

Edición digital de documentos antiguos:
marcación XML-TEI basada en los criterios CHARTA

CARMEN ISASI MARTÍNEZ (GRUPO GHEN),
LEYRE MARTÍN AIZPURU (UNIVERSIDAD DE SEVILLA -
UNIVERSIDAD DE SALAMANCA), SANTIAGO PÉREZ ISASI
(UNIVERSIDADE DE LISBOA), ELENA PIERAZZO
(UNIVERSITÉ DE TOURS) Y PAUL SPENCE
(KING'S COLLEGE LONDON - KCL)

Edición digital de documentos antiguos:
marcación XML-TEI basada en los
criterios CHARTA



Sevilla 2020

Colección: Lingüística
Núm.: 58

COMITÉ EDITORIAL:

José Beltrán Fortes
(Director de la Editorial Universidad de Sevilla)
Araceli López Serena
(Subdirectora)

Concepción Barrero Rodríguez
Rafael Fernández Chacón
María Gracia García Martín
Ana Ilundáin Larrañeta
María del Pópulo Pablo-Romero Gil-Delgado
Manuel Padilla Cruz
Marta Palenque Sánchez
María Eugenia Petit-Breuilh Sepúlveda
José-Leonardo Ruiz Sánchez
Antonio Tejedor Cabrera

Reservados todos los derechos. Ni la totalidad ni parte de este libro puede reproducirse o transmitirse por ningún procedimiento electrónico o mecánico, incluyendo fotocopia, grabación magnética o cualquier almacenamiento de información y sistema de recuperación, sin permiso escrito de la Editorial Universidad de Sevilla.

© Editorial Universidad de Sevilla 2020
C/ Porvenir, 27 - 41013 Sevilla.
Tlfs.: 954 487 447; 954 487 451; Fax: 954 487 443
Correo electrónico: eus4@us.es
Web: <<https://editorial.us.es>>

© De los textos, los autores 2020

Impreso en papel ecológico
Impreso en España-Printed in Spain
ISBN: 978-84-472-2973-4
Depósito Legal: SE 614-2020
Maquetación: Cuadratín Estudio
Impresión:

ÍNDICE

I. INTRODUCCIÓN.....	11
1. Historia: Investigación CHARTA-TEI.....	11
2. Objetivos.....	13
3. Decisiones metodológicas.....	14
4. Corpus empleados.....	19
4.1. CODCAR.....	19
4.2. CODOMME.....	20
4.3. CODOMSA.....	20
4.4. Corpus SAI.....	20
4.5. CODEA+ 2015: Corpus de Documentos Españoles Anteriores a 1800.....	21
4.6. CORHEN: Corpus Histórico del Español norteño.....	21
5. Destinatarios.....	22
6. Consideraciones complementarias.....	22
II. GUÍA DE ETIQUETADO.....	25
1. Estructura principal de un documento CHARTA-TEI.....	26
1.1. Metadatos de los documentos: la cabecera <teiHeader>.....	26
1.1.1. Estructura principal de la cabecera.....	27
1.1.1.1. <fileDesc>.....	27
1.1.1.1.1. <titleStmt>.....	28
1.1.1.1.2. <publicationStmt>.....	29
1.1.1.1.3. <sourceDesc>.....	30
1.1.1.2. <encodingDesc>.....	33
1.1.1.2.1. La tipología de documentos: <classDecl>.....	34
1.1.1.3. <profileDesc>.....	35
1.1.1.4. <revisionDesc>.....	36
1.2. Resumen de cómo convertir una cabecera CHARTA a cabecera TEI (<teiHeader>).....	37
2. Edición.....	38
2.1. Estructura principal del documento: <text>.....	38
2.1.1. Salto de página o folio.....	39
2.1.2. Salto de columna.....	41
2.1.3. Salto de línea.....	42
2.1.4. Título.....	43

2.2. Casos de edición	44
2.2.1. Desarrollo de abreviaturas	44
2.2.2. Intervenciones en el texto	46
2.2.2.1. Intervención del escribano.....	46
2.2.2.1.1. Tachado.....	46
2.2.2.1.2. Cancelado.....	48
2.2.2.1.3. Raspado	48
2.2.2.1.4. Sobre raspado.....	49
2.2.2.1.5. Sobrescrito.....	50
2.2.2.1.6. Interlineado.....	51
2.2.2.1.7. Margen.....	52
2.2.2.1.8. Cambio de mano	53
2.2.2.1.9. Blanco	54
2.2.2.2. Enmiendas de editor (actual).....	55
2.2.3. Deterioro del original y restitución de elementos en el texto por parte del editor actual.....	56
2.2.3.1. Deterioro que impide la lectura de fragmentos de texto escrito: Se saben cuántos caracteres faltan	56
2.2.3.2. Deterioro que impide la lectura de fragmentos de texto escrito: No se sabe cuántos caracteres faltan.....	58
2.2.3.3. Se desconoce si el deterioro afecta a texto escrito o no.	59
2.2.4. Signos o elementos especiales.....	60
2.2.4.1. Signo del notario	61
2.2.4.2. Firmas y rúbricas.....	62
2.2.4.2.1. Caso especial: firmas en otras lenguas.....	63
2.2.4.3. Crismón.....	64
2.2.4.4. Cruz.....	65
2.2.4.5. Calderón	66
2.2.4.6. Rueda	66
2.2.5. Normalización para la presentación crítica	70
2.2.6. Estilo directo y citas	74
2.2.7. Otras lenguas, el elemento <foreign>	75
2.3. Certeza	77
2.4. Enlace con imagen	77
2.4.1. Enlace a la imagen.....	77
2.4.2. Conservación y publicación de las imágenes	78
2.5. Caracteres reservados en XML.....	78
2.6. Casos complejos de etiquetado.....	79
2.6.1. Escenarios que incluyen la normalización.....	79
2.6.2. Los saltos de página dentro de otros casos	80
3. Unas reflexiones finales	80

APÉNDICE 1: REFLEXIÓN SOBRE ABREVIATURAS Y SU EXPANSIÓN	83
APÉNDICE 2: ESTRUCTURA DIPLOMÁTICA	85
1. Descripción	85
2. Edición según CHARTA	86
3. Marcación TEI	86
APÉNDICE 3: MARCACIÓN DE NOMBRES DE PERSONA Y LUGAR.....	87
Nombres de persona	87
Lugares	88
APÉNDICE 4: DOS DOCUMENTOS COMPLETOS ETIQUETADOS	91
BIBLIOGRAFÍA	107

I. INTRODUCCIÓN

1. HISTORIA: INVESTIGACIÓN CHARTA-TEI

Este trabajo¹ es el resultado de un desarrollo colaborativo llevado a cabo a lo largo de varios años por diversos grupos de investigación de diferente procedencia que han aportado sus conocimientos tanto desde el ámbito más filológico como desde las Humanidades Digitales para crear un trabajo interdisciplinario.

La investigación sobre la transformación de los criterios de edición de la Red CHARTA al metalenguaje XML-TEI se inició en el año 2009, entonces con el nombre provisional de “Proyecto piloto CHARTA-TEI”, a partir de una *Ayuda para la consolidación de grupos* concedida por el Gobierno Vasco-Eusko Jaurlaritz a un proyecto presentado por el equipo *Seminario Alfonso Irigoien* (SAI) de la Universidad de Deusto, e inicialmente ideado por Carmen Isasi, Pedro Sánchez-Prieto y Paul Spence. Esta ayuda permitió desarrollar durante ese mismo año 2009 y en 2010 tres reuniones de trabajo en la Universidad de Deusto (UD), King’s College London (KCL) y Universidad de Alcalá (UAH). Las mencionadas universidades estuvieron representadas en las reuniones por Carmen Isasi y Santiago Pérez Isasi (equipo SAI-UD)², Paul Spence y Elena Pierazzo³ (KCL) y Pedro Sánchez-Prieto e Irene Vicente

1. Este libro ha recibido financiación de los proyectos “Variación lingüística en la documentación de Castilla y León. II. Los documentos de Mombeltrán. Edición y estudio” (Ministerio de Ciencia e Innovación, ref. I*DFFI2014-57615-P) y “CHARTA 3.0: de la edición digital a la web semántica” (Comunidad de Madrid - Proyectos de I+D para Jóvenes Investigadores de la Universidad de Alcalá 2019, ref. CM/JIN/2019-008). Además, ha contado con el apoyo del proyecto “El castellano norteño en la Edad Media. Estudio lingüístico de la documentación palentina” (Ministerio de Economía, Industria y Competitividad, ref. FF2016-80230-P).

Agradecemos a Ricardo Pichel y Pedro J. Carrillo Gómez sus lecturas minuciosas y los comentarios hechos al borrador de este trabajo. Cualquier error es la sola responsabilidad de los autores.

2. En la actualidad es investigador del *Centro de Estudos Comparatistas* de la Universidade de Lisboa.

3. Actualmente, investigadora de la Universidad de Tours.

(grupo GITHE, UAH), con asistencia, en cada caso, de otros miembros de los respectivos departamentos.

El principal y último objetivo de estas reuniones fue abordar la representación de los criterios de edición de la Red CHARTA mediante un lenguaje de marcación conforme a los estándares de la propuesta TEI (*Text Encoding Initiative*)⁴. Ese análisis preliminar fue complementado y recogido en un documento por Santiago Pérez Isasi y Carmen Isasi, y algunos de los puntos de esta primera síntesis se presentaron en la reunión científica de la Red CHARTA de ese mismo año (Sigüenza, noviembre 2010), donde se discutió la utilidad y funcionalidad de las marcas propuestas. En definitiva, el trabajo con la guía al inicio del proceso de investigación consistió en el análisis de la estructura general del modelo de marcación atendiendo a los diferentes casos de edición considerados en los criterios de CHARTA.

El proyecto prosiguió con la estancia de Ana Lobo Puga y Leyre Martín Aizpuru (GEDHYTAS) en el King's College London bajo la tutoría de Paul Spence durante el primer trimestre del año 2012. Durante la misma el trabajo se centró en realizar pruebas de marcación sobre textos del corpus que aportaron las investigadoras de GEDHYTAS y en verificar la validez de las etiquetas propuestas. A lo largo de 2013, todo el equipo se encargó de anotar las posibles carencias del piloto y se sugirieron mejoras teniendo en cuenta fenómenos o elementos de edición que no se hubieran considerado aún. A continuación, los investigadores realizaron diferentes pruebas de etiquetado de textos en una serie de documentos de muestra, en algunos casos con diferentes etiquetas, sobre todo a partir de los corpus documentales de los grupos SAI y GEDHYTAS (para una descripción de los mismos, *cfr.* apartado 4.).

En 2014 se redactó la introducción teórica al trabajo, se completó la ejemplificación y se editó en la página web de la Red CHARTA, con el fin de hacerla accesible a los miembros de la propia red u otros investigadores interesados, y también con el propósito de promover el debate científico acerca de su aplicación y sus utilidades. Si bien el Corpus CHARTA⁵ no ha acogido la marcación TEI como criterio general de edición digital, este propósito se ha visto parcialmente cumplido dado que en el tiempo transcurrido hasta 2019 son varios los investigadores que han tenido en cuenta la existencia de una versión inicial del estudio, disponible en la página web de CHARTA. Entre ellos, destacamos las aportaciones de Pichel y Varela (2014), Isasi (2015), Allés y del Río Riande (2016), Torruella (2017), Rozas Candás (2017), Rojas Castro, (2017 y 2018), Araújo *et al.* (2018 y 2019), Calderón (2018), Fradejas Rueda (2019) y Martín Aizpuru *et al.* (2019) si bien, más allá de las aplicaciones parciales de los

4. Se puede consultar la página web oficial de este consorcio: <<http://www.tei-c.org>> [Última consulta: septiembre 2019].

5. <<http://www.corpuscharta.es/>> [Última consulta: septiembre de 2019].

propios autores de la propuesta, está aún pendiente su aplicación y aprovechamiento en una edición extensa.

En el momento actual, ha parecido necesario una actualización de la guía que ha implicado las siguientes tareas⁶:

1. Revisión de las etiquetas y *schema* para el etiquetado a la luz de las soluciones vigentes en las *TEI Guidelines*.
2. Inclusión de nuevos ejemplos.
3. Reestructuración de contenidos.
4. Actualización de referencias (corpus relacionados, bibliografía concerniente a la propia guía etc.).
5. Corrección de erratas.

2. OBJETIVOS

Como se ha señalado, el objetivo general de esta guía ha sido, desde su primer planteamiento, expresar los criterios de transcripción y de edición de la Red CHARTA (2013) mediante un lenguaje de marcación conforme a la propuesta *Text Encoding Initiative* (TEI), en su versión P5⁷. Atendiendo a las vías actuales de edición de documentos, definitivamente abiertas al horizonte digital, se pretende, así, partir de la consecución de los criterios de edición de CHARTA⁸ como modelo de tratamiento filológico y conjugarlo con el de TEI⁹, para optimizar su aprovechamiento en el ámbito de la edición digital.

Quedan al margen la revisión teórica de los debates sobre la propia edición digital o el uso de estándares de marcación, temas que cuentan con largo recorrido bibliográfico, de cuya actualidad son testimonio hasta el día de hoy publicaciones y reuniones congresuales, pero que exceden los límites de nuestros propósitos. Así, remitiéndonos, solo a modo de ejemplo, a algunas de las referencias más recientes podemos recordar la mirada sobre el ámbito de las Humanidades Digitales y el patrimonio medieval recogida en González y Bermúdez (2019), la aplicación de la marcación digital a corpus documentales, como

6. En este proceso han participado como coautores en el presente trabajo todos los miembros del equipo editor de la guía 2014, con la ausencia de Ana Lobo Puga, que renunció a participar en esta fase.

7. Para el manejo de esta guía se presupone un conocimiento básico sobre TEI así como de los principios filológicos para llevar a cabo una transcripción paleográfica y presentación crítica de textos antiguos.

8. El funcionamiento de la Red CHARTA así como los criterios propuestos han sido planteados en numerosos trabajos: vg. Sánchez-Prieto Borja (2012), Sánchez-Prieto Borja *et al.* (2012) Sánchez González de Herrero *et al.* (2013), Díez del Corral Areta y Martín Aizpuru (2014).

9. Entre las muchas publicaciones dedicadas a TEI, valgan como referencia Cummings (2008), TEI (2011a, 2011b), Vanhoutte (2004) y Pierazzo (2015). En el mundo hispánico podemos mencionar, entre muchas, las de Fradejas (2009-2010) e Isasi (2010 y 2011).

atestigua, entre otros, Vaamonde (2018), o bien las reflexiones sobre la misma coyuntura actual de las Humanidades Digitales de Fiormonte (2018) y Spence (2018).

Importa insistir en que el foco de nuestro interés es la codificación de los textos para ulteriores tratamientos como la extracción de datos o la publicación digital. A este respecto, aunque no vinculados a la propuesta CHARTA-TEI, se pueden mencionar los desarrollos generados en el marco de la propia Red CHARTA. Así, los resultados de la aplicación de los recursos diseñados por Francisco Javier Pueyo Mena y Bautista Horcajada, al corpus CODEA+¹⁰; la herramienta LYNEAL (*Letras y Números en Análisis Lingüísticos*), obra de Hiroto Ueda, ensayada en varios corpus, como por ejemplo CORHEN (Torrens y Ueda en línea¹¹), o el tratamiento con XML-TEI del corpus ODE (Calderón Campos y García-Godoy 2010-2019), explotado a través de TEITOK¹².

3. DECISIONES METODOLÓGICAS

Desde el inicio de la confección de esta propuesta se adoptaron varios acuerdos metodológicos que afectan (i) al uso de TEI, (ii) la marcación en fuente única (FU) y (iii) las visualizaciones.

(i) Se decidió hacer esta prueba en la aplicación de marcación TEI a los criterios de CHARTA por las ventajas ofrecidas por TEI y XML, ambos frecuentemente utilizados para representar, publicar y preservar textos en las Humanidades.

Como es bien sabido, el consorcio TEI ofrece un marco conceptual y práctico¹³ para marcar textos según un estándar generalizado para las Humanidades, que a la vez propone modelos comunes para facilitar el intercambio y la sostenibilidad de las “marcas digitales”, y un sistema mecanizado para la personalización de estos modelos donde se descubre casuística nueva cuando una comunidad científica determinada quiere dar cierta homogeneidad a sus opciones, como en nuestro caso. TEI emplea como base el XML (*Extensible Markup Language*), un metalenguaje estandarizado que facilita el intercambio de datos, el aprovechamiento del texto por el usuario o las máquinas y la publicación de textos, tanto en papel como en formatos digitales.

10. Este corpus se encuentra alojado en la siguiente página web: <<http://www.corpuscodea.es/>> [Última consulta: septiembre de 2019].

11. La versión del corpus CORHEN en Lyneal se encuentra en el siguiente enlace: <<http://shimoda.lllf.uam.es/ueda/lyneal/corhen.htm>> [Última consulta: septiembre de 2019].

12. <<http://teitok.corpuswiki.org/>> [Última consulta: septiembre de 2019].

13. <<http://www.tei-c.org/Guidelines/P5>> [Última consulta: septiembre de 2019].

En este sentido, aun cuando, como se ha dicho, está fuera de nuestros objetivos la revisión teórica de las soluciones contenidas en la *TEI Guideline P5*, conviene recordar aquí que sus propuestas ofrecen un amplio repertorio de etiquetas que rodean fragmentos de texto para su representación estructural y semántica y permiten separar los procesos de preparación y análisis de un texto de su posible o posibles presentaciones; por ejemplo, el título de una obra no se destaca en cursiva, sino que se enmarca con la etiqueta <title> para aplicar después la presentación deseada, es decir, la cursiva, en otro proceso. Esta separación facilita usos múltiples de una misma codificación, puede producir resultados distintos tanto en papel como en formato digital y, entre otras opciones, posibilitar el acceso simultáneo o alternativo a las versiones paleográficas o críticas generadas desde una FU. Así, podemos decir que TEI no entiende la marcación como un proceso *técnico*, en el sentido de la informática, sino como una extensión de las facultades del filólogo o editor, que emplea herramientas del siglo XXI, ya que, como señala Smith (2004), “marcar es editar y editar es marcar”.

(ii) En nuestro caso, acordamos que la investigación iba a probar este modelo de marcación digital según el concepto de la FU, es decir, a diferencia del modelo propuesto por los criterios de CHARTA, no se etiquetarían por separado la Transcripción Paleográfica (TP) y la Presentación crítica (PC), sino que se codificaría cada documento en un único texto XML-TEI, para luego separar las dos versiones en la fase de presentación.

Un aspecto muy importante es el hecho de que, mientras un documento creado según los criterios de CHARTA representa, a la vez, su fase de preparación y su presentación final¹⁴, un documento preparado en XML-TEI en FU marca una distinción muy clara entre su representación y su presentación. Se podría cuestionar, desde este punto de vista, si realmente un documento marcado según las opciones desarrolladas en esta guía corresponde a los criterios de CHARTA en su estado actual. Una respuesta práctica sería que, efectivamente, este modelo CHARTA-TEI permite la generación de dos documentos distintos para la publicación: uno siguiendo perfectamente los criterios de CHARTA para la TP y el otro siguiendo las normas para la PC; es decir, que se consigue el mismo resultado. En el plano científico se podría decir que este proyecto, diseñado en el ámbito de CHARTA, supone un cuestionamiento de cómo funcionarían sus criterios si se fusionaran con estándares técnicos de edición digital.

14. Los signos y convenciones usados apenas varían entre un texto preparado y un texto publicado; solo cambian pequeñas cuestiones de formato como la sustitución de los corchetes de abreviatura (< >) por cursiva en la visualización de la crítica o las llaves con numeración de líneas ({12}) por barra oblicua y número volado (/¹²).

Al margen de esta opción conceptual, finalmente decidimos que en el ejemplario se presentarían los diferentes casos de edición sobre el texto de la TP o de la PC, según se estimara pertinente, se renunciaría a la presentación en FU en función de la simplificación didáctica y para la intelección más clara del ejemplo que se está ofreciendo, solo marcamos el caso que se está ejemplificando. Recordamos que, entre los materiales complementarios de la presente guía, adjuntamos una serie de documentos etiquetados en fuente única con XML-TEI donde se puede observar la combinación de marcas que aquí presentamos por separado para cada caso.

Exponemos a continuación un resumen de los resultados concretos de la parte teórica y práctica de esta investigación:

- Se creó, basándose en los criterios de CHARTA, un modelo completo según las normas TEI, que reunía los siguientes puntos:
 - un análisis de casos filológicos de edición a todos los niveles (por ejemplo, el desarrollo de abreviaturas, el deterioro del original, la normalización gráfica).
 - su expresión según las normas de CHARTA, con referencia bibliográfica a sus criterios de edición.
 - ejemplos de la presentación crítica o de la transcripción paleográfica y presentación crítica de cada caso, si se considera pertinente.
 - fragmentos del facsímil para cada caso ejemplificado.
 - ejemplos de su marcación en TEI.

Caso	CHARTA TP	CHARTA PC	CHARTA TEI
Salto de página/folio de facs	{h 9r}	{h 9r}	<pb n="h 9r" facs="[título].jpg"/>
Salto de columna	{b}	{b}	<cb n="b"/>
Salto de línea	{8}	{8}	<lb n="8"/>

Imagen 1: Fragmento del modelo de casos que dio lugar a este documento.

Javier de choice orig Urbina orig reg Urbina reg choice ,

Imagen 2: Un ejemplo sencillo de una palabra normalizada con marcas TEI para la regularización gráfica.

- Para probar el modelo, se crearon varios recursos prácticos:
 - una **plantilla** para etiquetar un documento CHARTA (en TEI/XML).
 - varias **muestras** de documentos etiquetados según las normas de CHARTA y los principios TEI/XML.
 - un fichero **schema**¹⁵, que permite a un programa de edición XML fijar y aplicar las normas de edición y marcación (ejemplificados por este mismo documento) según un modelo formal y comprensible para un ordenador¹⁶.

(iii) Dado que una de las ventajas de la marcación XML es la flexibilidad en las posibles visualizaciones finales, se han realizado varias pruebas para demostrar que, efectivamente, este modelo podría recrear una edición conforme con los criterios de CHARTA, pero, en cualquier caso, no se pretende ofrecer soluciones cerradas. En esta misma línea, se sugieren, aunque de manera muy somera, algunas de las formas de visualizar aspectos de un documento archivístico a partir de la marcación TEI.

Murço (<sic>)

- `murço (<corr>)` [CORP-NUM]
- `murço (<corr>)` [CODCAR-0508]

de Gallizia de Gallizia (<sic>)

- `de Gallizia (<corr>)` [CODCAR-0508]

Imagen 3: Un fragmento de un índice de formas enmendadas.

Johan

- `<a>` [CODCAR-0359]

Juan

- `<ua>` [CODEA-1508]
- `<a>` [DARSXVIII-sn0001]
- `<a>` [DARSXVIII-sn0001]
- `<a>` [DARSXVIII-sn0001]
- `<a>` [DARSXVIII-sn0001]

Justiçias

- `<içias>` [CODEA-1508]

Imagen 4: Fragmento de un índice de formas abreviadas.

15. Allés (2018) define el esquema como el archivo XML cuya “función principal es la de establecer y definir los requisitos que debe respetar el documento XML”.

16. En el siguiente enlace ofrecemos el esquema CHARTA-TEI: <<https://github.com/CHARTA-TEI>>.

[cruz] {h 1r}
 {1} Mi E~~st~~imado Roque, h~~h~~e recivi
 {2}do tu C~~carta~~, e~~sta~~Está sin leer en la C~~com~~<unidad> p<or>q<ue>
 {2} tarde a~~ha~~ venido, e~~sta~~ te a~~ge~~hago p<ara> q<ue> no e~~ste~~isestéis
 {3} con C~~uid~~ado. Ayer
 {4} vino el hijo de M~~artin~~Martín.
 {5} q~~u~~andoCuando me dijo q<ue> no t~~eni~~atenia Carta~~carta~~, ya me
 {6} dio q<ue> pensar, e~~es~~ malo hacerse en todas
 {7} las S~~se~~manas. La M~~a~~ria>María Ant<onia> me a~~ha~~ enviado.
 {8} Ya han venido los A~~a~~mos, p~~pare~~ce q<ue>
 {9} no tendremos novedad en el I~~nv~~ierno.
 {10} n~~o~~ nos han vicitado p<or> el tiempo.
 {11} Al P~~p~~<adre> M~~m~~<astro>, q<ue> el Prior e~~sta~~está bastante
 {12} delicado, estos dos d~~ias~~días q<ue> e~~sta~~está mejor, q<ue>
 {13} E~~en~~comiende mucho, q<ue> si F~~al~~ta este, todo
 {14} F~~al~~to, este S~~s~~ostiene la C~~com~~<unidad>, m~~em~~orias a
 {15} todos, e~~sta~~ tu h~~at~~ía,
 [firma S<anto> D<omingo>]
 {16} 27.

Imagen 5: SAI. DG. Una visualización experimental que muestra la estructura de las líneas y contrapone formas originales / normalizadas¹⁸.

1 Sepan ~~qu~~antos[~~cu~~antos] esta carta ~~u~~ieren[~~v~~ieren]&[~~e~~] oyeren, ~~co~~mo[~~có~~mo] ~~no~~s[~~nó~~s,] don A[LFONSO]fonsol, por la gracia de ~~di~~os[~~di~~os] ~~re~~y[~~re~~y] de Castiella[~~de~~] de Toledo, de 2 Leon[~~le~~ón] de Gallizia, de ~~se~~uilla[~~se~~villa], de ~~co~~rdoua[~~co~~rdova], de Murcia, de ~~ja~~hen[~~ja~~én] &[~~e~~] del ~~al~~garue[~~al~~garve], ~~o~~torgamos[~~o~~torgamos] a ~~u~~os[~~vó~~s,] el ~~co~~ncejo[~~co~~ncejo] de Burgos, de 3 de ~~vi~~lla[~~vi~~lla] &[~~e~~] de aldeas, por muchos ~~se~~rucios[~~se~~rucios] &[~~e~~] buenos que nos siempre fiziestes, &[~~e~~] ~~por~~ que[~~por~~que] ~~u~~os[~~vó~~s] &[~~e~~] los otros ~~co~~ncejos[~~co~~ncejos] de ~~ca~~stiella[~~ca~~stiella] &[~~e~~] de ~~e~~stremadura nos prometiestes por ~~u~~uestras[~~u~~uestras] cartas abiertas de nos dar ~~ca~~danno[~~ca~~d año] ~~se~~rucio[~~se~~rucio] que montasse tanto como una moneda, ~~de~~los[~~d~~e]los por ~~an~~nos[~~a~~ños] ~~se~~ñalados[~~se~~ñalados] &[~~e~~] ~~de~~los[~~d~~e]los por ~~qu~~anto[~~cu~~anto] tiempo ~~no~~s[~~nó~~s] ~~tu~~viéssimos[~~tu~~viéssimos] por bien, &[~~e~~] ~~o~~trossi[~~o~~trossi] ~~por~~ que[~~por~~que] otorgastes que nos dariedes ~~lo~~gano[~~lo~~gano] el ~~se~~rucio[~~se~~rucio] de dos annos 6 bien &[~~e~~] ~~co~~mplidamientre, que era cosa que ~~au~~emos[~~au~~emos] mucho mester pora fecho del imperio. E ~~no~~s[~~nó~~s], entendiendo la ~~u~~estra[~~u~~estra] grand pobreza, prometemos de ~~u~~os[~~vó~~s] ~~nu~~nqua[~~nu~~nca] demandar ~~da~~qui[~~d~~aquí] adelante los ~~se~~rucios[~~se~~rucios] de los otros ~~an~~nos[~~a~~ños] &[~~e~~] ~~qu~~itamos ~~u~~os[~~vó~~s] ~~qu~~itamosvosos por siempre ~~ll~~amas[~~ll~~amás], 8 ~~u~~os[~~vó~~s] ~~da~~ndo nos[~~da~~ndonos] ~~lo~~gano[~~lo~~gano] el ~~se~~rucio[~~se~~rucio] como sobredicho es. E otorgamos que ~~no~~s[~~nó~~s] ni los otros reyes que ~~re~~gnaran[~~re~~gnarán] ~~de~~spues[~~de~~spués] de ~~no~~s[~~nó~~s] en Castiella &[~~e~~] en Leon[~~le~~ón], que ~~u~~os[~~vó~~s] lo no podamos demandar por fuero ni por uso. E ~~por~~ que[~~por~~que] ~~de~~sto[~~de~~sto] ~~de~~sto seades[~~de~~mas] ~~ma~~s[~~ma~~s] ~~se~~guros[~~se~~guros], ~~da~~mos ~~u~~os[~~da~~mosvos] ende esta nuestra 10 carta seellada con ~~nu~~estro[~~nu~~estro] seello de

Imagen 6: Visualización de revisión editorial que muestra formas originales/ normalizadas¹⁹.

17. EAH-AHE, Archivo de la casa de Hériz legajo 2,79.1838-04-07.

18. Ejemplo tomado de Archivo Municipal de Burgos, sección histórica, nº 117. 1274-05-05.

ha hecho entrega de los siete <ex>r</ex>eales<ex> que <choice><orig>costo</orig><reg>costó</reg></choice> la aba

Imagen 7: Una visualización de la marcación de una línea de un documento²⁰ según las recomendaciones de TEI. Se capta una palabra, ‘reales’, en su forma abreviada y con la parte desarrollada, con marcas semánticas que se ofrecen a múltiples visualizaciones.

{14} ha hecho entrega de los siete r<eales> que costo la aba

Imagen 8: La transformación (automática) del texto marcado en CHARTA / versión Transcripción Paleográfica.

{14} ha hecho entrega de los siete reales que costó la aba

Imagen 9: La transformación (automática) del texto marcado en CHARTA / versión Presentación Crítica.

4. CORPUS EMPLEADOS

Tanto el análisis de casos como el ejemplario aquí presentados se fundamentan en los materiales de los corpus que a continuación describiremos brevemente.

4.1. CODCAR

El *Corpus de documentación de cancillería real castellana* (CODCAR) desarrollado por el Grupo de Estudios de Documentos Históricos y Textos Antiguos de la Universidad de Salamanca (GEDHYTAS) incluye los documentos del siglo XIII y de la primera década del XIV, durante los reinados de Fernando III (rey de Castilla, 1217-1252, y de León, 1230-1252), Alfonso X (1252-1284), Sancho IV (1284-1295) y Fernando IV (1295-1312). Para preparar esta edición los miembros del grupo han trabajado en numerosos archivos nacionales —y uno portugués—, tanto eclesiásticos como de titularidad pública²⁰.

19. SAI. DG. EAH-AHE, Archivo de la casa de Zavala, sección Juan Manuel Olalde, subsección M. de Galarra 44.28, legajo 6750.

20. Cfr. Martín Aizpuru y Sánchez González de Herrero (2019).

4.2. CODOMME

El *Corpus de documentación medieval de Miranda de Ebro* (CODOMME) desarrollado por el Grupo de Estudios de Documentos Históricos y Textos Antiguos de la Universidad de Salamanca (GEDHYTAS) está formado por 203 testimonios que se custodian actualmente en el Archivo Histórico Provincial de Burgos (AHPB) y en el Archivo Municipal de Miranda de Ebro (AMME). Cronológicamente, el corpus abarca dos siglos y medio, entre la segunda mitad del siglo XIII (1254) y los años finales del XV (1495)²¹.

4.3. CODOMSA

El Corpus de Documentación Medieval del Sur de Ávila (CODOMSA) está formado por diversos documentos del siglo XV conservados en su mayoría en el Archivo Municipal de Mombeltrán (AMM), el único de la zona que contiene textos originales anteriores a 1500. Hasta este momento han sido localizados, digitalizados y transcritos 63 documentos de distinta temática y extensión. Asimismo, se conserva un número importante de documentos redactados en la zona, en el Archivo Histórico Municipal de Cuéllar (AHMC), Segovia. La documentación de concejo depositada en el Archivo Municipal de Mombeltrán y firmada en el sur de Ávila está escrita en papel, salvo un pergamino que traslada el privilegio de villa de Mombeltrán en 1402. En cuanto a la tipología, predominan los documentos de concejo, entre los que se incluyen cartas de poder, cartas de procuración, pleitos y sentencias, acuerdos entre vecinos y concejo, cartas de venta, cuentas y recaudación de impuestos, delimitación de territorios jurisdiccionales y pleitos homenaje de reconocimiento a los distintos señores de la villa. Hay también traslados de cartas regias realizadas en el siglo XV y cartas de poder.

4.4. Corpus SAI

El corpus SAI, denominado así por proceder de los trabajos del *Seminario Alfonso Irigoien*, grupo de investigación vinculado a la Universidad de Deusto hasta el año 2013 y hoy desaparecido, comprende documentos del País Vasco redactados en principio entre los siglos XVI y XIX. Se trata sobre todo de documentación municipal y de documentos privados, en especial cartas familiares,

21. Para una descripción completa del corpus: Grande López et al. (2017), Marcet Rodríguez y Sánchez González de Herrero (2016), Sánchez González de Herrero et al. (2014a y 2014b).

localizados y transcritos en su mayor parte en el marco del proyecto *Estudio pancrónico experimental y documental del seseo vasco* (SESEO)²² y editados, también parcialmente, en Isasi y Ramírez Luengo (eds.) (2013). Los documentos proceden de los siguientes archivos: Archivo Histórico Foral de Bilbao (AHFB), Archivo Histórico de Euskadi (AHE-EAH), Archivo Histórico Municipal de Durango (AMD-DUA) y Archivo Histórico Municipal de Aramaio (AMA-AUA).

4.5. CODEA+ 2015: Corpus de Documentos Españoles Anteriores a 1800

Elaborado por el Grupo de Investigación de Textos para la Historia del Español de la Universidad de Alcalá, actualmente el CODEA+ 2015²³ contiene documentos de entre los siglos XI y XVIII, seleccionados y transcritos para este corpus con criterios expresamente establecidos.

Su objetivo primordial es el desarrollo de una geografía lingüística diacrónica (por superposición de sincronías sucesivas). La cuantificación de los datos y el cruce de parámetros diversos permitirá desarrollos dialectométricos. Pero muestra igualmente su utilidad en cualquiera de los niveles de la historia de lengua (grafemática, fonética y fonología, morfología y sintaxis, léxico), así como de la paleografía, diplomática e historia de las mentalidades y de la vida privada. Este corpus se integra parcialmente en el de CHARTA (documentos 1 a 400, hasta ahora).

4.6. CORHEN: Corpus Histórico del Español norteño

El *Corpus Histórico del Español Norteño* (CORHEN)²⁴, realizado en el marco del proyecto *El castellano norteño en la Edad Media* del grupo GHEN²⁵ es un corpus creciente de documentación particular medieval de las variedades castellanas norteñas. Se concibe como una publicación seriada integrada por diferentes colecciones documentales.

Su objetivo principal es servir de base para el estudio y la reconstrucción de la historia del castellano desde sus primeras manifestaciones escritas, con un número representativo de antecedentes latinos e híbridos latino-romances.

22. *Estudio pancrónico experimental y documental del seseo vasco* (SVEP), financiado por el Ministerio de Ciencia e Innovación (FFI2008- 02377).

23. Consulta del corpus en: <<http://www.corpuscodea.es/>>, y más información sobre el mismo en: <<http://www.textoshispanicos.es/>> [Última consulta: septiembre de 2019].

24. Consulta del corpus en: <<http://corhen.es>> [Última consulta: septiembre de 2019].

25. La información sobre el grupo de investigación se encuentra en la siguiente página web: <<https://ghen.es>> [última consulta septiembre 2019].

5. DESTINATARIOS

Los usuarios ideales de esta investigación, en quienes se ha pensado al preparar el material incluido en la parte “II. Guía de etiquetado”, son investigadores y especialistas tanto en la edición como en el estudio filológico de documentación antigua manuscrita —especialmente del ámbito hispánico— que estén interesados en la aplicación de los criterios de CHARTA. Por este motivo, los casos ofrecidos como muestra en la guía proceden de los corpus descritos en el apartado anterior, incluidos en la Red CHARTA, de tipología, data y procedencia muy dispar, lo que ofrece un arco amplio de ejemplificaciones.

Asimismo, el uso de esta guía de marcación podría extrapolarse a documentación de otras áreas fuera del ámbito hispánico inicial, ya que muchos de los casos de edición serán comunes y así se reconocen, de hecho, en la propia *TEI Guideline P5*.

6. CONSIDERACIONES COMPLEMENTARIAS

Como se ha dicho al inicio, el objetivo inicial de este estudio se ha centrado en la codificación de los textos, en consonancia con los criterios de CHARTA, pero, como es sabido, la transferencia de parte de la investigación al entorno digital tiende a trasladar el enfoque desde una salida concreta (la publicación) hacia el ciclo de la investigación en su conjunto, dado que “the shift of paradigm (if we have one) is not from print to digital – it is from static to dynamic, and it is from output-driven to input- and user-driven” (Rehbein 2008: 2).

En este sentido, lo que aquí se propone es, según se verá, cómo editar un documento que pueda representar las decisiones editoriales según las normas de CHARTA e incluya tanto el contenido del documento (su estructura e interpretación) como la información general sobre el documento (metadatos) para ayudar en su preservación y su recuperación futura, por ejemplo, en procesos de búsqueda. Sin embargo, durante esta investigación también hemos empezado a experimentar con otros procesos, como se verá en los apéndices: las estructuras diplomáticas o el manejo de ‘entidades’ que sirvan para reseñar, por ejemplo, nombres de personas y lugares que se repiten a lo largo de un corpus, lo que podría generar índices y estudios prosopográficos o topográficos a partir de los mismos.

Este tipo de estudio debería acercarnos, por lo tanto, al ciclo de la investigación de manera integral, por ejemplo:

- Evaluar el proceso de creación de índices desde la misma marcación CHARTA-TEI.
- Conectar con otros proyectos de investigación que emplean métodos digitales.

- Considerar las implicaciones para el proceso de publicación: en otros dispositivos desde la misma FU, formatos distintos, opciones de personalización de la presentación en pantalla por parte del usuario, etc.
- Explorar los estándares digitales más allá de TEI (como METS²⁶ o RDF²⁷), que siguen culturas digitales distintas: de edición, de bibliotecas digitales, de webs semánticas o de redes comunes de conocimiento. Juntos, estos lenguajes representan intentos de compartir los resultados de la investigación bajo una perspectiva de apertura: acceso y datos abiertos.

Entre las cuestiones susceptibles de investigación en un futuro, conviene destacar las siguientes:

- Examinar más detenidamente las ventajas y desventajas del trabajo con FU —propuesta ahora como solución preferente— o con fuente doble.
- Sopesar la conveniencia de castellanizar el vocabulario de la marcación, acción limitada en esta guía al valor de los atributos (por ejemplo: `<del type="tachado">`²⁸).
- Analizar la gran variedad de usos de un corpus como CHARTA y su efecto en el logro de un marco común para la edición.
- Analizar los procesos generalizables y automatizables para una comunidad de investigación como CHARTA, que pretende compartir esfuerzos para crear recursos intelectuales que apuntan hacia una práctica común. Esta investigación podría analizar los distintos procesos analógicos y digitales dentro de la preparación de una edición, y compararlos con los procesos de otros ‘actores’ asociados, como los archivos, las bibliotecas o las editoriales.

26. METS: *Metadata Encoding & Transmission Standard*. Más información en: <http://www.loc.gov/standards/mets/> [Última consulta: septiembre de 2019].

27. RDF: *Resource Description Framework*. Más información en: <http://www.w3.org/RDF/> [Última consulta: septiembre de 2019].

28. Como se verá en la segunda parte de este trabajo, “Guía de etiquetado”, el elemento `` se emplea para cualquier forma de texto borrado.