

**Hacia la conformación de un corpus histórico
para el español colonial de Costa Rica**
*Paving the Way for a Historical Corpus
of Costa Rican Spanish during the Colonial Period*

Gabriela Cruz Volio
Universidad de Costa Rica, Costa Rica
gabriela.cruzvolio@ucr.ac.cr

RESUMEN

El objetivo de este artículo es proponer criterios teórico-metodológicos para conformar un corpus diacrónico del español colonial de Costa Rica. Si bien se han realizado numerosos trabajos de transcripción y edición de textos hispanoamericanos que han resultado en la conformación de corpus históricos, la representación del español de Costa Rica todavía es muy baja, pues hay una escasez de bases documentales constituidas por textos de interés lingüístico. Para conformar un corpus diacrónico de infraestructura filológica, consideramos que la selección de documentos debe seguir criterios cronológicos, geográficos y tipológicos según las particularidades de la documentación costarricense de la época colonial y que la representación de los textos debe ser por medio de la transcripción paleográfica y la edición filológica.

PALABRAS CLAVE: corpus diacrónico; historia del español de Costa Rica; español de Centroamérica.

ABSTRACT

The goal of this article is to propose a theoretical-methodological framework in order to build a diachronic corpus of Costa Rican Spanish during the colonial period. Although there are numerous works dedicated to the transcription and edition of Hispano-American texts, which have resulted in the conformation of historical corpora, the representation of Costa Rican Spanish is still very low. This is mainly due to the lack of textual databases oriented towards the study of Costa Rican Spanish and its history. In order to constitute a diachronic corpus, we insist that the texts should be selected taking into account chronological, geographical

and typological criteria according to the documentation that was produced during the colonial period in Costa Rica. Also, we maintain that the presentation of texts should follow both a paleographic transcription and a critical edition.

KEYWORDS: *Diachronic Corpus; History of Costa Rican Spanish; Central American Spanish.*

1. INTRODUCCIÓN¹

En este artículo presentamos una propuesta teórico-metodológica para constituir un corpus histórico del español colonial de Costa Rica. Puesto que la finalidad del corpus es proporcionar una base documental diseñada sistemáticamente, el corpus estará conformado por documentos de archivo que serán transcritos paleográficamente y editados filológicamente. De esta manera, contaremos con un repositorio textual que permitirá no solo la investigación de fenómenos lingüísticos, sino también el estudio de aspectos sociales y culturales de la historia de Costa Rica en particular y de Centroamérica en general.

Puesto que el presente estudio se inserta en el marco del interés y esfuerzo de otros investigadores preocupados por crear bases textuales que permitan realizar estudios diacrónicos sobre las variedades del español hispanoamericano, en la segunda sección nos referimos a los trabajos dedicados a la edición de textos y a la conformación de corpus históricos. Luego, como la intención es crear una base documental desarrollada científicamente, en la tercera sección establecemos los conceptos teóricos relevantes, detallamos la composición del corpus y aclaramos los criterios de transcripción y edición de los textos. Por último, sintetizamos las características generales del corpus a modo de conclusión.

¹ Este artículo se enmarca en el proyecto de investigación “Constitución de un corpus histórico para el español colonial de Costa Rica – I Etapa” (número C0094), el cual se encuentra inscrito en el Instituto de Investigaciones Lingüísticas de la Universidad de Costa Rica. La carga académica para realizar este proyecto ha sido asignada por la Escuela de Filología, Lingüística y Literatura de la Universidad de Costa Rica.

2. ANTECEDENTES: EDICIÓN DE TEXTOS Y CONFORMACIÓN DE CORPUS HISTÓRICOS HISPANOAMERICANOS

De acuerdo con Company Company (2001, p. 207), la ausencia de bases documentales fidedignas desde un punto de vista filológico ha dificultado los estudios en historia de la lengua. Ciertamente, en los últimos años se ha avanzado de manera considerable en la conformación de corpus diacrónicos hispanoamericanos, como veremos en el apartado 2.1. No obstante, a pesar del desarrollo alcanzado en esta área, todavía es necesario trabajar en transcripciones y en la preparación de fuentes documentales para cubrir la amplia diversidad textual que se encuentra en los archivos hispanoamericanos, especialmente en las zonas americanas que aún no han sido cubiertas en este sentido (Hernández *et al.*, 2019, pp. 209-214). En Costa Rica, como se detalla en los apartados 2.2. y 2.3., todavía nos hace falta contar con una estructura documental accesible que sea apta para realizar estudios lingüísticos.

2.1. EDICIÓN DE TEXTOS HISPANOAMERICANOS PARA LA HISTORIA DEL ESPAÑOL

La necesidad de conformar corpus históricos hispanoamericanos para realizar estudios lingüísticos se ha sentido desde la década de los años noventa, lo cual se evidencia en los esfuerzos de distintos investigadores por transcribir y editar textos que han resultado en varias colecciones y corpus.² Por ejemplo, la obra de Company Company (1994), *Documentos Lingüísticos de la Nueva España*, reúne la transcripción de 320 textos fechados entre 1525 y 1816. Todos los textos, en su mayoría cartas, denuncias, testimonios judiciales, juicios de residencia, inventarios, testamentos, así como informes y peticiones, pertenecen a la región del Altiplano Central y se caracterizan por sus rasgos coloquiales. Asimismo, Fontanella de Weinberg (1993), Rojas Mayer (2000, 2008, 2008b) y Guzmán Riverón (2015) han compilado y editado cinco tomos de *Documentos para la historia lingüística de Hispanoamérica* con el fin de realizar estudios diacrónicos sobre las variedades americanas del español. Mientras los primeros cuatro volúmenes comprenden textos oficiales como testamentos, inventarios, cartas, actas, denuncias, etc., especialmente de Argentina, Bolivia

² Valga aclarar que este apartado no comprende una lista exhaustiva de todos los trabajos que contemplan la edición de textos.

y Venezuela de los siglos XVI al XVIII, el quinto volumen recopila textos del Caribe escritos entre el siglo XV y el siglo XVII. En el marco de esta iniciativa, se estableció como prioridad proporcionar transcripciones paleográficas que sirvieran para emprender estudios lingüísticos y procurar que la selección del material incluyera no solo documentos oficiales, sino textos de carácter más informal.

Además, desde el año 2000 la editorial Iberoamericana / Vervuert ha publicado siete volúmenes en la serie *Textos y documentos españoles y americanos* que reúnen distintas ediciones diplomáticas y paleográficas de documentos de archivo provenientes sobre todo de México y Perú (Contreras Seitz, 2017, p. 180). Igualmente, se han realizado transcripciones inéditas de diversos documentos de Bolivia, Venezuela, Argentina, Colombia, Chile, el Altiplano central de México y de materiales de la Nueva España (Provincia de Campeche, Oaxaca, Chiapas y Guatemala) con el fin de estudiar la historia lingüística de estas regiones. Las referencias completas de los trabajos acotados pueden consultarse en Contreras Seitz (2017, pp. 182-183).

En cuanto a otros trabajos de edición recientes, seguidamente nos referimos a los trabajos de Contreras Seitz (2009), Bertolotti, Coll y Polakof (2012, 2010) y Arias Álvarez (2014). El trabajo de Contreras Seitz (2009, p. 117), surge debido a la poca representación del español en Chile en corpus hispánicos, a raíz de lo cual conformó el *Corpus Diacrónico del Español de Chile (CorDECh)*, que está inscrito en el *corde (Corpus Diacrónico del Español)*. En principio, el corpus *CorDECh* se caracteriza por ser monolingüe, grande, diacrónico, textual y simple,³ pero en una segunda etapa el investigador apunta hacia la conformación de un corpus electrónico codificado y anotado (Contreras Seitz, 2009, pp. 114-115). Luego, la obra de Bertolotti, Coll y Polakof (2012, 2010), *Documentos para la historia del español en el Uruguay*, comprende alrededor de 140 textos escritos entre el siglo XVIII y el siglo XIX en la zona del actual Uruguay que se obtuvieron a partir de varios fondos documentales. El corpus,

³ Con respecto a las características acotadas (Contreras Seitz, 2009, pp. 114-115), un corpus monolingüe implica que solo se toma en cuenta una lengua o variedad lingüística. De acuerdo con la cantidad de los textos y su especificidad, un corpus grande no cuenta con un número elevado de palabras, mientras que al ser diacrónico incluye textos de diferentes períodos de tiempo. El carácter textual tiene que ver con la modalidad de la lengua, en tanto se trata de documentos escritos, y con el nivel de representatividad, pues se incluyen textos enteros. El carácter simple se refiere a que los datos del corpus no han sido ni anotados ni codificados.

de carácter simple,⁴ está conformado por cartas y documentos oficiales como “causas criminales, memorias, expedientes matrimoniales, cartas de anatemas, tasaciones de bienes, demandas judiciales, querellas civiles, sumarios, testamentos, etc.” (Bertolotti, Coll & Polakof, 2010, p. 11), los cuales se presentan en transcripción paleográfica. Por su parte, la obra dirigida y editada por Arias Álvarez (2014) comprende una amplia selección de documentos notariales de la época colonial de México. Los textos escogidos se ofrecen en transcripción paleográfica y en edición crítica, de acuerdo con los criterios del grupo *charta* (*Corpus Hispánico y Americano en la Red: Textos Antiguos*). Es importante destacar que el trabajo de Arias Álvarez se inscribe dentro del proyecto del *Corpus Electrónico del Español Colonial Mexicano (coreecom)*, el cual incluye la transcripción y edición de documentos de toda la época colonial del virreinato de Nueva España.

En el caso de Centroamérica, el panorama de la disposición de fuentes documentales para realizar análisis lingüísticos se torna un poco más escaso, según Ramírez Luengo (2004-2005, p. 52). No obstante, destacan los esfuerzos del autor citado, quien ha trabajado en la conformación de corpus documentales para el estudio del español de esta parte del continente. Así, ha editado un conjunto de 22 documentos escritos en los centros coloniales más importantes de Nicaragua durante la primera parte del Siglo de las Luces, entre 1704 y 1756 (Ramírez Luengo, 2011). Además, ha emprendido la edición de un total de 19 textos de las capitales provinciales y de otras regiones de Guatemala, Nicaragua, Honduras y El Salvador fechados entre 1703 y 1758 (Ramírez Luengo, 2006) y la transcripción y edición filológica de cinco cartas de Comayagua, Honduras redactadas entre 1711 y 1750. Recientemente, ha publicado un libro que presenta la transcripción y edición de 67 documentos de Honduras y El Salvador datados entre 1650 y 1819, los cuales en su mayoría se acercan al polo de la inmediatez comunicativa, como denuncias inquisitoriales y cartas privadas, así como declaraciones y testimonios, entre otros (Ramírez Luengo, 2017). En cuanto a la situación de Costa Rica, contamos con los trabajos de Quesada Pacheco, los cuales detallamos en el apartado 2.2., a continuación.

⁴ Un corpus simple, como se mencionó en la nota anterior, es un corpus que no ha sido ni anotado ni codificado, el cual “ha sido guardado en formato neutro (ASCII, también llamado *plain text*), y sin codificación para ninguno de sus aspectos” (Toruella y Llisteri, 1999, p. 12).

2.2. MATERIALES PARA EL ESTUDIO DE LA HISTORIA DEL ESPAÑOL DE COSTA RICA

La historia del español de Costa Rica ha sido objeto de estudio de algunos investigadores, quienes han tratado temas fonéticos y fonológicos (León Fernández, 1988; Ulate Zúñiga, 1991; Cabal-Jiménez, 1997), cuestiones de pragmática y sociolingüística (Cabal-Jiménez, 2016, 2013; Quesada Pacheco, 2005) y aspectos morfosintácticos (Quesada Pacheco, 2013) a partir de diversos documentos de archivo para la época colonial y de textos de prensa para la época independiente. Además, se encuentran los trabajos de Quesada Pacheco, quien ha elaborado el *Diccionario histórico del español de Costa Rica* (1995), ha descrito la variedad del español costarricense durante la época colonial en el libro *El español colonial de Costa Rica* (1990) y ha escrito la obra de mayor alcance titulada *Historia de la lengua española en Costa Rica* (2009).

A pesar de que definitivamente hay un interés por estudiar la historia del español en Costa Rica, son pocos los trabajos que proporcionan bases documentales de materiales de archivo para profundizar en el análisis de fenómenos fonético-fonológicos, léxicos, morfosintácticos y pragmáticos, lo cual empieza con la confección de transcripciones y ediciones filológicas. Quienes se han preocupado por transcribir documentos de archivo han sido principalmente historiadores, como se evidencia en la labor de Fernández Bonilla, quien entre 1881 y 1907 publicó 10 volúmenes de documentación archivística costarricense con materiales que van desde 1576 hasta 1821. Ahora, puesto que su finalidad era ofrecer una serie de textos para conocer la historia de Costa Rica, la colección de Fernández Bonilla reúne más que nada documentos oficiales, como cédulas reales, títulos de tierras, cartas de ayuntamiento, u otros documentos como catálogos de aves, descripciones geográficas y relaciones de suceso, mientras que excluye otros materiales que podrían dar una idea del español de carácter más coloquial. Por otro lado, se trata de transcripciones realizadas con criterios históricos, en tanto se modernizan las grafías, la puntuación e incluso el léxico del documento, lo cual definitivamente no es útil para la investigación lingüística.

En cuanto a la edición de materiales con fines lingüísticos, hasta el momento solo contamos con los 27 documentos transcritos por Quesada Pacheco que se incluyen en el segundo tomo de los *Documentos para la historia lingüística de Hispanoamérica. Siglos xvi a xviii* (Rojas Mayer, 2000) y con la edición de 25 documentos realizada por Quesada Pacheco (1987). La finalidad de esta

última obra ha sido reunir una serie de documentos de archivo fechados entre 1562 y 1818 con el fin de brindar materiales para realizar estudios lingüísticos del español colonial de Costa Rica. Si bien Quesada Pacheco se basa en las *Normas del Congreso de Archivistas* de 1961, el autor aclara que no sigue todas las cláusulas por “razones de análisis lingüístico” (Quesada Pacheco, 1987, p. 9) y procede a detallar sus criterios de transcripción. Esto, junto con el hecho de que los textos fueron escogidos de acuerdo con criterios lingüísticos, hace que el destacado trabajo de Quesada Pacheco sea un punto de partida esencial en la conformación de bases documentales del español colonial de Costa Rica.

En esta misma línea, el presente proyecto pretende contribuir a ampliar la infraestructura documental para el estudio diacrónico del español de Costa Rica al proporcionar un conjunto de textos transcritos paleográficamente y editados filológicamente. Además, queremos ofrecer un corpus electrónico de libre acceso que podrá ser incluido en el Sistema de Corpus del Instituto de Investigaciones Lingüísticas (SCINIL).

2.3. REPRESENTACIÓN DEL ESPAÑOL DE COSTA RICA EN CORPUS DIACRÓNICOS

La recopilación, transcripción y edición de diversos textos procedentes de distintas partes de Hispanoamérica ha resultado en su representación dentro de corpus diacrónicos hispánicos. Por ejemplo, el *Corpus Diacrónico y Diatópico del Español de América (cordiam)*, dirigido por Concepción Company y Virginia Bertolotti y lanzado por la Academia Mexicana de la Lengua en 2015, toma sus fuentes del trabajo de transcripción y edición que han realizado diferentes equipos investigadores, como los mencionados en los apartados anteriores. El *cordiam* es un corpus electrónico y de libre acceso que consta de 12907 textos divididos en tres grandes categorías (4946 de documentos, 2436 de literatura y 5525 de prensa) procedentes de los 19 países americanos de habla hispana más el sur y el oeste de Estados Unidos, Jamaica, Haití y Guyana. De los 4946 textos clasificados como documentos, que es lo que nos interesa para el presente proyecto de investigación, tan solo 67 documentos son de Costa Rica.⁵

⁵ Los datos han sido obtenidos en la página de CORDIAM (<https://www.cordiam.org/doc/tres-corpus.html>) [Acceso: 29 de agosto de 2020]. El cálculo del total de documentos de Costa Rica se obtuvo por medio de la búsqueda de palabras en la plataforma.

Además, contamos con el *Corpus Hispánico y Americano en la Red: Textos Antiguos (charta)*, que se trata de un proyecto cuyo objetivo es editar y analizar documentos archivísticos en español de los siglos XII al XIX. En este proyecto se reúnen textos de toda clase de registros, por lo cual comprende tanto documentos oficiales que se caracterizan por un registro culto y elevado como documentos no oficiales que recogen el uso lingüístico de personas de diversos estratos sociales y en los que se pueden ver marcas de oralidad. En cuanto a la distribución geográfica de los materiales contenidos en *charta*, de un total de 2076 documentos, 89 de ellos son de países hispanoamericanos como Colombia, Cuba, Ecuador, El Salvador, México, Panamá, República Dominicana, Venezuela y Guatemala. En este corpus, todavía no hay textos procedentes de Costa Rica.

Por último, debemos mencionar el *Corpus Diacrónico del Español (corde)* de la Real Academia Española, el cual recopila textos de variada naturaleza que van desde los orígenes del español hasta el año 1974. De aproximadamente 300 millones de palabras que conforman el corpus, 19.48% de ellas se encuentran en textos hispanoamericanos, según Contreras Seitz (2009, p. 115). En cuanto a los documentos de Costa Rica correspondientes a la época colonial, en el *corde* hay 207 textos costarricenses que datan del año 1563 al año 1818. En total, estos documentos de Costa Rica contienen 356250 palabras.⁶ Si bien esta es una cantidad nada despreciable, se trata exclusivamente de documentos notariales y de ordenamientos y códigos legales, que además han sido transcritos sobre la base de criterios históricos, no lingüísticos.

Dada la baja representación del español de Costa Rica en los corpus diacrónicos disponibles, consideramos necesario emprender proyectos que contribuyan a aumentar la presencia del español costarricense a través de bases documentales filológicas, cuyos textos sean transcritos y editados de acuerdo con criterios lingüísticos. Esto permitirá dar una mayor continuidad a los estudios de historia del español de Costa Rica.

3. DISEÑO DEL CORPUS: CONSIDERACIONES TEÓRICO-METODOLÓGICAS

La conformación de un corpus diseñado científicamente implica una serie de aspectos teóricos y metodológicos. Por lo tanto, en primer lugar, hacemos

⁶ Estas cifras han sido obtenidas de la nómina de autores y obras del CORDE, en la página <http://ntlle.rae.es/nomina/jsp/NominaFor.jsp> [Acceso: 29 de agosto de 2020].

referencia a la relación entre la variación lingüística y la modalidad escrita del lenguaje. Luego, detallamos la composición del corpus y explicamos de qué manera se llevará a cabo la selección de los textos que conformarán el corpus. Finalmente, mencionamos las características de la transcripción y la edición de los textos.

3.1. VARIACIÓN LINGÜÍSTICA Y MODALIDAD ESCRITA DEL LENGUAJE

Puesto que el objetivo de este proyecto es proporcionar una base documental diseñada científicamente para realizar estudios lingüísticos, es necesario considerar el carácter dinámico del lenguaje, según el cual las lenguas siempre están en un proceso de variación que explicaría el cambio lingüístico a nivel diacrónico (Martinet, 1990). Dado que la lengua se encuentra en un proceso constante de “sincronía dinámica” (Martinet, 1984), es esperable asumir que la manifestación de variantes propias de la modalidad oral también se perciba en la modalidad escrita, por lo cual los textos que conforman la base documental son un medio para conocer el cambio lingüístico. Esto, claro está, sin ignorar el hecho de que el lenguaje escrito tiende a tener un carácter más conservador que el lenguaje oral, como lo señala Lass (1997, p. 58).

Consecuentemente, en la transcripción de los textos mantendremos las variaciones o anomalías gramaticales, ya que estas pueden ser significativas en tanto reflejan inestabilidad en el sistema lingüístico y pueden representar cambios gramaticales o pragmáticos en proceso, de acuerdo con Company Company (2001, p. 211). Ahora bien, la consideración anterior nunca estará libre de problemas, pues como señala la autora citada, siempre queda la duda de si las variaciones gráficas son efectivamente significativas o si se trata de un simple error o descuido del hablante-escritor. Sin dejar de lado esta observación, para las transcripciones emplearemos métodos paleográficos, ya que la paleografía es inseparable del estudio de la lengua, especialmente en sus niveles gráfico y fónico (Sánchez Prieto-Borja, 1998, p. 37). Como lo expresa Contreras Seitz (2009, p. 112), la paleografía “[...] comporta una relación con la conciencia lingüística y escrituraria del hablante, a quien las distinciones de este último tipo le eran más propicias en la medida en que el sistema fonológico que poseía resguardaba determinadas oposiciones del mismo”.

Para la conformación de corpus históricos, evidentemente, solo se cuenta con datos pertenecientes a la modalidad escrita. Si bien esto aparenta ser pro-

blemático porque se supone que solo los datos propios del medio oral pueden reflejar el habla, las investigaciones en pragmática histórica comprueban que es posible encontrar elementos propios de la oralidad en textos del medio escrito. Al retomar la propuesta de Koch y Oesterreicher (1985) en la que se plantea que la concepción oral y la concepción escrita del lenguaje forman parte de un *continuum*, consideramos que las características relacionadas con la inmediatez comunicativa, típicas de la concepción oral del lenguaje, no quedan restringidas al medio oral, sino que también se dan en el código gráfico o escrito. Así, los rasgos de la inmediatez comunicativa pueden encontrarse en textos escritos que reconstruyen el lenguaje oral, como en el caso de documentos jurídicos, por ejemplo, como señala Kytö (2010, p. 49).

Por otra parte, Conde-Silvestre (2007, p. 44) hace énfasis en que los datos escritos también son susceptibles de manifestar la variación estructurada propia de la lengua. Este autor indica que las fuentes escritas son válidas para el análisis sociolingüístico, ya que es posible correlacionar ciertas elecciones de expresión lingüística en documentos escritos con factores sociales determinados. Para una investigación de carácter sociolingüístico, eso sí, lo ideal es buscar textos escritos que de algún modo registren intercambios comunicativos que han sucedido en el medio oral (Conde-Silvestre, 2007, p. 45).

3.2. COMPOSICIÓN DEL CORPUS

El corpus que conformaremos en este proyecto será de carácter electrónico y simple en el sentido de que los datos presentados no estarán ni anotados ni codificados, al menos en un inicio, pues dado el estado manuscrito de los documentos de archivo se hace imperativo empezar con su transcripción. Se trata de un corpus diacrónico, que en esta primera etapa iniciará con la recopilación de textos del siglo XVIII. Más adelante, seguiremos con la transcripción y edición de documentos coloniales costarricenses de los siglos XVII y XVI.

El corpus estará constituido por documentos de archivo. Si bien en los archivos encontramos mucha documentación de carácter oficial, Company Company (2001b, p. 122) insiste en que es posible conseguir instancias de lenguaje coloquial en este tipo de documentos. Para esto, la autora citada recomienda fijarse en fondos que contengan correspondencia, pleitos por robo o asesinato, documentación legal sobre prohibiciones que tengan que ver con la regulación de la vida cotidiana, etc., en los que es más probable obtener muestras de lenguaje espontáneo. Cabe recalcar que los documentos

de archivo tienen sus ventajas, pues su procedencia geográfica y temporal suele estar detallada en el texto mismo. Además, a diferencia de los textos literarios, no siguen normas estéticas y pueden reflejar el modo de hablar de personas de distintos grupos sociales cuando presentan testimonios de informantes.

En un principio, recurriremos a la documentación del Archivo Nacional de Costa Rica.⁷ En particular, los materiales se obtendrán de los siguientes fondos del Archivo Histórico (Dirección General del Archivo Nacional, 2018), debido a que contienen documentos como mandatos, cartas, informes, testimonios de sucesión, litigios, demandas, quejas, peticiones, solicitudes, juicios, reclamos, querellas, gastos y peticiones de vecinos, inventarios de bienes, testamentos y juicios de sucesión de bienes, entre otros:

- a. Cartago, 1578-1830: contiene 1150 documentos textuales.
- b. Guatemala, 1539-1821: contiene 1181 documentos textuales.
- c. Complementario colonial, 1517-1829: contiene 7501 documentos textuales.
- d. Mortuales coloniales, 1641-1830: contiene 2485 documentos textuales.

3.2.1. CRITERIOS DE SELECCIÓN

Realizaremos la selección de documentos según tres criterios: a) cronológico, b) geográfico y c) tipológico. En cuanto al criterio cronológico, en una primera etapa trabajaremos con documentos redactados entre 1700 y 1821. En este primer período, distribuiremos los textos en tres cortes cronológicos de cuarenta años cada uno, de acuerdo con el supuesto de Company Company (1994, citada por Contreras Seitz, 2009, p. 117) de que en el lapso de unas dos generaciones los cambios lingüísticos podrían ser percibidos con mayor facilidad en la lengua escrita. Posteriormente, incluiremos textos de los siglos XVII y XVIII.

Con respecto al criterio geográfico, buscaremos documentación redactada en las dos zonas dialectales que distingue Quesada Pacheco (2009, pp. 488-491) para los períodos de 1590 a 1650 y de 1650 a 1850: a) regiones altas o interiores y b) regiones bajas o costeras. Las regiones altas o interiores corresponden al Valle Central, cuya documentación se produjo en Cartago, Ujarrás,

⁷ Valga decir que el Archivo Histórico Arquidiocesano Monseñor Bernardo Augusto Thiel, ubicado en San José, Costa Rica, conserva una gran cantidad de documentos coloniales. En otras etapas del proyecto, también recurriremos a este patrimonio documental.

Barva y Aserrí durante el período colonial. Las regiones bajas o costeras, por su lado, se dividen en el Pacífico (con Esparza, Nicoya y Caldera) y en el Atlántico (con Suerre y Matina).

En cuanto al criterio tipológico, nos concentraremos en tipos de textos que se acerquen, en la medida de lo posible, al polo de la inmediatez comunicativa. A su vez, para determinar qué textos pueden reflejar un uso coloquial y espontáneo del lenguaje, hemos establecido tres criterios que se aclaran en la subsección 3.2.2., a continuación.

En aras de buscar el balance, trataremos de que los períodos temporales, las zonas dialectales y los tipos textuales escogidos estén representados de manera más o menos equitativa. Ahora bien, somos conscientes de que la construcción de un corpus representativo para la historia de la lengua es imposible tanto en la teoría como en la práctica, pues, como señala Kabatek (2013, p. 9), “la lengua no es la suma de los textos sino algo distinto” y la “producción de textos archivados no corresponde a más que una mínima parte de la producción lingüística total”.⁸ Por lo tanto, siguiendo al autor citado (Kabatek, 2013), apuntaremos a una “representatividad cualitativa” por medio de un corpus que presente una amplia gama de las posibilidades de la lengua en un momento dado.

3.2.2. TEXTOS DE INTERÉS LINGÜÍSTICO

Como señala Company Company (2001, p. 214), la constitución de un corpus con fines lingüísticos no cuenta con textos que puedan ser considerados de interés lingüístico *a priori*, como sí sucede con la edición de textos literarios, que buscan la recuperación y la fijación de un texto en específico que se tiene de antemano. Como mencionamos anteriormente, entendemos que un texto es de interés lingüístico cuando presenta rasgos propios de la concepción oral del lenguaje y se aproxima a la inmediatez comunicativa. Consecuentemente, nos basaremos en tres criterios para seleccionar los textos que conformarán el corpus: a) la hechura del documento, b) el tipo textual y c) la competencia escrita de impronta oral.

⁸ Es importante aclarar que no se trata de una dificultad exclusiva de los corpus diacrónicos, sino que tanto la representatividad como el balance son metas que han sido cuestionadas también en los corpus sincrónicos y de variedades lingüísticas contemporáneas. Mientras que Mukherjee (citado por Kohnen, 2007) afirma que la representatividad absoluta es imposible, McEnery, Xiao y Tono (citados por Kohnen, 2007), indican que cualquier afirmación que tenga que ver con el balance de un corpus es un “acto de fe”.

Dado que en los archivos se encuentra una enorme cantidad de documentación oficial, debemos prestar atención a documentos que estén escritos con letras cursivas y rápidas, que no contengan sellos oficiales o cuyo tamaño de papel no sea el típico del usado en el ámbito administrativo. Estos factores, que tienen que ver con la hechura del documento, aumentan la probabilidad de que se trate de textos de valor lingüístico, como afirma Company Company (2001b, p. 126), ya que son indicadores de un carácter menos oficial.

Para el criterio de tipo textual, seguimos la tipología propuesta por Bertolotti y Company (2014) y su clasificación de textos jurídicos, administrativos y cronísticos. Los textos jurídicos, si bien contienen muchas fórmulas fijas y pueden ser de carácter muy oficial, también presentan acotaciones más espontáneas, especialmente en el cuerpo de la denuncia y en la transcripción de testimonios. En este sentido, Company Company (2001b, p. 135) aconseja escoger juicios de corte popular, como aquellos hechos por asesinato, adulterio, brujería o acusaciones de injuria, por ejemplo. En los textos de tipo administrativo se encuentran documentos como testamentos e inventarios de bienes, los cuales son fundamentales para los estudios de léxico, especialmente lo que tiene que ver con cuestiones de la vida cotidiana, por lo cual privilegiaremos este tipo de documentación. Los documentos cronísticos contienen secuencias descriptivas y narrativas, puesto que “describen paisajes, relatan sucesos raros o curiosos para el «cronista», describen acciones propias de ciertos grupos humanos, creencias, costumbres, festividades y comportamientos de esos grupos y tienen, muchas veces, aunque no necesariamente, una ordenación temporal” (Bertolotti y Company, 2014, p. 143). Este tipo textual es relevante por el criterio relacionado con el grado de competencia escrita.

De acuerdo con Oesterreicher (2005, p. 746), hay determinadas situaciones comunicativas que propician la presencia de lo “hablado en lo escrito”. Así, los textos que muestran una *competencia escrita de impronta oral* tienen que ver con la formación cultural deficiente del escribiente, tal como explica el autor citado:

Es decir, el autor que escribe (o dicta) un texto no conoce suficientemente ni la variedad lingüística exigida por el género respectivo, ni las reglas discursivas válidas para la estructuración del texto. Muchas veces, no sabe aprovechar las posibilidades de la comunicación escrita (apuntes para la redacción, preparación temática, corrección de lo escrito, etc.). En consecuencia, el texto producido contiene, por regla general, construcciones y elementos que normalmente sólo

se utilizan en el ámbito de la inmediatez comunicativa: por un lado, rasgos universales de la lengua hablada y, por otro, fenómenos que pertenecen claramente, bien a registros diafásicos o niveles diastráticos bajos, bien a variantes dialectales normalmente no admitidas en la escritura. Además, la inseguridad en el manejo de las normas lingüísticas y discursivo-textuales provoca ultracorrecciones y fórmulas fijas (Oesterreicher, 2005, p. 747).

De acuerdo con estos postulados, escogeremos documentos como cartas, informes y todo tipo de relaciones que estén redactadas en primera persona, que sean abundantes en pasajes narrativos y descriptivos y que tengan un carácter más personal que el de documentos de carácter más oficial.

En este punto, queremos mencionar que todos los tipos textuales acotados, así como aquellos que presentan las características deseables en cuanto a la hechura del documento, se encuentran en los fondos documentales del Archivo Nacional de Costa Rica, pues ya lo hemos constatado por medio de búsquedas a partir de los ficheros del Archivo y con el cotejo de los documentos obtenidos gracias a los servicios que brinda esta institución.

3.3. TRANSCRIPCIÓN Y EDICIÓN DEL CORPUS

En los archivos hispanoamericanos contamos con diversos documentos manuscritos en varios tipos de letra procesal, como la semejante a la cortesana y la de formas encadenadas con mezcla de bastardilla o itálica (Cavallini de Arauz, 1986, p. 16). Estos documentos deben hacerse accesibles para facilitar su lectura, pues en estado manuscrito, ya sea en soporte físico, en formato de imagen digitalizada o en soporte de microfilm, se hace muy difícil la realización de estudios lingüísticos de manera sistemática. Por lo tanto, después de seleccionar la documentación relevante el siguiente paso será transcribir y editar los textos, lo cual, en palabras de Contreras Seitz (2017, p. 176), resulta ser la cuestión de “[...] mayor relevancia, la más demandante y, también, la menos comprendida”. Tanto la transcripción como la edición de los textos deben fundamentarse en la coherencia, homogeneidad y continuidad, como acota Company Company (2001, p. 220), lo cual se procurará al establecer una serie de criterios previamente.

En la transcripción paleográfica, nos apegaremos estrictamente a la fidelidad con respecto a los usos gráficos de los textos, pero no con respecto a las diferentes formas que puede adoptar una misma letra, pues la finalidad

es examinar el valor y la evolución fonética de las grafías, de acuerdo con los postulados del proyecto *charta* (2013, pp. 6, 15). Puesto que nuestro objetivo es respetar el documento original en cuanto a su variación y a sus particularidades codicológicas, en la transcripción paleográfica de los textos representaremos las tachaduras y adiciones hechas por el hablante/escribiente, señalaremos los cambios de renglón en el original y respetaremos las grafías que pueden dar cuenta de distinciones fonéticas.

Para formular la edición filológica de los textos, también nos basaremos en los criterios de *charta* (2013), específicamente en los que atañen a la presentación crítica de los textos, de manera que realizaremos modificaciones como la modernización de mayúsculas y minúsculas y de la puntuación. También emplearemos las normas de acentuación de uso actual, ya que, como apunta Company Company (2001, p. 221), esto facilitará el uso del material editado para la elaboración de obras filológicas como léxicos y diccionarios históricos, pues la acentuación afecta la identificación y el reconocimiento de lemas en una computadora. De acuerdo con Contreras Seitz (2009, p. 130), contar con la transcripción paleográfica, así como con la edición filológica de cada texto abrirá más campos de investigación y permitirá la amplia difusión del corpus. Mientras que la transcripción paleográfica se enfoca en el sistema gráfico y permite estudiar los valores fonéticos de las grafías, la edición filológica sirve para facilitar la lectura y es más apta para investigar aspectos morfosintácticos y léxicos, así como aspectos históricos y sociales (*charta*, 2013, p. 6).

Por último, con el fin de conservar y compartir el trabajo realizado con las comunidades investigadoras, incluiremos las transcripciones paleográficas y las ediciones filológicas de los textos en el Sistema de Corpus del INIL (SCINIL). Al formar parte de un corpus electrónico, los textos no solo se podrán leer fácilmente, sino que podrán ser usados para estudios de lingüística histórica del español de Costa Rica.

4. CONCLUSIONES: CARACTERÍSTICAS DEL CORPUS

A modo de conclusión, sintetizamos las características que tendrá el corpus que proponemos:

- a. Tamaño: en una primera etapa, trabajaremos con 25 textos del siglo XVIII.
- b. Abierto: si bien se trata de un corpus pequeño, la idea es ampliarlo continuamente con la inclusión de más textos transcritos y editados, tanto del siglo XVIII como de los siglos XVII y XVI.
- c. Primario: en palabras de Sánchez-Prieto Borja (2012, p. 9), se trata de materiales textuales propios, ya que no estamos tomando ediciones ya formuladas, sino que somos responsables de todo el proceso de edición.
- d. Representatividad textual: incluiremos solo textos completos, no fragmentarios, por lo cual se tratará de un corpus “short and fat”, de acuerdo con Kohnen (2007).
- e. Infraestructura filológica: el proceso de conformación del corpus implica una serie de criterios de selección y de presentación rigurosos por medio de la transcripción paleográfica y la edición filológica.
- f. Formato electrónico: el corpus será de libre acceso y podrá revisarse y actualizarse continuamente. Si bien en un principio será simple, eventualmente se contemplará su anotación.

REFERENCIAS BIBLIOGRÁFICAS

- Arias Álvarez, B. (Ed.). (2014). *Documentos públicos y privados del siglo xvi. Textos para la historia colonial mexicana i*. México: Universidad Autónoma Nacional de México.
- Bertolotti, V., Coll, M. y Polakof, A.C. (2010). *Documentos para la historia del español en el Uruguay. Vol. 1. Cartas personales y documentos oficiales y privados del siglo xviii*. Montevideo: Facultad de Humanidades y Ciencias de la Educación.
- Bertolotti, V., Coll, M. y Polakof, A.C. (2012). *Documentos para la historia del español en el Uruguay. Vol. 2. Cartas personales y documentos oficiales y privados del siglo xix*. Montevideo: Facultad de Humanidades y Ciencias de la Educación.
- Bertolotti, V. y Company Company, C. (2014). “El Corpus Diacrónico y Diatópico del Español de América (CORDIAM). Propuesta de tipología textual”. *Cuadernos de la alfal*, 6, pp. 130-148.
- Cabal-Jiménez, M. (1997). *El español colonial de Costa Rica en el primer cuarto del siglo xix: análisis fonético-fonológico* (Tesis de maestría). Universidad de Costa Rica, San José.

- _____. (2013). *From deference to face: the evolution of usted in Costa Rican Spanish* (Tesis doctoral). Universidad de Illinois, Urbana-Champaign.
- Cavallini de Arauz, L. (1986). *Elementos de paleografía hispanoamericana*. San José: Universidad de Costa Rica.
- _____. (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”. *Kánina*, 40(1), pp. 43-70.
- [charta] *Corpus Hispánico y Americano en la Red: Textos Antiguos*. Recuperado de: www.corpuscharta.es
- [charta] (2013). *Criterios de edición de documentos hispánicos (Orígenes-Siglo xix) de la Red Internacional charta* [Versión de abril de 2013]. Recuperado de: <https://www.redcharta.es/criterios-de-edicion/>
- Company Company, C. (1994). *Documentos Lingüísticos de la Nueva España*. México: Universidad Nacional Autónoma de México.
- _____. (2001a). “Para una historia del español americano. La edición crítica de documentos coloniales de interés lingüístico”. En L. Funes y J. L. Moure (Eds.), *Studia in honorem Germán Orduna* (pp. 207-224). Alcalá de Henares: Universidad de Alcalá.
- _____. (2001b). “Aspectos metodológicos prácticos para una filología lingüística del español colonial de México”. En B. Clark de Lara y F. Curiel Defossé (Coords.), *Filología mexicana* (pp. 111-139). México: Universidad Autónoma de México.
- Conde Silvestre, J. C. (2007). *Sociolingüística histórica*. Madrid: Gredos.
- Contreras Seitz, M. (2008). “Cómo editar textos coloniales”. *Estudios Filológicos*, 43, pp. 63-82.
- _____. (2009). “Hacia la constitución de un corpus diacrónico del español de Chile”. *Revista de Lingüística Teórica y Aplicada, Concepción (Chile)*, 47(2), pp. 111-134.
- Contreras Seitz, M. (2017). “Lo que cuentan los documentos: Para una historia del español de Chile en el período Colonial”. *Atenea*, 515, pp. 173-188.
- [corde] Real Academia Española: Banco de datos (corde) [en línea]. *Corpus diacrónico del español*. Recuperado de: <http://corpus.rae.es/cordenet.html>
- [cordiam] Academia Mexicana de la Lengua, *Corpus Diacrónico y Diatópico del Español de América*. Recuperado de: www.cordiam.org.
- [coreecom] *Corpus Electrónico del Español Colonial Mexicano* (coordinado por B. Arias Álvarez). Recuperado de: <http://www.iifilologicas.unam.mx/coreecom/>

- Dirección General del Archivo Nacional. (2018). *Guía de información custodiada en el Departamento de Archivo Histórico*. Recuperado de: http://www.archivonacional.go.cr/index.php?option=com_content&view=article&id=217:guia-de-fondos&catid=69:archivo-historico&Itemid=91.
- Fernández Bonilla, L. (1881-1907). *Colección de documentos para la historia de Costa Rica, 10 tomos*. San José / Barcelona: Imprenta Nacional / Imprenta Viuda de Luis Tasso.
- Fontanella de Weinberg, M. B. (Comp. y ed.). (1993). *Documentos para la historia lingüística de Hispanoamérica. Siglos xvi-xviii*. Tomo 1, Anejo LIII del Boletín de la Real Academia Española. Madrid: Real Academia Española.
- Guzmán Riverón, M. (Comp. y ed.). (2015). *Documentos para la historia lingüística de Hispanoamérica. Textos del Caribe. Siglos xv-xvii*. Tomo 5, Anejo LXII del Boletín de la Real Academia Española. Madrid: Real Academia Española.
- Hernández, E., Bravo-García, E., Fernández Alcaide, M., Fernández de Castro, F. y Sánchez Méndez, J. P. (2019). “Morfosintaxis histórica hispanoamericana: estado de la cuestión, fuentes, métodos y perspectivas”. En V. Codita y M. de la Torre (Eds.), *Tendencias y perspectivas en el estudio de la morfosintaxis histórica hispanoamericana* (pp. 185-222). Madrid / Frankfurt: Iberoamericana / Vervuert.
- Kabatek, J. (2013). “¿Es posible una lingüística histórica basada en un corpus representativo?”. *Iberoromania: Revista dedicada a las lenguas y literaturas iberorrománicas de Europa y América*, (77), pp. 8-28.
- Koch, P. y Oesterreicher, W. (1985). “Sprache der Nähe - Sprache der Distanz. Mündlichkeit und Schriftlichkeit im Spannungsfeld von Sprachtheorie und Sprachgeschichte”. *Romanistisches Jahrbuch*, 36, pp. 15-43.
- Kohnen, T. (2007). “From Helsinki through the centuries: the design and development of English diachronic corpora”. *Studies in Variation, Contacts and Change in English 2: Towards Multimedia in Corpus Studies*. Recuperado de: <http://www.helsinki.fi/varieng/series/volumes/02/kohnen/>.
- Kytö, M. (2010). “Data in historical pragmatics”. En A. H. Jucker y I. Taaitsainen (Eds.), *Handbook of Historical Pragmatics* (pp. 33-67). Berlin: Mouton de Gruyter.
- Lass, R. (1997). *Historical Linguistics and Language Change*. Cambridge: Cambridge University Press.
- León Fernández, M. (1988). *El seseo durante el siglo xvi en Costa Rica: Zona de Cartago* (Tesis de licenciatura). Universidad de Costa Rica, San José.

- Martinet, A. (1984). “De la synchronie dynamique à la diachronie”. *Diachronica*, 1, pp. 53-64.
- _____. (1990). “La synchronie dynamique”. *La Linguistique. Fasc. Linguistique et “facteurs externes”*, 26, pp. 13-23.
- Oesterreicher, W. (2005). “Textos entre inmediatez y distancia comunicativas. El problema de lo hablado en lo escrito en el Siglo de Oro”. En R. Cano (Coord.), *Historia de la lengua española* (pp. 729-769). 2ª ed. Barcelona: Ariel.
- Quesada Pacheco, M.A. (1987). *Fuentes documentales para el estudio del español colonial de Costa Rica*. San José: Alma Mater.
- _____. (1990). *El español colonial de Costa Rica*. San José: Editorial de la Universidad de Costa Rica.
- _____. (1995). *Diccionario histórico del español de Costa Rica*. San José: Editorial de la Universidad Estatal a Distancia.
- _____. (2005). *Formas de tratamiento en Costa Rica: Visión histórica*. Bergen: University of Bergen.
- _____. (2009). *Historia de la lengua española en Costa Rica*. San José: Editorial de la Universidad de Costa Rica.
- _____. (2013). “El sistema verbal del español de Costa Rica en los albores de la época independiente”. *Signo y Seña*, (23), pp. 81-102.
- Ramírez Luengo, J. L. (2004-2005). “Contribución a la historia del español de Honduras: Edición y estudio de documentos hondureños del siglo XVIII”. *Anuario de Letras. Lingüística y Filología*, 42-42, pp. 51-75.
- _____. (2006). “Materiales para la historia de la lengua española en Centroamérica: Algunos documentos dieciochescos (1703-1758)”. *Ámbitos*, (16), pp. 119-136.
- _____. (2011). “Un corpus para la historia del español en Nicaragua: edición de documentos oficiales del siglo XVIII (1704-1756)”. *Moenia*, (17), pp. 333-366.
- _____. (2017). *Textos para la historia del español, xi, Honduras y El Salvador*. Alcalá de Henares: Universidad de Alcalá.
- Rojas Mayer, E. (Comp. y ed.) (2000). *Documentos para la historia lingüística de Hispanoamérica. Siglos xvi-xviii*. Tomo 2. Madrid: Real Academia Española.
- Rojas Mayer, E. (Comp. y ed.) (2008a). *Documentos para la historia lingüística de Hispanoamérica. Siglos xvi-xviii*. Tomo 3. Madrid: Real Academia Española.

- Rojas Mayer, E. (Comp. y ed.) (2008b). *Documentos para la historia lingüística de Hispanoamérica. Siglos xvi-xviii*. Tomo 4. Madrid: Real Academia Española.
- Sánchez-Prieto Borja, P. (1998). *Cómo editar los textos medievales. Criterios para su presentación gráfica*. Madrid: Arco / Libros.
- _____. (2012). “Desarrollo y explotación del «Corpus de Documentos Españoles Anteriores a 1700» (CODEA)”. *Scriptum Digital*, (1), pp. 5-35.
- Toruella, J. y Llisterri, J. (1999). “Diseño de corpus textuales y orales”. En J. M. Blecua, G. Clavería, C. Sánchez y J. Torruella (Eds.), *Filología e informática. Nuevas tecnologías en los estudios filológicos* (pp. 45-77). Barcelona: Milenio. Recuperado de: http://liceu.uab.es/~joaquim/publicacions/Torruella_Llisterri_99.pdf.
- Ulate Zúñiga, R. (1991). *El seseo en el español escrito en el Noroeste de Costa Rica durante el siglo xvi. Zonas de Esparza y Guanacaste* (Tesis de licenciatura). Universidad de Costa Rica, San José.