNe 978-84-472-2204-9), 105 pp.**

PILAR ARRABAL RODRÍGUEZ
Universidad de Granada
pilararrabal@ugr.es
ORCID-ID: 0000-0003-1199-7849

En los últimos años hemos asistido a la digitalización de un gran número de corpus históricos con un claro objetivo final: posibilitar la máxima explotación de los datos. En este sentido, el desarrollo de campos como la Lingüística computacional y de nuevas técnicas en el marco de las Humanidades digitales han permitido la aparición de corpus anotados y editados que facilitan la recuperación de información de múltiples formas. La presencia de herramientas que conjuguen el rigor filológico con un procesamiento avanzado de los datos se hace ineludible a fin de obtener estudios lingüísticos fiables.

Entre los corpus de calidad filológica indiscutible destaca el corpus CHARTA¹ por su amplia representatividad del español en textos archivísticos desde los siglos XII al XIX. El mérito, además, reside en la adopción de unos sólidos criterios de edición para la transcripción de documentos (CHARTA, 2013), que han sido aplicados por los numerosos grupos internacionales de trabajo que componen la Red CHARTA. El volumen que aquí se reseña propone, partiendo de esos criterios ya establecidos, aplicar un nuevo estándar técnico de edición digital con el fin de optimizar el aprovechamiento del corpus.

De acuerdo con las prácticas actuales de edición, los autores se basan en el estándar internacional para la codificación de textos en formato digital propuesto por el consorcio TEI (*Text Encoding Initiative*)². La formulación de estas directrices se lleva a cabo en el lenguaje informático XML (*Extensible Markup Language*) que codifica el texto por medio de etiquetas que posteriormente podrán ser procesadas por las máquinas³. Este volumen, por tanto, plantea un novedoso sistema de codificación de los textos y sirve de manual al usuario interesado, ya que en él encontrará, junto a cada caso de edición, descripciones, ejemplos y las etiquetas de uso correspondientes.

¹ (Corpus Hispánico y Americano en la Red: Textos Antiguos), de acceso en: <http://www.corpusCHARTA.es/>. El corpus está elaborado por diferentes grupos de investigación, los cuales conforman la Red CHARTA. Más información sobre la Red puede consultarse a su vez en: <https://www.redCHARTA.es/>.

² <https://TEI-c.org/>

³ Las etiquetas son instrucciones formuladas entre caracteres angulares < > que engloban un fragmento de texto. Una palabra tachada, por ejemplo, se señala con la etiqueta (del inglés *deletion*).

Sin embargo, la propuesta de incorporar el estándar TEI al corpus CHARTA no es del todo reciente. La idea vio la luz en el año 2009 como base de un proyecto liderado por Carmen Isasi, Pedro Sánchez-Prieto y Paul Spence. Fruto de este trabajo de investigación y de posteriores análisis y pruebas de marcación surgió una guía completa de etiquetado CHARTA-TEI (Isasi Martínez, *et al.*, 2014) cuyo objetivo era promover su implantación. Aunque en ese momento no se llegaron a aplicar estos criterios, actualmente CHARTA ha manifestado su interés por adoptar un sistema diferente de edición de sus documentos. Este reciente volumen surge, por tanto, en un momento de nuevo favorable y ofrece una versión mejorada y ampliada de la anterior guía que implica además una actualización en lo relativo a los siguientes aspectos:

1. Etiquetas.

Desde que se publicara la guía (2014) hasta el momento de la realización de este volumen (2020) se han sucedido distintas versiones de las *TEI Guidelines* que han podido afectar al conjunto de etiquetas empleadas. Por esta razón, se ha procedido a una revisión de las etiquetas conforme a los criterios actuales de las directrices TEI.

2. Ejemplos.

A lo largo del volumen se suceden multitud de ejemplos a los que se aplica la marcación TEI. Al anterior ejemplario fundamentado en textos del *Corpus de documentación de cancillería real castellana* (CODCAR), el *Corpus de documentación medieval de Miranda de Ebro* (CODOMME), el corpus SAI (procedente de los trabajos del *Seminario Alfonsí Irigoien*) y el *Corpus de documentos Españoles Anteriores a 1700* (CODEA)⁴, se añaden ahora nuevos ejemplos y análisis de casos tomados de otros tres corpus: el *Corpus de Documentación Medieval del Sur de Ávila* (CODOMSA), el *Corpus Histórico del Español Norteño* (CORHEN)⁵ y la nueva versión del CODEA (CODEA+ 2015), que contempla una ampliación cronológica hasta 1800. De todos estos corpus se ofrece además una breve descripción en lo que se refiere a tipología documental, ámbito geográfico o periodo cronológico representado, entre otros datos de interés.

3. Contenidos.

En cuanto a los contenidos se ofrece una reestructuración de los mismos que queda evidenciada primeramente en la disposición del índice. Esta reorganización afecta sobre todo a lo referente en la propuesta de etiquetado. Se reduce asimismo el número de apéndices y se incluyen ejemplos completos de dos textos etiquetados (apéndice 4), aspecto que en la guía anterior se reducía exclusivamente al desarrollo de la cabecera.

4. Referencias.

Se ofrece una actualización de referencias en cuanto a la bibliografía y los corpus relacionados. En este sentido, el volumen destaca especialmente

⁴ Acceso en: <http://corpuscodea.es/>.

⁵ Acceso en: <https://corhen.es/>.

por la inclusión de una bibliografía harto reciente entre la que figuran referencias actuales sobre la marcación TEI aplicada a diferentes tipos de corpus o en torno a las consideraciones teóricas de la edición digital como uno de los campos integrantes de las Humanidades digitales. No es objetivo de los autores, sin embargo, hacer una revisión teórica sobre los temas mencionados, aunque no faltan en la obra alusiones, notas al pie y referencias a este tipo de publicaciones estrechamente relacionadas con la propuesta CHARTA-TEI.

5. Erratas.

Se subsanan las erratas detectadas.

La publicación se divide en dos bloques. El primero presenta una parte introductoria donde se recogen cuestiones importantes acerca del origen y recorrido del trabajo realizado, objetivos y metodología empleada. Le sigue un apartado en el que se detallan los corpus consultados para la ejemplificación de los casos ofrecidos a lo largo de la propuesta de etiquetado y una breve mención a los usuarios a los que se destina esta publicación. Se cierra este bloque con un apartado de consideraciones complementarias en el que se alude a futuras vías de investigación, así como a la necesidad de examinar la viabilidad de algunas de las soluciones que se han adoptado o de otras alternativas existentes. Algunas de estas cuestiones se desarrollan en los apéndices finales, como la pluralidad de soluciones en cuanto a la anotación de abreviaturas y expansiones (apéndice 1) o nuevas propuestas de etiquetado en torno a la estructura diplomática y la marcación de antropónimos y topónimos (apéndices 2 y 3 respectivamente). Estos procesos de codificación permitirían estudios interdisciplinares a partir del propio corpus.

El segundo bloque recoge la guía de etiquetado propiamente dicha. Comienza con una descripción básica de la estructura de un documento, el cual debe constar de dos partes esenciales: una cabecera donde se recogen los metadatos y el cuerpo del texto.

A lo largo de estas páginas se ofrece una propuesta de marcación que da una salida TEI a todos los elementos codicológicos que se contemplan en las normas de transcripción de CHARTA para manuscritos. En esta parte se reúnen numerosos casos de edición a los que se aplica la codificación TEI y para los que se incluye además una imagen del facsímil que ilustra fácilmente los ejemplos propuestos. Para cada uno de ellos se sigue la misma estructura: 1. descripción breve del caso; 2. edición de uno o varios ejemplos según los criterios CHARTA; 3. solución en TEI acompañada del ejemplo debidamente etiquetado.

Una de las virtudes de la publicación radica precisamente en este punto. El usuario familiarizado con los criterios CHARTA encontrará fácilmente las equivalencias correspondientes. La ejemplificación de casos es amplia y variada y muy difícilmente se detectará algún caso de edición para el que la guía no ofrezca una salida aplicada al estándar TEI. Así, el volumen incluye una propuesta de marcado que abarca todo lo relacionado con la estructura, cuestiones paratextuales y de formato, así como el etiquetado de las características originales de contenido reflejadas en el manuscrito, algunas de las cuales se recogen en la siguiente tabla:

<i>Caso de edición</i>	<i>Criterio CHARTA (TP)</i>	<i>Solución CHARTA-TEI</i>
Inicio de página	{h 5v}	<pb n= "5v"/>
Salto de línea	{8}	<lb n= "8"/>
Salto de columna	{a} {b}	<cb n= "a"/> <cb n= "b"/>
Título	[<i>título: Presa</i>]	<div> <head>Presa</head> <ab>[...]</ab> </div>
Abreviaturas	p<ero>	<expan>p <ex>ero</ex> </expan>
Tachado	[<i>tachado: por</i>]	por
Sobrescrito	Pes[<i>sobrescrito: c</i>]etas	pe<subst> <del type= "sobrescrito"/>s <add>c</add> </subst>etas
Interlineado	yo [<i>interlineado: auia</i>]	yo <add place= "encima">auia</add>
Adiciones al margen	[<i>margen: texto</i>]	<add place= "margen">texto</add>
Cambio de mano	[<i>mano 2</i>]	<handShift new= "#h2"/>
Deterioro	[<i>roto</i>], [<i>mancha</i>]	<gap reason= "roto"/> <gap reason= "mancha"/>
Firma	[<i>firma: nombre</i>] [<i>rúbrica</i>]	<signed> <seg type= "firma">Nombre</seg> <seg type= "rúbrica"> </signed>
Estilo directo	texto	<quote>texto</quote>
Lengua extranjera	[<i>eusk.: Nagusi verri</i>]	<foreign xml:lang= "eu">Nagusi verri</foreign>

Tabla 1. Equivalencias de algunos de los criterios de transcripción CHARTA adaptados a TEI.

Tal y como se desprende de la Tabla 1, el uso del estándar de edición que aquí se presenta requiere de cierta técnica y conocimiento a la hora de construir las etiquetas que proponen las directrices CHARTA-TEI. De esta manera, la obra se enfoca a usuarios muy concretos que los autores describen como «investigadores y especialistas tanto en la edición como en el estudio filológico de documentación antigua manuscrita [...] que estén interesados en la aplicación de los criterios de CHARTA» (p. 18). Sin embargo, conviene precisar aún más que el público al que se dirige deberá estar necesariamente familiarizado con algunos conceptos clave. En primer lugar, con las reglas básicas de formación XML, lenguaje informático en el que se formula el estándar TEI, pero también con el manejo de algún editor XML que facilite el proceso de transcripción. Los autores del volumen mencionan para ello el programa *Oxygen*⁶, aunque hasta el momento solo se han realizado algunos ensayos y su aplicación se plantea como tarea futura dentro de un proceso más amplio. Por tanto, el lector que acuda a este manual con la intención de iniciarse en la edición digital no hallará algunas nociones elementales, no obstante, sí referencias externas que sin duda servirán de ayuda.

Como es sabido, los criterios actuales de CHARTA ofrecen un sistema doble de presentación (transcripción paleográfica y presentación crítica) para cada uno de los

⁶ <https://www.oxygenxml.com/>

documentos que componen el corpus. Esta edición doble, que sin duda es considerada como una de las virtudes del corpus, pues responde a las demandas de un público variado y posibilita la interdisciplinariedad, ha supuesto también uno de los grandes retos a los que se ha enfrentado la realización de una propuesta CHARTA-TEI. Este proyecto apuesta por un modelo de fuente única en el que se combinan las dos versiones en un solo archivo XML-TEI, así como todo lo relacionado con cuestiones extratextuales, cuestiones puramente lingüísticas y metadatos. Adicionalmente, la visualización del facsímil será posible por medio de un enlace a la imagen. A pesar de esta codificación en fuente única, nada impediría separar las dos versiones en una fase posterior de presentación (p. 11).

Precisamente, no debemos olvidar que la razón principal por la que aplicar un estándar de codificación con las características mencionadas reside en «ulteriores tratamientos como la extracción de datos o la publicación digital» (p. 10). No se encuentran en este volumen referencias a las posibles salidas o vías de publicación digitales del corpus tras la implantación de este sistema de edición, aunque sí se mencionan algunas de las soluciones técnicas adoptadas por otros corpus históricos: el sistema LYNEAL⁷, probado en algunos corpus como el CORHEN o el CODEA; o la herramienta TEITOK⁸, también aplicada a corpus como *Oralia diacrónica del español* (ODE).

En efecto, son muchas las vías futuras de investigación que quedan abiertas tras la implantación del modelo que se propone, muchas de las cuales ya se mencionan en la propia obra. A algunas de ellas se tratará de dar respuesta conforme el proyecto avance y se adopten los cambios pertinentes, pero para ello, el siguiente paso es la propia acogida de los criterios propuestos. Sin duda, el esfuerzo por plantear y aplicar este estándar a un gran corpus como CHARTA es loable y supone el primer paso necesario para conseguir un mejor aprovechamiento de los datos. Este volumen, por tanto, ofrece una propuesta sólida de acuerdo con las últimas técnicas de edición digital y con la que contribuir a la mejora tecnológica de CHARTA, un corpus que ya goza de la calidad y el rigor necesarios para los estudios lingüísticos de índole histórica.

REFERENCIAS BIBLIOGRÁFICAS

CHARTA. (2013): *Criterios de edición de documentos hispánicos (Orígenes – Siglo XX) de la Red Internacional CHARTA*. <http://www.redCHARTA.es/criterios-de-edicion/> [Consulta: 22/12/2020].

ISASI MARTÍNEZ, C. (coord.); Lobo Puga, A.; Martín Aizpuru, L.; Pérez Isasi, S.; Pierazzo, E.; Spence, P. (coord.). (2014): *Guía para editar textos CHARTA según el estándar TEI: una propuesta*. [Consulta: 20/12/2020].

⁷ En el siguiente enlace pueden consultarse aquellos corpus en su versión LYNEAL:
<http://shimoda.llf.uam.es/ueda/lyneal/>

⁸ Un listado de la totalidad de corpus que hoy en día utilizan TEITOK se encuentra disponible en:
<http://www.TEItok.org/index.php?action=projects>