



HAL
open science

Modelado de datos en TEI en proyectos de correspondencia digital en español

Marta López

► **To cite this version:**

Marta López. Modelado de datos en TEI en proyectos de correspondencia digital en español. Journal of the Text Encoding Initiative, 2023, 16, 10.4000/jtei.4730 . hal-03841148v2

HAL Id: hal-03841148

<https://hal.science/hal-03841148v2>

Submitted on 17 May 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - ShareAlike 4.0 International License



Modelado de datos en TEI en proyectos de correspondencia digital en español

TEI data modeling in Spanish-language digital correspondence projects

Marta López Izquierdo



Edición electrónica

URL: <https://journals.openedition.org/jtei/4730>

DOI: 10.4000/jtei.4730

ISSN: 2162-5603

Editor

TEI Consortium

Referencia electrónica

Marta López Izquierdo, "Modelado de datos en TEI en proyectos de correspondencia digital en español", *Journal of the Text Encoding Initiative* [Online], 16 | 2023, Online since 01 August 2023, connection on 18 August 2023. URL: <http://journals.openedition.org/jtei/4730> ; DOI: <https://doi.org/10.4000/jtei.4730>

For this publication a Creative Commons Attribution 4.0 International license has been granted by the author(s) who retain full copyright.

Modelado de datos en TEI en proyectos de correspondencia digital en español

TEI data modeling in Spanish-language digital correspondence projects

Marta López Izquierdo

RESUMEN

Este artículo analiza diferentes proyectos de edición digital de cartas en español que utilizan el marcado TEI, mostrando sus particularidades con respecto a otros proyectos epistolares del ámbito internacional. Se describen distintas prácticas tanto en el modelado de metadatos como de anotaciones en el texto y se valora su interés a la hora de extraer el máximo rendimiento de

las posibilidades que ofrecen las humanidades digitales hoy para el conocimiento de los textos epistolares en sus múltiples dimensiones y, en particular, para la explotación de su estructura reticular.

ABSTRACT

This paper analyzes different projects for the digital edition of letters in Spanish that use TEI markup, showing their specificity regarding other epistolary projects at global scale. I describe different practices, both in the metadata modeling and textual annotations. I evaluate how these methodologies, anchored in the disciplines of the digital humanities, contribute to the knowledge of epistolary texts in their multiple dimensions, in particular their reticular structure.

ÍNDICE

Palabras Clave: cartas en español, redes epistolares, metadatos, anotación lingüística, TEITOK

Keywords: letters in Spanish, epistolary networks, metadata, linguistic annotation, TEITOK

1. Proyectos hispánicos de correspondencia digital en el marco internacional

- 1 El género epistolar forma parte de la comunicación humana desde la Antigüedad, aunque en la actualidad experimenta importantes mutaciones, ligadas al desarrollo de la comunicación escrita por la vía digital. Por su amplia difusión, el texto epistolar conoce múltiples formatos y estilos y puede reunir tanto cartas de alta calidad literaria como escritas en una variedad alejada de los estándares lingüísticos de una época. Su interés trasciende el marco interpersonal que une al emisor y al destinatario, pues posee valor representativo de una época, un autor, un acontecimiento histórico o un territorio. De ahí, el interés que la edición y el estudio de cartas y epistolarios ha despertado entre los diversos estudiosos desde la Edad Media (Petrucci 2008).
- 2 En la actualidad, la edición digital de cartas es un terreno muy fértil en el campo de las humanidades digitales que irradia en muchas direcciones y disciplinas, con intereses comunes, pero también con sus propias especificidades: la filología, la literatura, la lingüística, la historia,

la sociología, entre otras. Son numerosos los proyectos internacionales que han surgido en los últimos veinte años experimentando con las potencialidades de la edición digital aplicada a las cartas y en particular, utilizando el lenguaje XML-TEI. De hecho, la edición digital epistolar ha servido como terreno de reflexión para el desarrollo de la TEI, a través de la creación de un grupo específico de trabajo dentro del consorcio para el marcado de los rasgos epistolares: TEI-Correspondence-SIG (Stadler, Illetschko y Seifert, 2016). Entre los proyectos pioneros, cabe citar “Model Editions Partnership: Historical Editions in the Digital Age” (MEP),¹ desarrollado en la década de los 1990 y que se interesó por las cartas en tanto que documentos históricos, o el proyecto *Digital Archive of Letters in Flanders (DALF)*,² que recoge desde 2002 cartas de escritores flamencos fechadas entre los siglos XIX y XX, y cuyo modelo editorial ha servido de base para el epistolario digital Vincent Van Gogh: The Letters (2009),³ de amplia repercusión en la esfera de la edición digital del género que nos ocupa.

- 3 Stadler, Illetschko y Seifert (2016) documentan ampliamente los epistolarios digitales de los quince primeros años del siglo XXI, con representación de proyectos europeos y americanos,⁴ aunque no se encuentran mencionados entre ellos proyectos realizados por equipos del ámbito hispánico ni sobre cartas en español.⁵
- 4 Sin embargo, en esta época ya estaba en marcha el proyecto *Post Scriptum* (Vaamonde et al. 2014), que contó con una importante financiación europea entre 2012–2017,⁶ y que representa a día de hoy el proyecto de edición digital epistolar más notable en lenguas ibéricas (portugués y español). Anteriormente, la Universidad de Lisboa había llevado a cabo otros dos proyectos de edición digital de cartas en portugués, precedentes directos del proyecto *CARDS: Unknown Letters* (Marquilhas 2006–2009) y *FLY: Forgotten Letters* (Marquilhas 2010–2013).
- 5 Es necesario también recordar otros proyectos de edición digital de cartas que se han llevado a cabo en el ámbito hispánico con otros sistemas de codificación: es el caso de corpus lingüísticos digitales que incluyen entre sus documentos cartas y epistolarios varios, como los corpus de la Real Academia Española CORDE y CREA o el más reciente CORPES XXI,⁷ aunque solo este último adopta el lenguaje XML-TEI para la codificación.⁸

- 6 Han de citarse igualmente los distintos corpus lingüísticos históricos de equipos vinculados a la Red CHARTA,⁹ que incluyen entre la documentación editada en línea numerosas cartas de diverso tipo y que llevan varios años realizando un importante trabajo de aclimatación del lenguaje TEI a la documentación histórica en español (Isasi et al. 2014, 2020).
- 7 En época aún más reciente arrancan otros proyectos de edición digital basados en TEI que incluyen cartas (como Gondomar¹⁰ o CoDiAJe,¹¹ entre otros) o, más específicamente, que se conciben como epistolarios digitales (CAREXIL-FR,¹² en construcción desde 2019).
- 8 Este artículo presenta algunos ejemplos representativos del uso de la TEI para el marcado de cartas en proyectos del ámbito hispánico. Comenzaremos nuestro análisis haciendo algunas observaciones sobre las particularidades de la edición digital de correspondencias en proyectos internacionales antes de pasar a presentar con mayor detalle algunas de las experiencias que incluyen cartas en español. Nos interesa en particular la manera en que los editores experimentan con las potencialidades del marcado TEI para anotar las características específicas de las cartas. También nos detendremos en el entorno TEITOK (Janssen 2014-),¹³ que es la herramienta hoy en día más utilizada para los proyectos de corpus epistolares, entre otras tipologías de corpus, en España. Tras exponer cómo se han modelado los metadatos de las cartas en TEI en diferentes proyectos internacionales e hispánicos (sección 2), estudiaremos el marcado de los textos epistolares, en función de los intereses de investigación (textuales, pragmáticos, lingüísticos) (sección 3), y terminaremos ofreciendo algunas conclusiones de nuestro trabajo en el proyecto *Cartas de Republicanos Españoles Refugiados y exiliados en Francia* (CAREXIL-FR) (sección 4).

2. Marcado de metadatos para la correspondencia en TEI: algunos ejemplos

- 9 Los primeros proyectos de edición digital de cartas que hemos mencionado previamente destacaron la importancia de desarrollar anotaciones específicas para diferenciar el género epistolar de otros tipos textuales. Como explican Stadler, Illetschko y Seifert (2016), los ya mencionados MEP y DALF introducen marcas para identificar nociones como el autor y el destinatario de las cartas, las fórmulas de despedida o las postdatas, que poco a poco irán configurando el módulo <correspDesc> *descripción de la correspondencia*, adoptado por

el consorcio TEI en 2015 (ejemplo 1). Como se puede observar, se utilizan dos bloques, el primero, <correspAction>, informa sobre la situación de emisión y recepción de la carta (quién, dónde, cuándo), mientras que el segundo, <correspContext>, enlaza la carta dentro de una correspondencia, identificando otras cartas enviadas o recibidas por las mismas personas.

Ejemplo 1. Codificación de <correspDesc>.

```
<correspDesc>
  <correspAction type="sent">
    <persName>Nombre del autor de la carta</persName>
    <placeName>Lugar donde se escribe la carta</placeName>
    <date when="1999-01-01"/>
  </correspAction>
  <correspAction type="received">
    <persName>Nombre del destinatario de la carta</persName>
    <placeName>Lugar donde se manda la carta</placeName>
    <date when="1999-01-02"/>
  </correspAction>
  <correspContext>
    <ref type="prev" target="ID de la carta"> Carta anterior de <persName>
      autor</persName> a <persName> destinatario </persName>: <date
when="1999-01-01"/>
    </ref>
  </correspContext>
</correspDesc>
```

- 10 El mismo grupo que ha creado el módulo epistolar en TEI ha diseñado el formato CMIF (Correspondance Metadata Interchange Format) y una herramienta <correspSearch>¹⁴ para integrar en una sola base de datos interrogable distintos corpus epistolares. Se utilizan los metadatos más importantes para la comparabilidad de cartas de orígenes diversos: nombre del emisor con su identificador, nombre del destinatario con su identificador, fecha de escritura y de recepción, lugar de escritura y de recepción, con identificador, número de carta en la edición de referencia (si la hay) y URL de la carta editada. La base de datos acoge cualquier colección de cartas que disponga de los índices con el formato indicado. Actualmente, enlaza 195.934 cartas con 11.839

personas identificadas.¹⁵ Se trata en su mayoría de cartas en alemán, fechadas entre el siglo XVI y el XX. Hay también colecciones de cartas noruegas, inglesas y francesas. La presencia de equipos españoles o que trabajan sobre cartas en español es anecdótica.

- 11 El interés de este tipo de bases de datos que reúnen diferentes corpus epistolares se ha puesto de manifiesto en proyectos como *Mapping the Republic of Letters*, repositorio que permite la visualización de las redes de correspondencia entre escritores, intelectuales y artistas, desde Erasmo hasta Franklin. Con base en la Universidad de Stanford, se compone de un conjunto de instituciones asociadas entre las que figuran la Universidad de Oxford, el CNRS (Centre National de la Recherche Scientifique) francés y otras de los Países Bajos e Italia. Se reúnen en él diversos estudios de casos particulares sobre diferentes personalidades del mundo de las letras.¹⁶ Para el ámbito hispánico, incluye el proyecto *An Intellectual Map of Science in the Spanish Empire, 1600–1810*, dirigido por Marcelo Aranda, basado en el *Spanish Scientist Database*, que a su vez fue compilada a partir del *Diccionario Histórico de la ciencia moderna en España* (López Piñero et al. 1983), que abarca el período de finales de la Edad Media hasta 1970. El proyecto incluye científicos activos en la época moderna y constituye una base de datos biográfica de cerca de 360 personalidades. Sin embargo, no está centrado en la compilación de cartas, como en el caso de los otros proyectos del repositorio, sino en interacciones de personas y organizaciones a través del imperio español.
- 12 Ha de mencionarse igualmente, por su interés en el ámbito hispánico, el catálogo digital *Spanish Republic of Letters*,¹⁷ que recoge las redes de correspondencia que se establecieron entre los humanistas del Renacimiento español y los humanistas de otros países europeos, con el objetivo de visibilizar la importancia de los eruditos españoles en el desarrollo de este movimiento cultural. Se recogen así las noticias bibliográficas de más de 5.500 cartas escritas entre el siglo XVI y el XVII, con los metadatos siguientes: tipo de documento, autor, destinatario, fecha, lugar de origen y de destino, repositorio donde se encuentra el original, edición moderna y lengua. La plataforma permite realizar búsquedas por nombres propios de persona en el texto de las cartas. No disponemos de información sobre el protocolo de marcado que se ha utilizado en este proyecto.¹⁸
- 13 Los editores de cartas en español que adoptan el marcado XML-TEI han adoptado en muy pocas ocasiones el módulo <correspDesc> y ello probablemente debido a que las cartas no se editan en proyectos especializados en correspondencias, sino que incluyen otros tipos de documentos. Ello hace que se haya propuesto una cabecera genérica para todas las categorías textuales, como ocurre

por ejemplo en CORPES XXI,¹⁹ en Gondomar, o en los proyectos de la Red CHARTA. En el corpus en línea de este último consorcio, solo figuran 35 cartas privadas sobre un total de 2.076 documentos (7 pertenecientes al corpus CODEA²⁰ y las demás al *Corpus del Seminario Alfonso Irigoyen (CSAI)*).²¹

- 14 El manual del proyecto CHARTA (Isasi Martínez et al. 2020) incluye el marcado de la tipología de documentos en dos fases: primero, creación de una taxonomía, anidada en `<classDecl>`, que engloba todas las clases reconocidas de documentos por el proyecto CHARTA en 2013,²² con 10 tipos en total (ejemplo 2), y en segundo lugar, aplicación de la tipología a cada documento con la etiqueta `<textClass>`, como se muestra en el ejemplo siguiente (ejemplo 3), aplicado a las cartas privadas, que se subdividen en cartas privadas, cartas semipúblicas y cartas secretas.

Ejemplo 2. Creación de la taxonomía en la Red CHARTA.

```
<classDecl>
  <taxonomy xml:id="Tip-CH">
    <bibl>Tipología CHARTA, propuesta de octubre de 2013</bibl> [...] <category
xml:id="car-pri">
  <category xml:id="car-pri-pri">
    <catDesc>Cartas privadas</catDesc>
  </category>
  <category xml:id="car-pri-spu">
    <catDesc>Cartas semipúblicas</catDesc>
  </category>
  <category xml:id="car-pri-sec">
    <catDesc>Cartas secretas</catDesc>
  </category>
</category> [...] </taxonomy>
</classDecl>
```

Ejemplo 3. Tipología de cada documento en la Red CHARTA.

```
<profileDesc>
  <textClass>
    <catRef target="#car-pri-spu"/>
  </textClass>
</profileDesc>
```

- 15 Las taxonomías se adaptan según el tipo de proyecto: así, el *Corpus Documental y Hemerográfico de la Cuba del Novecientos* (CODHECUN),²³ vinculado a esta misma red, recoge también cartas, entre otros tipos documentales, con una taxonomía algo diferente, pues distingue entre cartas oficiales, particulares, personales y comunicaciones.²⁴
- 16 Como vemos, aunque la categoría *carta* está contemplada en la taxonomía de la Red CHARTA, no se prevé un tratamiento específico para los metadatos epistolares o al menos no a través del módulo <correspDesc>. Por ejemplo, el *Corpus de documentos hispánicos de Suiza* (COSUIZA) igualmente asociado a esta red, utiliza la etiqueta <particDesc>, *descripción de participantes*, que dentro de TEI se utiliza para identificar los participantes de cualquier interacción lingüística.²⁵ En el caso de las cartas, este elemento permite identificar a emisores y destinatarios, como se puede ver en el ejemplo 4.²⁶

Ejemplo 4. <particDesc> en COSUIZA.

```
<particDesc>
  <person role="sent" sex="M">
    <persName>Fidel Andrés de Villalobos</persName>
  </person>
  <person role="received" sex="M">
    <persName>Armand François de Saint-Saphorin</persName>
  </person>
</particDesc>
```

- 17 Fuera ya de esta red, el proyecto CoDiAJe – Ladino Corpus (Quintana 2012-), que reúne documentos en judeo-español de los siglos XIV al XXI, clasificados por género textual, incluye seis cartas en un conjunto documental de 89 documentos. El proyecto da acceso a una serie muy completa de metadatos, entre los que figura el título, el autor, el género, el lugar de origen, la fecha, el lugar de nacimiento del autor, el lugar de residencia, la lengua, la descripción del contenido, palabras clave, la dirección de la escritura (puesto que se trata de un texto judeo-español escrito en alfabeto hebreo) y el alfabeto. Otros metadatos incluyen el repositorio, el género textual, y la traducción. El proyecto utiliza un marcado genérico para todas las tipologías textuales y no se contempla por el momento un tratamiento específico de corpus epistolares.²⁷

- 18 El módulo <correspDesc> sí aparece en algunos proyectos especializados en edición de cartas, como es el caso de *Post Scriptum* (CLUL 2014–2017). Se trata de una base de datos que reúne cartas privadas escritas en España y Portugal en la Edad moderna con objetivos de investigación pluridisciplinar. Desde el punto de vista de la edición, este proyecto siguió en sus primeros pasos el modelo del proyecto DALF, en una versión personalizada para responder a las necesidades de los investigadores de la Universidad de Lisboa. En una segunda fase, se adoptó el módulo TEI-CORRESP, acorde por consiguiente a los estándares del consorcio (Vaamonde 2018, 149) (ejemplo 5).

Ejemplo 5. <correspDesc> en *Post Scriptum*.²⁸

```
<correspDesc>
  <correspAction type="sent">
    <persName key="CDD.xml#FN8" xml:id="FN8" cert="low">
      <name>Fernando Nunes</name>
    </persName>
    <placeName confidence="0" type="origin"> España, Madrid <location>
      <geo>40.4167754 -3.7037902</geo>
    </location>
    </placeName>
    <date confidence="0" when="1630-11-04" from="1630" to="1630" when-
custom="1630.11.04"/>
  </correspAction>
  <correspAction type="received">
    <persName key="CDD.xml#FPS2" cert="low">
      <name>Francisco Pacheco da Silva</name>
    </persName>
    <placeName confidence="0" type="destination"> España, Madrid <location>
      <geo/>
    </location>
    </placeName>
  </correspAction>
</correspDesc>
```

- 19 Como se puede observar, el proyecto *Post Scriptum* reproduce el esquema principal del módulo `<correspDesc>` y personaliza algunos de los elementos por medio del atributo `@key`, que permite enlazar algunas informaciones con un fichero central que reúne todos los datos de cierto tipo, en este caso el fichero `CDD.xml` donde el identificador `#FN8` remite al autor. El mismo procedimiento se utiliza para el destinatario.
- 20 De manera similar, el proyecto CAREXIL-FR (López Izquierdo 2019-), dedicado a la edición de cartas de petición formal del exilio republicano español, utiliza el módulo específico para correspondencias, con diversos complementos de información a través de los atributos `@corresp` (ejemplo 6). Se recurre en este caso a un fichero central (`ner.xml`) para la identificación de varios contenidos: la persona (`#pers-MSE`) y su función como autor o destinatario (`@role="author"/@"addressee"`), el lugar (`#shelter-Ceilhes`), al que va asociado el tipo de centro en que se encuentra el refugiado autor de las cartas (refugio, campo de internamiento, colonia infantil, etc.), y por último, la identificación de otras cartas que pertenecen al mismo intercambio epistolar (respuesta del destinatario u otras cartas del mismo emisor).

Ejemplo 6. `<correspDesc>` en CAREXIL-FR.

```

<correspDesc>
  <correspAction type="sent">
    <persName corresp="ner.xml#pers-MSE" role="author"/>
    <location cert="1" type="origin">
      <placeName corresp="ner.xml#shelter-Ceilhes"/>
    </location>
    <date cert="1" when="1939-11-27"/>
  </correspAction>
  <correspAction type="received">
    <persName corresp="ner.xml#pers-RDM" role="addressee"/>
    <location cert="1">
      <placeName key="ner.xml#place-Paris"/>
    </location>
  </correspAction>
  <correspContext>
    <ref type="post" target="2051271"> Carta anterior de <persName> Margarita
      Suárez</persName> a <persName> Renée de Monbrison </persName>: <date
when="1939-12-22"/>
    </ref>
  </correspContext>
</correspDesc>

```

```
</correspContext>  
</correspDesc>
```

- 21 Como podemos ver, son todavía muy escasos los proyectos de ámbito hispánico que se han adueñado de este esquema para el modelado de metadatos. Volveremos sobre esta cuestión en las conclusiones de nuestro trabajo para explicar este relativo desinterés. Pero primero nos detendremos en las posibilidades que ofrece la TEI para la anotación del texto mismo de la carta, en función de los variados intereses de investigación.

3. TEI para la anotación pluridisciplinar de textos epistolares

- 22 Son diversos los elementos TEI que se han adoptado para caracterizar la estructura típica de las cartas. De nuevo, fue el proyecto DALF el que ofreció los primeros ejemplos en el panorama internacional, adaptando elementos como `<salute>`, `<opener>`, `<closer>`, para las distintas partes reconocibles en una carta, a partir de los elementos disponibles en TEI P5²⁹ para marcar la estructura de cualquier tipo de texto.
- 23 En los proyectos de ámbito hispánico, aquí también encontramos diferentes decisiones editoriales según si los corpus digitales contienen solo cartas o también otros tipos de documentos. Como ocurría con el modelado de metadatos, la anotación de las estructuras textuales en los proyectos generalistas propone una estructuración común para diferentes tipologías textuales, sin diferenciar en el marcado las particularidades epistolares.
- 24 Por el contrario, los proyectos especializados en corpus epistolares recogen la propuesta de DALF y otros proyectos internacionales (*Vincent Van Gogh: The Letters, Letters and Texts: Intellectual Berlin around 1800*,³⁰ entre otros), ajustándolos a la última versión del estándar TEI P5. Es el caso de los proyectos ya mencionados supra, Post Scriptum o CAREXIL-FR, como puede verse en el [ejemplo 7](#). Las tres principales partes de todo texto epistolar (encabezamiento, cuerpo y despedida) aparecen identificadas con los elementos `<opener>`, `<p>` y `<closer>`, respectivamente, anidados dentro de `<body>`. El cuerpo se compone de tantos párrafos como aparezcan en la carta que se edita. Además, dentro de estas unidades podemos encontrar nuevas subestructuras, para saludos, `<salute>`, tanto dentro de `<opener>` como de `<closer>`, fecha `<date>`, firma `<signed>`, dirección `<address>`, etc. Asimismo, puede anotarse el nombre del destinatario, cuando aparece, asociándole

un identificador único y una función con los atributos @key y @role. Por último, el elemento <postscript> permite introducir el texto de las eventuales postdatas de las cartas. Hemos de hacer notar que todos estos elementos están disponibles en las Guías Directrices de la TEI para textos de otros tipos, en particular para textos en prosa, de manera que no presentan soluciones de marcado ad hoc para las cartas, pero su combinación permite reflejar adecuadamente los rasgos estructurales epistolares.

Ejemplo 7. Plantilla para la anotación de la estructura textual de una carta en CAREXIL-FR.

```
<text>
  <body>
    <opener>
      <placeName> ...</placeName>
      <date>...</date>
      <persName key="ner.xml#pers-xxx" role="addressee"> ... </persName>
      <address>
        <addrLine>... </addrLine>
      </address>
      <salute>... </salute>
    </opener>
    <p> ...</p>
    <closer>
      <salute>... </salute>
      <signed>...</signed>
      <address>
        <addrLine>... </addrLine>
      </address>
    </closer>
    <postscript>
      <p>...</p>
    </postscript>
  </body>
</text>
```

- 25 A estas marcas de la estructura textual de la carta se han añadido otros elementos que permiten dar cuenta de formulaciones rituales en la sección de apertura o cierre de la carta. Se trata de partes opcionales, conocidas como exordio y conclusión, cuya aparición se introduce en *Post Scriptum* (y siguiendo su modelo, en CAREXIL-FR) con los elementos <seg> y el atributo @type generalmente anidados en el cuerpo (ejemplo 8).

Ejemplo 8. Elementos para las divisiones estructurales opcionales de las cartas, exordio y conclusión, en *Post Scriptum* (Vaamonde 2019).

```
<p>
  <seg type="harangue"> ...</seg>
</p>
<p>
  <seg type="peroration">... </seg>
</p>
```

- 26 Por el carácter pluridisciplinar del proyecto *Post Scriptum*, las anotaciones que se utilizan se destinan a lingüistas e historiadores. Este mismo objetivo pluridisciplinar anima el proyecto CAREXIL-FR, con una orientación pragmática, sociolingüística e histórica. Por ello, ambos proyectos desarrollan anotaciones que permiten la interrogación del corpus epistolar con fines específicos. Así por ejemplo, en CAREXIL-FR se ha creado un sistema de marcado de actos de habla para estudiar la estructura pragmática de los actos de petición, que se descomponen en una multitud de microactos (Blum-Kulka y Olshtain 1984, Blum-Kulka, House y Kasper 1989), cuya aparición puede estudiarse en las cartas por medio de un marcado con el elemento <seg> *segment*, especificado con los atributos @type y @subtype (ejemplo 9):

Ejemplo 9. Marcado de actos de habla en CAREXIL-FR.³¹

```
<p>
  <seg type="act" subtype="explanation">
    <lb/>hemos pasado como hemos <lb/>podido por no tener de molestarle pero me
    encuentra ahora en <lb/>una situacion muy apurada </seg>
  <seg type="act" subtype="request">
    <lb/>si Vd. pudiera hacer algo le agra- <lb/>deceria mucho </seg>
</p>
```

- 27 Por otro lado, es posible anotar formas lingüísticas que se deseen estudiar en relación con los objetivos de los investigadores. En las recomendaciones de la TEI, se deja abierta la posibilidad de añadir anotaciones lingüísticas en función de las necesidades de los estudiosos, sin limitar las modalidades.³² Se definen como anotaciones lingüísticas aquellas que permiten el análisis de los rasgos lingüísticos de un texto, distintos de las propiedades estructurales del mismo o de la información descriptiva sobre su contexto de producción o uso.³³ La etiqueta puede consistir en asociar un código a cada palabra o unidad (por ejemplo, para la categoría morfosintáctica de una palabra) o a un grupo de palabras. La adscripción de estos códigos puede ser automática, manual o mixta. El manual TEI ofrece ejemplos de este tipo de anotación lingüística, que se reproduce aquí (ejemplo 10), donde el elemento `<s>` *sentence* delimita la oración, `<w>` *word*, cada palabra, y el atributo `@ana` "analysis", proporciona la interpretación del código. TEI propone también otros elementos para distintos niveles sintácticos, como sintagma `<phr>` *phrase*, cláusula `<cl>` *clause* o morfema `<m>` *morph*. Los códigos usados por cada proyecto deben ser explicitados adecuadamente e incorporados en los metadatos del documento, dentro del elemento `<interpretation>`,³⁴ en la descripción de codificación (`<encodingDesc>`) de la cabecera (`<teiHeader>`).

Ejemplo 10. Anotación lingüística en TEI (TEI Consortium 2023).³⁵

```

<s>
  <w ana="#AT0">The </w>
  <w ana="#NN1">victim</w>
  <w ana="#POS">'s</w>
  <w ana="#NN2">friends </w>
  <w ana="#VVD">told </w>
  <w ana="#NN2">police </w>
  <w ana="#CJT">that </w>
  <w ana="#NP0">Kruger </w>
  <w ana="#VVD">drove </w>
  <w ana="#PRP">into </w>
  <w ana="#AT0">the </w>
  <w ana="#NN1">quarry </w>
  <w ana="#CJC">and </w>
  <w ana="#AV0">never </w>
  <w ana="#VVD">surfaced</w>
</s>

```

- 28 Asimismo, la clase `att.linguistics` proporciona una lista de atributos que pueden utilizarse para caracterizar las propiedades lingüísticas de las unidades (token), en particular las palabras (`<w>`): `@lemma`, `@pos part-of-speech`, `@msd morphosyntactic description`, etc.³⁶
- 29 Muchos de los proyectos de correspondencia aparecidos en los últimos años en el ámbito hispánico se basan en el entorno TEITOK creado por Maarten Janssen (2014-). Es el caso de los dos proyectos de correspondencias que hemos citado aquí, *Post Scriptum* y CAREXIL-FR.³⁷ Además, la Red CHARTA ha adoptado recientemente esta herramienta y se está procediendo a la migración de los corpus.³⁸ Aunque TEITOK es un entorno que utiliza el marcado XML-TEI, sus especificidades lo alejan en algunos casos del estándar TEI P5, como por ejemplo en la identificación de cada token, como se ve en la tabla 1:

Tabla 1. Anotación TEITOK/TEI P5 de la secuencia: "la tenemos enferma con una polmunia."³⁹.

<code><tok id="w-87" lemma="la" mfs="PP3FSD00">la</tok></code>	<code><w lemma="la" xml:id="w-87">la</w></code>
<code><tok id="w-88" lemma="tener" mfs="VMIP1P0">tenemos</tok></code>	<code><w lemma="tener" xml:id="w-88">tenemos</w></code>
<code><tok id="w-89" form="enferma" mfs="AQ0FS0" lemma="enfermo">enfer-<lb n="10" id="e-13"/>ma</tok></code>	<code><w lemma="enfermo" xml:id="w-89">enfer-<lb n="10" id="e-13"/> ma</w></code>
<code><tok id="w-92" lemma="con" mfs="SPS00">con</tok></code>	<code><w lemma="con" xml:id="w-92">con</w></code>
<code><tok id="w-93" lemma="uno" mfs="DI0FS0">una</tok></code>	<code><w lemma="uno" xml:id="w-93">una</w></code>
<code><hi rend="underlined" resp="CAEERF"> <tok form="polmunia" id="w-94" nform="pulmonía" lemma="pulmonía" mfs="NCFS000">polmunia</tok></hi></code>	<code><hi rend="underlined" resp="CAEERF"> <choice> <orig> <w lemma="pulmonía "</code>

	<pre>xml:id="w-94">polmunia</w> </ orig> <reg> <w>pulmonía</w> </ reg> </choice> </hi></pre>
<pre><tok id="w-95" lemma="," mfs="Fc">,</tok></pre>	<pre><pc lemma="," xml:id="w-95">,</ pc></pre>

- 30 La versión TEI P5 de la [tabla 1](#) es la que se obtiene aplicando la herramienta de transformación integrada en la plataforma TEITOK, gracias al cual se asegura la interoperabilidad con otros proyectos en TEI. Observemos algunas diferencias que se dan entre las dos versiones: los elementos `<w>` *word* y `<pc>` *punctuation character* se marcan con un mismo elemento `<tok>` en la plataforma TEITOK. Además de los atributos `@xml:id` (en TEITOK `@id`) y `@lemma`, TEITOK ofrece para cada token un atributo de *categoría morfosintáctica* `@mfs`, al que se asocia un código lingüístico que remite a las etiquetas del consorcio EAGLES.⁴⁰ En la conversión a TEI P5 disponible en TEITOK, no se generan automáticamente las etiquetas morfosintácticas, aunque es posible implementar esta versión con atributos TEI ya mencionados (`@pos` y `@msd`), para garantizar una perfecta interoperabilidad. Destaca por último el diferente tratamiento que se ofrece para las unidades normalizadas: en TEITOK, la forma original aparece identificada con el atributo `@form` y la forma normalizada con el atributo `@nform`, mientras que en TEI P5, se utiliza el elemento `<orig>` *original form* y `<reg>` *regularization*, anidados en `<choice>`, elemento que permite introducir variantes.⁴¹
- 31 El principal interés que ofrece el marcado de estos elementos para el estudio de los textos epistolares (ya sea en TEITOK o en TEI P5) radica en la correlación que puede establecerse entre la presencia de rasgos subestándares (como el empleo de la metátesis ilustrada en *polmunia*) y otros rasgos textuales específicos de las cartas (inventario de las fórmulas de saludo o despedida, variación en los temas tratados dentro del cuerpo, fórmulas de tratamiento, estrategias de mitigación para el acto de la petición epistolar, etc.). La anotación posible, tanto en TEITOK como en TEI P5, de estos diferentes contenidos debe permitir la identificación de recurrencias significativas entre el perfil sociolectal de los autores de las cartas (nivel de escolarización, género, edad, procedencia) y las técnicas epistolares reflejadas en las cartas.

- 32 Asimismo, fuera de las etiquetas de lema, categoría morfosintáctica o normalizaciones, TEITOK ofrece la posibilidad de introducir atributos con información lingüística específica. Dichos atributos toman la forma @ling y sus valores responden a códigos definidos libremente por cada proyecto. En el caso de CAREXIL-FR, se asocia un código para algunas de las construcciones sintácticas que se desean estudiar, como las completivas asindéticas, que se asocian al código sint-*asind* *sintaxis asindética*. Esta misma anotación puede aparecer en lenguaje TEI P5 como se muestra en la segunda columna de la [tabla 2](#), en contraste con la anotación en TEITOK (columna 1). Nótese que la anotación lingüística aparece en CAREXIL-FR asociada al nivel del token, en este caso, el verbo de la subordinada.⁴² Se trata de una construcción sintáctica marcada que puede estudiarse en relación a su aparición dentro de la estructura del texto epistolar (aquí, secuencia ritual que prepara la despedida) y de los sociolectos de los autores (en este caso, se trata de una autora con nivel de escolarización medio).

Tabla 2. Anotación lingüística en TEITOK y en TEI P5.⁴³

TEITOK	TEI P5
<pre><tok>Rogando</tok> <tok>me</tok> <tok ling="sint-asind"> perdonen </ tok> <tok>lo</tok> <tok>estensa</tok> <tok>que</tok> <tok>soy</tok></pre>	<pre><s> <w>Rogando</w> <cl ana="#asind"> <w>me</w> <w>perdonen</w> <w>lo</w> <w>estensa</w> <w>que</w> <w>soy</w> </ cl> </s></pre>

- 33 Por último, TEITOK facilita, gracias a su motor de búsqueda incorporado, la combinación de criterios de búsqueda de tipo documental (metadatos) y textual (estructuras y formas lingüísticas), algo que es crucial a la hora de considerar los textos epistolares en la dimensión reticular que los caracteriza, como textos que se insertan en series de correspondencia, con corresponales diversos o recurrentes, desde lugares donde están circulando determinados patrones compositivos (por ejemplo, los centros de refugiados desde donde diversos autores envían cartas que responden a modelos similares) y en los que intervienen diversos factores de variación (perfil social, origen geográfico, edad, género).

4. Conclusiones

- 34 Los párrafos anteriores han permitido ofrecer algunas calas en los diferentes usos que se están haciendo del marcado TEI en proyectos digitales de edición de cartas en español, poniéndolos en perspectiva con otros proyectos internacionales. Como se ha podido comprobar, los proyectos del ámbito ibérico (España y Portugal) se caracterizan por emanar de una fuerte tradición filológica, muy vinculada a la lingüística histórica. Así, vemos aparecer proyectos de edición digital preocupados por la fidelidad a las fuentes textuales y, a la vez, por la aplicación de metodologías propias de la lingüística de corpus. Con razón, se ha reconocido el carácter pionero de *Post Scriptum* a la hora de asociar una plataforma de edición digital y un corpus lingüístico interrogable (Henny-Krahmer 2019). Esta doble necesidad se ha visto satisfecha gracias al entorno TEITOK de Janssen (2014), que fue creado precisamente en el marco de este proyecto. No es de extrañar, por consiguiente, que sea esta herramienta la que está extendiéndose de manera muy amplia para otros proyectos hispánicos de edición digital asociados a búsquedas en el corpus. Aunque el uso de TEI en TEITOK no responde exactamente al estándar TEI P5, la puesta a disposición de una herramienta de transformación permite asegurar la interoperabilidad con otras plataformas.
- 35 Son escasos sin embargo los proyectos que en esta área se especializan en la edición y explotación de correspondencias, con pocas excepciones (entre ellas, las dos mencionadas aquí: *Post Scriptum* y *CAREXIL-FR*, dos proyectos pluridisciplinarios, con un fuerte componente a la vez histórico y lingüístico), pues en la mayoría de los casos los corpus editados incluyen diversos tipos de documentación para la investigación en lingüística histórica.
- 36 Este objeto de estudio más amplio no favorece la utilización de herramientas que permitirían sacar el máximo rendimiento de los corpus epistolares, en particular aquellos aspectos basados en la estructura reticular de este tipo de textos que han servido en proyectos internacionales para la exploración y visualización de diferentes tipos de redes (epistolares, sociales, geográficas, etc.).⁴⁴
- 37 Por otro lado, estos proyectos internacionales, concebidos muchas veces como repositorios de cartas, se han desarrollado gracias a la anotación de metadatos específicos para la correspondencia, en particular el módulo TEI `<correspDesc>`. Gracias a él, se han podido comenzar a enlazar fondos epistolares diferentes y a relacionar diversas redes de correspondencias.

- 38 Cabe pensar que los proyectos hispánicos podrían beneficiarse de estos desarrollos si se incluyera en ellos, como han empezado haciendo *Post Scriptum* y CAREXIL-FR, el módulo <correspDesc>. Su interés se extiende, a nuestro modo de ver, tanto a aquellos proyectos que únicamente editan cartas como a los que las incluyen entre otras tipologías de documentos, ya que, como hemos visto, permite dar cuenta de la estructura reticular característica del género epistolar: la carta no existe por sí sola, sino en el interior de redes de cartas, y en relación con redes de corresponsales así como de redes de puntos geográficos (Walter 2018). Su modelado como metadatos específicos facilitaría su posterior visualización gracias a herramientas *ad hoc* como cartografías y grafos. Por otro lado, este módulo ofrece una sólida estructura para proyectar bases de datos futuras en que se enlacen los corpus epistolares disponibles de manera interoperable, como la herramienta <correspSearch> ya mencionada.
- 39 A su vez, la tradición filológica y el perfil lingüístico de los corpus de ámbito hispánico pueden aportar un desarrollo más completo a los proyectos epistolares internacionales que ofrecen pocas posibilidades de búsquedas en los textos, pues, aparte de la indexación, la anotación cubre generalmente, en el mejor de los casos, algunas palabras clave y no llega al detalle que se ha mostrado en los corpus españoles o sobre el español.
- 40 Una de las grandes ventajas del marcado TEI es que puede completarse progresivamente, por medio de capas sucesivas de anotación, a medida que se van vislumbrando posibilidades de explotación y estudio de los textos. Apostemos por que los textos epistolares en español puedan beneficiarse en un futuro próximo de todas las posibilidades de su tratamiento digital por medio de bases de datos, ediciones filológicas en línea, motores de búsqueda, cartografías y herramientas de visualización.

BIBLIOGRAFÍA

- Álvarez, Rosario y Ernesto González Seoane, eds. 2017. *Gondomar. Corpus dixital de textos galegos da Idade Moderna*. Santiago de Compostela: Instituto da Lingua Galega. <http://ilg.usc.gal/gondomar/>.
- Aranda, Marcelo, dir. *An Intellectual Map of Science in the Spanish Empire, 1600-1810*. <http://republicofletters.stanford.edu/casestudies/spanishempire.html>.

- Baillet, Anne, ed. 2012. *Letters and Texts: Intellectual Berlin around 1800*. Berlin: Humboldt-Universität zu Berlin. <http://www.berliner-intellektuelle.eu/>.
- EAGLES. 1996. *Recommendations for the Morphosyntactic Annotation of Corpora, EAG—TCWG—MAC/R. Consejo de Europa*. <http://www.ilc.cnr.it/EAGLES96/annotate/annotate.html>.
- Blum-Kulka, Shoshana, Juliane House y Gabriele Kasper. 1989. *Cross-Cultural Pragmatics: Requests and Apologies*. Advances in Discourse Processes. Norwood, N.J.: Ablex Pub. Corp.
- Blum-Kulka, Soshana y Elite Olshtain. 1984. "Requests and Apologies: A Cross-Cultural Study of Speech Act Realization Pattern." *Applied Linguistics* 5(3): 196–213.
- Bravo-García, Eva, Ana Mancera Rueda y Leyre Martín Aizpuru, dirs. *Corpus Documental y Hemerográfico de la Cuba del Novecientos (CODHECUN)*. <http://cuba19.us.es/>.
- Castillo Lluch, Mónica y Elena Díez del Corral, dirs. 2013–. *COSUIZA: Corpus de documentos hispánicos de Suiza*. <https://grafila.unil.ch/cosuiza/>.
- Centrum voor Teksteditie en Bronnenstudie, Koninklijke Academie voor Nederlandse Taal- en Letterkunde (Centre for Scholarly Editing and Document Studies, Royal Academy of Dutch Language and Literature). 2002–. *DALF: Digital Archive of Letters in Flanders*. <http://ctb.kantl.be/project/dalf/index.htm>.
- CHARTA. *Corpus Hispánico y Americano en la Red: Textos Antiguos*. <http://www.corpuscharta.es>.
- CLUL (ed.). 2014. "P.S. Post Scriptum. Arquivo Digital de Escrita Quotidiana em Portugal e Espanhana Época Moderna." <http://ps.clul.ul.pt>.
- Enrique-Arias, Andrés, dir. s. f. *Corpus Mallorca*. <https://www.corpusmallorca.es/>.
- Finden, Paula y Dan Edelstein, dirs. 2013. *Mapping the Republic of Letters*. <http://republicofletters.stanford.edu/>.
- Grupo de Investigación Textos para la Historia del Español [GITHE], s.f. *CODEA + 2022 (Corpus de documentos españoles anteriores a 1900)*. <http://corpuscodea.es> DOI:10.37536/CODEA.2015 .
- Henny-Krahmer, Ulrike. 2019. "Bridging Edition and Corpus: A Review of P. S. Post Scriptum: A Digital Archive of Ordinary Writing (Early Modern Portugal and Spain)." *RIDE*. <https://ride.i-d-e.de/issues/issue-10/post-scriptum-digital-archive/?hilite=post-scriptum>.
- Isasi Martínez, Carmen dir. s. f. *Corpus SAI (Seminario Alfonso Irigoyen)*. No disponible en línea.
- Isasi Martínez, Carmen, Leyre Martín Aizpuru, Santiago Pérez Isasi, Elena Pierazzo y Paul Spence. 2020. *Edición digital de documentos antiguos: marcación XML_TEI basada en los criterios CHARTA*. Sevilla: Editorial Universidad de Sevilla.
- Isasi Martínez, Carmen y Paul Spence, eds. 2014. *Guía Para Editar Textos CHARTA Según El Estándar TEI: Una Propuesta*. <http://www.charta.es/investigacion/charta-tei/>.
- Jansen, Leo, Hans Luijten y Nienke Bakker, eds. 2009. *Vincent van Gogh. The Letters*. <http://vangoghletters.org/>.
- Janssen, Maarten. 2014–. *TEITOK. A tokenized TEI environment*. <http://www.teitok.org>.

- Lazure, Guy, dir. s. f. *Spanish Republic of Letters*. University of Windsor (Canada). <http://cdigs.uwindsor.ca/srl/>.
- Leech, Geoffrey y Andrew Wilson. 1996. *Recommendations for the Morphosyntactic Annotation of Corpora*. EAG—TCWG—MAC/R. <http://www.ilc.cnr.it/EAGLES96/annotate/annotate.html>.
- López Izquierdo, Marta, coord. 2019–. *CAREXIL-FR: Cartas de Republicanos Españoles Refugiados y Exiliados en Francia*. <http://carexil.huma-num.fr>.
- López Piñero, José M., Thomas Glick, Víctor Navarro Brotón y Eugenio Portela Marcos, eds. 1983. *Diccionario histórico de la ciencia moderna en España*. 1985. Madrid. Ediciones Península. 2 vols.
- Marquilhas, Rita, dir. 2006–2009. *CARDS: Unknown Letters*. Universidade de Lisboa.
- Marquilhas, Rita, coord. 2010–2013. *FLY: Forgotten Letters. 1900–1974*. Universidade de Lisboa. <http://fly.clul.ul.pt/>.
- Morvan, Jossiane. 2023. “Éditer une correspondance en XML-TEI.” *Inventorier les correspondances des Lumières. Analyse matérielle et traitements numériques*, dirigido por Claire Bustarret, con la colaboración de Emmanuelle de Champs y Nicolas Rieucou, Ferney, Centre international d’étude du XVIIIe siècle, 161–167.
- Pelayo, Javier Antón, dir. 2023. *EPICAT. Cartes familiars de Catalunya*. <https://epicat.uab.cat/>.
- Petrucchi, Armando. 2008. *Scrivere Lettere. Una Storia Plurimillennaria*. Roma-Bari: Laterza.
- Quintana, Aldina. 2012–. *CoDiAJe - The Annotated Diachronic Corpus of Judeo-Spanish*. <http://corptediglif.upf.edu/teitok/codiaje/>.
- . 2020. *CoDiAJe - The annotated Diachronic Corpus of Judeo-Spanish. Description of a Multi-alphabetic corpus and its textual and linguistic annotations*. *Scriptum Digital* 9: 209–236.
- REAL ACADEMIA ESPAÑOLA. Banco de datos (CORDE) [en línea]. *Corpus diacrónico del español*. <http://www.rae.es>.
- REAL ACADEMIA ESPAÑOLA. Banco de datos (CREA) [en línea]. *Corpus de referencia del español actual*. <http://www.rae.es>.
- REAL ACADEMIA ESPAÑOLA. Banco de datos (CORPES XXI) [en línea]. *Corpus del español del Siglo XXI*. <http://www.rae.es>.
- REAL ACADEMIA ESPAÑOLA. 2013 (revisado 2020). *Corpus del español del siglo XXI. Descripción del sistema de codificación. Libros y prensa*. <https://www.rae.es/banco-de-datos/corpes-xxi>.
- Rieucou, Nicolas, dir. 2016–2021. *Inventaire Condorcet*. <https://www.inventaire-condorcet.com/>.
- TEI Consortium. 2023. *TEI P5: Guidelines for Electronic Text Encoding and Interchange*. Versión 4.6.0. <https://tei-c.org/Vault/P5/4.6.0/release/doc/tei-p5-doc/en/html/index.html>.
- Stadler, Peter, Marcel Illetschko y Sabine Seifert. 2016. “Towards a Model for Encoding Correspondence in the TEI: Developing and Implementing <correspDesc>,” *Journal of the Text Encoding Initiative* 9. <http://journals.openedition.org/jtei/1433>. DOI:10.4000/jtei.1433.

- Vaamonde, Gael. 2018. “Escritura Epistolar, Edición Digital y Anotación de Corpus.” *Quadernos Del Instituto Historia de La Lengua*, 11: 139–64.
- . 2019. *P. S. Post Scriptum. Archivo digital de escritura cotidiana. Personalización del esquema TEI y documentación*. http://teitok.clul.ul.pt/postscriptum/files/ps_doc.html#ps_formulaic.
- Vaamonde, Gael, Ana Luísa Costa, Rita, Marquilhas, Clara Pinto y Fernanda Pratas. 2014. “Post Scriptum: Archivo Digital de Escritura Cotidiana,” *Humanidades Digitales: desafíos, logros y perspectivas de futuro*, editado por Sagrario López Poza y Nieves Pena Sueiro, Janus [en línea], Anexo 1: 473–482. <https://www.janusdigital.es/anexos/contribucion.htm?id=41>.
- Walter, Richard. 2018. “L’éditition Numérique de Correspondances. Guide Méthodologique.” <https://cahier.hypotheses.org/guides/guide-correspondance>.

NOTAS

- 1 Puede consultarse una descripción en, fecha de consulta 16 de julio de 2023, <https://tei-c.org/activities/projects/model-editions-partnership-historical-editions-in-the-digital-age/>. El proyecto, albergado en la Universidad de South-Carolina, no está ya accesible en línea.
- 2 Fecha de consulta 16 de julio de 2023, <https://ctb.kantl.be/project/dalf/index.htm>
- 3 Fecha de consulta 16 de julio de 2023, <http://vangoghletters.org/>. Véase Jansen, Luijten, y Bakker 2009.
- 4 No es nuestro objetivo aquí describir de manera exhaustiva la totalidad de estos proyectos. Véase el artículo citado de Stadler, Illetschko y Seifert (2016), así como el catálogo elaborado por el Institut für Dokumentologie und Editorik e.V. (IDE), fecha de consulta 16 de julio de 2023, https://v3.digitale-edition.de/vlet_letters.html. En este catálogo, se recogen 121 proyectos de correspondencia digital, entre los cuales solo figura un proyecto del ámbito ibérico: FLY/CARDS, al que más adelante me referiré, realizado por el equipo de la Universidad de Lisboa.
- 5 Tampoco se recogen proyectos españoles o sobre el español en, fecha de consulta 16 de julio de 2023, https://www.zotero.org/groups/2160280/digital_correspondence_projects/library, a excepción de *Post Scriptum*.
- 6 Un informe del proyecto en: fecha de consulta 16 de julio de 2023, <https://cordis.europa.eu/project/id/295562/reporting>.

- 7 Se trata de CORDE (Corpus Diacrónico del Español, fecha de consulta 16 de julio de 2023, <https://www.rae.es/banco-de-datos/corde>); CREA (Corpus de Referencia del Español Actual, fecha de consulta 16 de julio de 2023, <https://www.rae.es/banco-de-datos/crea>) y CORPES XXI (Corpus del Español del siglo XXI, fecha de consulta 16 de julio de 2023, <https://www.rae.es/banco-de-datos/corpes-xxi>).
- 8 El corpus CORDE presenta entre los textos históricos la categoría “Cartas y relaciones,” que reúne textos epistolares de naturaleza muy diversa (cartas públicas, privadas, administrativas...) y textos no epistolares; el corpus CREA ofrece la categoría Cartas: personales, empresa, propaganda; el CORPES XXI detalla más específicamente “carta oficial Circular” y “carta particular,” además de “carta al Director.” Sin embargo, las búsquedas efectuadas seleccionando la tipología “Carta particular” remiten a un solo documento, publicado en prensa.
- 9 *Corpus Hispánico y Americano en la Red: Textos Antiguos*, fecha de consulta 16 de julio de 2023, <http://www.corpuscharta.es/>
- 10 *Gondomar. Corpus dixital de textos galegos da Idade Moderna.*, fecha de consulta 16 de julio de 2023, <http://ilg.usc.es/teitok/gondomar/en/index.php?action=home>. Véase Álvarez y González Seoane (2017).
- 11 *Corpus diacrónico anotado del judeo-español*. Fecha de consulta 16 de julio de 2023, <http://corptedig-glif.upf.edu/teitok/codiaje/>. Véase Quintana (2020)
- 12 *CARTas de REpublicanos Españoles REFugiados y EXILIados en FRancia*, fecha de consulta 16 de julio de 2023, <http://carexil.huma-num.fr/>. Véase López Izquierdo (2019-).
- 13 Sitio web de TEITOK: fecha de consulta 16 de julio de 2023, <http://www.teitok.org>.
- 14 Se trata de una API (Application Programming Interface) en una página web que permite el intercambio de datos entre proyectos epistolares a través de la transmisión de ficheros XML en formato CMIF. Fecha de consulta 16 de julio de 2023, <https://correspsearch.net/en/home.html>
- 15 Datos disponibles a la fecha de 27 de enero de 2023.
- 16 Este proyecto incluye metadatos codificados con TEI, pero no exclusivamente. Para una correspondencia dentro de este proyecto en que se hace uso de la TEI, puede consultarse el *Inventaire Condorcet*. Fecha de consulta 16 de julio de 2023, <https://www.inventaire-condorcet.com/> (Rieucou 2016–2021) y el artículo de Morvan (2023), que detalla la personalización del marcado para este epistolario.

- 17 Fecha de consulta 16 de julio de 2023, <http://cdigs.uwindsor.ca/srl/>. Véase Lazure s.f..
- 18 Existen otros proyectos de correspondencia digital, o que incluyen correspondencia digital de ámbito ibérico, que se han construido sin emplear la TEI. No los podemos describir in extenso aquí, bástenos citar, a modo de ejemplo, el corpus *Cartes familiars de Catalunya (segles XVI-XIX)* (EPICAT), fecha de consulta 16 de julio de 2023 <https://epicat.uab.cat/>, que edita más de 2500 cartas familiares de Cataluña, en catalán, español, francés o italiano, entre los siglos XVI a XIX (Pelayo 2023), o el *Corpus Mallorca* (Enrique-Arias s.f., fecha de consulta 16 de julio de 2023, <https://www.corpusmallorca.es/>), que incluye entre la documentación editada cerca de 500 cartas, en su mayoría privadas.
- 19 Aunque los ficheros XML de cada documento no están disponibles para la consulta, esta información puede deducirse del documento que describe el sistema de anotación con TEI en este corpus.
- 20 CODEA pertenece al Grupo de Investigación Textos para la Historia del Español (GITHE, s. f.).
- 21 En la actualidad, este corpus está disponible en línea solo a través del corpus CHARTA.
- 22 La tipología de documentos del proyecto CHARTA está accesible en la página del proyecto: fecha de consulta 16 de julio de 2023, [https://75debdcd3c.cbaul-cdnwnd.com/a0324a9da168ff19cd0d0e5a8d92deb8/200000025-edbc7eeb54/Propuesta%20tipolog%C3%ADa%20documental%20CHARTA%20\(1\).pdf](https://75debdcd3c.cbaul-cdnwnd.com/a0324a9da168ff19cd0d0e5a8d92deb8/200000025-edbc7eeb54/Propuesta%20tipolog%C3%ADa%20documental%20CHARTA%20(1).pdf) En ella, se proponen las siguientes categorías de cartas: *Cartas de compraventas y contratos* y *Cartas privadas*, definidas como *Documentos de carácter privado o semiprivado*, que engloban otras subcategorías: *Cartas privadas*, *Cartas semipúblicas*, *Cartas secretas*, etc.
- 23 CODHECUN es un corpus dedicado al estudio de relaciones entre Cuba y Andalucía en el siglo XIX dentro del proyecto “Lengua, identidad y memoria a través de las cartas y la prensa de Andalucía y Cuba (s. XIX),” dirigido por Eva Bravo-García (Universidad de Sevilla), disponible en, fecha de consulta 16 de julio de 2023, <http://cuba19.us.es/cuba19/index.php?action=home>.
- 24 Agradecemos a Leyre Martínez que nos haya proporcionado información sobre el modelado de los datos del proyecto CODHECUN antes de su difusión pública.
- 25 Fecha de consulta 16 de julio de 2023, <https://www.tei-c.org/Vault/P5/4.6.0/doc/tei-p5-doc/es/html/ref-particDesc.html>. Véase Castillo Lluch y Díez del Corral (2013-).
- 26 Carta COSUIZA-0015.xml, acceso al fichero TEI-XML por gentileza de Mónica Castillo Lluch.

- 27 Aldina Quintana, comunicación personal. Para una descripción más completa de este corpus y el desafío de trabajar con una lengua sin estándar y por consiguiente sujeta a gran variación interna, véase Quintana (2020).
- 28 Extracto del documento editado en la plataforma: PSCR5285.xml.
- 29 Fecha de consulta 16 de julio de 2023, <https://ctb.kantl.be/project/dalf/dalfdoc/DALFstructure.html>.
- 30 Para este proyecto, véase Baillot 2012.
- 31 Documento CAREXIL_4140036: fecha de consulta 16 de julio de 2023, http://carexil.humanum.fr/index.php?action=file&id=CAREXIL_4140036.xml.
- 32 “The present Guidelines do not advocate any particular approach to linguistic annotation (or *tagging*); instead a number of general analytic facilities are provided which support the representation of most forms of annotation in a standard and self-documenting manner” (“15.4. Linguistic annotation of corpora,” TEI Consortium 2023).
- 33 Fecha de consulta 16 de julio de 2023, <https://www.tei-c.org/Vault/P5/4.6.0/doc/tei-p5-doc/en/html/CC.html#CCAN1>.
- 34 Fecha de consulta 16 de julio de 2023, <https://www.tei-c.org/Vault/P5/4.6.0/doc/tei-p5-doc/en/html/ref-interpretation.html>.
- 35 En este ejemplo, los códigos representan la categoría morfosintáctica de las palabras anotadas: “#AT0” artículo, “#NN1” nombre común singular, “#POS” posesivo, etc.
- 36 Existen otros módulos de anotación lingüística en TEI para las dependencias sintácticas (Graphs, Networks and Trees, fecha de consulta 16 de julio de 2023, <https://tei-c.org/Vault/P5/4.6.0/release/doc/tei-p5-doc/en/html/GD.html>) o rasgos estructurales (<fs> feature structure, fecha de consulta 16 de julio de 2023, <https://tei-c.org/Vault/P5/4.6.0/release/doc/tei-p5-doc/en/html/FS.html>), sobre los que no podemos extendernos aquí.
- 37 Una lista no exhaustiva de los proyectos que utilizan en la actualidad TEITOK está disponible en la página de la plataforma: fecha de consulta 16 de julio de 2023, <http://www.teitok.org/index.php?action=projects>.
- 38 Puede consultarse una breve descripción del proyecto CHARTA 3.0, que se encarga de esta migración, en la página siguiente: fecha de consulta 16 de julio de 2023, http://textoshispanicos.es/index.php?title=CHARTA_3.0

- 39 Documento CAREXIL_4144042.
- 40 Expert Advisory Group on Language Engineering Standards (EAGLES), Consejo de Europa. Véanse sus principales recomendaciones en Leech y Wilson (1996).
- 41 Para una explicación de las principales diferencias entre TEITOK y TEI P5, puede consultarse la plataforma TEITOK, rúbrica “TEITOK vs TEI/XML” en, fecha de consulta 16 de julio de 2023, <http://www.teitok.org/index.php?action=help&id=teixml>.
- 42 TEITOK ofrece la posibilidad de anotar también oraciones y cláusulas, pero no se ha activado todavía esta posibilidad en CAREXIL-FR.
- 43 Documento CAREXIL_7240010.
- 44 Sobre este punto en particular, véanse las recomendaciones del grupo Correspondance CAHIER en Walter 2018.

AUTOR

MARTA LÓPEZ IZQUIERDO

Marta López Izquierdo es catedrática de Lingüística española y Humanidades digitales en la Universidad París 8. En sus últimos trabajos se interesa por la variación (socio)lingüística y el contacto de lenguas en corpus escritos y orales de migrantes españoles en Francia. Es coordinadora del proyecto CAREXIL-FR. ORCID: <https://orcid.org/0000-0003-2487-0716>