

El Corpus histórico del español de Costa Rica (COHIECOS) Camino a la digitalización¹

The Corpus histórico del español de Costa Rica (COHIECOS)
On the path to digitization

Gabriela Cruz Volio

Universidad de Costa Rica (San José, Costa Rica)

gabriela.cruzvolio@ucr.ac.cr

<https://orcid.org/0000-0003-4341-5217>

Recibido el 17/10/2023, aceptado el 11/1/2024, publicado el 18/10/2024

Creative Commons Attribution 4.0 International

© 2024 Gabriela Cruz Volio

Cómo citar este artículo

Cruz Volio, Gabriela 2024. El *Corpus histórico del español de Costa Rica (COHIECOS)*. Camino a la digitalización. *Studia linguistica romanica* 2024.12, 107-135. <https://doi.org/10.25364/19.2024.12.6>.

Resumen

El objetivo de este artículo es describir el desarrollo del *Corpus histórico del español de Costa Rica (COHIECOS)*, un proyecto que busca compilar y analizar documentos coloniales costarricenses con el fin de investigar sobre la historia del español en esta región. Además de sistematizar las características del corpus, se detallan las ventajas y las desventajas de las plataformas empleadas para el procesamiento de los documentos, como el uso de *Transkribus* para la transcripción y el etiquetado. En esta línea, se destaca la importancia de visualizar y acceder a los documentos de manera eficiente, lo cual se aborda mediante la digitalización del corpus en *TEITOK*. A modo de conclusión, se presenta un balance del trabajo realizado hasta el momento, del aprovechamiento de las plataformas digitales y de las posibilidades de inclusión de este corpus en proyectos más amplios.

Abstract

The objective of this article is to describe the development of the *Corpus histórico del español de Costa Rica (COHIECOS)*, a project aimed at compiling and analyzing colonial Costa Rican documents to investigate the history of Spanish in this region. Along with outlining the features of the corpus, the article details the advantages and disadvantages of the platforms used for document processing, such as the use of *Transkribus* for transcription and tagging. In this regard, the importance of efficiently visualizing and accessing the documents is emphasized, which is addressed through the digitization of the corpus in *TEITOK*. The conclusion provides an overview of the work carried out so far, the use of digital platforms, and the potential for integrating this corpus into larger projects.

¹ Este artículo se enmarca en los proyectos de investigación *Constitución de un corpus histórico para el español colonial de Costa Rica, I Etapa* (número C0094) y *Constitución de un corpus histórico para el español colonial de Costa Rica, II Etapa* (número C3047), que se desarrollan en el seno del Instituto de investigaciones lingüísticas de la Universidad de Costa Rica. Ambos proyectos se adscriben al Programa de *Estudios coloniales centroamericanos* (número C0904) del Instituto de investigaciones lingüísticas. La carga académica para realizar estos proyectos ha sido asignada por la Escuela de filología, lingüística y literatura de la Universidad de Costa Rica.

Índice

1 Introducción.....	109
2 Características y conformación del <i>COHIECOS</i>	111
3 Transcripción y enriquecimiento de los documentos con <i>Transkribus</i>	115
3.1 Criterios de transcripción.....	115
3.2 Transcripción y enriquecimiento de los documentos con <i>Transkribus</i>	119
4 Digitalización y visualización del corpus en <i>TEITOK</i>	124
5 Conclusiones.....	128
Abreviaturas y referencias bibliográficas.....	131

1 Introducción

[1] Desde la última década del siglo XX, ha habido investigaciones sobre la historia del español de Costa Rica. Según Quesada Pacheco (1990), en la segunda edición de 1919 del *Diccionario de costarriqueñismos*, Carlos Gagini, su ilustre autor, hace referencia a la diacronía de voces costarricenses. Asimismo, Quesada Pacheco (1990) apunta que Arturo Agüero Chaves, en su libro titulado *El español de América y Costa Rica*, publicado en 1962, comenta ciertas cuestiones generales relacionadas con el habla de Costa Rica en la época colonial. A pesar de este interés por la historia de la variedad costarricense del español, no es sino hasta la publicación del libro de Quesada Pacheco (1990), *El español colonial de Costa Rica*, que se cuenta con una investigación sistemática del español de Costa Rica en los niveles fonético-fonológico, morfosintáctico y léxico-semántico de una época determinada. Más adelante, Quesada Pacheco (1995) publica el *Diccionario histórico del español de Costa Rica*, que constituye una de las pocas obras lexicográficas que se han realizado hasta el momento sobre la historia de una variedad específica del español americano. Luego, en el libro *Historia de la lengua española en Costa Rica*, Quesada Pacheco (2009) presenta un estudio histórico y abarcador del español costarricense que cubre desde el siglo XVI hasta el siglo XX.

[2] Sumado a esto, desde finales de la década de los ochenta se han realizado estudios que se concentran en aspectos más puntuales de la historia del español de Costa Rica, como el artículo de Quesada Pacheco (1987a) sobre el léxico ganadero en el español colonial de Costa Rica, los trabajos de Ulate Zúñiga (1991) y de León Fernández (1988) sobre el seseo en zonas específicas de Costa Rica durante el siglo XVI y las investigaciones de Cabal Jiménez (1997, 2021) sobre los rasgos fonético-fonológicos del español de Costa Rica a inicios del siglo XIX. Igualmente, se ha indagado acerca de aspectos morfosintácticos del verbo en el español de Costa Rica durante el siglo XIX (Quesada Pacheco 2013) y se ha investigado sobre las formas de tratamiento en Costa Rica desde una perspectiva histórica (Quesada Pacheco 2010; Cabal Jiménez 2016). Más recientemente, se cuenta con el estudio de Cruz Volio (2023) sobre usos gráficos en documentos coloniales costarricenses del siglo XVIII.

[3] A pesar del considerable interés por la historia del español de Costa Rica, ha habido poca preocupación por preparar materiales para el estudio de la lengua, excepto el libro de Quesada Pacheco (1987b), que presenta 25 documentos de archivo de 1562 a 1812 transcritos con criterios lingüísticos, y los documentos coloniales costarricenses, también editados por Quesada Pacheco, que se encuentran en Fontanella de Weinberg, Rojas Mayer & Guzmán Riverón (2000). Valga mencionar que esta escasez de materiales para la historia de la lengua se repite en Centroamérica, a excepción del laborioso trabajo de edición que ha venido realizando Ramírez Luengo (2005, 2006, 2011, 2017). Esta situación ha tenido como consecuencia que el español de Costa Rica, específicamente, tenga muy poca represen-

Documentos de Costa Rica en corpus diacrónicos

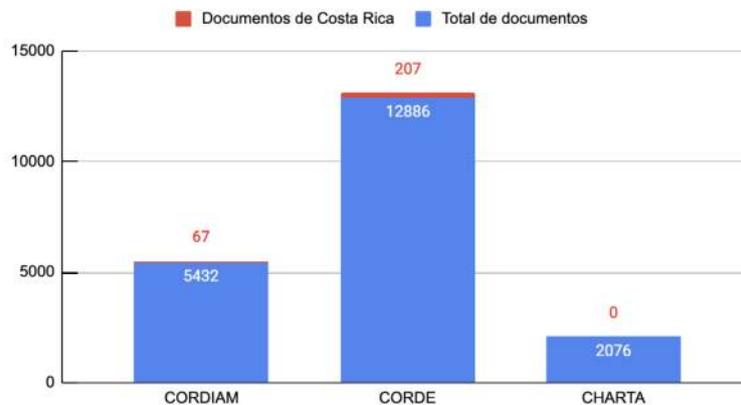


Figura 1: Documentos de Costa Rica en corpus diacrónicos (según datos de octubre de 2023)²

En la figura 1, se contrasta la cantidad de documentos de Costa Rica en relación con la totalidad de documentos disponibles en tres corpus históricos del español: el *Corpus diacrónico y diatópico del español de América (CORDIAM)*, el *Corpus diacrónico del español (CORDE)* y el *Corpus hispánico y americano en la red (CHARTA)*. En el *CORDIAM*, los datos de la figura 1 corresponden a las búsquedas realizadas en el parámetro de *tipo textual*, en el que se ha seleccionado la opción *documentos*, mientras que en los resultados del *CORDE* se han excluido los textos de prensa. En todos los casos, el periodo consultado es de 1500 a 1821. En cuanto a los 207 documentos de Costa Rica que aparecen en el *CORDE*, es importante aclarar que se trata de ediciones de documentos notariales cuya transcripción descansa sobre la base de criterios históricos, no lingüísticos³. En el *CHARTA*, aún no hay documentos de Costa Rica, pero sí hay documentación centroamericana de Guatemala, El Salvador y Panamá⁴.

[4] A raíz de la escasez de materiales preparados específicamente para la historia del español de Costa Rica y de la consiguiente baja representatividad de la variedad del español costarricense en corpus diacrónicos en línea, se tomó la deci-

² Estas consultas fueron realizadas el 13 de octubre de 2023. Valga mencionar que a finales de 2023 habrá más documentos de Costa Rica en el *CORDIAM*, que se aportaron en el marco del presente proyecto.

³ Se trata de textos incluidos en los 10 tomos de la *Colección de documentos para la historia de Costa Rica*, publicados entre 1881 y 1907, disponibles en la *Biblioteca digital AECID*.

⁴ En Cruz Volio (2021a) se ofrece un estado de la cuestión acerca del trabajo de edición de textos y de la conformación de corpus históricos hispanoamericanos y se brinda un recuento de la representación del español de Costa Rica en corpus históricos en línea.

sión de conformar un corpus de documentos coloniales costarricenses editados con criterios lingüísticos y en formato electrónico (Cruz Volio 2021a, 2021b). Por lo tanto, el objetivo de este artículo es presentar el avance del *Corpus histórico del español de Costa Rica (COHIECOS)*, detallar las ventajas y las desventajas de los programas empleados para su elaboración y ponderar el camino a seguir, particularmente en lo que atañe a su digitalización. En la siguiente sección, se describen las características y la conformación del *COHIECOS*. Luego, en la tercera sección, se especifican los criterios de transcripción y el uso de la herramienta *Transkribus* para enriquecer los documentos. La cuarta sección trata acerca del paso del corpus a formato digital y su visualización por medio de *TEITOK*. Por último, en las conclusiones se brinda un balance del trabajo realizado hasta el momento y se detallan los próximos pasos para continuar con la construcción del corpus.

2 Características y conformación del *COHIECOS*

[5] La finalidad del *COHIECOS* es crear una base documental cuyo diseño sistemático y filológico permita continuar con el estudio de la historia del español de Costa Rica (Cruz Volio 2021a). El diseño del corpus comprende la búsqueda y selección de los documentos, la descripción documental por medio de metadatos, los criterios de transcripción y el empleo de plataformas informáticas para el procesamiento de los textos.

[6] Dado el valor de la documentación archivística para el estudio de la historia del español (Ramírez Luengo 2016), el *COHIECOS* está conformado de documentos jurídico-administrativos obtenidos del Archivo nacional de Costa Rica (ANCR) y del Archivo histórico arquidiocesano Bernardo Augusto Thiel (AHABAT). En la tabla 1 se presentan los fondos que resguardan los documentos coloniales en Costa Rica y que, por la documentación que custodian, se toman en cuenta para el *COHIECOS*:

Archivo	Fondo	Periodo comprendido	Documentación
ANCR	<i>Mortuales coloniales</i>	1641-1830	Testamentos, inventarios de bienes de difuntos, etc.
ANCR	<i>Complementario colonial</i>	1517-1821	Informes, causas criminales, cartas, etc.
ANCR	<i>Cartago</i>	1578-1830	Cartas, autos, informes, relaciones, etc.
ANCR	<i>Guatemala</i>	1539-1821	Autos, informes, causas criminales, pedimentos, testamentos, demandas, quejas, castas, juicio, querellas, cuentas de bienes de difuntos, etc.
ANCR	<i>Protocolos</i>	1602-1961	Cartas de compraventa, cartas de dote, testamentos, etc.
AHABAT	<i>Fondos antiguos</i>	1582-1920	Peticiones, demandas, causas criminales, acusaciones, denuncias, inventarios, cartas, etc.

Tabla 1: Fondos de archivo con documentación colonial de Costa Rica

Puesto que el *COHIECOS* es un repositorio de documentos para la historia del español de Costa Rica, hay que aclarar que en este corpus se considera que un *texto colonial costarricense* es un texto producido en Costa Rica, aunque no necesariamente por una persona nacida en territorio costarricense, ya que no siempre se tiene certeza sobre la identidad de los escribientes (Cruz Volio 2021b). Asimismo, no se incluyen documentos de archivo completos, sino que se seleccionan una o varias unidades textuales que se encuentran en documentos de archivo de gran extensión (Cruz Volio 2021a; Cruz Volio 2021b). Por lo general, una unidad textual se inicia con la mención del lugar y de la fecha de emisión del texto y se finaliza con las firmas del escribiente y de los testigos.

[7] Para la selección de los textos se han establecido criterios geográficos, cronológicos y tipológicos, de acuerdo con los postulados de Company Company (2001a). Con respecto a los criterios geográficos, se sigue la delimitación propuesta por Quesada Pacheco (2009) para el español de Costa Rica durante los siglos XVII y XVIII, según la cual se hace una división entre las zonas altas o del Valle Central (Cartago, Ujarrás, Barva y Aserrí) y las zonas bajas o costeras (Esparza, Nicoya y Caldera en el Pacífico; Suerre y Matina en el Atlántico) a partir de una serie de rasgos fónicos.

[8] En cuanto a la delimitación cronológica, por ahora se ha trabajado con documentos del siglo XVIII y de inicios del siglo XIX, hasta 1821 (Cruz Volio 2021b). Por razones históricas, el lapso de 1700 hasta 1821 se está tratando como un solo periodo para incluir documentos de finales de la época colonial en el corpus. A su vez, los documentos se han distribuido en tres bloques de cuatro décadas

cada uno. La finalidad de esta división interna es lograr, en la medida de lo posible, un corpus más balanceado en el que no haya una cantidad de documentos concentrada en las primeras o en las últimas décadas de un periodo, por ejemplo. Claro está, esto depende de la disponibilidad de los documentos. Asimismo, se tiene previsto ampliar la extensión temporal del corpus, por lo cual ya se ha iniciado con la selección de documentos del siglo XVII, cuya distribución se hará en dos cortes cronológicos debido a la cantidad limitada y a las características tan formales de los documentos coloniales disponibles en los archivos. Si eventualmente se incluyeran documentos del siglo XVI, no sería pertinente delimitar un corte cronológico interno, dada la escasez de documentos producidos en suelo costarricense en esta época. Aunque la exploración del suelo costarricense comenzó en las primeras décadas del siglo XVI, en esta primera fase de la conquista los asentamientos fueron de carácter efímero, especialmente en el Caribe (Molina & Palmer 2007 [1997]). No es sino hasta después de 1561, en la segunda fase de la conquista, que se tomaron los territorios del Valle Central y, en 1562, Juan Vázquez de Coronado fundó la capital colonial de Cartago (Molina & Palmer 2007 [1997]). Así, el Valle Central cayó bajo el dominio español hacia finales del siglo XVI (Molina & Palmer 2007 [1997]). Aunado a ello, Fournier García (2004) indica que la parroquia de Cartago, que es el archivo de mayor antigüedad de la Iglesia en Costa Rica, se establece con la fundación de Cartago. Por lo tanto, la producción de documentos coloniales en suelo costarricense habría comenzado en fecha bastante tardía y la cantidad de documentos del siglo XVI que se conserva seguramente es ínfima⁵.

[9] El último criterio tiene que ver con el tipo textual y la temática de los textos. Si bien los textos jurídico-administrativos suelen tener un carácter bastante oficial, especialmente en las partes formulaicas al inicio de los documentos, en los archivos también se encuentran textos de tono coloquial. En este sentido, se sigue el juicio de Company Company (2001b) y se prefieren todos aquellos textos que traten acerca de temas más populares, como causas criminales, pleitos o acusaciones por injurias, especialmente si contienen declaraciones de testigos. Además, se incluyen documentos ricos para la historia del léxico, como los testamentos y los inventarios de bienes de difuntos, que contienen mucho vocabulario de la vida cotidiana. Más adelante, se pretende agregar otros tipos de documentos, como informes y relaciones de diversos asuntos. Los documentos se han clasificado de acuerdo con la tipología de *CHARTA* (2014), por su carácter más general. En la tabla 2, se resumen los documentos editados hasta el momento⁶:

5 En el *CORDIAM* hay dos documentos costarricenses producidos en 1562.

6 En la transcripción de documentos, han colaborado las filólogas Mariángel Jiménez Castro, Jimena Solano y, especialmente, Valeria Solís Lemus.

Criterios		Documentos editados	Total de documentos
Geográficos	Zona alta	34	49
	Zona baja	15	
Cronológicos	1700-1739	11	
	1740-1779	11	
	1780-1821	27	
Tipológicos	Actas y declaraciones	26	
	Testamentos e inventarios	22	
	Informes y relaciones	1	

Tabla 2: Resumen de 49 documentos editados en triple presentación y su distribución de acuerdo con los criterios de selección

La tabla 2 presenta la información de los textos del siglo XVIII transcritos paleográficamente y editados críticamente. En total, los textos cuentan con alrededor de 43000 palabras. Ya se comenzó con el proceso de búsqueda y selección de documentos del siglo XVII, pero aún no se han transcrito. Por otro lado, es importante mencionar que, además de estos 49 documentos, hay por lo menos 30 documentos del siglo XVIII que ya cuentan con la transcripción paleográfica, pero no se incluyen en la tabla 2 porque no se han editado críticamente. Como es evidente, hay más documentos pertenecientes a las zonas altas que a las zonas bajas, lo cual tiene que ver con la realidad de los archivos costarricenses, por cuanto albergan más documentos producidos en Cartago (zona alta), así como en otras regiones altas del Valle Central de Costa Rica. Lamentablemente, la documentación emanada de las zonas bajas o costeras es mucho menor, sobre todo en el Atlántico, por lo cual alcanzar un corpus balanceado del español colonial de Costa Rica es imposible (Cruz Volio 2021b). En materia de corpus históricos, se trabaja con lo que se tiene.

[10] Adicionalmente, toda la información contenida en el *COHIECOS* se ha sistematizado mediante el uso de metadatos, indispensables para la descripción, la organización y la identificación de los documentos que conforman el corpus. En la tabla 3 se presenta un ejemplo de un texto del *COHIECOS* descrito por metadatos, basados en las características de la cabecera de los documentos en *CHARTA* (2013):

Metadatos	Descripción	Ejemplo
Transcriptora	Persona o personas responsables de la edición	Gabriela Cruz Volio
Tipo	Tipo de documento según las categorías amplias de <i>CHARTA</i>	Actas y declaraciones
Archivo	Archivo que custodia el documento original	ANCR
Identificación	Fondo del archivo, signatura y folios del texto transcrito	<i>Complementario colonial</i> , expediente 6727, ff. 2r-3v
Regesto	Breve resumen del contenido del texto	Petición de Bárbara García por injurias y agresiones
Fecha	Data del texto	1801 setiembre 12
Localización	Localidad, provincia y país actual del texto	San José, San José, Costa Rica
Geolocalización	Coordenadas geográficas del texto	9.927426852716067 - 84.08845489977347
Escriptor	Persona, profesional o no, que escribe el texto. Si no se sabe o no hay seguridad sobre esta información, se consigna «???/» antes	???/ Juan Ramírez

Tabla 3: Ejemplo de un texto del *COHIECOS* descrito por metadatos (*CHARTA* 2013: 10)

A continuación, se detallan los criterios de transcripción de los documentos, el enriquecimiento con la plataforma *Transkribus* y el paso a la digitalización por medio de la máquina virtual *TEITOK*.

3 Transcripción y enriquecimiento de los documentos con *Transkribus*

3.1 Criterios de transcripción

[11] Como se mencionó en la sección 2, el propósito del *COHIECOS* es ofrecer un corpus diacrónico de soporte filológico de material fidedigno que permita realizar estudios lingüísticos. Por esta razón, los documentos se transcriben y se editan de acuerdo con los criterios de *CHARTA*, con lo cual los textos se presentan en tres versiones:

1. Facsímil.
2. Transcripción paleográfica.
3. Presentación crítica.

La versión del facsímil consiste en la reproducción del documento en soporte digital de imagen. Para garantizar una copia de alta calidad, la reproducción de los documentos jurídicos se obtiene por medio de los servicios de digitalización del ANCR y del AHABAT. En el caso de los testamentos e inventarios de bienes de difuntos, las imágenes se descargan directamente de las colecciones de *FamilySearch*, pues esta organización de interés genealógico ha digitalizado gran parte del

fondo de *Mortuales coloniales* del ANCR, que contiene testamentos e inventarios de bienes costarricenses fechados entre 1641 y 1830.

[12] En la transcripción paleográfica se siguen los lineamientos de *CHARTA* (2013) y se respetan todas las grafías que aparecen en los documentos, ya que el propósito de esta edición es posibilitar el análisis gráfico-fónico del español de Costa Rica a través de sus textos. Además, se conservan las mayúsculas y las minúsculas, las marcas de acentuación, la separación y unión de palabras, y la puntuación original del texto. Puesto que la transcripción paleográfica se mantiene fiel a los usos gráficos, resulta un tipo de edición ideal para estudiar la historia de las grafías y examinar su evolución fónica (*CHARTA* 2013). El valor de este tipo de edición consta en el estudio de Cruz Volio (2023) sobre los usos gráficos en documentos coloniales costarricenses del siglo XVIII que fueron transcritos paleográficamente. En el trabajo citado, se comprueba la existencia de una relación entre el grado de competencia de los escribientes y el empleo de normas más tradicionales que competían con las normas innovadoras. Ejemplo de ello son las grafías con <i> para la representación de la consonante fricativa palatal en palabras como *plaia* (1703, Esparza), *oio* (1710, Cartago) y *Nicoia* (1773, Bagaces), en oposición a las grafías con <y>, más innovadoras, así como los usos de las grafías con <R> y <r> en posición interior de palabra para la vibrante múltiple, como en *eRores* (1733, Barva)⁷. Es importante mencionar que estos resultados refuerzan cuestiones señaladas anteriormente por Frago (2010) en cuanto al vínculo entre grafías más conservadoras y bajos niveles de instrucción de los escribientes presente en documentos hispanoamericanos de los siglos XVIII y XIX.

[13] Por otro lado, en la presentación crítica (*CHARTA* 2013) se modernizan los usos gráficos que no representan una evolución fónica, pero se mantienen las grafías cuya variación puede ser relevante fónicamente. Asimismo, se normaliza el uso de mayúsculas y minúsculas, la acentuación, la unión y separación de palabras, y la puntuación. En el caso de elementos susceptibles de haber pasado por procesos de gramaticalización, como *aunque* y *porque*, se mantiene la separación o la unión de palabras, según aparezca en los documentos. Por estas características, la presentación crítica facilita la investigación sobre cuestiones morfosintácticas y léxicas (*CHARTA* 2013), como se ha puesto de relieve en el estudio de la historia del léxico en documentos coloniales costarricenses del siglo XVIII y principios del XIX en Cruz Volio (2024). A partir de la edición crítica de una serie de documentos costarricenses, en el trabajo mencionado se ofrece una clasificación onomasiológica del léxico cotidiano presente en inventarios de bienes de difuntos de Costa Rica. Más allá de los resultados obtenidos de la clasificación propuesta, valga mencionar algunas cuestiones puntuales con respecto a ciertos usos léxicos presentes en los documentos editados, como por ejemplo el empleo de la palabra

⁷ Los cuatro ejemplos citados con el año y el lugar de producción pertenecen a tres documentos costarricenses disponibles en el fondo de *Complementario colonial* del ANCR, con los números de expediente 5840, 6398, 5748 y 5833, respectivamente.

achota (1815, Heredia) en función de adjetivo (se refiere a una vaca de color achioté) en la categoría de bienes semovientes (Cruz Volio 2024). Si bien hay obras lexicográficas de carácter histórico que registran otros derivados de la palabra *achiote*, como el costarriqueñismo *achiotal* 'campo sembrado de achioté' (Quesada Pacheco 1995, s. v. *achiotal*), hasta donde se sabe no hay otros registros de *achota* como adjetivo en documentación colonial. Además, en los documentos también se registran costarriqueñismos como *peine* (1759, Barva)⁸, referido a una herramienta de labranza, que no aparecen en otros estudios de carácter histórico, con lo cual la edición de estos documentos⁹ permite precisar diacrónicamente y diatópicamente el léxico contenido en ellos (Cruz Volio 2023, 2024).

[14] Con respecto al tratamiento de las abreviaturas, estas se desarrollan tanto en la transcripción paleográfica como en la presentación crítica, excepto cuando se trata de formas de tratamiento como *vmd*, *vm* y *vd*, ya que no es posible saber con certeza si estas formas realmente correspondían a *vuestra merced* o a *usted* en el siglo XVIII (Obediente Sosa 2012; Ramírez Luengo 2012; Cruz Volio 2021b). En la transcripción paleográfica, el desarrollo de las abreviaturas se consigna entre corchetes angulares (por ejemplo, *d<ic>ho*), mientras que en la presentación crítica no se deja rastro del desarrollo de las abreviaturas (por ejemplo, *dicho*).

[15] En los criterios de transcripción también se toman en cuenta los aspectos codicológicos del manuscrito (*CHARTA* 2013), entre los que se incluyen indicaciones sobre el número de hojas y de líneas, fragmentos reconstruidos o ilegibles por deterioro del documento, signos o elementos especiales e intervenciones en el texto. En la tabla 4 se resumen los elementos codicológicos más frecuentes en la edición de documentos coloniales costarricenses llevada a cabo hasta el momento:

8 El vocablo *peine* se señala como costarriqueñismo en Agüero Chaves (1996, s. v. *peine*).

9 Los dos ejemplos citados con el año y el lugar de producción pertenecen a dos documentos costarricenses disponibles en el fondo de *Mortuales coloniales* del ANCR, con los números de expediente 2000 y 2043, respectivamente, y están digitalizados en *FamilySearch*.

Elementos codicológicos	Indicación	Ejemplos
Número de hoja	Se consigna entre llaves. Se indica si es recto (r) o verso (v) inmediatamente después del número de la hoja sin dejar espacio	Hoja número 1, recto: {h 1r}
Número de línea	Se consigna entre llaves en el lugar que corresponda en el texto. Si va en mitad de una palabra, no se dejan espacios	Línea 1: {1} Línea 5 en medio de palabra: testa{5}mento
Signos o elementos especiales	Su presencia se indica entre corchetes	[cruz]
Firma y rúbrica	Se indica entre corchetes. El texto de la firma queda en los corchetes, después de la indicación de «firma», seguida de dos puntos. La rúbrica corresponde al signo que le sigue a la firma	[firma: nombre] [rúbrica]
Intervenciones en el texto	Se especifica si hay texto tachado, interlineado o sobrescrito, en caso de que pueda leerse. Estas indicaciones se integran en el texto y van entre corchetes. El texto al que se refieren se incluye en los corchetes	[tachado: texto] [interlineado: texto] [sobrescrito: texto]

Tabla 4: Indicación de elementos codicológicos según los criterios de *CHARTA* (2013)

Con respecto a las indicaciones de los fragmentos reconstruidos o ilegibles en los lineamientos de *CHARTA* (2013), estas dependen de si se trata de la transcripción paleográfica o de la presentación crítica. En la edición paleográfica, cuando se sabe el número exacto de letras ilegibles, se usa un asterisco por cada letra. Si el número no se puede determinar con exactitud, el fragmento ilegible se marca con tres asteriscos entre corchetes [** * **], con cada asterisco separado por un espacio. En la presentación crítica, los fragmentos imposibles de reconstruir se indican con puntos suspensivos entre corchetes angulares <...>. En cuanto a los fragmentos reconstruidos, estos se señalan entre corchetes angulares.

[16] No obstante, la indicación de los elementos codicológicos y de las marcas de las abreviaturas puede resultar problemática para el procesamiento posterior de los textos en programas informáticos (Spence 2014), en los que las marcas de estos elementos no pueden interpretarse adecuadamente (Calderón Campos 2019). Ejemplo de ello es cuando las marcas de cambio de línea interrumpen una palabra, ya que al haber un número entre llaves en medio de la palabra los programas para la extracción de concordancias la interpretan como dos palabras independientes; esto también sucede con el desarrollo de abreviaturas por medio de corchetes angulares. Igualmente, las indicaciones de las firmas, de las rúbricas y del deterioro del documento implican la inclusión de palabras, como *tachado* o *interlineado*, que no forman parte de los textos. Por ello, al pensar en un corpus lin-

güístico como un producto digital, lo mejor sería tratar esta información como parte del enriquecimiento del texto.

3.2 Transcripción y enriquecimiento de los documentos con *Transkribus*

[17] Para transcribir los documentos del *COHIECOS*, se usa la herramienta *Transkribus*, que es una máquina de inteligencia artificial diseñada para la digitalización, el reconocimiento y la transcripción de textos. Específicamente, se ha recurrido a la versión de *Transkribus Expert Client*, que es una aplicación de escritorio que debe ser descargada e instalada en la computadora¹⁰. En el procesamiento de los documentos que conforman el *COHIECOS*, esta herramienta se ha empleado para llevar a cabo la transcripción, hasta ahora manual, y la exportación del material en diversos formatos. En principio, se escogió esta plataforma porque *Transkribus* permite reconocer las regiones de texto del documento y segmentar automáticamente las líneas de la imagen (Muehlberger et al. 2019), de modo que se establece una correspondencia entre la imagen y las líneas del editor de texto, que es el espacio en donde se agrega la transcripción, ya sea manual o automáticamente, como se observa en la figura 2:

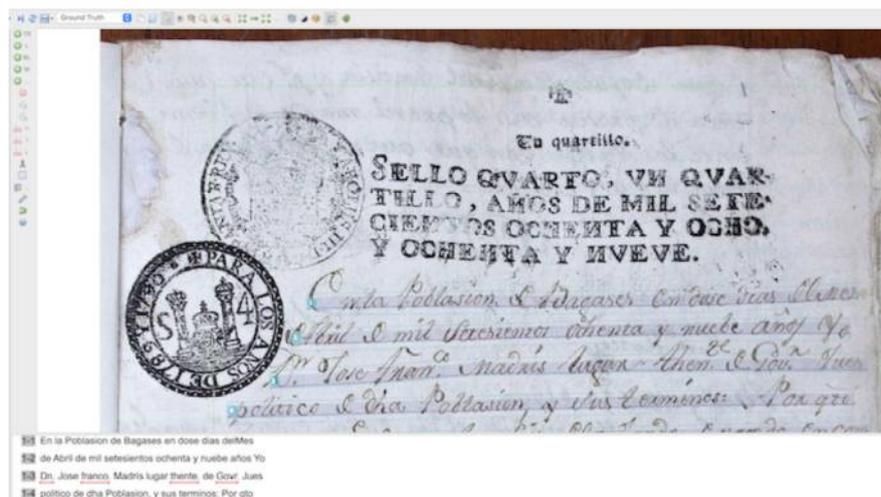


Figura 2: Alineación de la imagen con el editor de texto en *Transkribus*

El reconocimiento de regiones y la segmentación se ha hecho de forma automática con la herramienta para el análisis del diseño del texto, que detecta las regiones del texto, las líneas que lo rodean y las líneas de base que pasan justo por debajo de este (Muehlberger et al. 2019). En los documentos del *COHIECOS*, la herramienta de reconocimiento de la estructura del texto ha funcionado bastante bien.

¹⁰ Actualmente existe la aplicación de *Transkribus* en línea, que no requiere ser instalada y puede usarse directamente en un navegador.

No obstante, a veces hay que hacer algunos ajustes manuales en la detección de regiones y líneas de texto, como cuando la herramienta marca los sellos o las anotaciones al margen, que no se consideran como parte integral del documento por incluir en el corpus. Al menos por ahora, no ha sido necesario entrenar un modelo nuevo para la identificación de la estructura de estos textos. El editor de texto de *Transkribus* y su correspondencia con las líneas detectadas en las regiones del manuscrito permiten prescindir de casi todos los elementos codicológicos de *CHAR-TA*, lo cual es evidente en el caso de la numeración de líneas, como se muestra en la figura 1, más arriba.

[18] Por otro lado, en *Transkribus* es posible enriquecer los documentos con las herramientas de etiquetado estructurales y textuales (Jander 2016). Las etiquetas estructurales sirven para anotar elementos de la estructura del documento, como párrafos, encabezados, pies de página y *marginalia*, entre otros. En cambio, las etiquetas textuales se emplean para agregar otro tipo de información relacionada con el contenido, como nombres de personas y lugares, así como información que tiene que ver con las características de los documentos, como las abreviaturas, el tachado del texto, los fragmentos ilegibles y los fragmentos reconstruidos por quien edita. Los documentos del *COHIECOS* se anotan solamente con etiquetas textuales, que se resumen en la tabla 5:

Etiquetas de <i>Transkribus</i>	Función	Uso en documentos del <i>COHIECOS</i>
<i>abbrev</i>	Marcar palabras abreviadas	Se etiqueta la abreviación de la palabra en el editor de texto y se agrega la expansión de la abreviatura como una propiedad.
<i>gap</i>	Señalar fragmentos del documento imposibles de leer	Se usa cuando es imposible leer el documento y reconstruir el texto, ya sea por roturas, manchas, dobleces, etc.
<i>sic</i>	Marcar errores del escribiente	Se usa cuando es evidente que el escribiente cometió un error que no se relaciona con un aspecto de la lengua. Es posible agregar la forma correcta como una de las propiedades de esta etiqueta.
<i>unclear</i>	Señalar fragmentos de texto ilegibles	Se usa cuando no hay certeza en la transcripción porque no se puede leer bien el fragmento, ya sea por deterioro del manuscrito, porque está tachado o por la letra del escribiente. Es posible agregar una sugerencia de transcripción como parte de las propiedades de esta etiqueta.
<i>supplied</i>	Marcar texto reconstruido	Se usa cuando no se puede leer el texto, pero sí es posible deducir lo que dice. Se usa frecuentemente en los finales de palabras cuyas letras no se pueden ver bien por el doblez del manuscrito.

Tabla 5: Etiquetas textuales de *Transkribus* empleadas en la colección de *COHIECOS*

[19] En el caso de las abreviaturas, además, se agrega su desarrollo por medio de la función *expansion*, que es una de las propiedades del etiquetado y que además puede usarse para el entrenamiento de modelos de reconocimiento automático, ya sea que se presenten abreviadas o desarrolladas con la expansión (Cou-ture et al. 2022). Esta información adicional, al igual que en el resto del etiquetado, queda almacenada en el sistema de *Transkribus* y puede usarse para realizar búsquedas dentro de los textos y así asignar valores a las etiquetas recuperadas mediante el buscador, lo cual se ilustra en la figura 3:

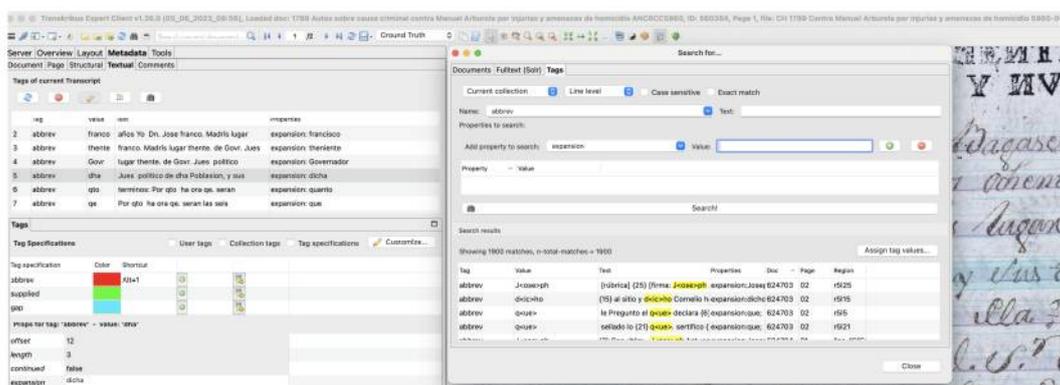


Figura 3: Etiquetado y búsqueda por etiquetas en *Transkribus*

[20] *Transkribus* permite la exportación en diversos formatos, como imágenes o archivos en PDF, XML, DOCX y TEI, entre otros (Schlagdenhauffen 2020). Para cada uno de estos formatos es posible seleccionar opciones de exportación más específicas. En la exportación en formato Microsoft Word, por ejemplo, está la opción de mantener, expandir o sustituir las abreviaturas por la forma desarrollada, así como de señalar el texto reconstruido en corchetes cuadrados, como se muestra en la figura 4:

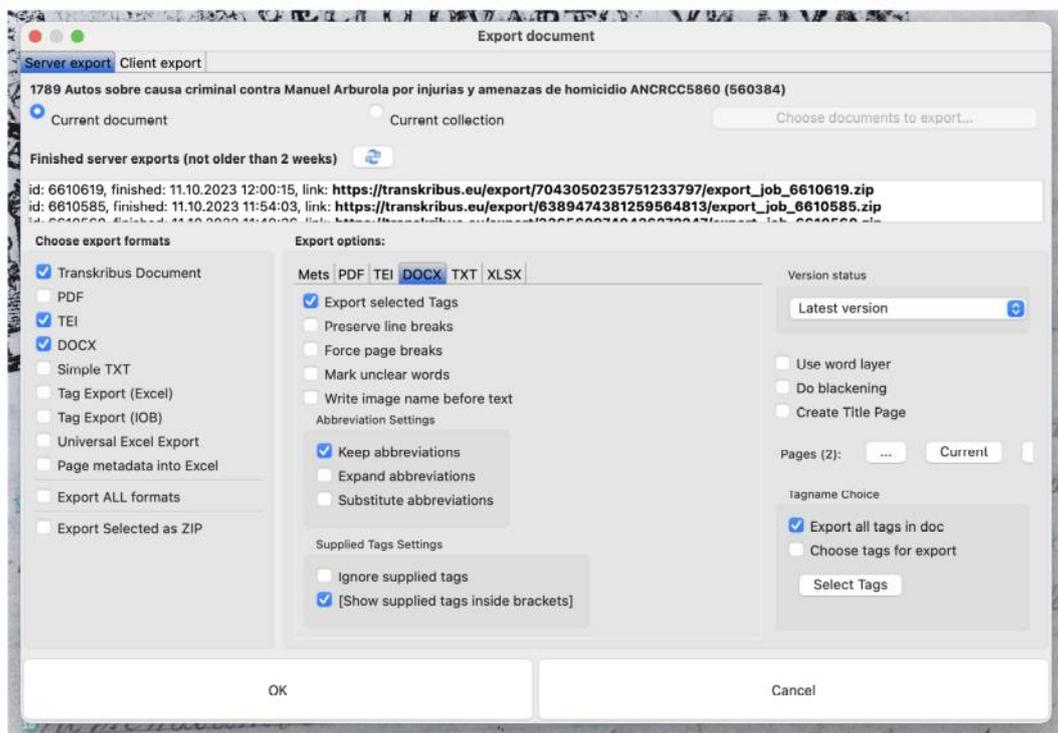


Figura 4: Opciones de exportación en diferentes formatos en *Transkribus*

[21] La codificación de los documentos mediante etiquetas se observa en la exportación del documento en el formato TEI, esquema de codificación estándar, con base en el metalenguaje XML, para la representación de textos en formato digital (Alcaraz Martínez & Vázquez Puig 2016). A modo de ejemplo, en la figura 5 se incluye un fragmento de la transcripción de un documento en formato TEI adaptado en *Transkribus*, en el cual las abreviaturas están codificadas con la etiqueta correspondiente `<abbr>` y la propiedad agregada de expansión `<expand>`:

```
<lb facs="#facs_2_r217" n="N006"/>Cerca de <choice><expand>dicho</expand><abbr>dho</abbr></choice> hatto, saliendo a buscar el otro
<lb facs="#facs_2_r218" n="N007"/>dia Vestias ael Campo hallaron al <choice><expand>dicho</expand><abbr>dho</abbr></choice> Yndio
<lb facs="#facs_2_r219" n="N008"/>dentro de Vn Pajuelar, en el Monte, aparta
<lb facs="#facs_2_r2110" n="N009"/>do Vn Poco del Camino <choice><expand>Real</expand><abbr>Rl</abbr></choice>, muerto, Y Podri
<lb facs="#facs_2_r2111" n="N010"/>do, el Cuerpo desuerte, queno Pudieron reconocerle
<lb facs="#facs_2_r2112" n="N011"/>herida ni golpe alguno, Y hauiendo hallado <choice><expand>dicho</expand><abbr>dho</abbr></choice>
```

Figura 5: Fragmento de un texto exportado en formato TEI con *Transkribus*

Así, en la primera línea de la transcripción, que se delimita de forma estándar con `<lb>` (*line beginning*), se observa la forma abreviada *dho* con su respectiva expansión *dicho*, todo dentro de la selección de la etiqueta de abreviatura `<abbr>` (*abbreviation*): `<choice><expand>dicho</expand><abbr>dho</abbr></choice>`. En este caso, la etiqueta `<choice>` que rodea a la etiqueta `<abbr>` indica que la pala-

bra tiene dos lecturas alternativas, ya sea como la forma abreviada o como la forma desarrollada, la cual se especifica en la propiedad de la etiqueta. Además de las etiquetas textuales, en TEI se codifica la estructura completa del documento y sus metadatos (Burnard 2014). Por razones de espacio, en este artículo no se abordan estas cuestiones; baste solo mencionar que la correspondencia entre la imagen y las regiones de texto se representan en la codificación de TEI. Esto se ilustra en la figura 5, más arriba, en la que la leyenda `"#fac_2_r217" n="N006"` remite al número de folio del facsímil (`#fac_2`), la ubicación de la línea en la imagen (`r217`) y el número de línea de la transcripción (`n="N006"`).

[22] Las ventajas de contar con documentos enriquecidos con anotaciones, especialmente para las búsquedas dentro de las colecciones documentales en *Transkribus*, en las que se puede buscar por palabra, por documento y por etiqueta (Muehlberger et al. 2019), son incuestionables. No obstante, la transcripción y el etiquetado de textos en *Transkribus* en ocasiones es incompatible con los criterios de transcripción de *CHARTA*, ya que persiguen objetivos distintos. La finalidad de *Transkribus* es «revolucionar el acceso a documentos históricos» por medio de la transcripción, la anotación y el entrenamiento de modelos con tecnologías diseñadas para el reconocimiento automático de textos manuscritos (Muehlberger et al. 2019: 955)¹¹, mientras que los objetivos de *CHARTA* giran alrededor del desarrollo de una metodología filológica para la edición de textos y la elaboración de corpus en línea que sirvan para el análisis lingüístico de textos (Diez del Corral Areta & Martín Aizpuru 2014). Como se pretende que el *COHIECOS* sea un corpus de estructura filológica, después de la transcripción y el procesamiento de los documentos en *Transkribus*, se exporta un documento en formato Word que sigue todos los criterios de la transcripción paleográfica de *CHARTA*. En *Transkribus*, se conserva el documento editado sin los elementos codicológicos y solo con las etiquetas textuales mencionadas anteriormente. En el caso de las abreviaturas, en *Transkribus* se mantienen sin desarrollar y su expansión queda codificada en la plataforma.

[23] Evidentemente, esto representa doble trabajo y hace falta resolver la visualización del corpus, lo cual se aborda en la siguiente sección. Sin embargo, merece la pena contar con el material transcrito y enriquecido en *Transkribus*, por cuanto esta plataforma, más allá de facilitar la transcripción, la búsqueda en documentos con herramientas informáticas y el etiquetado de documentos, es una máquina de inteligencia artificial que posibilita la transcripción automática de manuscritos con la tecnología de HTR (*hand-written text recognition*) (Muehlberger et al. 2019; Aranda García 2022). Para ello, primero hay que entrenar un modelo por medio de la transcripción manual de textos escritos con tipos de letras simila-

11 READ-COOP declara la misión perseguida con *Transkribus* en su página web: «We want to revolutionise the access to historical documents (handwritten and printed). For this, we rely on the support of cutting-edge technology such as Handwritten Text Recognition (HTR) and Keyword Spotting (KWS)».

res. La plataforma está diseñada para transcribir textos escritos en cualquier lengua. En español, por ejemplo, ya se han desarrollado algunos modelos, entre los que destacan uno para textos del Siglo de Oro en letra gótica y redonda (Bazzaco 2020; Bazzaco et al. 2022) y uno para impresos del siglo XVI (Cuéllar 2023).

[24] El problema con los documentos de archivo del periodo colonial es que son textos escritos por muchas personas diferentes, cuyos trazos de escritura tienden a variar considerablemente. En cambio, los modelos de reconocimiento automático que es posible crear en *Transkribus* se entrenan de manera más rápida y eficiente, con un margen de error inferior, cuando el material es más uniforme. Por ejemplo, los modelos para texto impreso pueden ser entrenados con tan solo 5000 palabras, mientras que para los documentos escritos a mano por una sola persona basta con 15000 palabras (Muehlberger et al. 2019). En el caso de los documentos coloniales de Costa Rica, cuya escritura no es uniforme, primero habría que transcribir una enorme cantidad de documentos para entrenar un modelo, pues es necesario que la máquina aprenda a reconocer todas las formas de las letras. De momento, esto no es posible, dadas las condiciones y los recursos limitados que se tienen para desarrollar esta empresa. No obstante, resulta alentador que otras instituciones están trabajando en proyectos similares, pero de gran envergadura, como el proyecto *Unlocking the colonial archive (UCA)*, en el que ya se está avanzando en el entrenamiento de un modelo para la letra itálica cursiva del siglo XVIII en documentos hispanoamericanos. En vista de la similitud de los materiales y de los objetivos compartidos, las transcripciones de los documentos del *COHIECOS* en *Transkribus* siguen las mismas convenciones que se usan en el *UCA* (como la transcripción de las letras voladas de las abreviaturas a nivel del texto)¹², por lo cual perfectamente podrían emplearse como material de entrenamiento para mejorar los modelos desarrollados.

4 Digitalización y visualización del corpus en *TEITOK*

[25] El trabajo de transcripción y edición de los documentos que conforman el *COHIECOS* se encuentra en el *Repositorio centroamericano de patrimonio cultural* de la Universidad de Costa Rica. Allí los documentos están disponibles en la triple presentación: facsímil, transcripción paleográfica y presentación crítica, según los criterios de *CHARTA*. Si bien se trata de una colección en formato electrónico, es un corpus 'simple' (Contreras Seitz 2008: 66), pues no está anotado lin-

12 El modelo entrenado en *Transkribus* y las convenciones de transcripción empleadas se conocieron a raíz de la participación en el curso en línea *Instituto NEH-AHRC de paleografía y humanidades digitales*, a cargo de la cooperación LLILAS Benson colecciones y Estudios latinoamericanos de la Universidad de Texas, Austin. Este curso virtual, que tuvo lugar del 23 de enero al 10 de marzo de 2023, brindó formación práctica en la lectura y visualización de manuscritos coloniales hispánicos. Debo agradecer a Albert Palacios, coordinador de investigaciones digitales de LLILAS Benson colecciones y Estudios latinoamericanos, quien tuvo la generosidad de compartir el modelo y usar una recopilación de textos tomados del *COHIECOS* para mejorar la transcripción automática de documentos coloniales costarricenses.

güísticamente. Además, puesto que las transcripciones están en dos archivos de PDF independientes, tampoco es posible hacer búsquedas complejas de información lingüística (Calderón Campos 2019). La importancia de esta colección especializada radica en el valor histórico y social de los documentos como patrimonio cultural costarricense y su difusión en un medio de acceso libre.

[26] Ahora bien, para aprovechar mejor el trabajo filológico que se ha realizado en la selección y en la transcripción de los documentos del *COHIECOS*, sobre todo en lo que atañe a la triple presentación del texto en facsímil, transcripción paleográfica y presentación crítica, hay que apuntar hacia una presentación digital del texto que permita su visualización desde varias perspectivas (Aranda García 2022). Esto significa el paso a un corpus digital en el que los textos cuenten con dos capas de información que reúnen la materialización del texto escrito en un formato legible por seres humanos y la información codificada a través de tecnologías informáticas para el procesamiento por computadoras (Aranda García 2022; Lucía Megías 2019). La articulación de estas dos capas de información es lo que permite una interacción más dinámica con el texto digital (Lucía Megías 2019), como la posibilidad de visualizar un texto en sus tres presentaciones en un único archivo digital.

[27] En lo que respecta a la creación de corpus digitales con fines lingüísticos, el proyecto *Oralia diacrónica del español (ODE)* ofrece una solución a la integración de estas dos capas de información en un corpus diseñado para el estudio diacrónico y diatópico de la lengua (Calderón Campos & Díaz-Bravo 2021). Por medio de un enfoque filológico y de procesamiento textual, así como de lingüística de corpus, en tanto los textos están codificados, el *ODE* permite la visualización y la interacción con el corpus en diferentes formatos. Esto se logra por medio de un único archivo en XML en el que están todos los niveles de la anotación y del etiquetado con TEI (Calderón Campos & Díaz-Bravo 2021)¹³. Para que los documentos del corpus puedan ser leídos sin etiquetas, pero que a la vez contengan todos los datos de codificación informática, en el *ODE* se recurre a la máquina virtual *TEITOK* (Calderón Campos 2019), de Maarten Janssen.

[28] El antecedente fundamental de corpus diacrónicos iberorrománicos que emplean la plataforma de *TEITOK* es el corpus del proyecto *Post Scriptum* (Vaamonde Dos Santos 2015). Este corpus está conformado por «cartas particulares escritas en portugués y en español durante la Edad Moderna (desde el s. XVI hasta el primer tercio del s. XIX) por personas pertenecientes a diferentes estratos sociales» (Vaamonde Dos Santos et al. 2014: 473). Las cartas, transcritas paleográficamente, cuentan con ediciones digitales codificadas con XML y cuya anotación sigue los estándares de TEI (Vaamonde Dos Santos et al. 2014). Para el procesamiento lingüístico de los textos, que comienza por la tokenización, se emplea el

¹³ Véase también Arrabal Rodríguez (2020), que detalla la elaboración de un corpus digital de inventarios de bienes de los siglos XVIII y XIX anotado con *TEITOK*. Su trabajo forma parte de *ODE*.

sistema *TEITOK*¹⁴ desarrollado por Maarten Janssen en 2014 para poder integrar las capas de información de las ediciones paleográficas del *Post Scriptum* y su respectiva anotación «en un mismo conjunto de datos XML» (Vaamonde Dos Santos 2015: 60).

[29] El sistema de *TEITOK* es una plataforma basada en la web que sirve para la creación, el alojamiento y la publicación de corpus anotados, en donde los archivos en XML están en formato TEI, pero con un sistema de tokenización modificado (Janssen 2016). En *TEITOK*, lo que permite la visualización múltiple de los textos a partir de una fuente única es el empleo de la etiqueta <tok> en un proceso automático de tokenización, que consiste en la asignación de esta etiqueta a cada una de las formas transcritas (Janssen 2016; Calderón Campos 2019; Arrabal Rodríguez 2020). Así, cada forma es identificada por medio de un solo elemento <tok> al que se le agregan distintos atributos con la información de la anotación lingüística (Janssen 2016). De esta manera, los distintos tipos de edición relacionados con cada forma transcrita se conservan vinculados en la interfaz (Vaamonde Dos Santos 2015). Además, *TEITOK* sirve para la lematización semiautomática del corpus y el etiquetado morfosintáctico de los documentos, con lo cual es posible realizar búsquedas complejas (Janssen 2016; Calderón Campos 2019).

[30] Sin lugar a duda, la plataforma de *TEITOK* de Janssen ofrece un sistema ideal para la visualización, la anotación y la búsqueda en documentos históricos diseñados para realizar estudios lingüísticos. Por eso, el *ODE* ha sido un punto de referencia fundamental para tomar la decisión de alojar los documentos del *COHIECOS* en *TEITOK*¹⁵. Otro punto importante es que el corpus *CHARTA*, dentro del cual se inscribe el *COHIECOS*, busca migrar a la plataforma *TEITOK*, como consta en la descripción del proyecto *CHARTA 3.0*. De esta manera, al transcribir los documentos con los criterios de *CHARTA* y al ser procesados con *TEITOK*, los documentos del *COHIECOS* podrán estar vinculados con *CHARTA* en su versión 3.0. El proceso de inclusión de los documentos en *TEITOK* ya se inició, si bien lentamente, y aún está en proceso. Para ello, ha sido indispensable contar con el apoyo del proyecto *CHARTA 3.0* y del Grupo de investigación de textos para la historia del español (GITHE) de la Universidad de Alcalá, gracias a cuya ayuda se obtuvieron las herramientas requeridas para convertir las transcripciones en texto plano a archivos XML compatibles con *TEITOK*¹⁶. Por ahora, el trabajo que sigue

14 El proceso de tokenización en *TEITOK* se realiza sobre un documento en formato de TEI en el cual cada token, es decir, cada una de las formas identificadas, se codifica con un elemento <tok> (Janssen 2016).

15 Además, los corpus tienen características en común, ya que ambos están conformados de documentos jurídico-administrativos, en su mayor parte inventarios y declaraciones de testigos (Calderón Campos 2019; Calderón Campos & Vaamonde Dos Santos 2020). El *ODE* incluye también certificaciones periciales de cirujanos relacionados con las declaraciones de testigos en juicios criminales (Calderón Campos & Vaamonde Dos Santos 2020). Igualmente, en el *COHIECOS* también se pretende incluir otros tipos textuales (Cruz Volio 2021a; Cruz Volio 2021b) más adelante.

16 Esto no habría sido posible sin la valiosa ayuda de Víctor Caballero Gómez, del equipo GITHE,

es continuar con la edición digital de los documentos del *COHIECOS* en *TEITOK*. Eventualmente sería deseable lograr una manera de integrar el trabajo que se ha realizado en *Transkribus* con la visualización de los documentos en *TEITOK*, pero de momento no es posible abordar esta inquietud.

[31] A continuación se muestran dos capturas de pantalla con la información y con la visualización de uno de los documentos del *COHIECOS* editados en *TEITOK*¹⁷. La figura 6 contiene el encabezado con los metadatos del documento, cuya información corresponde a la que se incluye en la cabecera de los documentos en *CHARTA* (2013: 7-10), como se había visto en la tabla 3:

COHIECOS-0002

COHIECOS-0002

Transcriber	Gabriela Cruz Volio
Corpus-Number	COHIECOS-0002
Repository	Archivo Nacional de Costa Rica
Identifier	Complementario Colonial, expediente 6727, ff. 2r-3v
Summary	Petición de Bárbara García por injurias y agresiones
Date	1801 setiembre 12
Place	San José, San José, Costa Rica
Province	San José
Country	Costa Rica
Scriptor	???/ Juan Ramírez

- [less header data](#)

Figura 6: Encabezado del documento *COHIECOS-0002* en *TEITOK*

La figura 7 presenta la visualización del documento *COHIECOS-0002* en *TEITOK*. Contiene tanto la imagen del manuscrito como la transcripción del texto, cuyos niveles de edición y de anotación pueden consultarse por medio de las opciones de visualización en la parte superior de la captura de pantalla:

a quien le agradezco profundamente su dedicada atención. Asimismo, agradezco las recomendaciones de José Luis Ramírez Luengo, del Grupo de investigación en el Seminario queretano de historia de la lengua, de Belén Almeida Cabrejas y de Ricardo Pichel, también del equipo GITHE, y de Miguel Calderón Campos, director del *ODE*, en este proceso.

¹⁷ La incorporación de los documentos del *COHIECOS* en *TEITOK* está en una etapa muy inicial, y todavía hay varias cuestiones por resolver. El *COHIECOS* es un corpus de acceso libre en línea.

View options

Text: [Paleographic form](#) [Transcribed form](#) [Full form](#) [Critical form](#) - Show: [Colors](#) [Formatting](#) [<pb>](#) [<lb>](#) [Images](#) - Tags: [Lemma](#)Edit the information about each word of this file by clicking on the word in the text below, or click [here](#) to edit the raw XML

[1r]

1 Señor **governador** y **comandante general**:
 2 Doña Barvara **García** del
 3 **vecindario** de **Villa Nueva**, en la mejor forma que ha lugar en
 4 **derecho** y en virtud a la licencia que **he impetrado** de mi
 5 **marido don Francisco Ramírez**, como consta de el **expediente**:
 6 que devidamente presento ante Vd, paresco diciendo que
 7 estando en mi casa el **factor** de **tavaco don Mariano**
 8 **Valenzuela**, **don Manuel Albarado**, **don Feliz de Bonilla**, **don**
 9 **Marcos Angulo** y **don Cipriano Fernández**, todos en
 10 **combersación** en el corredor de mi casa la noche de el **día ocho**
 11 del que rije, como entre las siete y las ocho, con mis **combersacion**
 12 **llegó a ella doña Rosa Morales**, mujer del espresado **combersacion**
 13 **Mariano**, embosada en un **cabriolés**, y armada de un **combersación**
 14 o navaja, **acompañándola** el **indio chino** que les sirv,
 15 y arremetiendo sobre su marido con un grande **escándalo**
 16 **hablándole** en estos **términos**, **habló** con respeto: "**Ahora lo**
 17 **berás**, demonio, **cómo** me la pagas **tú** y todos esos diablos.
 18 **Camina** para tu casa. **¿No** te tengo dicho que no bengas
 19 a esta casa?" **A lo** que respondió **don Mariano** que no
 20 **quería irse**, a lo que le **replicó doña Rosa**: "**¿Con** que te
 21 **quieres quedar** con todas esas **putonas**?" **A esta sasón**
 22 sacó de la bolsa la nabaja o **puñal** que **traía**



Figura 7: Visualización del documento COHIECOS-0002 en TEITOK

En la figura 7 se ha seleccionado la visualización de la presentación crítica (*critical form*), de modo que las formas editadas aparecen en color verde. Las formas marcadas con rojo corresponden a palabras que tienen una ruptura a final de línea, pero que en la visualización de formas transcritas (*transcribed form*) aparecen sin ruptura. Por otro lado, las formas en azul (*full form*) son abreviaturas que han sido desarrolladas. Al posicionar el cursor sobre cualquiera de estas formas se abre una ventana emergente en donde se muestran los niveles de edición, como la forma transcrita paleográficamente y la forma editada críticamente, como puede verse en la figura 7 con el ejemplo de la forma transcrita *combersacion* y su respectiva forma crítica *combersación*. Después del proceso de lematización y de anotación morfosintáctica, esta información también aparecería allí, pero esto corresponde a una etapa más avanzada. Por ahora se agregó manualmente el lema de la palabra *conversación* para mostrar esta posibilidad de codificación y de visualización en TEITOK.

5 Conclusiones

[32] El COHIECOS es un corpus conformado por documentos de archivo costarricenses de la época colonial, cuya finalidad es ofrecer una base documental de estructura filológica y en formato digital para el estudio de la historia del espa-

ñol de Costa Rica. Debido a la cantidad de trabajo que implica la creación de un corpus histórico de este tipo, se trata de un corpus modesto en cuanto a su tamaño. No obstante, puesto que está científicamente diseñado, este corpus podrá usarse para realizar investigaciones lingüísticas rigurosas sobre la base de datos fidedignos. El carácter científico del corpus pasa por los criterios de selección de los documentos, la transcripción y la edición en triple presentación y el uso de las plataformas empleadas para su enriquecimiento, visualización y anotación. Aunque otros países de habla hispana han avanzado en la conformación de corpus históricos, en Costa Rica todavía no se cuenta con un repositorio de este tipo para darle continuidad a los estudios diacrónicos de la variedad del español propia, de modo que este es un proyecto innovador y necesario.

[33] Por otro lado, no se trata de un proyecto aislado, sino que se inserta en el marco de la preocupación por crear bases textuales para realizar investigaciones diacrónicas del español, como el *ODE*, el *CODEA+ 2022* y el *CHARTA*¹⁸. Con respecto a corpus documentales del español americano, se encuentran el *CHEM*, el *COREECOM*¹⁹ y el *CORDIAM*. Por los criterios de transcripción empleados en el *COHIECOS*, los textos pueden ser incluidos en otros corpus diacrónicos. De esta manera, se contribuye con la representación de documentos americanos en corpus de mayor alcance, como el *CORDIAM* y el *CHARTA*. De hecho, los documentos que se han editado hasta el momento ya se enviaron al *CORDIAM*, por lo cual pronto estarán disponibles también en esta plataforma. Esto es especialmente relevante para la historia del español de Centroamérica, ya que los países centroamericanos, al menos por ahora, están muy poco representados.

[34] Si bien el paso de los documentos transcritos al sistema de *TEITOK* apenas está comenzando, ya se cuenta con las herramientas para hacerlo, lo cual significa que será posible visualizar los textos en triple presentación, lematizar el corpus y anotar los documentos, lo cual permitirá realizar búsquedas complejas. Además, el corpus podrá formar parte en el proyecto *CHARTA 3.0*. Más adelante, sería ideal lograr vincular, de forma más eficiente, el uso de *Transkribus* con *TEITOK*, pero por ahora eso no es posible, al menos en este proyecto.

[35] Puesto que el trabajo de transcripción se ha realizado con *Transkribus*, poco a poco se ha ido aumentando la cantidad de materiales necesarios para el entrenamiento de un modelo de reconocimiento automático con la tecnología que ofrece esta plataforma. Además, como los textos se han enriquecido con el etiquetado de *Transkribus*, esta información, sobre todo la del desarrollo de abreviatu-

18 En el *CORHIBER* puede consultarse la información sobre otros corpus documentales del español.

19 Al igual que otros corpus digitales que se han mencionado en este trabajo, el *COREECOM* permite realizar consultas lingüísticas de acuerdo con parámetros diacrónicos (según distintos cortes cronológicos), diatópicos (según la zona geográfica), diastráticos (según el sexo, el origen dialectal y el origen étnico-social del amanuense) y diafásicos (según el tipo de documento y la variedad textual).

ras, también puede aprovecharse para el entrenamiento de un modelo con etiquetas. De nuevo, no se trata de un esfuerzo aislado, sino que este trabajo puede ser aprovechado por proyectos como el de *Unlocking the colonial archive*, ya sea para enriquecer los modelos que allí se están desarrollando o para la creación de un modelo basado en los que ya existen.

[36] En definitiva, al poner estos aspectos sobre la balanza, las ventajas del *COHIECOS* tienen más peso que su modesto tamaño, que de todas formas seguirá creciendo, lenta pero firmemente.

Abreviaturas y referencias bibliográficas

- Agüero Chaves 1996 = Arturo Agüero Chaves 1996. *El español de Costa Rica: Léxico*. Asamblea legislativa de la República de Costa Rica.
- AHABAT = Archivo histórico arquidiocesano Bernardo Augusto Thiel. <https://ahabatcr.org/>.
- Alcaraz Martínez & Vázquez Puig 2016 = Rubén Alcaraz Martínez, Elisabet Vázquez Puig 2016. TEI: un estàndard per codificar textos en l'àmbit de les humanitats digitals. *BiD. Textos universitaris de biblioteconomia i documentació* 37.
- ANCR = Archivo nacional de Costa Rica. <https://www.archivonacional.go.cr/>.
- Aranda García 2022 = Nuria Aranda García 2022. Humanidades digitales y literatura medieval española: la integración de Transkribus en la base de datos COMEDIC. *Historias fingidas. Número especial 1. Humanidades digitales y estudios literarios hispánicos*, 127-149. <https://doi.org/10.13136/2284-2667/1107>.
- Arrabal Rodríguez 2020 = Pilar Arrabal Rodríguez 2020. Edición de un corpus digital de inventarios de bienes. *Procesamiento del lenguaje natural* 65, 67-74. <http://journal.sepln.org/sepln/ojs/ojs/index.php/pln/article/view/6278>.
- Bazzaco 2020 = Stefano Bazzaco 2020. El reconocimiento automático de textos en letra gótica del Siglo de Oro: creación de un modelo HTR basado en libros de caballerías del siglo XVI en la plataforma Transkribus. *Janus. Estudios sobre el Siglo de Oro* 9, 534-561. <https://www.-janusdigital.es/articulo.htm?id=160>.
- Bazzaco et al. 2022 = Stefano Bazzaco, Ana M. Jiménez Ruiz, Ángela Torralba Ruberte, Mónica Martín Molares 2022. Sistemas de reconocimiento de textos e impresos hispánicos de la Edad Moderna. La creación de unos modelos de HTR para la transcripción automatizada de documentos en gótica y redonda (s. XV-XVII). *Historias fingidas, Número especial 1. Humanidades digitales y estudios literarios hispánicos*, 67-125. <https://historiasfingidas.dlss.univr.it/article/view/1190>.
- Biblioteca digital AECID* = Agencia española de cooperación internacional para el desarrollo (AECID). *Biblioteca digital AECID*. <https://bibliotecadigital.aecid.es>.
- Burnard 2014 = Lou Burnard 2014. *What is the Text encoding initiative? How to add intelligent markup to digital resources*. OpenEdition Press. <https://doi.org/10.4000/books.oep.426>.
- Cabal Jiménez 1997 = Munia Cabal Jiménez 1997. *El español colonial de Costa Rica en el primer cuarto del siglo XIX: análisis fonético-fonológico*. Tesis de maestría, Universidad de Costa Rica.
- Cabal Jiménez 2016 = Munia Cabal Jiménez 2016. «Ya puede echar de ver quan complacido quedo, así porque gosas de salud como porque pones en práctica mis consejos»: variación pronominal en el español colonial de Costa Rica. *Káñina. Revista de artes y letras* 40.1, 43-70. <https://doi.org/10.15517/rk.v40i1.24139>.
- Cabal Jiménez 2021 = Munia Cabal Jiménez 2021. El español de inicios del siglo XIX en Costa Rica: nivel fonético-fonológico. Alexander Sánchez Mora, Gabriela Cruz Volio, José L. Ramírez Luengo (eds.). *La palabra olvidada: la lengua y la literatura de Centroamérica entre la colonia y la independencia. Vol. 1. Lengua*. Encino, 59-102.
- Calderón Campos 2019 = Miguel Calderón Campos 2019. La edición de corpus históricos en la plataforma TEITOK. El caso de Oralia diacrónica del español (ODE). *CHIMERA. Revista de corpus de lenguas romances y estudios lingüísticos* 6, 21-36. <https://revistas.uam.es/chimera/article/view/10999>.
- Calderón Campos & Díaz-Bravo 2021 = Miguel Calderón Campos, Rocío Díaz-Bravo 2021. An online corpus for the study of historical dialectology: *Oralia diacrónica del español. Digital scholarship in the humanities* 36.2, ii30-ii48. <https://doi.org/10.1093/dlc/fqaa066>.
- Calderón Campos & Vaamonde Dos Santos 2020 = Miguel Calderón Campos, Gael Vaamonde

- Dos Santos 2020. *Oralia diacrónica del español: Un nuevo corpus de la Edad Moderna. Scriptum digital* 9, 167-189. <https://doi.org/10.5565/rev/scriptum.108>.
- CHARTA = Pedro Sánchez-Prieto Borja (ed.) 2011-. *Corpus hispánico y americano en la red*. <https://www.redcharta.es>.
- CHARTA 2013 = CHARTA 2013. *Criterios de edición de documentos hispánicos (orígenes-siglo XIX) de la red internacional CHARTA*. <http://www.redcharta.es/criterios-de-edicion/>.
- CHARTA 2014 = CHARTA 2014. Propuesta de tipología documental (enero 2014). <http://www.red-charta.es/criterios-de-edicion/>.
- CHARTA 3.0 = Proyecto CHARTA 3.0. http://textoshispanicos.es/index.php?title=CHARTA_3.0.
- CHEM = Alfonso Medina Urrea et al. (eds.) 2024. *Corpus histórico del español en México*. Universidad Nacional Autónoma de México. <http://www.corpus.unam.mx:8080/unificado/index.jsp?c=chem>.
- CODEA+ 2022 = Pedro Sánchez-Prieto Borja (ed.) 2022. *CODEA+ 2022. Corpus de documentos españoles anteriores a 1900*. <https://www.corpuscodea.es>.
- COHIECOS = Gabriela Cruz Volio (ed.) 2020-. *Corpus histórico del español de Costa Rica*. Universidad de Costa Rica. <https://teitok.ucr.ac.cr>.
- Company Company 2001a = Concepción Company Company 2001a. Para una historia del español americano. La edición crítica de documentos coloniales de interés lingüístico. Leonardo Funes, José L. Moure (eds.). *Studia in honorem Germán Orduna*. Universidad de Alcalá, 207-224.
- Company Company 2001b = Concepción Company Company 2001b. Aspectos metodológicos prácticos para una filología lingüística del español colonial de México. Guadalupe B. Clark de Lara, Fernando Curiel Defossé (eds.). *Filología mexicana*. Universidad Autónoma de México, 111-139.
- Contreras Seitz 2008 = Manuel Contreras Seitz 2008. Cómo editar textos coloniales. *Estudios filológicos* 43, 63-82. <http://dx.doi.org/10.4067/S0071-17132008000100005>.
- CORDE = Real Academia Española (ed.) 2008. *Corpus diacrónico del español*. <http://corpus.rae.es/cordenet.html>.
- CORDIAM = Concepción Company Company, Virginia Bertolotti (eds.) 2016-. *Corpus diacrónico y diatópico del español de América*. <https://www.cordiam.org/>.
- COREECOM = Beatriz Arias Álvarez (ed.) 2019-. *Corpus electrónico del español colonial mexicano*. Universidad Nacional Autónoma de México. <https://www.iifilologicas.unam.mx/coreecom/>.
- CORHIBER = Joan Torruella, Johannes Kabatek (eds.) 2023. *Portal de corpus históricos iberorrománicos. Versión 5.0*. <http://www.corhiber.org/>.
- Couture et al. 2023 = Beatrice Couture, Farah Verret, Maxime Gohier, Dominique Deslandres 2022. The challenges of HTR model training: Feedbacks from the project *Donner le gout de l'archive a l'ere numerique*. *Journal of data mining & digital humanities. Special issue Historical documents and automatic text recognition*. <https://jdm.dh.episciences.org/12556>.
- Cruz Volio 2021a = Gabriela Cruz Volio 2021a. Hacia la conformación de un corpus histórico para el español colonial de Costa Rica. *Diseminaciones. Revista de investigación y crítica en humanidades y ciencias sociales* 4.7, 79-98. <https://revistas.uaq.mx/index.php/diseminaciones/article/view/272>.
- Cruz Volio 2021b = Gabriela Cruz Volio 2021b. Cuestiones sobre la selección y la edición de documentos coloniales para un corpus histórico del español de Costa Rica. Alexander Sánchez Mora, Gabriela Cruz Volio, José L. Ramírez Luengo (eds.). *La palabra olvidada: la lengua y la literatura de Centroamérica entre la colonia y la independencia. Vol. 1. Lengua*. Encino, 17-57.
- Cruz Volio 2023 = Gabriela Cruz Volio 2023. Usos gráficos en documentos coloniales costarricen-

- ses del siglo XVIII. Róbinson Grajales Alzate, Lirian A. Ciro (eds.). *Estudios lingüísticos e interdisciplinarios en Latinoamérica*. Peter Lang, 41-60.
- Cruz Volio 2024 = Gabriela Cruz Volio 2024. Clasificación del léxico cotidiano en inventarios de bienes costarricenses del siglo XVIII. *Káñina. Revista de artes y letras* 48.1, 1-24. <https://doi.org/10.15517/rk.v48i1.59580>.
- Cuéllar 2023 = Álvaro Cuéllar 2023. La inteligencia artificial al rescate del Siglo de Oro: transcripción y modernización automática de mil trescientos impresos y manuscritos teatrales. *Hipogrifo* 11.1, 101-115. <http://dx.doi.org/10.13035/H.2023.11.01.08>.
- Diez del Corral Areta & Martín Aizpuru 2014 = Elena Diez del Corral Areta, Leyre Martín Aizpuru 2014. Sin corpus no hay historia: La red CHARTA como un proyecto de edición común. *Cuadernos de lingüística de El Colegio de México* 2, 287-314. <https://doi.org/10.24201/lecm.v2i0.20>.
- FamilySearch = The Church of Jesus Christ of Latter-day saints. *FamilySearch*. <https://www.familysearch.org>.
- Fontanella de Weinberg, Rojas Mayer & Guzmán Riverón 2000 = María B. Fontanella de Weinberg, Elena Rojas Mayer, Martha Guzmán Riverón (eds.) 2000. *Documentos para la historia lingüística de Hispanoamérica. Siglos XVI a XVIII. Vol. 2*. Real Academia Española.
- Fournier García 2004 = Eduardo Fournier García 2004. Los archivos de la Iglesia Católica, pasado y futuro: El caso de Costa Rica. *Revista de historia* 49-50, 221-242. <https://www.revistas.una.ac.cr/index.php/historia/article/view/1787>.
- Frago 2010 = Juan A. Frago 2010. *El español de América en la independencia*. Taurus.
- GITHE = Grupo de investigación de textos para la historia del español. <https://corpora.uah.es>.
- Jander 2016 = Melina Jander 2016. *Handwritten text recognition - Transkribus: A user report*. The electronic text reuse acquisition project (eTRAP). https://www.etrapp.eu/wp-content/uploads/2016/11/TrAIN-Transkribus_User_Report-2016.pdf.
- Janssen 2016 = Maarten Janssen 2016. TEITOK: Text-faithful annotated corpora. Nicoletta Calzolari et al. (eds.). *Proceedings of the Tenth international conference on language resources and evaluation (LREC'16). May 2016. Portorož, Slovenia*. European language resources association, 4037-4043. <https://aclanthology.org/L16-1637>.
- León Fernández 1988 = Maribel León Fernández 1988. *El seseo en el español colonial escrito de Costa Rica (1561-1600)*. Tesis de grado, Universidad de Costa Rica.
- Lucía Megías 2019 = José M. Lucía Megías 2019. El editor de texto ante el reto digital: elogio de la edición 2.0. *Revista de humanidades digitales* 4, 93-114. <https://revistas.uned.es/index.php/RHD/article/view/25188>.
- Molina & Palmer 2007 [1997] = Iván Molina, Steven Palmer 2007 [1997]. *Historia de Costa Rica: breve, actualizada y con ilustraciones*. 2a edición. Universidad de Costa Rica.
- Muehlberger et al. 2019 = Guenter Muehlberger et al. 2019. Transforming scholarship in the archives through handwritten text recognition. Transkribus as a case study. *Journal of Documentation* 75.5, 954-976. <https://doi.org/10.1108/JD-07-2018-0114>.
- Obediente Sosa 2012 = Enrique Obediente Sosa 2012. El documento americano: problemas de definición y de edición. María J. Torrens Álvarez, Pedro Sánchez-Prieto Borja (eds.). *Nuevas perspectivas para la edición y el estudio de documentos hispánicos antiguos*. Peter Lang, 270-281.
- ODE = Miguel Calderón Campos, María Teresa García-Godoy 2019-. *Oralia Diacrónica del Español (ODE)*. <http://corpora.ugr.es/ode>.
- Quesada Pacheco 1987a = Miguel Á. Quesada Pacheco 1987b. Léxico ganadero de la Costa Rica colonial. *Filología y lingüística* 13.2, 147-156.
- Quesada Pacheco 1987b = Miguel Á. Quesada Pacheco 1987b. *Fuentes documentales para el estudio del español colonial de Costa Rica*. Alma Mater.

- Quesada Pacheco 1990 = Miguel Á. Quesada Pacheco 1990. *El español colonial de Costa Rica*. Universidad de Costa Rica.
- Quesada Pacheco 1995 = Miguel Á. Quesada Pacheco 1995. *Diccionario histórico del español de Costa Rica*. Universidad Estatal a Distancia.
- Quesada Pacheco 2009 = Miguel Á. Quesada Pacheco 2009. *Historia de la lengua española en Costa Rica*. Universidad de Costa Rica.
- Quesada Pacheco 2010 = Miguel Á. Quesada Pacheco 2010. Formas de tratamiento en Costa Rica y su evolución (1561-2000). Martin Hummel, Bettina Kluge, María E. Vázquez Laslop (eds.). *Formas y fórmulas de tratamiento en el mundo hispánico*. El Colegio de México, Karl-Franzens-Universität Graz, 649-670.
- Quesada Pacheco 2013 = Miguel Á. Quesada Pacheco 2013. El sistema verbal del español de Costa Rica en los albores de la época independiente. *Signo y seña* 23, 81-102. <http://revistas-cientificas.filo.uba.ar/index.php/sys/article/view/3228>.
- Post Scriptum = Centro de lingüística da Universidade de Lisboa (ed.) 2014. *Post Scriptum. Archivo digital de escritura cotidiana en Portugal y España en la Edad moderna*. <http://teitok-clul.ul.pt/postscriptum/index.php>.
- Ramírez Luengo 2005 = José L. Ramírez Luengo 2005. Contribución a la historia del español de Honduras: Edición y estudio de documentos hondureños del siglo XVIII. *Anuario de letras. lingüística y filología* 42, 51-75. <https://revistas-filologicas.unam.mx/anuario-letras/index.php/al/article/view/945>.
- Ramírez Luengo 2006 = José L. Ramírez Luengo 2006. Materiales para la historia de la lengua española en Centroamérica: Algunos documentos dieciochescos (1703-1758). *Ámbitos. Revista de estudios de ciencias sociales y humanidades* 16, 119-135. <https://helvia.uco.es/xmlui/handle/10396/11364>.
- Ramírez Luengo 2011 = José L. Ramírez Luengo 2011. Un corpus para la historia del español en Nicaragua: edición de documentos oficiales del siglo XVIII (1704-1756). *Moenia* 17, 333-366. <https://revistas.usc.gal/index.php/moenia/article/view/215>.
- Ramírez Luengo 2012 = José L. Ramírez Luengo 2012. Algunas cuestiones teóricas acerca de la edición de documentos lingüísticos americanos. María J. Torrens Álvarez, Pedro Sánchez-Prieto Borja (eds.). *Nuevas perspectivas para la edición y el estudio de documentos hispánicos antiguos*. Peter Lang, 297-306.
- Ramírez Luengo 2016 = José L. Ramírez Luengo 2016. Documentación de archivo e historia de la lengua: una reflexión desde el caso colombiano. *Lingüística y literatura* 70, 87-117. <https://doi.org/10.17533/udea.lyl.n70a04>.
- Ramírez Luengo 2017 = José L. Ramírez Luengo (ed.) 2017. *Textos para la historia del español. XI. Honduras y El Salvador*. Universidad de Alcalá.
- READ-COOP = READ-COOP SCE. Recognition and enrichment of archival documents. Societas cooperativa europaea. <https://readcoop.eu>.
- Repositorio centroamericano de patrimonio cultural = Universidad de Costa Rica. *Repositorio centroamericano de patrimonio cultural*. <https://repositorio.iiarte.ucr.ac.cr>.
- Schlagdenhauffen 2020 = Régis Schlagdenhauffen 2020. Optical recognition assisted transcription with Transkribus: The experiment concerning Eugène Wilhelm's personal diary (1885-1951). *Journal of data mining & digital humanities*. <https://dx.doi.org/10.46298/jdmdh.6249>.
- Spence 2014 = Paul Spence 2014. Siete retos en edición digital para las fuentes documentales. *Scriptum digital* 3, 153-181. <https://doi.org/10.5565/rev/scriptum.53>.
- TEI = Text encoding initiative. <http://www.tei-c.org>.
- TEITOK = Maarten Janssen 2014. *TEITOK - a tokenized TEI environment*. <http://www.teitok.org>.
- Transkribus = READ-COOP 2023. *Transkribus*. <https://www.transkribus.org/>.

UCA = *Unlocking the colonial archive*. <https://unlockingarchives.com/>.

Ulate Zúñiga 1991 = Renán Ulate Zúñiga 1991. *El seseo en el español escrito en el noroeste de Costa Rica durante el siglo XVI: zonas de Esparza y Guanacaste*. Tesis de grado, Universidad de Costa Rica.

Vaamonde Dos Santos 2015 = Gael Vaamonde Dos Santos 2015. P. S. Post Scriptum. Dos corpus diacrónicos de escritura cotidiana. *Procesamiento del lenguaje natural* 55, 57-64. <http://journal.sepln.org/sepln/ojs/ojs/index.php/pln/article/view/5216/3020>.

Vaamonde Dos Santos et al. 2014 = Gael Vaamonde Dos Santos, Ana L. Costa, Rita Marquilhas, Clara Pinto, Fernanda Pratas 2014. Post Scriptum: Archivo digital de escritura cotidiana. *Janus. Estudios sobre el Siglo de Oro*, Anexo 1, 473-482. <https://www.janusdigital.es/anexos/contribucion.htm?id=41>.