

Diez años de *CORDIAM*
Reflexiones sobre la configuración de un (mejor) corpus para la historia del español de América

Ten years of *CORDIAM*
Some reflections on the configuration of a (better) corpus for the history of Latin American Spanish

José Luis Ramírez Luengo
Universidad Complutense de Madrid (Madrid, España)
joseluis.ramirezluengo@gmail.com
<https://orcid.org/0000-0002-5564-2372>

Recibido el 17/10/2023, aceptado el 14/2/2024, publicado el 18/10/2024

Creative Commons Attribution 4.0 International
© 2024 José Luis Ramírez Luengo

Cómo citar este artículo

Ramírez Luengo, José Luis 2024. Diez años de *CORDIAM*. Reflexiones sobre la configuración de un (mejor) corpus para la historia del español de América. *Studia linguistica romanica* 2024.12, 34-57. <https://doi.org/10.25364/19.2024.12.3>.

Resumen

Aunque son varios los corpus existentes hoy para el estudio de la historia del español de América (*CORDE*, *CHARTA*), no cabe duda de que *CORDIAM* se erige, desde hace ya diez años, en una referencia básica para esta área de investigación filológica. En este sentido, este trabajo pretende describir los méritos con los que hoy cuenta el corpus y señalar, al mismo tiempo, posibles líneas de actuación futuras que lo mejoren. Así, tras revisar la historia de la edición de textos americanos de carácter lingüístico y señalar el papel de *CORDIAM* en ella, se procederá a describir las principales características del corpus, para terminar señalando una serie de actuaciones que puedan solucionar algunas de sus carencias, todo ello con el propósito de calibrar sus aportes y su relevancia diez años después de su puesta en marcha.

Abstract

Although several corpora exist for studying the history of Latin American Spanish (*CORDE*, *CHARTA*), there is no doubt that, since 2013, *CORDIAM* has become a fundamental reference in this philological area. This paper aims to describe the advantages of this corpus and to outline future actions that could further enhance it. Accordingly, the history of the linguistic editing of Latin American texts, the role of *CORDIAM* in this history and the main features of the corpus will be described. Finally, I will discuss potential actions to address its shortcomings, with the aim of evaluating its contribution and relevance ten years after its implementation.

Índice

1 Introducción.....	36
2 El <i>CORDIAM</i> en su contexto: las fuentes para la historia del español americano y su valoración a través del tiempo.....	37
3 El <i>CORDIAM</i> de hoy y el <i>CORDIAM</i> de mañana, I: unas reflexiones desde lo cuantitativo.....	40
4 El <i>CORDIAM</i> de hoy y el <i>CORDIAM</i> de mañana, II: algunas cuestiones más allá de lo numérico.....	48
5 Unas conclusiones provisionales: las carencias actuales de un corpus como futuro horizonte de trabajo.....	51
Abreviaturas y referencias bibliográficas.....	54

1 Introducción

[1] En un contexto como el que caracteriza en las últimas décadas a los estudios del español – marcado por un creciente interés por la dialectología histórica y, a resultas de ello, por la diacronía de las variedades americanas de esta lengua –, es inevitable que cada vez se preste más atención a las fuentes documentales que sustentan estos análisis, pues, como bien recordaba Company Company (2001: 208) hace ya muchos años, «sólo cuando tengamos la documentación filológica adecuada, podremos conocer la idiosincrasia dialectal del español que arribó a las distintas zonas del continente americano, así como su posterior evolución». Este interés creciente por las fuentes da como resultado dos tipos de actuaciones independientes pero mutuamente relacionadas: por un lado, la aparición de nuevos corpus lingüísticos de muy distinta naturaleza y alcance (entre otros, *CORDE*, *CODEA*, *CorLexIn* o *Léxico hispanoamericano*)¹; por otro, el desarrollo de reflexiones sobre cómo configurar un repositorio de este tipo y qué características debe presentar para cumplir con los necesarios requisitos de representatividad, variación y rigurosidad filológica que exige la tarea de reconstruir de la forma más precisa y completa posible la situación que presenta en el pasado cualquier variedad diatópica del español (Isasi Martínez 2012; Ramírez Luengo 2012a; Kabatek 2013, 2016), en el convencimiento de que solo así, a partir del conocimiento del devenir diacrónico de las diversas hablas que componen este idioma, es posible comprender la historia del español en su conjunto y, con ello, los resultados que actualmente se descubren en el mundo hispánico.

[2] Tomando como eje, por tanto, las dos cuestiones que se acaban de mencionar, la pretensión de este trabajo es describir la situación actual y las perspectivas de futuro que ofrece el que probablemente constituye a día de hoy el repositorio fundamental para todos aquellos investigadores que se interesan por la historia del español americano, el *Corpus diatópico y diacrónico del español de América (CORDIAM)*, cuya presentación como proyecto tuvo lugar en la Universidad de Passau en 2011, en el XI Congreso de la Asociación Alemana de Hispanistas, y de cuya puesta en marcha se cumple ahora una década. De este modo, estas páginas se organizan de la siguiente manera: tras esta breve introducción general (§ 1), se comenzará por situar, desde un punto de vista historiográfico, la aparición de *CORDIAM* dentro del proceso de creación de corpus lingüísticos americanos (§ 2); después – y a partir de lo expuesto en Bertolotti & Company Company (2018) –, se pasará a analizar críticamente el estado que muestra hoy *CORDIAM*, describiendo cuantitativa y cualitativamente sus fondos y señalando una serie de cuestiones que, desde el punto de vista filológico, o bien constituyen claras carencias en este momento o bien son susceptibles de mejora (§ 3 y 4); finalmente, el traba-

¹ Para una revisión de estos y otros proyectos, de ámbito americano o no, véase Corbella Díaz, Fajardo Aguirre & Langenbacher-Liebott (2018); en cuanto a *CODEA*, sus objetivos y lineamientos generales se describen en Sánchez-Prieto Borja (2012).

jo se cierra con unas reflexiones que, retomando lo dicho anteriormente, señalen posibles líneas de trabajo que habrá que desarrollar en el futuro con vistas a la consecución de un corpus más representativo, más variado y más riguroso filológicamente que permita, en consecuencia, un mejor conocimiento de la diacronía del español americano (§ 5).

2 El *CORDIAM* en su contexto: las fuentes para la historia del español americano y su valoración a través del tiempo

[3] Como suele acontecer siempre en la ciencia, la aparición de una nueva infraestructura de investigación no constituye un hecho casual y fortuito, sino que tiene su origen en los trabajos desarrollados previamente y guarda relación con los cambios que, con el paso del tiempo, se imponen en la forma de comprender el objeto de estudio, en este caso la historia del español de América; así, parece oportuno comenzar el análisis de *CORDIAM* revisando desde un punto de vista historiográfico las fuentes que tradicionalmente se han utilizado para investigar el devenir diacrónico de las variedades de esta lengua en el Nuevo Mundo y que, por tanto, se pueden entender como antecedentes de este corpus. En este sentido, se indicaba ya en Ramírez Luengo (2009: 174-176) que es posible establecer una serie de etapas – consecutivas en cuanto a sus inicios, si bien muchas de ellas se solapan e incluso conviven hasta el momento actual – que se caracterizan no tanto por la documentación editada², sino más bien por la forma como se enfrentan a la cuestión de la propia edición de textos y por la importancia que se le concede a esta tarea.

[4] De este modo, el primer periodo – que abarca desde los inicios de la investigación hasta los años setenta del siglo XX y cuenta con obras de autores tan relevantes como Henríquez Ureña, Guitarte, Rosenblat, o incluso el mismo Cuervo³ – se caracteriza por un evidente desinterés respecto a la preparación de corpus de estudio y, a raíz de ello, por el aprovechamiento preferente de dos tipos de fuentes lingüísticas: la literatura colonial hispanoamericana y los textos editados por los historiadores. Se trata, sin embargo, de documentación que presenta serios problemas en cuanto a su fiabilidad y rigor filológico, pues si «la literatura colonial siguió hasta bien entrado el siglo XIX modelos peninsulares, opacando así, o también estereotipando, las peculiaridades dialectales americanas», las transcripciones preparadas para el estudio de la historia presentan un «carácter muy oficial» y se centran «en el qué se dice y descuidan el cómo se dice, pasando por alto los detalles formales imprescindibles para realizar la diacronía de cualquier lengua» (Company Company 2001: 208), de manera que su empleo en los estudios sobre la historia del español americano solo se entiende – además de por la facili-

2 Pues se debe tener en cuenta que muchas veces los tipos documentales que conforman los corpus se repiten.

3 Para un estudio de este periodo y de sus aportaciones a la cuestión, véase el profundo y minucioso trabajo de Medina López (2022: 84-89).

dad que supone acceder a ella – por la escasa trascendencia que se concede en esta época a la edición, considerada poco más que como tarea auxiliar o preparatoria para el trabajo filológico propiamente dicho⁴.

[5] Ahora bien, esta situación experimenta un importante cambio a partir de los años setenta, cuando el rigor que se exige en estos momentos en la investigación filológica determina una radical transformación en los materiales que se van a utilizar para investigar la diacronía de las variedades hispánicas del Nuevo Mundo. Este cambio de tendencia, inaugurado y ejemplificado magistralmente por el estudio de Cock Hincapié (1969) sobre las sibilantes en el Nuevo Reino de Granada⁵, se caracteriza según Medina López (1995: 29) por factores como el análisis de la evolución de un rasgo específico, la limitación de un periodo determinado, la utilización de una metodología rigurosa y – en lo que se refiere a las fuentes – el empleo de documentación transcrita de la época, o más concretamente, documentación transcrita con el propósito de servir de base para estudios de tipo lingüístico. Con todo, este nuevo modelo – que supone un claro avance respecto a la etapa anterior y cuenta con trabajos de la importancia de Álvarez Nazario (1982), Fontanella de Weinberg (1987) o Quesada Pacheco (1990) – presenta una serie de deficiencias en relación con el corpus de estudio, entre las que destacan que no se facilite la lista de textos analizados, que no se expliciten los criterios de edición con que estos se han transcrito y muy especialmente que no se ponga a disposición del estudioso la documentación base del trabajo, algo que impide desarrollar otros estudios sobre los mismos textos o incluso aplicar nuevas perspectivas de análisis a las temáticas ya tratadas con anterioridad.

[6] Como forma de superar las limitaciones anteriores, a partir de la década de los ochenta se puede constatar el comienzo de otra etapa que se caracteriza por la aparición de investigaciones – como, entre otras, Lope Blanch (1985), García Carrillo (1988) y Parodi (1995) para México, Elizaincín, Malcuori & Bertolotti (1997) para la banda Oriental o Choy López (1999) para el caso de Cuba – en las que el análisis lingüístico se acompaña de la edición del corpus utilizado, lo que sirve para solucionar las deficiencias que se han planteado en el párrafo anterior. Con todo, una revisión de tales obras demuestra cómo en ellas la edición de los documentos se sigue considerando en general un aspecto «subsidiario y subordinado al estudio lingüístico en sí», esto es, «apenas como un paso previo necesario para poder desarrollar el análisis del estado de lengua que interesa» (Ramírez

4 Junto a esta primera razón, es probable que la propia consideración de la historia de la lengua en la época no como un fin en sí mismo, sino más bien como una ayuda para el mejor análisis y la más profunda comprensión de la literatura y la cultura de un pueblo, contribuya también a este estado de cosas.

5 Cabe indicar que, como bien señala Medina López (2022: 90), el estudio de Cock Hincapié (1969) tiene un claro antecedente en el trabajo de Amado Alonso, quien ya en 1951 utiliza la documentación de archivo para sus investigaciones sobre la historia del seseo y defiende el interés de estos materiales para la historia del español de América.

Luengo 2009: 175), lo que determina que aquellos aspectos más estrictamente relacionados con la manipulación y transmisión del documento – por ejemplo, los criterios empleados a la hora de transcribir los textos – queden en un segundo plano.

[7] En esta línea, no cabe duda de que el momento de cambio fundamental en todo este proceso lo constituye la cuarta etapa, cuyo inicio se puede localizar en la primera parte de la década siguiente: en efecto, es a partir de los años noventa – y sobre todo en los inicios del nuevo siglo – cuando la edición de documentos americanos cobra una entidad propia, reflejada en la aparición de antologías de textos históricos que pretenden dotar a los estudiosos de fuentes para el análisis⁶ y que se caracterizan, en palabras de Medina López (1995: 42-43), por conformarse por documentos de archivo de tipología muy variada – reflejo del *giro archivístico* que, según Bertolotti & Company Company (2018: 102), identifica a esta época en los estudios de lingüística histórica –, por centrarse desde el punto de vista cronológico en la época colonial y por emplear unos criterios de transcripción notablemente conservadores. De este modo, si hasta ahora la preparación de corpus se había entendido como una cuestión secundaria, trabajos como Fontanella de Weinberg (1993), Company Company (1994), Rojas Mayer (2000), Obediente Sosa (2002), Rivarola (2009), Bertolotti, Coll & Polakof (2012) o Ramírez Luengo (2017) sitúan esta cuestión en el centro mismo de la labor investigadora, al centrarse en la creación de infraestructura de investigación que se caracterice por la rigurosidad en la transcripción y, por tanto, por su fiabilidad para la investigación de cualquier aspecto de la diacronía del español americano.

[8] Por último, el muy notable desarrollo que experimenta la lingüística del corpus durante estas primeras décadas del siglo XXI y la incorporación de las herramientas informáticas al quehacer filológico permite establecer una quinta etapa que – si bien continúa los presupuestos de la anterior en lo que se refiere a la importancia concedida a la edición de textos y al carácter autónomo de esta tarea – supone una transformación tanto en lo que se refiere al tamaño de los corpus, conformados ahora por muchos millones de palabras, como a sus posibilidades de explotación, pues la naturaleza digital de estos y la existencia de motores de búsqueda permiten – además de una saludable democratización del conocimiento – llevar a cabo análisis más robustos y más fiables de una forma mucho más rápida y cómoda de lo que podría imaginarse hace apenas algunas décadas. Este nuevo estado de cosas se refleja en los grandes repositorios digitales como *CORDE*, *CHARTA* o naturalmente *CORDIAM*, que aúna los enfoques más aceptados en estos momentos sobre lo que supone editar un texto – esto es, su consideración como labor científica por sí misma – con todas las posibilidades que ofrece la

6 En esta línea, es fundamental la tarea desarrollada por la Comisión de estudio histórico del español de América de la Asociación de lingüística y filología de América Latina (ALFAL), que da como resultado la aparición de varios volúmenes de textos que ofrecen al investigador documentación filológicamente fiable de la mayor parte de Hispanoamérica (Medina López 2022: 94-95).

tecnología actual, y que precisamente por eso, por recoger y aprovechar todos los avances que se han producido en este ámbito, se ha convertido en el corpus de referencia para la obtención de datos sobre el devenir del español de América.

[9] Así pues, la revisión historiográfica que se acaba de llevar a cabo demuestra que a lo largo del tiempo se ha producido un paulatino aumento de la importancia concedida a las fuentes empleadas para el estudio histórico de la lengua española en el Nuevo Mundo, algo que se refleja en el conjunto de transformaciones que se señala a continuación: a) preparación de estudios (con fuentes subsidiarias) > preparación de fuentes; b) menor presencia de la edición > mayor presencia de la edición; c) menor atención a la labor de edición > mayor atención a la labor de edición; d) tipologías más restringidas (literatura) > tipologías más amplias (literatura, documentación de archivo, prensa); e) cronología más acotada (época colonial) > cronología más amplia (siglos XVI-XIX); f) corpus más pequeños > corpus más amplios > *macrocorpus* informatizados. Se descubre, así, una dinámica de trabajo que deja de considerar la preparación de documentos lingüísticos como una tarea subsidiaria o auxiliar para verla como una labor científica, eminentemente filológica e indispensable para llevar a cabo los trabajos posteriores, lo que permite concluir, en definitiva, que con el paso del tiempo la edición de textos adquiere valor por sí misma y termina por transformarse en una disciplina autónoma y fundamental para quienes se interesan por la historia de la lengua, tal y como de hecho demuestra el trabajo conjunto, colaborativo y de largo aliento que representa precisamente el corpus *CORDIAM*.

3 El *CORDIAM* de hoy y el *CORDIAM* de mañana, I: unas reflexiones desde lo cuantitativo

[10] Retomando las palabras con que se cierra al apartado anterior, no cabe duda de que uno de los rasgos que mejor define a este corpus – junto a su naturaleza estrictamente americana y su amplitud cronológica – lo constituye su carácter colaborativo, es decir, el aprovechamiento para su configuración de la labor desarrollada previamente por múltiples investigadores, que de manera generosa han aportado sus antologías para configurar el punto de encuentro común y compartido que pretende ser *CORDIAM*⁷.

[11] Como se puede advertir fácilmente, esta decisión⁸ posee unas claras ventajas, entre las que es posible mencionar, por ejemplo, la rapidez con la que se

7 Para un listado completo de los investigadores participantes hasta el momento, véanse las páginas *Los corpus/Documentos/Colaboradores*, *Los corpus/Literatura/Colaboradores* y *Los corpus/Prensa/Colaboradores* de *CORDIAM*.

8 Que, cabe decir, en modo alguno afecta a la rigurosidad filológica del repositorio, pues solo se incorporan «materiales extraídos directamente de archivo, paleografiados directamente por especialistas en lengua y editados con criterios ecdóticos explícitos» (Bertolotti & Company Company 2018: 103); restricciones semejantes – si bien adaptadas a sus propias especificidades – se aplican también a los subcorpus de prensa y literatura.

ha conformado un corpus cuantitativamente muy sólido, el ahorro de esfuerzos y recursos que conlleva su creación a partir de materiales previamente editados o la accesibilidad que supone contar en un solo lugar con documentos lingüísticos aún inéditos y con antologías de difícil acceso, habida cuenta de la escasa circulación que en ocasiones presentan sus versiones en papel (Bertolotti & Company Company 2014: 133). Ahora bien, no se oculta a nadie que estas ventajas se acompañan también de ciertos inconvenientes, entre los que se pueden citar fundamentalmente dos: por un lado, el empleo de diferentes criterios de edición – ya que «los corpus fueron subidos a *CORDIAM* tal como fueron autorizados y enviados por los colaboradores» (*CORDIAM*) –, lo que en ocasiones produce ciertas dificultades a la hora de comparar los datos⁹; por otro, la existencia de una muy marcada disparidad cuantitativa desde el punto de vista diatópico y diacrónico en los fondos del repositorio, algo que no carece de interés historiográfico – por cuanto refleja la relevancia que, dentro de su tradición filológica, los distintos países americanos han concedido a los estudios históricos – pero que termina por configurar un corpus descompensado que impide obtener una visión precisa de los distintos procesos evolutivos que tienen lugar en las diferentes variedades que conforman el denominado *español de/en América*.

[12] De este modo, y a la luz de este último aserto, quizá no esté de más revisar en primer lugar cuáles son los fondos que atesora actualmente *CORDIAM*, pues la constatación de las zonas y los momentos que están más escasamente representados sin duda permitirá esbozar líneas de trabajo futuro que estratégicamente vayan completando tales espacios vacíos y contribuyan, por tanto, a construir un corpus mucho más homogéneo y representativo y, en consecuencia, mucho más útil para la mejor y más efectiva reconstrucción del desarrollo del español en el continente. A este respecto, y desde un punto de vista cronológico, la situación que presenta *CORDIAM* en estos momentos¹⁰ es la siguiente (figura 1):

9 Esto se hace especialmente claro, por ejemplo, cuando se pretende emplear *CORDIAM* para analizar la historia de las tendencias gráficas o de la puntuación (Ramírez Luengo 2012b, 2015), pues tales análisis exigen un escrupuloso respeto de la grafía original que en ocasiones no se ha producido; con todo, el problema queda parcialmente solucionado gracias a la remisión a los criterios editoriales específicos de cada una de las antologías consignadas, así como por la aplicación de ciertas normas generales de edición y sistematización – especificadas, para cada uno de los subcorpus, en la propia página web de *CORDIAM*, en concreto en la pestaña *Los corpus* – que aseguran un alto grado de coherencia interna.

10 En concreto, el 12 de noviembre de 2022, cuando se extraen y se organizan todos los datos que se expondrán en estas páginas.

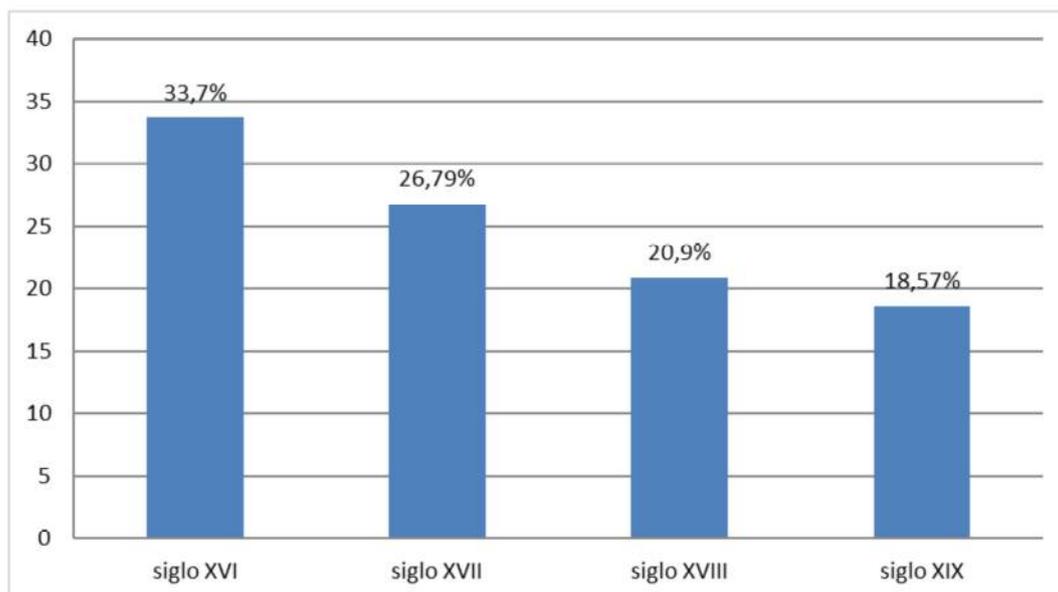


Figura 1: Distribución cronológica de los materiales de *CORDIAM*

[13] Como se puede apreciar, los materiales recogidos en *CORDIAM* se van reduciendo de manera progresiva conforme se avanza cronológicamente, de manera que si los siglos XVI y XVII equivalen a un 33,70 % y un 26,79 % del total del corpus respectivamente – es decir, unos porcentajes superiores al 25 % que supondría, en principio, una distribución equilibrada –, en el caso de la centuria ilustrada tal porcentaje desciende al 21 %, en una tendencia que se hace especialmente evidente para el Ochocientos, el periodo cronológico sin duda peor representado (18,5 %). De hecho, una revisión por segmentos temporales de cincuenta años pone de manifiesto cómo, en realidad, son dos los momentos específicos que muestran una alarmante falta de representación en el corpus – en concreto, la primera mitad del siglo XVIII y la segunda mitad del siglo XIX, con porcentajes del 6,32 % y del 5,63 % respectivamente –, mientras que el corte cronológico 1551-1600 aparece claramente sobrerrepresentado, al concentrar por sí mismo más del 20 % de todas las palabras de *CORDIAM*¹¹.

[14] Por supuesto, esta situación que se acaba de describir no es en modo alguno casual, y refleja el escaso interés que, hasta hace poco tiempo, han despertado en la investigación sobre historia del español los momentos más cercanos a la sincronía actual – muy especialmente el siglo XIX (Buzek & Šincová 2015: 7-8)

¹¹ En concreto, los datos específicos – tanto el porcentaje dentro de *CORDIAM* como el número de palabras y la cantidad de textos – son los que se ofrecen a continuación: 1501-1550: 11,24 % (1586693; 557); 1551-1600: 22,46 % (3170268; 3051); 1601-1650: 13,06 % (1844149; 1485); 1651-1700: 13,73 % (1938967; 1636); 1701-1750: 6,32 % (893300; 1361); 1751-1800: 14,58 % (2058321; 2901); 1801-1850: 12,94 % (1826670; 4345); 1851-1900: 5,63 % (795575; 2329).

–, pero lo cierto es que supone un problema de notable gravedad si se tiene en cuenta la indudable relevancia que, para la historia de las variedades americanas de la lengua, presentan tanto el Setecientos como el Siglo de las Independencias. Aunque no es este el momento de profundizar en esta cuestión, no está de más recordar que, a la luz de determinados estudios sobre regiones específicas como el Altiplano mexicano, el siglo XVIII parece constituir el momento en el que, en el campo de la morfosintaxis, «tomó carta de naturaleza, esto es, se volvió parte del habla cotidiana del pueblo, un buen número de formas de expresión que constituyen caracterizadores dialectales del español de México hoy en día» (Company Company 2007: 18), algo que es también aplicable desde el punto de vista fónico y/o léxico, por ejemplo, a diversas áreas centroamericanas como El Salvador, Honduras o Guatemala (Ramírez Luengo 2019, 2021a, 2022); en cuanto al Ocho-cientos, algunos fenómenos propios de esta centuria como son el fraccionamiento de los virreinos coloniales en diferentes repúblicas independientes y, a resultas de ello, la normativización de sus usos dialectales como estándar culto regional, así como la masiva hispanización de la población autóctona que tiene lugar en estos momentos y la consecuente aparición y/o generalización de nuevas variedades de *español indigenizado* (Ramírez Luengo 2011: 15-17) demuestran la importancia fundamental que tiene esta época para la mejor comprensión del devenir diacrónico del español americano.

[15] Pasando ahora a la distribución geográfica de sus fondos, se puede sostener que desde este punto de vista *CORDIAM* evidencia también una situación de desproporción paralela a la que se ha señalado en la cronología, con unos países bien representados y otros que, por el contrario, tienen una presencia marcadamente más restringida, lo que de nuevo supone un problema si se pretende esbozar una imagen realista – o al menos lo menos sesgada posible – de la situación dialectal de Hispanoamérica en su historia. A este respecto, los datos que ofrece el corpus a día de hoy son los siguientes (tabla 1):

País	Palabras	Textos	Porcentaje <i>CORDIAM</i>
Argentina	366282	799	2,59 %
Bolivia	423028	259	2,99 %
Chile	1006634	913	7,13 %
Colombia	796736	975	5,64 %
Costa Rica	50579	67	0,35 %
Cuba	93136	80	0,65 %
Ecuador	149259	75	0,80 %
El Salvador	18605	34	0,13 %
Estados Unidos	342331	359	2,42 %
Guatemala	113543	139	0,80 %
Honduras	30664	41	0,21 %
México	4408816	5633	31,23 %
Nicaragua	55299	108	0,39 %
Panamá	39201	59	0,27 %
Paraguay	15002	22	0,10 %
Perú	3465641	2759	24,25 %
Puerto Rico	56152	33	0,39 %
Rep. Dominicana	697278	387	4,94 %
Uruguay	907972	2528	6,43 %
Venezuela	1077785	2359	7,63 %
Total	14113943	17665	100,00 %

Tabla 1: Distribución de los materiales de *CORDIAM* por zona geográfica

[16] De este modo, una primera revisión de los porcentajes que se facilitan en la tabla anterior demuestran que, si cronológicamente el corpus presenta diferencias de cierta relevancia, estas resultan mucho más marcadas desde el punto de vista geográfico, con zonas que ofrecen una sólida base documental y otras que, por el contrario, aparecen poco o mínimamente representadas; de hecho, esta dispar situación que muestra *CORDIAM* en lo que se refiere a la procedencia geográfica de sus fondos permite clasificar los países americanos en cuatro grandes grupos, tal y como refleja la siguiente tabla (tabla 2)¹²:

12 Cabe decir que los porcentajes utilizados para el establecimiento de los cuatro grupos propuestos no son casuales, sino que parten de la idea de que, para estar representadas de igual manera en el corpus las veinte naciones americanas, cada una de ellas debería aportar un 5 % de sus fondos; aunque este se puede considerar un criterio cuanto menos discutible – habida cuenta del diferente peso demográfico que hoy, y también en el pasado, muestran los diversos países americanos –, parece una forma sencilla y objetiva de evidenciar la dispar distribución de los fondos que atesora el corpus.

Situación en <i>CORDIAM</i>	País
Áreas sobrerrepresentadas (> 8 %)	México, Perú
Áreas representadas (7,99 % - 3 %)	Venezuela, Chile, Uruguay, Colombia, Rep. Dominicana
Áreas infrarrepresentadas (2,99 % - 1 %)	Bolivia, Argentina, Estados Unidos
Áreas muy infrarrepresentadas (< 1 %)	Ecuador, Guatemala, Cuba, Puerto Rico, Nicaragua, Costa Rica, Panamá, Honduras, El Salvador, Paraguay

Tabla 2: Situación de las distintas áreas geográficas en *CORDIAM*

[17] Así, salta a la vista que, mientras que zonas como Venezuela, Chile, Uruguay, Colombia y República Dominicana tienen en *CORDIAM* una presencia que – en principio – se puede calificar como adecuada¹³, México y Perú resultan claramente sobrerrepresentados, al concentrar, con casi ocho millones de palabras, más de la mitad de la documentación del corpus; frente a estas naciones, el resto de Hispanoamérica se muestra por el momento infrarrepresentado, con un conjunto de países (Ecuador, casi todas las Antillas, Centroamérica en su conjunto y Paraguay) en una situación especialmente grave, habida cuenta de que su presencia en *CORDIAM* no llega ni siquiera al 1 % de sus fondos. Por supuesto, no es difícil percibir los serios problemas que esta dispar distribución geográfica conlleva, desde el punto de vista de la dialectología histórica: por un lado, es evidente que la situación actual del corpus no permite describir con una base empírica suficiente el devenir diacrónico del español de zonas como por ejemplo Ecuador, Guatemala, Costa Rica o Paraguay; por otro, la escasa representación que algunas variedades tienen en el corpus frente al claro predominio estadístico de otras lleva a preguntarse si las generalizaciones que se extraen de sus datos se pueden considerar reflejo del español de América o más bien de *algunas* variedades de este diasistema, con lo que esto supone de riesgo de falseamiento de la evolución histórica, al generalizar como americano lo que en realidad es propio únicamente de alguna zona concreta.

[18] En esta misma línea, es importante señalar que los problemas que tienen que ver con la representatividad dialectal de *CORDIAM* no terminan aquí, sino que se ven acrecentados en muchos casos por la concentración geográfica que se descubre en el caso de algunos países muy fragmentados diatópicamente:

13 De hecho, es probable que Uruguay deba incorporarse a las áreas sobrerrepresentadas, dado que la peculiar historia del país – que, como es sabido, se incorpora al mundo hispánico colonial a partir prácticamente de la fundación de Montevideo en las primeras décadas del siglo XVIII – obliga a analizar su presencia en el corpus teniendo en cuenta únicamente las dos últimas centurias del mismo; en este sentido, los datos demuestran que los fondos uruguayos constituyen el 16,28 % de los materiales dieciochescos y decimonónicos del corpus – ascienden, incluso, al 52,68 % en el caso del Ochocientos –, es decir, se encuentran muy por encima de los porcentajes que corresponden a las áreas representadas.

en el caso de México, por ejemplo, mientras que el Altiplano Central, la Península de Yucatán, las costas veracruzanas, Oaxaca o Chiapas aparecen mejor o peor representados en el corpus, otras zonas dialectales señaladas por Lope Blanch (1996: 88) – como la costa de Guerrero, Michoacán y especialmente el norte – ofrecen una muy pobre o incluso nula documentación; una situación paralela se detecta también en Colombia – cuyos fondos reflejan lo que Montes Giraldo (1996: 142) denomina las variedades costeña atlántica (Cartagena, Santa Marta), andino oriental (Bogotá, Popayán) y en menor medida andino occidental (Antioquia), pero que prácticamente no cuentan con textos procedentes de zonas con una identidad lingüística tan marcada como son Nariño, Santander o el Chocó¹⁴ –, así como en Argentina, donde la relativamente buena representación de áreas dialectales como la litoral-bonaerense o el noroeste contrasta con la ausencia prácticamente total de otras como la región guaraníca o Cuyo (Fontanella de Weinberg 2004), o en el caso ecuatoriano, cuya documentación presente en *CORDIAM* procede en un 90 % de la zona serrana, con apenas ocho documentos que representan el español costeño del país¹⁵.

[19] Teniendo en cuenta, por tanto, todo lo expuesto hasta el momento, puede resultar interesante combinar ahora los datos cronológicos y geográficos ya presentados como forma de cartografiar la situación que ofrece en estos momentos el corpus y de señalar de forma más precisa así las carencias que, desde ambos puntos de vista, es necesario solucionar con cierta urgencia. Esto es, precisamente, lo que se registra en la tabla que se presenta a continuación (tabla 3):

14 En este último caso, naturalmente, por la dificultad que supone encontrar documentación de la región, habida cuenta de su particular historia y las dificultades que el mismo clima supone para la conservación de los documentos.

15 En concreto, se trata de seis textos de Guayaquil, uno de Babahoyo, capital de Los Ríos, y otro de la provincia de Esmeraldas; los 63 documentos restantes, todos ellos serranos (o de zonas amazónicas del piedemonte andino), se redactan en Chimbo, Cuenca, Guaranda, Latacunga, Lita, Macas, Riobamba, Zamora, Zaruma y muy especialmente Quito y alrededores, en concreto las parroquias de Cumbayá y Chillogallo.

País	1501-1550	1551-1600	1601-1650	1651-1700	1701-1750	1751-1800	1801-1850	1851-1900	Total
------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-----------	-------

AR	-	37962	84247	43400	26444	63884	73078	37267	366282
BO	2880	259695	35554	22237	24098	75476	3088	-	423028
CH	1892	109128	99148	425156	102929	217723	50658	-	1006634
CO	49575	111414	218025	43297	98183	211570	56547	8125	796736
CR	-	10947	5851	7393	13718	10790	1880	-	50579
CU	20765	50755	12507	1827	4373	902	-	2007	93136
EC	2543	36886	37002	13354	22922	13530	23022	-	149259
ES	-	1287	4403	4670	2784	5461	-	-	18605
EU	7129	47378	80470	72647	89724	44547	-	436	342331
GU	2845	6760	5601	22001	24066	48572	3698	-	113543
HO	-	4877	5624	8181	2289	8962	731	-	30664
MX	266987	1488029	262101	1126358	379485	796191	68360	21305	4408816
NI	-	2787	-	5468	13999	20278	12767	-	55299
PA	4070	27093	-	-	2289	5749	-	-	39201
PE	1196388	248283	950749	130318	47786	383384	247183	261550	3465641
PR	2729	17252	1895	581	-	1250	-	32445	56152
PY	-	860	-	1764	1600	10778	-	-	15002
RD	11912	675086	6563	2733	-	-	-	984	697278
UY	-	-	-	-	22956	97388	368470	419158	907972
VE	16978	33789	34409	7582	13655	41886	917188	12298	1077785
Total	1586693	3170268	1844149	1938967	893300	2058321	1826670	795575	14113943

Tabla 3: Distribución de los materiales de *CORDIAM* por periodo y zona geográfica

[20] De este modo, la tabla que se acaba de ofrecer permite detectar tres claros problemas de los que adolece en este momento *CORDIAM* y que será necesario solucionar en un futuro más o menos cercano. El primero lo constituye sin duda la existencia de *casillas vacías*, esto es, de periodos que por ahora no cuentan con ningún documento en el repositorio (entre otros, Argentina, 1501-1550; Paraguay, 1601-1650; Panamá, 1651-1700; Puerto Rico, 1701-1750; o El Salvador, 1801-1850), hecho especialmente grave en el caso del siglo XIX pero también en lo que se refiere a los primeros años de la colonización, ya que solo a partir de un profundo conocimiento del español que llega al Nuevo Mundo en estos momentos será posible entender los procesos de koineización que van a establecer las bases de las distintas variedades americanas de la lengua. Junto a esto, es también destacable el mal reparto cronológico que presenta la documentación de algunos países, cuya concentración en momentos concretos impide observar los cambios diacrónicos que experimenta tal variedad lingüística; a este respecto, quizá sea especialmente claro el caso de República Dominicana, que si bien pertenece cuantitativamente a las áreas representadas en el corpus, concentra el 96 % de su

documentación en la segunda mitad del siglo XVI. Por último, también supone un problema – mencionado anteriormente de forma general, pero más evidente aún a la luz de estos datos – la dispar distribución geográfica en periodos cronológicos concretos, algo que se evidencia en el caso de la segunda mitad del siglo XIX – cuando el 90 % del texto procede de tres únicos países: Argentina, Perú y Uruguay – y que conlleva, como se dijo ya, el riesgo de generalizar como propia del continente americano alguna cuestión en realidad circunscrita a tales variedades diatópicas.

[21] Ante tal situación, no cabe duda de que una de las tareas más urgentes que tiene por delante *CORDIAM* es la de ir ampliando la representación de las regiones y los momentos todavía infrarrepresentados en el corpus, lo que obliga a desarrollar nuevos proyectos que, bien con la ayuda de nuevos investigadores o bien echando mano del espíritu colaborativo que conforma e identifica a sus participantes, vayan completando los vacíos que se acaban de señalar y configuren, por tanto, ese corpus homogéneo, representativo y comparable con el que es necesario contar para seguir escribiendo la historia del español americano con el rigor y la seriedad que esta tarea exige.

4 El *CORDIAM* de hoy y el *CORDIAM* de mañana, II: algunas cuestiones más allá de lo numérico

[22] Ahora bien, es importante mencionar que no todos los proyectos de expansión futura del repositorio pasan por – o se quedan en – lo estrictamente cronológico y geográfico; muy al contrario, parece necesario reflexionar también sobre otras cuestiones que, como complemento de lo anterior, es preciso tener en cuenta a la hora de incorporar nueva documentación que sirva para construir un corpus que cumpla, como se ha dicho ya anteriormente, con la función de representar de un modo más fidedigno la(s) forma(s) de hablar español que se dan en América a lo largo de su historia.

[23] A este respecto, conviene señalar que la falta de representatividad diatópica o temporal que se viene indicando se ve acompañada también por la de carácter social, de manera que quizá una de las primeras cuestiones que haya que mencionar en este punto sea la necesidad de incorporar documentación generada por los grupos de la sociedad colonial que aparecen menos representados en *CORDIAM*, tales como las mujeres¹⁶, los indígenas y la población afrodescendiente, pues solo de esta manera será posible avanzar hacia lo que a partir de ahora debe ser un objetivo fundamental en los estudios sobre la historia del español de América: la incorporación a ellos de la perspectiva sociolingüística, por medio de la cual sea posible describir de manera más precisa la variación intrínseca a cualquier es-

16 Esta necesidad de incorporar la documentación femenina a los estudios de historia del español – así como de atender a sus propias especificidades – ha sido apuntada en múltiples ocasiones por Belén Almeida Cabrejas; a este respecto, consúltense las atinadas reflexiones que esta autora expone en su trabajo más reciente (Almeida Cabrejas 2023).

tado de lengua. Por supuesto, es evidente que las restricciones sociales que, en lo que se refiere a la escritura, existen durante la época colonial y el alto grado de analfabetismo que afecta a los grupos mencionados (Newland 1991) conlleva que por norma general no sea sencillo conseguir textos generados por ellos y que, por lo tanto, en cualquier corpus predomine necesariamente la documentación producida por hombres criollos de ascendencia hispánica; sin embargo, esta constatación no es óbice para que no se intente, en la medida de lo posible, ir incorporando escritos de los grupos mencionados a partir del expurgo y de la consulta sistemática de repositorios que resultan propensos a resguardarlos, tales como los archivos conventuales y familiares para el caso de las mujeres o los fondos pertenecientes a las cofradías religiosas para los indígenas y – con mayor dificultad – la población de origen africano.

[24] En esta misma línea – y partiendo del ejemplo que suponen los grupos indígenas –, quizá se pueda discutir también la posibilidad de incorporar a *CORDIAM* documentación de hablantes bilingües que cuentan con un idioma europeo como lengua materna, muy especialmente en contextos concretos como son las situaciones de aluvión migratorio – como el Río de la Plata decimonónico o Texas y California durante este mismo periodo, con una presencia masiva de italo-parlantes y hablantes anglosajones respectivamente – o los casos de contacto estable y continuado en el tiempo que se dan entre el portugués y el español en zonas como la franja norte del Uruguay, las fronteras argentinas y paraguayas con Brasil o el departamento boliviano de Pando (Ramírez Luengo 2010: 30-38)¹⁷. Cabe alegar al respecto, es verdad, que los usos de tales hablantes no forma parte del español de América como tal – habida cuenta de que constituyen más bien los resultados idiolectales de un proceso de adquisición –, pero lo cierto es que existen al menos tres argumentos que justifican su incorporación a los corpus: en primer lugar, porque su no inclusión supone una evidente distorsión de la realidad lingüística de las zonas ya mencionadas, al ignorar en la práctica el español en el que se expresa una parte muy importante de la sociedad del momento¹⁸; además, porque solo a

17 Aunque «para el caso de Argentina, Chile y Uruguay, se incorporan también algunos textos en español de hispanohablantes no nacidos en América», dado que sería difícil comprender «la constitución e identidad lingüística de estos tres países sin los fuertes movimientos migratorios de europeos a esas zonas en el siglo XIX» (Bertolotti & Company Company 2014: 134), no queda claro si se trata de hispanohablantes nativos o individuos que tienen el español como segunda lengua; en todo caso, parece oportuno no circunscribir este hecho a la zona y el segmento temporal mencionados, por cuanto puede ser de interés para otras coordenadas espacio-temporales.

18 A manera de ejemplo, a partir de los datos del censo de Buenos Aires de 1887 Fontanella de Weinberg (1987: 133) señala que el 32,1 % de la población de la ciudad en este momento tiene nacionalidad italiana, lo que – sumado a que «una buena cantidad de los que tenían nacionalidad argentina (...) eran hijos de hogares italianos» (Fontanella de Weinberg 1987: 136) – lleva a esta investigadora a concluir que en estos momentos «los hablantes de italiano debían superar en holgura la tercera parte de la población porteña» (Fontanella de Weinberg 1987: 136); así las cosas, no incorporar al corpus la documentación generada por estos hablantes supone cercenar de forma muy

partir de esta documentación se pueden comprender el origen y desarrollo temprano de fenómenos lingüísticos que aparecerán posteriormente en los monolingües y que indudablemente se deben al contacto¹⁹; por último, por su propio interés intrínseco, pues permite analizar unas hablas que, a pesar de su carácter muchas veces transitorio, ayudan a comprender mejor y desde un punto de vista histórico los procesos que produce la convivencia de lenguas.

[25] Finalmente, también puede ser ahora, tras diez años de trabajo, el momento de discutir la oportunidad de sumar un nuevo subcorpus a los tres ya presentes – dedicados respectivamente a los documentos de archivo, a la prensa y a la literatura (Bertolotti & Company Company 2018: 79) – que esté conformado por obras pertenecientes al universo de los discursos técnicos y científicos, ausentes hasta ahora del corpus o subsumidos, en el mejor de los casos, junto a otros tipos textuales. Como es fácil de comprender, la incorporación de esta documentación resulta de enorme interés para los estudios históricos sobre el léxico especializado, todavía poco desarrollados en el continente (Frias Núñez 1998; Torres Montes 1998; Gómez de Enterría 2014), pero también para otras cuestiones como son, por ejemplo, la incorporación de indigenismos – abundantes, como es sabido, en los tratados médicos (Montero Lazcano 2020; Ramírez Luengo 2021b) – o el análisis de determinadas estructuras morfosintácticas, que en ocasiones presentan una especial predilección por, o un mayor empleo en, estos géneros textuales; además, este subcorpus supondría también la incorporación de nuevo material impreso a los fondos del corpus, algo que resulta de suma relevancia para poder llevar a cabo estudios relacionados con los procesos de estandarización ortográfica, la cual – como bien ha demostrado Cervelli (2021) para el siglo XVIII – presenta relevantes diferencias según el tipo de producción textual.

[26] De este modo, si el propósito de *CORDIAM* es, tal y como se dijo más arriba, representar del modo más preciso posible la multiplicidad de *formas de hablar* que conforman históricamente el español de América, salta a la vista que es necesario que en sus fondos tengan también cabida aquellos materiales que, desde lo autorial y desde lo tipológico, representan determinadas variedades que por motivos muy diversos han quedado más o menos relegadas en los tradicionales estudios diacrónicos sobre la lengua, si bien no por eso dejan de constituir, qué duda cabe, una parte fundamental de la rica variación que caracteriza, tanto hoy como en el pasado, al español propio del Nuevo Mundo.

5 Unas conclusiones provisionales: las carencias actuales de un corpus co-

notable la comprensión de la situación que, durante los últimos años del Ochocientos, presenta el español en la capital argentina.

19 A este respecto, no cabe duda de que un estudio detallado de aspectos como la imposición de ciertos usos preposicionales en determinados grupos sociales porteños (*voy del médico*) o la adopción y generalización de italianismos – tanto puros (*nona, laburo, raviolo*) como semánticos (*augurar* 'desear') – en esta variedad lingüística se puede ver favorecido por el análisis del habla de los bilingües que constituyen su origen. Esta misma idea es aplicable, naturalmente, al sur de los Estados Unidos y a la influencia del inglés en su español.

mo futuro horizonte de trabajo

[27] Una vez en este punto, parece importante retomar lo que se ha ido exponiendo a lo largo de estas páginas para concretar unas primeras conclusiones que, como se ha dicho ya en varias ocasiones, se deben entender en realidad como posibles líneas de trabajo que tendrán que ser desarrolladas en el futuro para conseguir la conformación de un corpus más completo y, por tanto, más útil para la investigación sobre el pasado lingüístico de América.

[28] Por lo que se refiere a las necesidades que ha mostrado esta rápida revisión, es obvio que lo primero que se debe hacer es articular proyectos y colaboraciones que permitan ir cubriendo los vacíos cronológicos – muy especialmente el siglo XIX, pero también, por ejemplo, la primera mitad del siglo XVI, momento fundacional imprescindible para poder entender la base lingüística del español del continente – y sobre todo la escasez geográfica que muestran países y zonas como Ecuador, Centroamérica en su totalidad, Puerto Rico o Paraguay, claramente infra-representados en el corpus; por lo que se refiere a esta última cuestión, además, es importante hacer hincapié en la necesidad de consultar de manera preferente los archivos de proximidad y locales, pues solo de esta manera será posible evitar la concentración documental en torno a los principales centros urbanos coloniales que en ocasiones muestra el corpus, permitiendo con ello una auténtica geografía lingüística histórica²⁰.

[29] Por otro lado, se hace necesario señalar que la mejora de la representatividad de *CORDIAM*, que debe constituir el objetivo de los próximos esfuerzos, no se circunscribe exclusivamente a lo diatópico y a lo cronológico, sino que se tiene que ampliar también a otros parámetros, tales como los sociales – con la incorporación, en la medida de lo posible, de documentación de grupos todavía insuficientemente representados, como las mujeres, los indígenas o muy especialmente la población afrodescendiente – o los tipológicos, en este caso por medio de la creación de un nuevo subcorpus dedicado al discurso técnico y científico. Como es fácil de percibir, las ventajas de contar con estos nuevos materiales son indudables: por lo que se refiere a la primera cuestión, porque permite llevar a cabo aproximaciones de tipo sociolingüístico a la diacronía del español, perspectiva sin duda fundamental para ofrecer una visión más completa y realista de los procesos que dan lugar a la situación actual; en cuanto a lo tipológico, porque supone un enriquecimiento de los registros documentados que aumenta la posibilidad de desarrollar estudios léxicos o de describir mejor la distribución textual de determinadas estructuras morfosintácticas. Así mismo, quizá estas ampliaciones se puedan completar con la toma en consideración en ciertos casos muy concretos de los es-

20 De hecho, el interés que encierran los archivos de proximidad no se reduce exclusivamente a lo geográfico, sino que se extiende también a los propios documentos que resguardan, pues en numerosas ocasiones son precisamente estos repositorios los que conservan textos más cercanos a la inmediatez comunicativa (Oesterreicher 1996) que permiten, por tanto, una mejor aproximación a la realidad dialectal de la época.

critos que generan los bilingües que tienen el italiano, el inglés o el portugués como lengua materna, pues no cabe duda de que su análisis puede ser un aporte de gran ayuda a la hora de comprender mejor la historia del español rioplatense, de las hablas del sur de los Estados Unidos o de las variedades lusitanizadas que se desarrollan en las fronteras circunbrasileñas.

[30] Finalmente, es evidente para cualquier usuario de *CORDIAM* que las características informáticas que posee el corpus constituyen una de sus grandes ventajas, dada la amplia gama de posibilidades de trabajo que, gracias a los metadatos que incorpora cada uno de los documentos (Bertolotti & Company Company 2018: 92-95), permite su buscador. Teniendo esto en cuenta, parece posible mencionar ahora determinados ajustes que, a partir de todo lo expuesto con anterioridad, se deberían implementar en el mencionado buscador como manera de aumentar tales posibilidades de trabajo: a manera de ejemplo, no estaría de más un nuevo metadato que describa la forma de producción textual (*manuscrito/impreso*) – importante, según se dijo ya, para la comprensión de determinados fenómenos gráficos –, así como otro que, desde lo geográfico, posibilite las búsquedas por localidad (*localidad*)²¹, pues este hecho permite establecer no solo áreas restringidas dentro de un país (la zona chiapaneca frente al resto de México), sino también regiones dialectales que se salen de los límites políticos nacionales (Nariño y la sierra ecuatoriana frente al resto de Colombia o la costa del Ecuador), permitiendo así un desarrollo mucho más efectivo de los estudios dedicados a la dialectología histórica; al mismo tiempo – y en caso de incorporar la documentación producida por bilingües –, quizá pueda ser oportuno contar con otro metadato que especifique esta condición (*monolingüe/bilingüe*), pues la toma en consideración de este criterio puede ser de notable interés para el análisis de ciertas cuestiones específicas²².

[31] En definitiva, se puede decir que, si el objetivo al crear *CORDIAM* era que este se erigiera como «el corpus de referencia para el español de América» (Bertolotti & Company Company 2018: 104), no cabe duda de que diez años después es esta una misión cumplida, tanto por su calidad intrínseca como por su importante y frecuente empleo por parte de los investigadores. Ahora bien, esta constatación no significa que, como corpus ampliable y perfectible que es, no se pueda

21 Aunque puede parecer tarea compleja, en realidad no lo es tanto, pues el metadato – ya presente en el corpus – *topónimo actual* permite establecer rápidamente un listado de las localidades donde se redactan los documentos; así las cosas, quizá lo único que se deba decidir sea si en esa nueva opción del buscador se permite introducir libremente el topónimo o, por el contrario, se ofrece al investigador el listado en cuestión para que él mismo seleccione los puntos que le interesa analizar.

22 A manera de ejemplo, no cabe duda de que esta circunstancia es fundamental para un estudio de las vacilaciones del vocalismo átono en el español andino del siglo XVII, cuya incidencia va a ser probablemente muy diferente según se tenga en cuenta textos generados por monolingües hispanófonos o por quechuahablantes bilingües; en el caso del léxico, la relevancia de este aspecto para el estudio de las estrategias de americanización léxica en el Perú del Seiscientos se ha puesto ya de manifiesto en Ramírez Luengo (2018).

-y no se deba- seguir mejorando a partir de un análisis de lo (mucho) que ya se ha logrado, pero sobre todo a partir de los objetivos que se pretenden conseguir en el futuro; es el momento, por tanto, de seguir trabajando con el mismo espíritu colaborativo que caracteriza a este proyecto desde sus mismos orígenes para conseguir un corpus todavía mejor, pues no cabe duda de que *CORDIAM*, en el fondo, no es sino el reflejo de ese mismo afán que compartimos todos los que contribuimos y nos servimos de él: conocer más profundamente y de forma más precisa el devenir histórico que ha dado como resultado el español en el que hoy en día se expresa y con el que se identifica la inmensa mayoría de los hispanohablantes.

Abreviaturas y referencias bibliográficas

- ALFAL = Asociación de lingüística y filología de América Latina. <https://mundoalfal.org/>.
- Almeida Cabrejas 2023 = Belén Almeida Cabrejas 2023. Escritura de mujeres en un corpus diacrónico: el caso de CODEA. Belén Almeida Cabrejas, Pedro Sánchez-Prieto Borja (eds.). *Varia lección de la lengua española. Estudios sobre el corpus CODEA*. Tirant lo Blanch, 391-438.
- Álvarez Nazario 1982 = Manuel Álvarez Nazario 1982. *Orígenes y desarrollo del español en Puerto Rico. Siglos XVI y XVII*. Universidad de Puerto Rico.
- Bertolotti & Company Company 2014 = Virginia Bertolotti, Concepción Company Company 2014. El *Corpus diacrónico y diatópico del español de América (CORDIAM)*. Propuesta de tipología textual. *Cuadernos de la ALFAL* 6, 130-148. https://www.mundoalfal.org/sites/default/files/revista/06_cuaderno_011.pdf.
- Bertolotti & Company Company 2018 = Virginia Bertolotti, Concepción Company Company 2018. El corpus para América: CORDIAM. Dolores Corbella Díaz, Alejandro Fajardo Aguirre, Jutta Langenbacher-Liebott (eds.). *Historia del léxico español y humanidades digitales*. Peter Lang, 75-105.
- Bertolotti, Coll & Polakof 2012 = Virginia Bertolotti, Magdalena Coll, Ana C. Polakof 2012. *Documentos para la historia del español en el Uruguay. Vol. 2. Cartas personales y documentos oficiales y privados del siglo XIX*. Universidad de la República.
- Buzek & Šincová 2015 = Ivo Buzek, Monika Šincová 2014. Introducción: *Una cercana sincronía opaca: estudios sobre el español del siglo XIX* (parte primera). *Études romanes de Brno* 36.1, 7-10. <https://hdl.handle.net/11222.digilib/134030>.
- Cervelli 2021 = Nicolò Cervelli 2021. *El uso de la tildación en el español del siglo XVIII: el caso de Guatemala (1750-1810)*. Tesis de maestría, Universidad Autónoma de Querétaro. <http://ri-ng.uaq.mx/handle/123456789/2858>.
- CHARTA = Pedro Sánchez-Prieto Borja (ed.) 2011-. *Corpus hispánico y americano en la red*. <https://www.redcharta.es>.
- Choy López 1999 = Luis R. Choy López 1999. *Periodización y orígenes en la historia del español de Cuba*. Tirant lo Blanch.
- Cock Hincapié 1969 = Olga Cock Hincapié 1969. *El seseo en el nuevo reino de Granada. 1550-1650*. Instituto Caro y Cuervo.
- Company Company 1994 = Concepción Company Company 1994. *Documentos lingüísticos de la Nueva España. Altiplano central*. Universidad Nacional Autónoma de México.
- Company Company 2001 = Concepción Company Company 2001. 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 2007 = Concepción Company Company 2007. *El siglo XVIII y la identidad lingüística de México. Discurso de ingreso a la Academia mexicana de la lengua. 10 de noviembre de 2005*. Universidad Nacional Autónoma de México.
- Corbella Díaz, Fajardo Aguirre & Langenbacher-Liebott 2018 = Dolores Corbella Díaz, Alejandro Fajardo Aguirre, Jutta Langenbacher-Liebott (eds.) 2018. *Historia del léxico español y humanidades digitales*. Peter Lang.
- CODEA = Pedro Sánchez-Prieto Borja (ed.) 2022. *CODEA. Corpus de documentos españoles anteriores a 1900*. <https://www.corpuscodea.es>.
- 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/>.
- CorLexIn = José R. Morala Rodríguez (ed.) 2024. *Corpus léxico de inventarios (CorLexIn)*. <https://apps2.rae.es/CORLEXIN.html>.
- Elizaincín, Malcuori & Bertolotti 1997 = Adolfo Elizaincín, Marisa Malcuori, Virginia Bertolotti 1997. *El español de la Banda Oriental en el siglo XVIII*. Universidad de la República.
- Fontanella de Weinberg 1987 = María B. Fontanella de Weinberg 1987. *El español bonaerense. Cuatro siglos de evolución lingüística (1580-1980)*. Hachette.
- Fontanella de Weinberg 1993 = María B. Fontanella de Weinberg (ed.) 1993. *Documentos para la historia lingüística de Hispanoamérica. Siglos XVI a XVIII. Vol. 1*. Real Academia Española.
- Fontanella de Weinberg 2004 = María B. Fontanella de Weinberg (ed.) 2004. *El español de la Argentina y sus variedades regionales*. Asociación Bernardino Rivadavia. Proyecto cultural Weinberg/Fontanella.
- Frías Núñez 1998 = Marcelo Frías Núñez 1998. Problemas terminológicos en la identificación de *La quina americana (1764-1828)*. José M. Urkia Etxabe (ed.). *Médicos vascos en América y Filipinas. II mesa redonda sobre historia de la medicina iberoamericana*. Real Sociedad Bascongada de Amigos del País, 53-61.
- García Carrillo 1988 = Antonio García Carrillo 1988. *El español en México en el siglo XVI. Estudio lingüístico de un documento judicial de la Audiencia de Guadalajara (Nueva España) del año 1578*. Alfar.
- Gómez de Enterría 2014 = Josefa Gómez de Enterría 2014. El vocabulario de las fiebres epidémicas en el español del siglo XVIII. España y México. José Luis Ramírez Luengo, Eva P. Velásquez Upegui (eds.). *La historia del español hoy. Estudios y perspectivas*. Axac, 199-216.
- Isasi Martínez 2012 = Carmen Isasi Martínez 2012. La edición electrónica y los corpus documentales: la innovación al servicio de la tradición. María Jesús Torrens, Pedro Sánchez-Prieto Borja (eds.). *Nuevas perspectivas para la edición y el estudio de documentos hispánicos antiguos*. Peter Lang, 351-371.
- Kabatek 2013 = Johannes Kabatek 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, 8-28.
- Kabatek 2016 = Johannes Kabatek 2016. Un nuevo capítulo en la lingüística histórica iberorrománica: el trabajo crítico con los corpus. Johannes Kabatek (ed.). *Lingüística de corpus y lingüística histórica iberorrománica. Introducción a este volumen*. De Gruyter, 1-17.
- Léxico hispanoamericano* = Ray Harris-Northall, John J. Nitti (eds.) 2007. *Peter Boyd-Bowman's Léxico hispanoamericano 1493-1993*. https://textred.spanport.wisc.edu/lexico_hispanoamericano/index.html.
- Lope Blanch 1985 = Juan M. Lope Blanch 1985. *El habla de Diego de Ordaz. Contribución a la historia del español americano*. Universidad Nacional Autónoma de México.
- Lope Blanch 1996 = Juan M. Lope Blanch 1996. México. Manuel Alvar (ed.). *Manual de dialectología hispánica. El español de América*. Ariel, 81-89.
- Medina López 1995 = Javier Medina López 1995. *El español de América y Canarias desde una perspectiva histórica*. Verbum.
- Medina López 2022 = Javier Medina López 2022. De Cuervo al CORDIAM: los corpus lingüísticos en el contexto de la historia del español de América. *Scriptum digital* 11, 83-103. <https://doi.org/10.5565/rev/scriptum.123>.
- Montero Lazcano 2020 = Mara Y. Montero Lazcano 2020. Indigenismos en el discurso médico de Guatemala del siglo XVIII: el caso de la *Instrucción sobre el modo de practicar la inoculación de las viruelas* de José Felipe Flores. *Études romanes de Brno* 41.2, 41-51. <https://hd->

- [l.handle.net/11222.digilib/143262](https://hdl.handle.net/11222.digilib/143262).
- Montes Giraldo 1996 = José J. Montes Giraldo 1996. Colombia. Manuel Alvar (ed.). *Manual de dialectología hispánica. El español de América*. Ariel, 134-145.
- Newland 1991 = Carlos Newland 1991. Spanish American elementary education before independence. Continuity and change in a colonial environment. *Itinerario* 15.2, 79-95.
- Obediente Sosa 2002 = Enrique Obediente Sosa (ed.) 2002. *Documentos para la historia lingüística de Mérida (Venezuela). (Siglos XVI-XVII)*. Universidad de Los Andes. https://www.human.ula.ve/linguisticahispanica/documentos/corpus_cdc.pdf.
- Oesterreicher 1996 = Wulf Oesterreicher 1996. Lo hablado en lo escrito. Reflexiones metodológicas y aproximación a una tipología. Thomas Kotschi, Wulf Oesterreicher, Klaus Zimmermann (eds). *El español hablado y la cultura oral en España e Hispanoamérica*. Iberoamericana, Vervuert, 317-340.
- Parodi 1995 = Claudia Parodi 1995. *Orígenes del español americano. Reconstrucción de la pronunciación*. Universidad Nacional Autónoma de México.
- Quesada Pacheco 1990 = Miguel Á. Quesada Pacheco 1990. *El español colonial de Costa Rica*. Universidad de Costa Rica.
- Ramírez Luengo 2009 = José Luis Ramírez Luengo 2009. La edición de textos americanos de carácter lingüístico: logros y necesidades. Cristina Castillo Martínez, José Luis Ramírez Luengo (eds.). *Lecturas y textos en el siglo XXI. Nuevos caminos en la edición textual*. Axac, 173-182.
- Ramírez Luengo 2010 = José Luis Ramírez Luengo 2010. El contacto del español y el portugués en la historia: situaciones y resultados americanos. *Letr@ Viv@* 10.1, 13-48. <https://drive.google.com/file/d/1-vjWy9y96ojTxIZLolVA6ZXyVse3Ob6k/view>.
- Ramírez Luengo 2011 = José Luis Ramírez Luengo 2011. *La lengua que hablaban los próceres. El español de América en la época de las independencias*. Voces del Sur.
- Ramírez Luengo 2012a = José Luis Ramírez Luengo 2012a. Algunas cuestiones teóricas acerca de la edición de documentos lingüísticos americanos. María Jesús Torrens, Pedro Sánchez-Prieto (eds.). *Nuevas perspectivas para la edición y el estudio de documentos lingüísticos antiguos*. Peter Lang, 301-310.
- Ramírez Luengo 2012b = José Luis Ramírez Luengo 2012b. Notas sobre las tendencias gráficas del español colombiano en la época de las independencias (1830). José Luis Ramírez Luengo (ed.). *Por sendas ignoradas. Estudios sobre la lengua española en el siglo XIX*. Axac, 167-182.
- Ramírez Luengo 2015 = José Luis Ramírez Luengo 2015. La muerte de una letra: empleo y decadencia de la <ç> en la escritura del español (siglos XVI-XVIII). *Diálogo de la lengua. Revista de filología y lingüística españolas* 7, 22-35. https://www.dialogodelalengua.com/articulo/pdf/7/2_Ramirez%20Luengo_DL_2015.pdf.
- Ramírez Luengo 2017 = José L. Ramírez Luengo (ed.) 2017. *Textos para la historia del español. XI. Honduras y El Salvador*. Universidad de Alcalá.
- Ramírez Luengo 2018 = José Luis Ramírez Luengo 2018. Contacto de lenguas e historia del léxico en el Perú: algunas notas sobre el vocabulario de monolingües y bilingües en el siglo XVII. Roxana Risco (ed.). *Estudios de variación y contacto lingüístico en el español peruano*. Universidad Nacional de La Plata, 41-58. <https://www.libros.fahce.unlp.edu.ar/index.php/libros/catalog/view/108/136/1033-1>.
- Ramírez Luengo 2019 = José Luis Ramírez Luengo 2019. La configuración fónica del español salvadoreño en la época colonial (1650-1803). *Boletín de la Real Academia Española* 99.320, 817-834. <https://revistas.rae.es/brae/article/view/213>.
- Ramírez Luengo 2021a = José Luis Ramírez Luengo 2021a. La americanización léxica del español guatemalteco de la Ilustración a partir de los *Apuntamientos sobre la agricultura y el co-*

- mercio del Reyno de Guatemala (1811)*. Guillermina Herrera Peña (ed.). *Publicación conmemorativa. Bicentenario de la Independencia 1821-2021. La lengua española en Guatemala*. Academia Guatemalteca de la Lengua, 67-111. <https://agl.org.gt/obras-conmemorativas/>.
- Ramírez Luengo 2021b = José Luis Ramírez Luengo 2021b. Los indigenismos en el *Florilegio medicinal* de Juan de Esteyneffer (1712). *Nuevas Glosas. Estudios lingüísticos y literarios* 1, 7-25. <https://revistas.filos.unam.mx/index.php/nuevasglosas/article/view/1417>.
- Ramírez Luengo 2022 = José Luis Ramírez Luengo 2022. La historia fónica del español hondureño: una aproximación a la época tardocolonial (1650-1800). *Revista de filología española* 102.1, 245-257. <https://doi.org/10.3989/rfe.2022.010>.
- Rivarola 2009 = José Luis Rivarola 2009. *Documentos lingüísticos del Perú. Siglos XVI y XVII. Edición y comentario*. Consejo superior de investigaciones científicas.
- Rojas Mayer 2000 = Elena Rojas Mayer (ed.) 2000. *Documentos para la historia lingüística de Hispanoamérica. Siglos XVI a XVIII. Vol. 2*. Real Academia Española.
- Sánchez-Prieto Borja 2012 = Pedro Sánchez-Prieto Borja 2012. Desarrollo y explotación del corpus de documentos españoles anteriores a 1700 (CODEA). *Scriptum digital* 1, 5-35. <https://doi.org/10.5565/rev/scriptum.31>.
- Torres Montes 1998 = Francisco Torres Montes 1998. Fitónimos amerindios recogidos por José Celestino Mutis. *Thesaurus. Boletín del Instituto Caro y Cuervo* 53.2, 242-270. https://cvc.cervantes.es/lengua/thesaurus/pdf/53/TH_53_002_026_0.pdf.