

# En busca de lo hablado en lo escrito en los corpus diacrónicos del español: una comparativa con los corpus anglosajones<sup>1</sup>

*In search of the spoken word in Spanish written diachronic corpora: A comparison with English corpora.*

**Paula Rodríguez Puente**  
Universidad de Oviedo, España

**Resumen:** El presente trabajo indaga en los corpus diacrónicos del español disponibles hasta la fecha, prestando especial atención a la escasa representación en los mismos de textos que se aproximan al plano oral. Si bien es cierto que la creación de corpus diacrónicos del español ha avanzado significativamente durante los últimos cuarenta años, todavía hay cabida para mejoras, especialmente en el ámbito del español peninsular y en lo que se refiere a la inclusión de textos que permitan estudios de las variedades diatópicas, diafásicas y diastráticas desde la perspectiva diacrónica. Esta investigación ofrece además datos sobre tres corpus diacrónicos en lengua inglesa centrados en esta línea, que se podrían tomar como modelo para la creación de herramientas similares para el español.

**Palabras clave:** corpus diacrónicos del español, corpus diacrónicos del inglés, oralidad, escrituralidad, habla coloquial.

**Abstract:** The present work analyses the diachronic corpora of Spanish available to date, paying especial attention to the scarce representation of the spoken language in them. Although it is true that the creation of diachronic corpora of Spanish has progressed significantly over the last forty years, there is still room for improvement, especially as regards peninsular Spanish, as well as the inclusion of documents that allow diatopic, diaphasic and diastratic analyses from the diachronic perspective. This piece of research provides data from three English diachronic corpora along these lines, which might be taken as a point of departure for the creation of similar tools in Spanish.

**Keywords:** diachronic corpora of Spanish, diachronic corpora of English, spokenness, writtenness, colloquial language.

## 1. Introducción

Los corpus, colecciones de textos representativas del lenguaje natural compiladas de forma sistemática o aleatoria y almacenadas y procesadas electrónicamente, han sido y son empleados como una importante fuente de

---

<sup>1</sup> Este estudio ha sido en parte financiado gracias a ayudas del Fondo Europeo de Desarrollo Regional y el Ministerio de Economía y Competitividad (becas FFI2014-52188-P y FFI2014-51873-REDT). La autora también quiere hacer constar su agradecimiento a los dos revisores anónimos por sus sugerencias de corrección y mejora.

información para el estudio histórico y sincrónico de las lenguas. La aplicación de las metodologías de la lingüística de corpus como recurso para la obtención y análisis de datos reales ofrece una visión única de la dinámica de las lenguas, permitiendo dar respuesta a numerosas cuestiones lexicográficas, gramaticales, de adquisición de segundas lenguas, así como aspectos relativos a la traducción, sociolingüística, dialectología o historia de la lengua, por mencionar algunas. Aunque la lingüística de corpus tuvo unos inicios un tanto truncados, debido en parte a las numerosas críticas recibidas especialmente de detractores generativistas, hoy en día cualquier investigación en el ámbito de la diacronía de las lenguas es prácticamente impensable sin recurrir a los corpus.

Los pioneros estudios realizados por Kädling (1887), Firth (1951) y Fries (1952) suelen señalarse como los primeros análisis de corpus propiamente dichos, aunque estos poco tienen que ver con la definición de corpus que manejamos actualmente, sintetizada en las palabras de Leech como "a helluva lot of text, stored on a computer" (Leech, 1992: 106).<sup>2</sup> Dicho de otro modo, la introducción de las nuevas tecnologías en el ámbito de la lingüística abrió la posibilidad de procesar y analizar datos lingüísticos de forma masiva, algo que no era posible en la denominada era pre-computacional. Hoy en día la investigación en lingüística diacrónica depende de los corpus electrónicos como fuente primaria de datos o para validar hipótesis. Sin embargo, sus aplicaciones no deben ser sobrevaloradas, ya que éstos no son representativos de toda una lengua, por lo que los investigadores no deben caer en lo que Rissanen denomina 'God's true fallacy' (1989: 17) o la falacia de creer que los datos de un corpus son extrapolables a toda una lengua.

En el ámbito hispánico los trabajos basados en corpus mantuvieron cierta relevancia especialmente en el estudio desde la perspectiva histórica, a pesar de las críticas de detractores que abogaban en contra del uso de corpus lingüísticos para el estudio de lenguas.<sup>3</sup> Los corpus diacrónicos, de hecho, permiten acceder y analizar enunciados producidos por hablantes del pasado, si bien con limitaciones, perfectamente resumidas en la famosa cita de Labov, quien argumentó que "historical linguistics can [...] be thought of the art of making the best of bad data" (1994: 11). Tal afirmación tiene su origen en el hecho de que los corpus diacrónicos no son representativos de los sistemas lingüísticos del pasado, sino que recogen solamente aquellos textos que han podido sobrevivir a los sucesos de la historia.

Los recursos electrónicos disponibles para el estudio de la historia del español han proliferado a lo largo de los últimos cuarenta años y hoy en día son muchas las herramientas de las que el investigador dispone para rastrear los orígenes de nuestro idioma. Sin embargo, todo corpus diacrónico tiene, por naturaleza, una limitación importante: la ausencia de testimonio de la lengua hablada. Varios trabajos recientes abogan por la inclusión en los corpus de tipos de textos más cercanos a o representativos de la lengua hablada (vid., entre otros, Culpeper y Kytö, 2010), como pueden ser las cartas privadas, las

---

<sup>2</sup> Para una visión más detallada sobre los antecedentes de la lingüística de corpus, vid. Enrique Arias (2009) y Rojo (2015).

<sup>3</sup> Vid. los pioneros trabajos del HSMS y ADMYTE que se relatan en la Sección 3 de este trabajo.

obras de teatro o las declaraciones de testigos. A pesar de ello, los corpus diacrónicos de español que hoy en día incluyen este tipo de textos entre sus documentos son muy escasos en comparación con otras lenguas como el inglés.

La finalidad de este trabajo es realizar un recorrido por las herramientas disponibles para los análisis basados en corpus en el ámbito de la lingüística histórica española, haciendo hincapié en el hecho de que la lengua oral está escasamente representada en los corpus diacrónicos del español, especialmente en lo referido a los estratos sociales más bajos y a la representación de autoras femeninas, y comparando su tradición con la anglosajona, donde sí existen ya algunas herramientas de esta índole. Si bien es cierto que se están realizando esfuerzos para solventar esta insuficiencia por parte de varios proyectos, especialmente en Hispanoamérica (vid. Sección 3), todavía queda mucho por hacer en este ámbito. El trabajo presenta datos sobre tres corpus diacrónicos de lengua inglesa (vid. Sección 4) que podrían tomarse como punto de partida para la creación de herramientas similares en lengua española, siempre desde una perspectiva de crítica constructiva y teniendo en cuenta que la elaboración de cualquier corpus electrónico no es, ni mucho menos, una tarea sencilla.

## **2. Representación del plano oral en los corpus diacrónicos**

En las últimas décadas el enriquecimiento de la lingüística histórica a través del contacto con otras disciplinas, así como la creación y proliferación de numerosos recursos electrónicos para este propósito, ha permitido perfeccionar las metodologías aplicadas en el análisis del cambio lingüístico y ha facilitado la apertura de nuevas vías de estudio a través de los corpus que, utilizados correctamente, poseen un valor incalculable para explicar el cambio lingüístico desde la perspectiva diacrónica. No obstante, como es lógico, todo corpus diacrónico presenta por su naturaleza una limitación forzosa, que es la ausencia de transcripciones directas de la lengua hablada, a pesar del reconocimiento de que todo cambio lingüístico se origina en su primer estadio en el plano oral, tal como reflejan las palabras de Rissanen (1999: 188):

It is a constant source of frustration for the language historian that all observations and analyses of the early periods have to be based on written evidence only, while the importance of speech in the development of languages is self-evident.

Aunque la inclusión de transcripciones de la lengua oral en corpus diacrónicos es impensable debido a la ausencia de dispositivos de grabación en épocas anteriores, sí es posible minimizar de algún modo esta carencia dando cabida en los corpus a textos más próximos al plano oral y de carácter informal. Como prosigue Rissanen (1999: 188):

It is, of course, a truism that no written text, be it dialogue in a comedy or novel, a sermon or a record of a debate or discussion, will ever give a faithful reproduction of spoken language. But by careful comparison of texts which stand at different distances from spoken language [...], it is possible to present hypotheses about whether a certain construction is favoured or avoided in the spoken language of the period.

Oesterreicher (2004: 735-736) refuerza la idea de Rissanen, tildando de “paradójico” el método adoptado para analizar el espacio discursivo-variacional de una lengua, pues al carecer de acceso a los fenómenos de inmediatez comunicativa en el medio oral, nos vemos obligados a buscar evidencia en su forma gráfica. Por otra parte, “es muy difícil captar las manifestaciones espontáneas y extremadamente contextualizadas de la inmediatez comunicativa [lo cual] representa un problema gravísimo, sobre todo para la investigación lingüística diacrónica” (Oesterreicher, 1996: 319). Si bien es cierto que, al reproducir textos orales en la escritura, “es obligado suprimir numerosos signos no verbales del diálogo, al mismo tiempo que otros de naturaleza verbal” (Bustos Tovar, 2001: 192) y que “la incorporación de lo coloquial en la literatura [...] nunca puede lograrse con total autenticidad [pues], aunque en grado diverso, siempre hay manipulación por parte del autor” (Narbona Jiménez, 1989: 24), son muchos los estudios que han indagado en las diversas tipologías textuales para determinar su grado de oralidad.<sup>4</sup> Tradicionalmente, las cartas personales han sido consideradas como testimonios de la interacción oral del pasado (vid., entre otros, Cano Aguilar, 1996 y los trabajos de Nevalainen y Raumolin-Brunberg con el *Corpus of Early English Correspondence*; Nevalainen et al., 1998). Éstas y los diarios personales, entre otros, representan estilos de comunicación más personal (Biber y Finegan, 1997: 255), y es en ellos donde debemos buscar características típicamente coloquiales.<sup>5</sup> Como señala Tieken-Boon van Ostade (2005: 131), en los contextos privados e íntimos las personas se sienten más relajadas y pueden bajar la guardia, incluso en lo referido a aspectos lingüísticos. Sin embargo, no debemos caer en el error de considerar rasgos orales toda aquella desviación de la norma que se presente en los textos escritos, puesto que los errores o las anomalías también existen, aunque a veces es difícil distinguirlas (Oesterreicher, 2004: 747).

Analizadas con referencia a la dimensión *Involved vs. Informational production* (Biber, 1988), las cartas personales pueden caracterizarse como orales en tanto en cuanto sus características formales las aproximan hacia el extremo más involucrado (*involved*). En la terminología de Koch y Oesterreicher (1985, 1990), basada en una escala de inmediatez y distancia comunicativa o, lo que es lo mismo, un continuo entre lo hablado y lo escrito prototípico, las cartas privadas de un semiculto (o las declaraciones de un testigo de origen humilde) constituyen formas cercanas al polo de la inmediatez (Oesterreicher, 2004: 734). Sin embargo, como señalan Culpeper y Kytö (2010: 16), las cartas personales presentan tres desventajas: 1) no representan una interacción activa, 2) no representan una interacción cara a cara entre los interlocutores, y 3) la representación de los interlocutores está condicionada por su grado de alfabetización, por lo que los miembros de estratos inferiores, así como las mujeres, tienen una representación menor. Por otra parte, en el ámbito anglófono los diversos estudios de análisis de variables múltiples llevados a cabo por Biber et al. (vid., entre otros, Biber, 1988, 2001;

---

<sup>4</sup> Para una amplia y exhaustiva revisión de los trabajos de este tipo, vid. López Serena (2007: 29-190).

<sup>5</sup> Es importante señalar, sin embargo, que los aspectos coloquiales también pueden encontrarse en textos no orales, como es el caso de los pasajes en estilo indirecto de *El Corbacho*, tal como muestra Bustos Tovar (2001: 198-200).

Biber y Finegan, 1988, 1989, 1992, 1997) en el corpus ARCHER (*A Representative Corpus of Historical English Registers*) han demostrado que los tipos de textos no pueden describirse a través de parámetros fijos, sino que son variables en el tiempo. Así, mientras las cartas privadas tal como las conocemos hoy en día podrían caracterizarse por un tono informal, de interacción y con una notoria involucración personal (Biber, 2001: 105) que las aproxima a la lengua oral, no ocurre así con la correspondencia privada del s. XVIII, de carácter eminentemente formal, expositivo, descriptivo y argumentativo (Biber, 2001; Biber y Finegan, 1989, 1997) y obedeciendo a una serie de convenciones postuladas en manuales de escritura. En definitiva, las cartas formales del s. XVIII no eran "thoughtless outpourings" (Anderson y Ehrenpreis, 1966: 273), sino el resultado de un complejo proceso de redacción (vid. Austin, 1998: 323; Nevalainen, 2001: 219-220). A pesar de estos inconvenientes, los documentos personales tienen un alto potencial para representar auténticas muestras de oralidad histórica (Elpass, 2012: 159), pues antes del advenimiento de las nuevas tecnologías, "los intentos de reproducción mimética de la oralidad [...] en determinadas obras literarias conformaban el único material con que contaban los analistas de la conversación coloquial" (López Serena, 2007:193). Otros ejemplos de documentos personales en los que también se puede discernir la impronta oral además de las cartas privadas y/o íntimas son, como apunta Oesterreicher (2004: 747-752), los libros de familia, diarios, documentos autobiográficos, las relaciones y crónicas soldadescas,<sup>6</sup> así como los apuntes y notas personales. Fuera del ámbito personal, por otra parte, también es posible encontrar rasgos de la lengua hablada en transcripciones de situaciones comunicativas reales, como las actas de tribunales, donde incluso se recogen blasfemias, maldiciones, insultos y ultrajes (Oesterreicher, 2004: 752).<sup>7</sup>

Como señala Oesterreicher (1996: 320), la representación de lo oral en el medio escrito representa un problema muy antiguo al que se le ha prestado escasa e imprecisa atención. Las primeras aproximaciones a la distinción entre lo oral y lo escrito desde una perspectiva multidimensional cobraron fuerza a partir de la publicación de los trabajos de Koch y Oesterreicher (1985, 1990), a los que siguieron numerosas publicaciones al respecto. López Serena (2007: 141-148) hace una minuciosa revisión de estos trabajos que resulta innecesaria e irrelevante repetir aquí. Tomando como punto de partida las apreciaciones sobre inmediatez y distancia comunicativa de Koch y Oesterreicher (1985: 19-21; 1990: 8-10) y Koch (1999), una aportación más reciente a aquellos documentos escritos que se aproximan al plano de la oralidad proviene del ámbito anglófono a través de la propuesta de Culpeper y Kytö (2010), que abogan por incluir dentro del grupo de documentos que se aproximan al plano de la oralidad textos como los procesos judiciales, las declaraciones de testigos, obras teatrales, manuales o instrucciones de tipo dialógico y los textos ficticios en prosa que contienen representaciones de diálogos. Las ventajas de estos documentos son, según Culpeper y Kytö (2010: 16), las siguientes:

- (Re)presentan la lengua hablada en curso.

---

<sup>6</sup> Para una valoración sobre los rasgos orales en las crónicas soldadescas, vid. Stoll (1996).

<sup>7</sup> Para un listado de aquellos textos en los que es posible discernir rasgos de lo que Oesterreicher (2004) denomina inmediatez comunicativa, vid. Oesterreicher (1996: 324-332; 2004: 747-756).

- (Re)presentan un lenguaje altamente interactivo y cara a cara entre los interlocutores.
- Cubren un amplio espectro de grupos sociales.
- Tienden a estar disponibles en cantidades razonablemente amplias, siempre dependiendo del género y el período.

Por otra parte, Culpeper y Kytö (2010: 17) reconocen las limitaciones de estos documentos en cuanto a su fiabilidad de representación, así como en el grado de exactitud con el que se puede llegar a conocer el entorno social y las relaciones entre los participantes. Aun así, su clasificación de tipos de textos resulta novedosa, pues abarca un amplio número de tipologías textuales que analizan atendiendo a diversas características para determinar su grado de oralidad. El estudio de Culpeper y Kytö (2010) propone una división fundamental entre tipos de textos: por un lado, los llamados *written-based and written-purposed text types* y, por otro lado, los denominados *speech-related texts*. Estos últimos se dividen además en tres subgrupos:

- *Speech-like*, un término concebido mediante una escala de inmediatez comunicativa (vid. Koch, 1999). La correspondencia personal, por ejemplo, representa este subgrupo en el sentido de que ni está basada en ni es diseñada para ser lenguaje oral, y la afirmación de que representa el lenguaje oral o coloquial se basa exclusivamente en que contiene características que son típicas de la lengua oral.
- *Speech-based*, como los procesos judiciales o las declaraciones de testigos, son textos basados en un acto real de lengua hablada. La exactitud con la que estos textos representan la lengua hablada es, sin embargo, cuestionable, dado que en ausencia de aparatos de grabación, como mucho se trata de reproducciones asistidas por notas taquigráficas.<sup>8</sup>
- *Speech-purposed*, textos diseñados para ser articulados de forma oral. Los sermones y las proclamaciones, por ejemplo, se elaboran como monólogos, mientras que las obras teatrales representan la interacción oral.

Las líneas divisorias de estos dos grandes grupos y sus ramificaciones no son ni mucho menos estrictas, por lo que es posible adscribir un texto concreto a más de una de ellas. De hecho, Culpeper y Kytö representan las intersecciones entre todos los tipos de texto tal como se muestra en la figura 1 (Culpeper y Kytö, 2010: 18).

---

<sup>8</sup> Para mayor información sobre la fiabilidad de los métodos taquigráficos para la representación de las transcripciones de los juicios, vid. Grund (2007) sobre los juicios de las brujas de Salem y Huber (2007) sobre transcripciones de los procesos judiciales del *Old Bailey*.

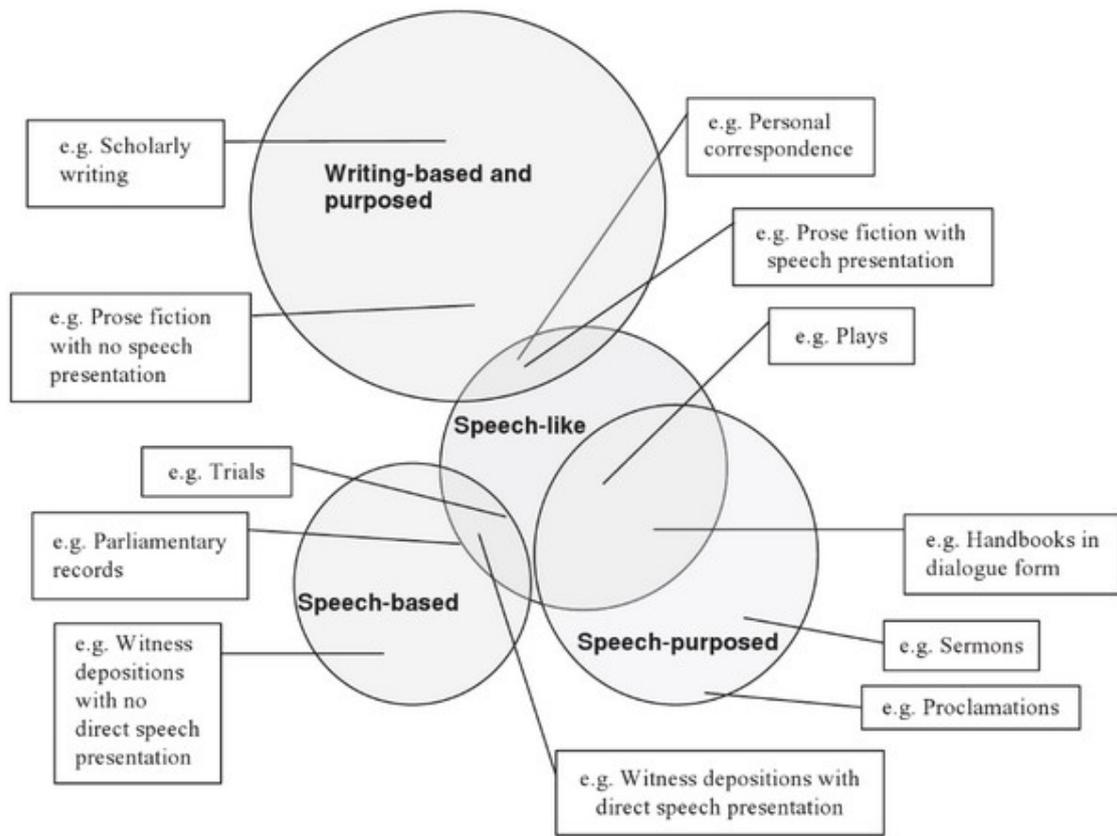


Figura 1: Interrelación entre los diversos tipos de textos escritos y orales [reproducción de Figure 1.2 en Culpeper y Kytö (2010: 18)]

En definitiva, aunque un corpus diacrónico no puede ni mucho menos contener muestras de lengua hablada, sí que existen ciertos tipos de textos que pueden describirse como más cercanos a ésta. La inclusión de los llamados *speech-related texts* en un corpus abre la oportunidad a los investigadores de conocer más de cerca la lengua oral del pasado e indagar en más profundidad en los distintos cambios lingüísticos sufridos por nuestro idioma con el paso del tiempo, siempre sin olvidar que tales textos no constituyen sino aproximaciones a la lengua oral (no representaciones fidedignas).

### 3. Corpus diacrónicos en lengua española y su representación de la oralidad

En la sección anterior describimos formas en que la lengua oral puede ser representada en los corpus diacrónicos, aunque con limitaciones. Aunque la gama de corpus electrónicos que facilitan el estudio de la diacronía del español ha proliferado sustanciosamente en los últimos cuarenta años, todavía queda mucho trabajo por hacer en este ámbito. Son varios los aspectos mejorables, entre los que destacan los tipos de textos incluidos en ellos, los niveles de acceso a los documentos (paleográfico, textos normalizados, en facsímil, etc.), la lematización y etiquetado a distintos niveles, así como el control de factores extratextuales.<sup>9</sup> En esta sección daremos un repaso a los

<sup>9</sup> Para un completo resumen de varios inconvenientes de los corpus diacrónicos actuales en español, vid. García-Salido y Vázquez-Rozas (2012).

corpus diacrónicos del español disponibles hasta la fecha, centrándonos en los aspectos mejorables en lo que concierne a los tipos de textos que incluyen. Como se ha señalado anteriormente, es en la lengua oral donde los cambios lingüísticos tienen lugar en primera instancia y tenemos a nuestro alcance un gran número de textos escritos que representan o reproducen la lengua hablada y coloquial (*speech-related text types*) que, aun con reconocidas limitaciones, resultan de valor incalculable para aproximarnos a la lengua de nuestros ancestros. A pesar de ello, son pocos los corpus lingüísticos que incluyen entre sus documentos textos de esta índole en el ámbito de la diacronía del español. En los últimos años, especialmente en el ámbito hispanoamericano, pero también en el del español peninsular, se ha comenzado a trabajar para incrementar la disponibilidad de este tipo de textos.<sup>10</sup>

### **3.1. Hispanic Seminary of Medieval Studies<sup>11</sup>**

Tan solo una década después de la creación del *Brown Corpus*<sup>12</sup> (1961) el Hispanic Seminary of Medieval Studies (HSMS) de la Universidad de Wisconsin, dirigido por los profesores Lloyd A. Kasten y John J. Nitti, comenzaba a utilizar los ordenadores para la compilación de diccionarios y el análisis textual mediante un sistema de transcripción semipaleográfica que utilizaba los códigos ASCII. Su primer gran proyecto fue el *Dictionary of the Old Spanish Language* (DOSL), que contenía principalmente textos de la variedad castellana (especialmente los vinculados a Alfonso X), si bien posteriormente se incorporaron otras variedades a través de otro gran proyecto desarrollado en la década de los 70: *The Concordances and Texts of the Royal Scriptorium Manuscripts of Alfonso X* (1978). Éste daría lugar años más tarde a la serie de *Textos and Concordancias* que, a través de los años, ha añadido a los archivos del DOSL una gran cantidad de textos adicionales. Desde 2005 el HSMS comenzó a trabajar en la *Biblioteca Digital de Textos del Español Antiguo* con intención de ofrecer acceso a los usuarios en línea. Su primer trabajo vio la luz en 2011 con la publicación de la *Obra en prosa de Alfonso X el sabio* y, tal como se recoge en la página web de la asociación, en la actualidad existen un total de siete corpus digitales disponibles resumidos en la tabla 1.

---

<sup>10</sup> La información de los diversos corpus que se mencionan en este trabajo se ha elaborado a partir de datos recabados en la bibliografía disponible, así como en las diferentes páginas web de los corpus. Aunque el análisis pretende ser exhaustivo, dada la existencia de múltiples trabajos en curso, es posible que inconscientemente e involuntariamente algunos corpus diacrónicos del español no se mencionen en este artículo. Si bien es cierto que Briz y Albelda (2009) y Rojo (2016) ofrecen un amplio y detallado listado de recursos lingüísticos disponibles, la labor de una investigación de estas características se habría visto beneficiada de la existencia de una base de datos que recoja información sobre todos los corpus disponibles hasta la fecha, del modo que lo hace el *Corpus Resource Database* (CoRD, <http://www.helsinki.fi/varieng/CoRD/>) con los corpus de lengua inglesa.

<sup>11</sup> La información de esta sección ha sido obtenida a través de la página web del Hispanic Seminary of Medieval Studies: <http://www.hispanicseminary.org/index-es.htm> (fecha de acceso 11/03/2016).

<sup>12</sup> El primer corpus electrónico del mundo que consiste en un millón de palabras de inglés estadounidense organizadas en textos de quince categorías diferentes.

<b>Corpus</b>	<b>Número de archivos</b>	<b>Número de palabras</b>
<i>Obra en prosa de Alfonso X el sabio</i>	20	3.255.914
<i>Textos médicos españoles</i>	55	2.642.403
<i>Textos navarro-aragoneses</i>	36	3.336.942
<i>Textos legales españoles</i>	49	2.779.914
<i>Textos bíblicos españoles</i>	19	4.936.036
<i>Textos poéticos españoles</i>	37	1.094.260
<i>Textos tempranos de Celestina</i>	21	1.277.649
<b>Total</b>	<b>237</b>	<b>193.231.118</b>

Tabla 1: Corpus digitales y distribución de número de textos y palabras de los corpus del HSMS

En el apéndice I se recoge un listado con información más detallada sobre los textos específicos incluidos en cada uno de los corpus del HSMS. Dicho listado pone de manifiesto que el lenguaje formal y literario es claramente predominante.

### **3.2. Archivo Digital de Manuscritos y Textos Españoles (ADMYTE)<sup>13</sup>**

Aproximadamente dos décadas después de que el HSMS comenzara su trabajo, aparecía en España el *Archivo Digital de Manuscritos y Textos Españoles* (ADMYTE), un proyecto desarrollado por iniciativa de los profesores Francisco Marcos Marín, Charles Faulhaber y Ángel Gómez Moreno (Marcos Marín, 1994). ADMYTE unía, por una parte, la reproducción completa de libros antiguos en facsímil anteriores al siglo XVI junto con su transcripción electrónica y, por otra parte, la transcripción electrónica con un complejo sistema de recuperación de la información en texto e imagen y con un diccionario formal para el español pre-clásico. En la actualidad, la base de datos de ADMYTE suma más de trescientos textos transcritos. El listado detallado de los textos de ADMYTE se recoge en el apéndice II que, muestra las referencias ordenadas en once grandes bloques temáticos (libros, textos científicos, crónicas y biografías, diccionarios y gramáticas, enciclopedias, textos legales, textos poéticos, prosa literaria, libros sapienciales y libros de viajes) que, al igual que ocurre con los textos del HSMS (apéndice I), ponen de manifiesto que se trata de documentos escritos de carácter formal y con una nula representación de la oralidad.

En la época en que ADMYTE publicó sus primeros trabajos pocos eran los usuarios que tenían acceso a las tecnologías de la información o estaban familiarizados con los medios informáticos, por lo que la difusión de sus trabajos tuvo escasa repercusión. En el ámbito de los corpus diacrónicos en lengua hispánica, habría que esperar hasta el siglo XXI a que se publicasen obras de libre acceso para los usuarios, siendo las más conocidas, si bien no las únicas, el *Corpus Diacrónico del Español* (CORDE) de la Real Academia Española y el *Corpus del Español* (Davies, 2002a).

<sup>13</sup> La información de esta sección ha sido extraída a través de la web <http://www.llf.uam.es/~fmarcos/informes/admyte/admytei.htm> (fecha de acceso 11/03/2016).

### 3.3. Corpus Diacrónico del Español (CORDE)<sup>14</sup>

En el año 1995, la Real Academia Española comenzó a aplicar el uso de las nuevas tecnologías para desarrollar un banco de datos que desde 1998 se encuentra a disposición del público bajo el nombre de *Corpus Diacrónico del Español* (CORDE), un corpus textual de todas las épocas y lugares en que se habló español, desde los inicios del idioma hasta el año 1975, cuando limita con el *Corpus de Referencia del Español Actual* (CREA). En la actualidad el CORDE cuenta con 250 millones de registros correspondientes a textos escritos de diversos géneros. Dichos textos contienen una amplia representación diatópica (74% de español peninsular y 26% para el resto), histórica y genérica (ficción en verso y prosa, y no ficción), ofreciendo un conjunto suficientemente representativo. Cronológicamente, el CORDE se divide en tres grandes etapas (Edad Media, Siglos de Oro y Época Contemporánea), que a su vez se pueden agrupar en períodos menores según criterios histórico-lingüísticos. Las figuras 2 y 3 muestran la división genérica, mientras que la figura 4 recoge la división cronológica. En la tabla 2 se resumen los tipos de géneros y distribución por palabras de los mismos.

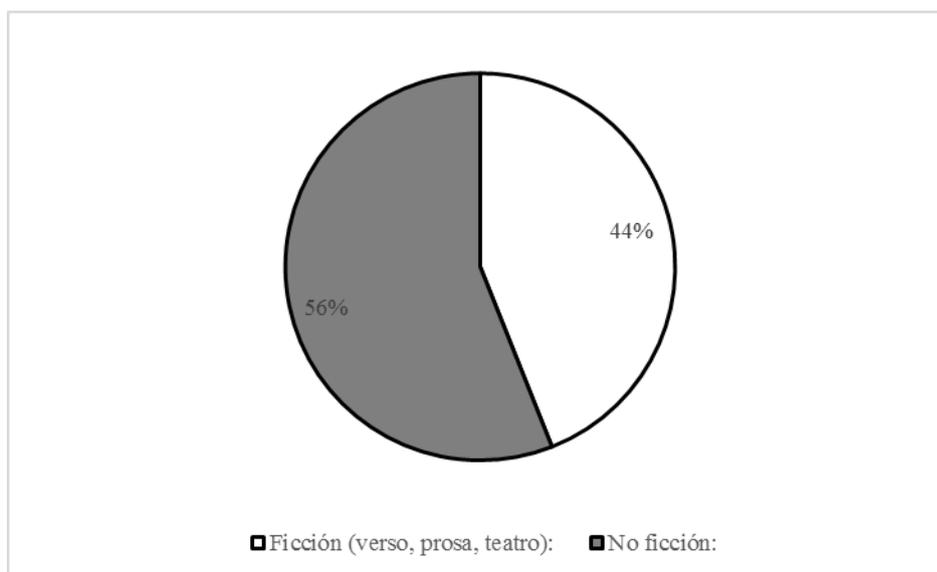


Figura 2: División genérica del CORDE

<sup>14</sup> Toda la información sobre el CORDE ha sido extraída de la web oficial de la Real Academia Española (<http://www.rae.es/recursos/banco-de-datos/corde>), de la sección de ayuda del CORDE ([http://corpus.rae.es/ayuda\\_c.htm](http://corpus.rae.es/ayuda_c.htm)) (fecha de acceso 14/03/2016) y de Rojo (en prensa).

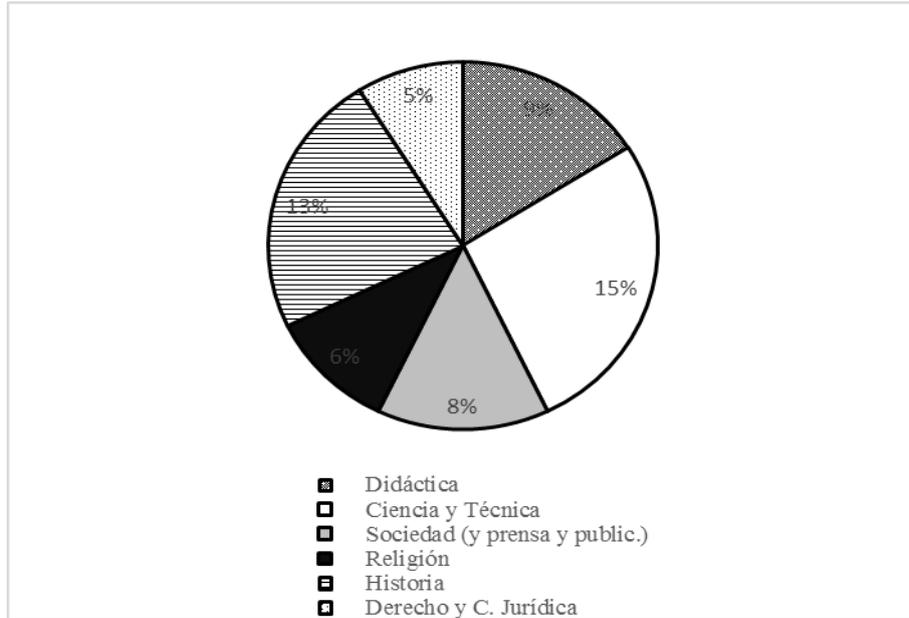


Figura 3: División de los textos no ficticios del CORDE

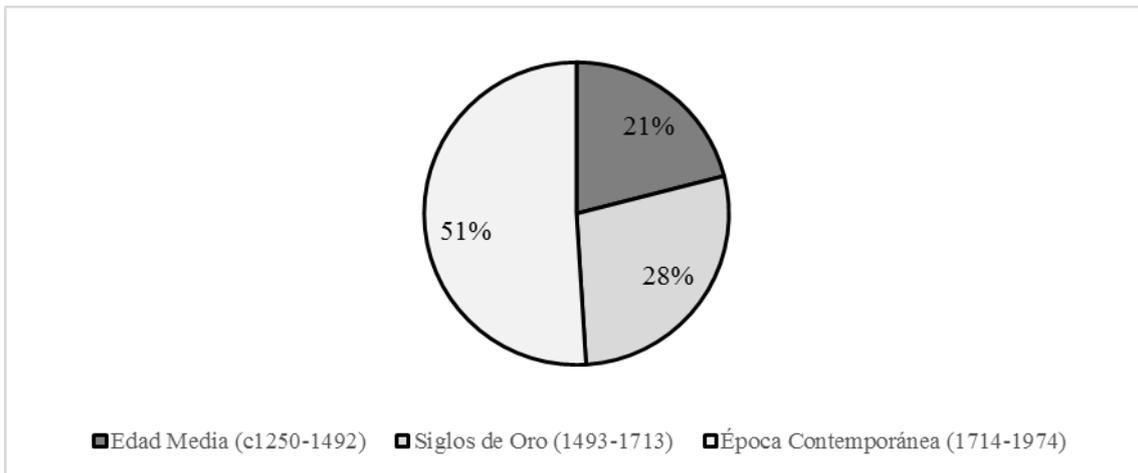


Figura 4: Distribución cronológica de los textos del CORDE

Géneros	Edad Media	Siglo de Oro	Contemporáneo	Total	%
Verso lírico	2.000.000	2.500.000	3.000.000	7.500.000	6%
Verso épico	3.250.000	1.000.000	750.000	5.000.000	4%
Verso dramático	750.000	2.812.500	1.625.000	5.187.500	4,15%
Prosa lírica	-	-	500.000	500.000	0,4%
Prosa narrativa	3.125.000	8.750.000	19.562.500	31.437.500	25,15%
Prosa dramática	250.000	1.250.000	3.375.000	4.875.000	3,9%
Prosa didáctica	2.125.000	2.625.000	6.625.000	11.375.000	9,1%

<b>Prosa científica</b>	4.200.000	4.637.500	9.537.500	18.375.000	14,7%
<b>Prosa de sociedad</b>	2.250.000.	2.875.000.	5.000.000	10.125.000	8,1%
<b>Periódicos</b>	-	375.000	3.000.000	3.375.000	2,7%
<b>Prosa religiosa</b>	2.500.000	2.250.000	2.875.000	7.625.000	6,1%
<b>Prosa histórica</b>	3.500.000	4.625.000	8.125.000	16.250.000	1,3%
<b>Prosa jurídica</b>	2.000.000	1.750.000	3.000.000	6.750.000	5,4%
<b>TOTALES</b>	<b>25.950.000</b>	<b>35.075.000</b>	<b>63.975.000</b>	<b>125.000.000</b>	<b>100%</b>

Tabla 2: Tipos de géneros y distribución de número de palabras por períodos en el CORDE

En cuanto a los tipos de textos, el CORDE destaca sobre los proyectos del HSMS y ADMYTE por la inclusión de diversas colecciones de cartas. Una simple búsqueda por "Nómina de autores y obras" permite conocer que bajo la categoría "cartas y relaciones" existen más de 4.000 documentos que suman casi seis millones de palabras. La principal desventaja es que el motor de búsqueda del CORDE y la organización de los metadatos del mismo no permiten distinguir si se trata de cartas privadas (más próximas a la lengua hablada; vid. Sección 2) o de otro tipo. Por otra parte, la gran mayoría de estas cartas, sino todas, fueron escritas por personajes letrados y eruditos de su época: políticos, nobles, eclesiásticos o escritores como Manuel Azaña, Sor María de Jesús o Miguel de Unamuno, entre muchos otros. Sin embargo, la voz del pueblo llano no está representada en estas cartas. Algo similar ocurre con la categoría "memorias y diarios", en la que cabría esperar documentos de índole personal caracterizados por un lenguaje íntimo y diseñados para no ser publicados, pero que incluye cartas (por ejemplo, de Colón a la Reina o a Fray Gaspar Gorricio) y memorias de viajes, entre otros documentos.<sup>15</sup> La prosa dramática breve y extensa, esta última incluyendo las categorías comedia, drama, tragedia y otros, sí están bien representadas en el CORDE (más de 34.000 documentos que sobrepasan los dos millones de palabras) y contienen abundantes diálogos, especialmente durante los Siglos de Oro. Por último, y también relacionada con la lengua oral, el CORDE incluye una categoría dentro del grupo de prosa narrativa extensa que se denomina "diálogo y miscelánea" que recoge 48 documentos que ascienden a más de 2,5 millones de palabras. La tabla 3 muestra las categorías del CORDE que recogen testimonio de la lengua oral y el número de palabras exacto de las mismas, obtenidas a través de búsqueda por "Nómina de autores y obras".

<b>Textos representativos de la lengua oral en el CORDE</b>	<b>Número de palabras</b>
Cartas y relaciones	5.776.202
Memorias y diarios	3.697.600
Prosa dramática breve y	2.391.914

<sup>15</sup> Sobre la importancia de la distinción entre diarios personales (*diaries*) y diarios de viajes, guerras o memorias (*journals*), vid. Yáñez-Bouza (2015).

extensa	
Prosa narrativa: diálogo y miscelánea	2.529.927
<b>Total</b>	<b>14.395.643</b>

Tabla 3: Textos representativos de la lengua oral en el CORDE y número total de palabras

La tabla 3 hace evidente que, si bien es cierto que existe una representación de la lengua oral del pasado en el CORDE, su proporción con respecto a los textos de tipo escrito y formal es ínfima, puesto que estos textos suponen un escaso 6% del total de número de palabras de este corpus. Por otra parte, como se ha mencionado anteriormente, los autores de estos textos son principalmente hombres eruditos y letrados, mientras que la voz del pueblo llano no se ve reflejada en ellos.

### 3.4. Corpus del Español<sup>16</sup>

En la actualidad el CORDE es fuente obligada para cualquier estudio diacrónico relacionado con la lengua española, aunque en los últimos años ha visto la luz su gran competidor el *Corpus del Español* (Davies, 2002a), un corpus de 100 millones de palabras (20 millones entre 1200-1400 y 40 millones para cada uno de los períodos entre 1500-1700 y 1800-1900) distribuidas en más de 20.000 textos del español de los siglos XIII al XX, compilado bajo la dirección de Mark Davies de la Brigham Young University. Los textos incluidos en el *Corpus de Español* provienen de distintas fuentes, incluyendo ADMYTE, el HSMS, la Biblioteca Virtual,<sup>17</sup> Comedia,<sup>18</sup> el Proyecto filosofía en español<sup>19</sup> y varias fuentes del español moderno: literatura (novelas, cuentos, obras dramáticas), textos orales (Habla Culta, el Corpus Oral,<sup>20</sup> transcripciones de congresos y entrevistas periodísticas), y miscelánea (enciclopedias, periódicos, etc.) (Davies, 2002b: 22). En la actualidad, se está trabajando en la ampliación de este corpus que, una vez finalizado, multiplicará su número de palabras por 100, ampliando la fecha de los textos más recientes hasta la época actual (tres o cuatro últimos años). El corpus, que ha sido anotado y lematizado, se presenta como una herramienta online con una interfaz que permite distintas posibilidades de búsqueda. Sin embargo, no ofrece hasta ahora una relación completa de los textos incluidos en él por lo que, aunque es posible saber que los diálogos de las obras teatrales están representados en el corpus, la herramienta de búsqueda actual no permite obtener un listado de estos textos como sí es posible en el CORDE.

### 3.5. Corpus del Nuevo Diccionario Histórico del Español (CDH)<sup>21</sup>

El *Corpus del Nuevo Diccionario Histórico del Español* reúne más de 350 millones de registros de textos procedentes de otras bases de datos como el HSMS, el CORDE y ADMYTE, por lo que aporta poco al plano de la oralidad

<sup>16</sup> Información extraída de la web oficial del *Corpus del Español*: <http://www.corpusdelespanol.org/x.asp> (fecha de acceso 14/03/2016).

<sup>17</sup> [www.cervantesvirtual.com](http://www.cervantesvirtual.com)

<sup>18</sup> [www.coh.arizona.edu/spanish/comedia/escomedi.html](http://www.coh.arizona.edu/spanish/comedia/escomedi.html)

<sup>19</sup> [www.filosofia.org](http://www.filosofia.org)

<sup>20</sup> [elvira.llf.uam.es/docs\\_es/corpus/corpus.html](http://elvira.llf.uam.es/docs_es/corpus/corpus.html)

<sup>21</sup> <http://www.rae.es/recursos/banco-de-datos/cdh>

diacrónica que nos atañe en este trabajo. Destaca, sin embargo, por la inclusión de lematización y anotación morfológica en la interfaz de búsqueda, algo que es posible en el *Corpus de Español*, pero no en CORDE.

### **3.6. Corpus Biblia Medieval**

El proyecto *Biblia Medieval* (Enrique-Arias, 2012), un corpus paralelo de casi seis millones y medio de palabras de uso libre en la red que contiene transcripciones paleográficas de los manuscritos de las versiones bíblicas medievales en castellano junto a su fuente hebrea o latina.<sup>22</sup> La gran ventaja de este corpus sobre otros es que permite cotejar varias versiones de un mismo texto, evitando así una visión sesgada de los fenómenos que se estudien en él. Sin embargo, los datos de *Biblia Medieval* aportan más bien poco al plano de la oralidad pues, como es bien sabido, el lenguaje bíblico se caracteriza por su formalidad, carácter prescriptivo y por la inclusión de arcaísmos poco frecuentes en la lengua oral.

### **3.7. Corpus de Documentos Españoles anteriores a 1700 (CODEA)<sup>23</sup>**

El grupo de Investigación de Textos para la Historia del Español (GITHE) de la Universidad de Alcalá se ha encargado desde hace unos años de la elaboración del *Corpus de Documentos Españoles anteriores a 1700* (CODEA 2011; Sánchez Prieto et al., 2009), que contiene más de 1.500 documentos transcritos con criterios filológicos (transcripción paleográfica, presentación crítica y facsímil) procedentes de diversos archivos de las provincias de España. El corpus recoge documentos legislativos, cartas de compraventa y contratos, actas y declaraciones, cartas privadas, testamentos e inventarios, informes y relaciones, estatutos, certificaciones y notas y breves, entre otros, dando así cobertura a un hueco existente en la documentación diacrónica del español. La distribución diacrónica por número de palabras del corpus se muestra en la tabla 4.

<b>Siglo</b>	<b>Número de palabras</b>	<b>%</b>
XI	0	0%
XII	1.241	0,12%
XII	205.548	19,94%
XIV	225.978	21,93%
XV	254.536	24,70%
XVI	235.580	22,86%
XVII	70.579	6,85%
-	37.187	3,61%
<b>Total</b>	<b>1.030.649</b>	<b>100%</b>

Tabla 4: Distribución diacrónica por número de palabras del CODEA (2011)

<sup>22</sup> Información extraída en la página web del proyecto *Biblia medieval* en <http://www.bibliamedieval.es/> (fecha de acceso 16/03/2016).

<sup>23</sup> Esta información ha sido tomada principalmente de la página web del CODEA: <http://www.textoshispanicos.es/> (fecha de acceso 14/03/2016).

En la actualidad el equipo de investigación está trabajando en la elaboración del CODEA+ 2015, que pretende ampliar de forma significativa el número de documentos original (tanto en lo cronológico, como en lo geográfico y diastrático) que lo convierta en un corpus más representativo, así como incorporar una interfaz de navegación y búsquedas que faciliten el análisis de los textos. Entre los avances más significativos se encuentra el de ampliar el límite temporal a 1800, así como la incorporación de piezas de las comunidades bilingües (Galicia, Cataluña, Valencia y País Vasco), que habían sido excluidas de la versión inicial. Otra novedad del CODEA+ 2015 será la ampliación del espectro social mediante la inclusión de expedientes de la inquisición y epistolarios femeninos, entre otros. La intención del grupo de investigación de la Universidad de Alcalá es que CODEA+ 2015 sea un corpus primario de acceso libre e inmediato, gratuito, avalado por diversas instituciones y fuente obligada para estudios lingüísticos o de otros ámbitos.

### **3.8. Post-Scriptum**

*Post Scriptum* (Vaamonde, 2015) es un proyecto en curso que comprende dos corpus de cartas privadas, uno para el español y otro para el portugués, junto con su edición crítica digital. El material epistolar abarca unas 3.500 cartas (un millón de palabras para cada lengua) producidas entre el siglo XVI hasta el primer tercio del siglo XIX por gente de diversa procedencia social y, por tanto, suelen reflejar una retórica próxima a la oralidad, contrastando así con el tipo de cartas que predomina en el CORDE (vid. Sección 3.3).

### **3.9. Corpus Diacrónico del Español del Reino de Granada (CORDEREGRA)<sup>24</sup>**

El *Corpus Diacrónico del Español del Reino de Granada* (Calderón Campos y García-Godoy 2015) difiere de los anteriores en que recoge una serie de documentos inéditos provenientes de esta antigua demarcación administrativa que comprendía las provincias de Granada, Málaga y Almería, y que fueron producidos entre 1492 y 1833. El corpus supone una aportación novedosa en cuanto a que es representativo de la variedad lingüística de esta región cuya castellanización coincide con la de Hispanoamérica. Los textos del corpus se enmarcan dentro de dos principales tipos: declaraciones de testigos e inventarios de bienes.<sup>25</sup> La inclusión de este tipo de documentos en el CORDEREGRA desmarca a este corpus de los anteriores por recoger textos de un carácter más informal. Tal como se destaca en la web oficial del CORDEREGRA, "[l]as declaraciones de testigos son interesantes por su carácter espontáneo, próximo en ocasiones a lo coloquial, [mientras que] los inventarios de bienes [...] ofrecen datos léxicos de enorme valor para el estudio de la vida cotidiana en esta región andaluza".

### **3.10. Corpus Histórico del Español de México (CHEM)**

En el ámbito hispanoamericano destaca el *Corpus Histórico del Español de México* (CHEM), puesto en marcha a partir del 2005. El CHEM está constituido por una colección de documentos diacrónicos que representan diversos

---

<sup>24</sup> Esta información ha sido extraída de la web del corpus: <http://www.corderegra.es/> (fecha de acceso 14/03/2016).

<sup>25</sup> La antología completa de textos del CORDEREGRA puede consultarse en Calderón Campos (2015).

géneros textuales tratando de privilegiar el carácter coloquial de los textos. Así, el corpus abarca cartas privadas de diversos tipos (de emigrantes, misioneros, al Rey, al Consejo, etc.), denuncias y testimonios en juicios, inventarios y testamentos y peticiones e informes (Contreras Seitz, 2009).

### **3.11. Corpus Diacrónico y Diatópico del Español de América (CORDIAM)<sup>26</sup>**

Por otra parte, el *Corpus Diacrónico y Diatópico del Español de América* (CORDIAM) es un recurso electrónico de libre acceso que reúne más de tres mil documentos escritos en español y en América entre los años 1494 y 1905. El CORDIAM contiene textos no literarios y no periodísticos de una amplia variedad discursiva que pueden dividirse en cuatro grandes grupos: cronísticos, administrativos, jurídicos y privados (cartas y otros). Los documentos cronísticos reflejan descripciones de paisajes y relatan sucesos, describen acciones, creencias, costumbres y comportamientos de grupos humanos. Su formato es próximo al texto literario en tanto a que tienen una ordenación temporal, aunque, al contrario del texto literario, éstos no poseen una finalidad estética. Los documentos administrativos, por otra parte, contienen facetas de la vida cotidiana (bienes materiales de vivos y difuntos, testimonio de la genealogía de los individuos, etc.) y registran y regulan la interacción no privada entre personas. Los documentos jurídicos son textos complejos y heterogéneos que contienen denuncias, querellas, postulación de preguntas, interrogatorios y respuestas, sentencias, segundas instancias de juicios, traslados, etc. Su característica más destacada es su carácter altamente dialógico, que los sitúa próximos a la oralidad. Por último, los documentos privados (cartas, notas, recados, etc.) son producto de la comunicación entre personas particulares que contienen temas personales, de carácter íntimo y afectivo. En definitiva, el CORDIAM destaca sobre los anteriores corpus por el hecho de incluir textos procedentes de una gran diversidad geográfica dentro del continente americano, muchos de los cuales se caracterizan por su inmediatez comunicativa.

### **3.12. Corpus Diacrónico del Español de Chile (CorDECh)**

Por último, el *Corpus Diacrónico del Español de Chile* (CorDECh) es un proyecto en desarrollo que Contreras Seitz describe como un "corpus monolingüe, grande, diacrónico, textual y simple" (Contreras Seitz, 2009: 115) que contendrá textos públicos y privados publicados entre 1548 y 1799. Al tratarse de un corpus en desarrollo, los datos disponibles del mismo son escasos más allá de lo que describe Contreras Seitz (2009).

## **4. Nuevos horizontes en los corpus diacrónicos del español: perspectivas desde el modelo anglosajón**

Como apuntamos en la sección anterior, en los últimos años han proliferado numerosos recursos electrónicos (algunos de ellos de acceso libre) que han optado por incluir entre sus muestras textuales documentos que se adscriben a registros informales y que podrían describirse como más cercanos a la lengua hablada. Sin embargo, todavía queda mucho por hacer, especialmente en el ámbito del español peninsular, pues no existe hasta la fecha un corpus diacrónico del español que incluya exclusivamente representaciones de textos

---

<sup>26</sup> Información extraída de la web: <http://www.cordiam.org/> (fecha de acceso 14/03/2016).

hablados. A continuación, se describen tres corpus del ámbito anglófono que se podrían tener en cuenta para la elaboración de herramientas similares en español: *Corpus of Early English Correspondence* (Nevalainen et al., 1998), el *Corpus of English Dialogues 1560-1760* (Kytö y Culpeper, 2006; vid. también Culpeper y Kytö, 1997; Kytö y Walker, 2006) y el *Old Bailey Corpus* (Huber et al., 2012).

El **Corpus of Early English Correspondence** (CEEC)<sup>27</sup> se completó en 1998, aunque hoy en día se usa esta denominación para referirse a una completa familia de corpus (con adición de etiquetados y ampliaciones de la versión original) que suman más de cinco millones de palabras contenidas en cartas personales recopiladas desde 1410 hasta 1800 y escritas por hombres y mujeres de diversa procedencia social. La tabla 5 proporciona datos de los distintos corpus de esta familia.

Corpus	Período	Número de palabras	Cartas	Escritores	Colecciones	Año de publicación
CEEC	1410?-1681	2.7 mill.	6039	778	96	--
CEECs	1418-1680	0.45 mill.	1147	194	23	1998
PCEEC	1410?-1681	2.2 mill.	4979	657	84	2006
CEECE	1681-1800	c. 2.2 mill.	c. 4900	> 300	74	--
CEECsU	1402-1663	c. 0.44 mill.	c. 900	> 100	20	--

Tabla 5: Datos de los corpus recogidos bajo la denominación de *Corpus of Early English Correspondence*

El CEEC se compiló con la intención de facilitar el análisis sociolingüístico en la historia del inglés y hasta la fecha ha proporcionado datos relevantes sobre la relación del cambio lingüístico y las variables sociales, reflejados en los numerosos trabajos publicados, entre los que destacan Nurmi (1999), Palander-Collin (1999), Nevalainen y Raumolin-Brunberg (2003), Nevala (2004) y Laitinen (2007).

Por su parte, el **Corpus of English Dialogues 1560-1760** (CED) reúne más de un millón de palabras distribuidas entre 177 archivos que contienen textos relacionados con la lengua oral.<sup>28</sup> El CED se compiló como una herramienta para el estudio de la lengua del período Inglés Moderno Temprano con especial atención a los diálogos, dada la relevancia de la interacción cara a cara como factor determinante en el cambio lingüístico. Además del tipo de texto, el corpus contiene información detallada de otras variables como la localización temporal, el sexo de los autores y el estatus social. Los tipos de texto incluidos en el CED así como el número de palabras se resumen en la

<sup>27</sup> Esta información ha sido obtenida de la web oficial del CEEC: <http://www.helsinki.fi/varieng/domains/CEEC.html>.

<sup>28</sup> Esta información ha sido extraída de la web oficial del CED: [http://www.engelska.uu.se/Research/English\\_Language/Research\\_Areas/Electronic\\_Resource\\_Projects/A\\_Corpus\\_of\\_English\\_Dialogues/](http://www.engelska.uu.se/Research/English_Language/Research_Areas/Electronic_Resource_Projects/A_Corpus_of_English_Dialogues/).

tabla 6, que también recoge el grado de intervención narrativa en los varios tipos de textos.

	<b>Diálogo auténtico</b>	<b>Diálogo construido</b>
<b>Mínima intervención narrativa</b>	Procesos judiciales 285.660 palabras	Comedias teatrales 238.590 palabras
		Trabajos didácticos Otros 162.250 palabras Enseñanza de idiomas 74.390 palabras
		Miscelánea 25.970 palabras
<b>Considerable intervención narrativa</b>	Declaraciones de testigos  172.940 palabras	Ficción  223.890 palabras
<b>Número total de palabras</b>	<b>172.940</b>	<b>725.090</b>

Tabla 6: Tipos de texto y número total de palabras del CED

El CED es parte de un proyecto de investigación titulado "Exploring spoken interaction of the Early Modern English period (1560-1760)" (vid., entre otros, Culpeper y Kytö, 1997, 2000, 2010). Por su configuración en torno a la interacción oral, el CED es una herramienta de gran valor para los estudios de estructura y/o cortesía conversacional, algo que hasta hace poco no se podía imaginar en estudios de lingüística histórica.

Por último, el **Old Bailey Corpus** (OBC)<sup>29</sup> incluye catorce millones de palabras extraídas de los procesos judiciales del Old Bailey, la corte criminal central de Londres, entre los años 1720 y 1913. Los procesos judiciales del OBC destacan por recoger las palabras literales de los participantes en los juicios mediante narraciones en primera persona lo que, según Huber (2007), los convierte en los documentos más cercanos a la lengua oral de este período. El análisis de Huber (2007) demuestra que, aunque los procesos judiciales no pueden describirse como totalmente fieles a la lengua hablada, "they are still among the few and best sources we have of spoken language before the advent of mechanical recording" (Huber 2007). Además, el OBC posee datos biográficos de los participantes (sexo, edad, profesión, estatus social), información pragmática (el rol de los participantes en los juicios: acusado, juez, testigo, víctima, etc.) e información textual (datos del escriba, del impresor y de la entidad de publicación), lo que lo convierte en una herramienta única para estudios de variación lingüística y sociolingüística, entre muchos otros.

En definitiva, el CEEC, el CED y el OBC son ejemplos de corpus en lengua inglesa que podrían imitarse en el ámbito de la lengua española, lo que abriría nuevas vías de investigación en el cambio lingüístico desde la perspectiva diacrónica que, como se ha mencionado en diversas ocasiones a lo largo de

<sup>29</sup> Disponible en: <http://www1.uni-giessen.de/oldbaileycorpus/>.

este trabajo, ocurre en primera instancia en la lengua oral. Ahora bien, para poder elaborar materiales de estas características para estudiar la diacronía del español, primero habrá que saber si existen y están disponibles estos textos para nuestra lengua. Seguramente sí, dada la rica tradición escrita y hablada del español, que ya en el siglo XVI se había convertido en la lengua corriente de gran parte de la población, un fenómeno que Oesterreicher describe como "singular y llamativo" (2004: 730) en comparación con otras lenguas romances. Sin embargo, la localización y procesamiento de tales documentos conlleva un duro esfuerzo humano y de recursos. La elaboración de cualquier corpus lingüístico no es una tarea fácil, rápida ni barata, aunque todo esfuerzo se ve recompensado por la utilidad y beneficio que conlleva poder disponer de ellos.

## 5. Conclusiones

La lingüística de corpus ha vivido una transformación en las últimas décadas en parte gracias al advenimiento de las nuevas tecnologías que en los últimos cuarenta años ha propiciado la creación y proliferación de numerosos corpus lingüísticos. Aunque los progresos han sido muchos, en el ámbito de la diacronía de la lengua española todavía queda mucho por avanzar, especialmente en lo que se refiere a la lengua hablada de nuestros ancestros. En este trabajo hemos dado una visión general de los principales corpus disponibles hasta la fecha actual para el estudio de la historia del español, destacando la escasez de documentos en los mismos que puedan describirse como más próximos a la lengua hablada, tales como las obras teatrales, las declaraciones de testigos, los procesos judiciales, las cartas personales o los diálogos ficticios, entre muchos otros. Además, cuando la lengua oral está representada en los corpus, suele ser de mano de autores cultos y letrados, mientras que la voz de los estratos inferiores y especialmente la de las mujeres no suele tener cabida. Especialmente en el ámbito del español peninsular, no existe ningún corpus del español equiparable a otros del mundo anglófono, como el CEEC, el CED o el OBC. El presente trabajo no pretende ser una crítica, sino una aportación constructiva que promueva la creación de herramientas similares en el ámbito de la lengua española, un idioma con una tradición escrita de un valor incalculable que debe ser aprovechada para enriquecer nuestros conocimientos sobre la lengua de nuestros antepasados.

## Bibliografía

- ANDERSON, H. & I. EHRENPREIS (1966): "The familiar letter in the eighteenth century: Some generalizations". *The familiar letter in the eighteenth century*. H. ANDERSON, P. B. DAGHLIAN & I. EHRENPREIS (eds.). Lawrence: University of Kansas Press, pp. 269-282.
- AUSTIN, F. (1998): "Epistolary conventions in *The Clift Family Correspondence*". *A reader in Early Modern English*. M. RYDÉN, I. TIEKEN-BOON VAN OSTADE & M. KYTÖ (eds.). Frankfurt am Main: Peter Lang, pp. 319-347.
- ARCHER = *A Representative Corpus of Historical English Registers* 3.1. 1990-1993/2002/2007/2010. Compilado bajo la supervisión de Douglas Biber y Edward Finegan (Northern Arizona University and University of Southern California). Modificado y expandido por el consorcio de universidades

- que comprende las siguientes: Northern Arizona, Southern California, Freiburg, Heidelberg, Helsinki, Uppsala, Michigan, Manchester, Lancaster, Bamberg, Zurich, Trier, Salford y Santiago de Compostela. Web oficial: <http://www.alc.manchester.ac.uk/subjects/lel/research/projects/archer/>
- BIBER, D. (1988): *Variation across speech and writing*. Cambridge: Cambridge University Press.
- BIBER, D. (2001): "Dimensions of variation among eighteenth-century speech-based and written registers". *Towards a history of English as a history of genres*. H. DILLER, & M. GÖRLACH (eds.). Heidelberg: Universitätsverlag C. Winter, pp. 89-109.
- BIBER, D. & E. FINEGAN (1988): "Drift in three English genres from the eighteenth to the twentieth centuries: A multidimensional approach". *Corpus linguistics, hard and soft. Proceedings of the 8th International Conference on English Language Research on Computerized Corpora*. K. MERJA, O. IHALAINEN & M. RISSANEN (eds.). Amsterdam: Rodopi, pp. 83-101.
- (1989): "Drift and the evolution of English style: A history of three genres". *Language*, 65,3, pp. 487-517.
- (1992): "The linguistic evolution of five written and speech-based genres from the seventeenth to the twentieth centuries". *History of Englishes: New methods and interpretations in historical linguistics*. M. RISSANEN, O. IHALAINEN, T. NEVALAINEN & I. TAAVITSAINEN (eds.). Berlin y New York: Mouton de Gruyter, pp. 688-704.
- (1997): "Diachronic relations among speech-based and written registers in English". *To explain the present: Studies in the changing English language in honour of Matti Rissanen*. T. NEVALAINEN & L. KAHLAS-TARKKA, (eds.). Helsinki: Société Néophilologique, pp. 253-275.
- BRIZ, A. Y M. ALBELDA (2009): "Estado actual de los corpus de lengua española hablada y escrita: I+D". *El español en el mundo. Anuario del Instituto Cervantes 2009*. Madrid: Instituto Cervantes, pp. 165-226.
- BUSTOS TOVAR, J. J. de. (2001): "De la oralidad a la escritura en la transición de la Edad Media al Renacimiento: la textualización del diálogo conversacional". *Críticón*, 81- 82, pp. 191-206.
- CALDERÓN CAMPOS, M. (2015): *El español del reino de Granada en sus documentos (1492-1833). Oralidad y escritura*. Bern: Peter Lang.
- CALDERÓN CAMPOS, M. & M. T. GARCÍA-GODOY (2015): *Corpus Diacrónico del Español del Reino de Granada. 1492-1833*. En línea en <http://www.corderegra.es>.
- CANO AGUILAR, R. (1996): "Lenguaje 'espontáneo' y retórica epistolar en cartas de emigrantes españoles a Indias". *El español hablado y la cultura oral en España e Hispanoamérica*. T. KOTSCHI, W. OESTERREICHER & K. ZIMMERMANN (eds.). Madrid: Iberoamericana Vervuert, pp. 375-404.
- CONTRERAS SEITZ, M. (2009): "Hacia la construcción de un Corpus Diacrónico del Español del Chile". *RLA, Revista de Lingüística Teórica y Aplicada*, 47,2, pp. 111-134.
- CULPEPER, J. & M. KYTÖ (1997): "Towards a corpus of dialogues, 1550-1750". *Language in time and space. Studies in honour of Wolfgang Viereck on the occasion of his 60th birthday*. H. RAMISCH, & K. WYNNE (eds.). Stuttgart: Franz Steiner Verlag, pp. 60-73.

- (2000): "Data in historical pragmatics: Spoken interaction (re)cast as writing". *Journal of Historical Pragmatics*, 1,2, pp. 175-199.
- (2010): *Early Modern English dialogues. Spoken interaction as writing*. Cambridge: Cambridge University Press.
- DAVIES, M. (2002a): *Corpus del Español: 100 million words, 1200s-1900s*. En línea en <http://www.corpusdelespanol.org>.
- (2002b): "Un corpus anotado de 100.000.000 palabras del español histórico y moderno". *Procesamiento del Lenguaje Natural*, 29, pp. 21-27.
- ELPASS, S. (2012): The use of private letters and diaries in sociolinguistic investigation. *The handbook of historical sociolinguistics*. J. M. HERNÁNDEZ-CAMPOY & J. C. CONDE-SILVESTRE (eds.). Oxford: Wiley-Blackwell, pp. 157-168.
- ENRIQUE-ARIAS, A. (2009): "Introducción lingüística de corpus y diacronía de las lenguas iberorrománicas". *Diacronía de las lenguas iberorrománicas: nuevas aportaciones desde la lingüística de corpus*. A. ENRIQUE ARIAS (ed.). Madrid y Frankfurt: Iberoamericana Vervuert, pp. 11-24.
- (2012): "Biblia medieval: diseño y aplicaciones de un corpus paralelo y alineado del español medieval". *Actas del VIII Congreso Internacional de Historia de la Lengua española: Santiago de Compostela, 14-18 de septiembre de 2009. Vol. I*. E. MONTERO CARTELLE Y C. MANZANO ROVIRA (eds.). Santiago de Compostela: Universidad de Santiago de Compostela, pp. 421-432.
- FIRTH, J. R. (1951): *Papers in linguistics, 1934-1951*. Oxford: Oxford University Press.
- FRIES, C. C. (1952): *The structure of English: An introduction to the construction of English sentences*. New York: Harcourt Brace.
- GARCÍA-SALIDO, M. & V. VÁZQUEZ-ROZAS (2012): "Los corpus diacrónicos como instrumento para el estudio del origen y distribución de la concordancia de objeto en español". *Scriptum Digital*, 1, pp. 67-84.
- GRUND, P. (2007): "From tongue to text: The transmission of the Salem witchcraft examination records". *American Speech*, 82, 2, pp. 119-150.
- HUBER, M. (2007): "The Old Bailey Proceedings, 1764-1834. Evaluating and annotating a corpus of 18th- and 19th-century spoken English". *Annotating variation and change (Studies in Variation, Contacts and Change in English 1)*. A. MEURMAN-SOLIN & A. NURMI (eds.). Helsinki: University of Helsinki. Disponible en línea: <http://www.helsinki.fi/varieng/series/volumes/01/huber/> (fecha de acceso 17/03/2016).
- HUBER, M., M. NISSEL, P. MAIWALD & B. WIDLITZKI (2012): *The Old Bailey Corpus. Spoken English in the 18th and 19th centuries*. Disponible en línea: [www.uni-giessen.de/oldbaileycorpus](http://www.uni-giessen.de/oldbaileycorpus).
- KÄDING, F. W. (1897): *Häufigkeitwörterbuch der deutschen Sprache*. Festgestellt durch einen Arbeitsausschuß der deutschen Stenographie-Systeme. Steglitz: no public.
- KOCH, P. (1999): "Court records and cartoons: Reflections of spontaneous dialogue in early Romance texts". *Historical dialogue analysis*. A. JUCKER, G. FRITZ Y F. LEBSANFT (eds.). Amsterdam y Philadelphia: John Benjamins, pp. 399-429.
- 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.
- (1990): *Gesprochene Sprache in der Romania: Französisch, Italienisch, Spanisch*. Tübinga: Niemeyer.
- KYTÖ, M. Y CULPEPER, J. (2006): *Corpus of English Dialogues 1560–1760*.
- KYTÖ, M. Y WALKER, T. (2006): "Guide to A *Corpus of English Dialogues 1560-1760*". *Studia Anglistica Upsaliensia /Acta Universitatis Upsaliensis*, 130.
- LAITINEN, M. (2007): *Agreement patterns in English: Diachronic corpus studies on common-number pronouns*. (Mémoires de la Société Néophilologique, 71). Helsinki: Société Néophilologique.
- LAVOB, W. (1994): *Principles of linguistic change: Internal factors*. Oxford: Blackwell.
- LEECH, G. (1992): "Corpora and theories of linguistic performance". *Directions in corpus linguistics. Proceedings of Nobel Symposium 82, Stockholm, 4-8 August 1991*. J. SVARTVIK (ed.). Berlín: De Gruyter, pp. 105-134.
- LÓPEZ SERENA, A. (2007): *Oralidad y escrituralidad en la recreación literaria del español coloquial*. Madrid: Gredos.
- MARCOS MARÍN, F. A. (1994): *Informática y humanidades*. Madrid: Gredos.
- NARBONA JIMÉNEZ, A. (1989): *Sintaxis española: nuevos y viejos enfoques*. Barcelona: Ariel.
- NEVALA, M. (2004): *Address in Early English correspondence. Its forms and socio-pragmatic functions*. (Mémoires de la Société Néophilologique de Helsinki 64). Helsinki: Société Néophilologique.
- NEVALAINEN, T. (2001): "Continental conventions in Early English correspondence". *Towards a history of English as a history of genres*. En H. Diller & M. Görlach (eds.). Heidelberg: Universitätsverlag C. Winter, pp. 203-225.
- NEVALAINEN, T. & H. RAUMOLIN-BRUNBERG (2003): *Historical sociolinguistics: Language change in Tudor and Stuart England*. London: Pearson Education.
- NEVALAINEN, T., H. RAUMOLIN-BRUNBERG, J. KERÄNEN, M. NEVALA, A. NURMI & M. PALANDER-COLLIN (1998): *Corpus of Early English Correspondence*. Department of Modern Languages, University of Helsinki.
- NURMI, A. (1999): *A social history of periphrastic DO*. (Mémoires de la Société Néophilologique de Helsinki, 56). Helsinki: Société Néophilologique.
- OESTERREICHER, W. (1996): "Lo hablado en lo escrito". *El español hablado y la cultura oral en España e Hispanoamérica*. T. KOTSCHI, W. OESTERREICHER & K. ZIMMERMANN (eds.). Madrid: Iberoamericana Vervuert, pp. 317-340.
- (2004): *Textos entre inmediatez y distancia comunicativas. El problema de lo hablado escrito en el Siglo de Oro*. *Historia de la lengua española*. R. CANO (ed.). Barcelona: Ariel, pp. 729-769.
- PALANDER-COLLIN, M. (1999): "Male and female styles in seventeenth-century correspondence: I THINK". *Language Variation and Change*, 11, pp. 123-141.
- REAL ACADEMIA ESPAÑOLA: Banco de datos (CORDE) [en línea]. *Corpus diacrónico del español*. <http://www.rae.es>
- REAL ACADEMIA ESPAÑOLA: Banco de datos (CREA) [en línea]. *Corpus de referencia del español actual*. <http://www.rae.es>

- RISSANEN, M. (1989): "Three problems connected with the use of diachronic corpora". *ICAME Journal*, 13, pp. 16-19.
- RISSANEN, M. (1999): "Syntax". *The Cambridge history of the English language*. R. LASS (ed.). Cambridge: Cambridge University Press, pp. 187-331.
- ROJO, G. (2015): Sobre los antecedentes de la lingüística de corpus. *Studium grammaticae. Homenaje al Profesor José Antonio Martínez*. A. ÁLVAREZ MENÉNDEZ et alii (eds.). Oviedo: Universidad de Oviedo, pp. 675-689.
- (2016): "Los corpus textuales del español". *Enciclopedia lingüística hispánica*. J. GUTIÉRREZ-REXACH (ed.). Oxon: Routledge, pp. 285-296.
- (en prensa): "Citius, maius, melius: del CREA al CORPES XXI". *Lingüística de corpus y lingüística histórica iberorrománica*. J. KABATEK (ed.). Berlín: de Gruyter.
- SÁNCHEZ-PRieto, B., F. PAREDES-GARCÍA, R. MARTÍNEZ-SÁNCHEZ, R. MIGUEL-FRANCO, M. SIMÓN-PARRA & I. VICENTE-MIGUEL (2009): "El Corpus de Documentos Españoles anteriores a 1700 (CODEA)". *Diacronía de las lenguas iberorrománicas: nuevas aportaciones desde la lingüística de corpus*. A. ENRIQUE ARIAS (ed.). Madrid y Frankfurt am Main: Iberoamericana, pp. 25-38.
- STOLL, E. (1996): "Competencia escrita de impronta oral en la crónica soldadesca de Pedro Pizarro". *El español hablado y la cultura oral en España e Hispanoamérica*. T. Kotschi, W. Oesterreicher & K. Zimmermann (eds.). Madrid: Iberoamericana Vervuert, pp. 427-446.
- TIEKEN-BOON VAN OSTADE, I. (2005): "Eighteenth-century English letters: In search of the vernacular". *Linguística e Filologia*, 21, pp. 113-146.
- VAAMONDE, G. (2015): "P.S: Post Scriptum: dos corpus diacrónicos de escritura cotidiana". *Procesamiento del Lenguaje Natural*, 55, pp.57-64.
- YÁÑEZ-BOUZA, N. (2015): "Have you ever written a diary or a journal? Diurnal prose and register variation". *Neuphilologische Mitteilungen*, 116,2, pp. 639-664.

### Apéndice I: Relación detallada de los textos incluidos en los corpus del Hispanic Seminary of Medieval Studies

<b>Obra en prosa de Alfonso X el sabio</b>		
<b>Fecha</b>	<b>Título</b>	<b>Biblioteca</b>
1250-1250	<i>Moamyn-Libro de las animalias</i>	Madrid: Nacional RES/270
1254 - 1254	<i>Judizios de las estrellas</i>	Madrid: Nacional MSS/3065
1256-1256	<i>Picatrix de Alfonso X</i>	Roma: Vaticana Reg. Lat. 1283
1259-1259	<i>Libro de las cruces</i>	Madrid: Nacional MSS/9294
1250-1279	<i>Lapidarios de Alfonso X</i>	Escorial: Monasterio h.I.15
1254-1260	<i>Tablas de Zarquiel</i>	Paris: Arsenal 8322
1254-1260	<i>Cánones de Albateni</i>	Paris: Arsenal 8322
1256-1265	<i>Libro de las leyes</i>	London: British Library Add. 20787
1270-1284	<i>Estoria de España I</i>	Escorial: Monasterio Y.I.2
1272-1275	<i>General Estoria I</i>	Madrid: Nacional MSS/816
1276-1279	<i>Libro de las formas y de las imágenes</i>	Escorial: Monasterio h.I.16
1277-1277	<i>Libros del saber de astronomía</i>	Madrid: Universidad Complutense BH MSS 156
1277-1277	<i>Libro del cuadrante señero</i>	Paris: Arsenal 8322
1280-1280	<i>General estoria IV</i>	Roma: Vaticana Urb. Lat. 539
1283-1283	<i>Libros de ajedrez, dados y tablas</i>	Escorial: Monasterio T.I.6
1284-1347	<i>Estoria de España II</i>	Escorial: Monasterio XI.I.4
1201-1400	<i>General estoria V</i>	Escorial: Monasterio I.I.2
1301-1400	<i>General estoria II</i>	Madrid: Nacional MSS/10237
1301-1400	<i>General estoria VI</i>	Toledo: Catedral 43-20
1401-1500	<i>General estoria V</i>	Escorial: Monasterio R.I.10
<b>Textos médicos españoles</b>		
<b>Fecha</b>	<b>Título</b>	<b>Biblioteca</b>
1250-1279	<i>Lapidarios de Alfonso X</i>	Escorial: Monasterio h.I.15
1381-1381	<i>Visita y consejo de médicos</i>	Madrid: Nacional MSS/18052
1401-1500	<i>Suma de la flor de cirugía</i>	Madrid: Nacional MSS/3338
1401-1500	<i>Macer herbolario</i>	Sevilla: Colombina 7-6-27
1401-1500	<i>De las melecinas</i>	Salamanca: Universitaria 1743
1401-1500	<i>Tesoro de la medicina</i>	Sevilla: Colombina 5-1-17
1401-1500	<i>Tratado de la patología general</i>	Madrid: Nacional MSS/10051
1440-1460	<i>Espejo de medicina</i>	Madrid: Nacional MSS/3384
1440-1460	<i>Cirugía</i>	Escorial: Monasterio h.III.17
1450-1500	<i>Tratado de las fiebres</i>	Escorial: Monasterio M.I.28
1471-1471	<i>El libro de recetas</i>	Madrid: Palacio II/3063
1471-1471	<i>Secretos de la medicina</i>	Madrid: Palacio II/3063
1481-1481	<i>Cirugía mayor</i>	Madrid: Nacional MSS/2147
1490-1500	<i>Menor daño de la medicina</i>	Escorial: Monasterio b.IV.34

1490-1500	<i>Compendio de la medicina</i>	Salamanca: Universitaria 2262
1493-1493	<i>Cirugía rimada</i>	Madrid: Nacional MSS/2153
1494-1494	<i>Compendio de la humana salud</i>	Madrid: Nacional INC/51
1494-1494	<i>Tratado de phisonomía</i>	Madrid: Nacional INC/51
1494-1494	<i>Tratado de la epidemia y pestilencia</i>	Madrid: Nacional INC/51
1495-1495	<i>Tratado de la generación de la criatura</i>	Madrid: Nacional INC/1335
1495-1495	<i>Lilio de medicina</i>	Madrid: Nacional INC/2438
1498-1498	<i>Tratado de cirugía</i>	Madrid: Nacional INC/196
1498-1498	<i>Sumario de la medicina</i>	Madrid: Nacional INC/1350
1500-1500	<i>Libro de los olios</i>	Salamanca: Universitaria 2262
1500-1500	<i>Recetas</i>	Salamanca: Universitaria 2262
1500-1500	<i>Propiedades del romero</i>	Salamanca: Universitaria 2262
1506-1506	<i>Tratado nuevo</i>	Toledo: Biblioteca de Castilla-La Mancha R/1010-2
1507-1507	<i>Tratado útil</i>	Toledo: Biblioteca de Castilla-La Mancha R/1010-4
1509-1509	<i>Cirugía</i>	Granada: Universitaria B-11
1515-1515	<i>Compendio de los boticarios</i>	New York: Hispanic Society of America
1516-1520	<i>Regimiento contra la peste</i>	Madrid: Nacional INC/2182(2)
1518-1518	<i>Libro de medecina llamado macer</i>	Madrid: Nacional R/9017
1519-1519	<i>Libro de medecina llamado macer</i>	Bethesda: National Library of Medicine WZ 240 L697
1541-1541	<i>Libro del arte de las comadres</i>	Madrid: Nacional R/1322
1542-1542	<i>Tractado contra el mal serpentino</i>	Madrid: Nacional R/13034
1545-1545	<i>Sevillana medicina</i>	Madrid: Nacional R/30652
1551-1551	<i>Libro de la anathomía del hombre</i>	Madrid: Nacional R/3398
1553-1553	<i>Libro del exercicio corporal</i>	Madrid: Nacional R/12270
1567-1567	<i>Secretos de chirurgia</i>	Madrid: Nacional R/4277
1570-1570	<i>Tratado breve y compendioso</i>	Madrid: Nacional R/5434
1580-1580	<i>Libro del parto humano</i>	Madrid: "Marqués de Valdecilla" BH MED 1930
1583-1583	<i>Tratado de la utilidad de la sangría</i>	Madrid: Nacional R/5735
1586-1586	<i>Tractado del uso de las mugeres</i>	Madrid: Complutense BH MED 1105
1589-1589	<i>Manual contra la gota</i>	Madrid: Nacional R/27991
1594-1594	<i>Discurso y despertador preservativo</i>	Madrid: Nacional R/432(2)
1594-1594	<i>Discurso particular preservativo</i>	Madrid: Nacional V/109/12
1598-1598	<i>Breve tratado de peste</i>	Madrid: Nacional R/4276
1598-1598	<i>Libro de experimentos médicos</i>	Madrid: "Marqués de Valdecilla" BH MED 1961
1598-1598	<i>Tratado de la peste</i>	Madrid: Nacional R/26692
1599-1599	<i>Instituciones para los algebristas</i>	Madrid: Nacional R/14133
1599-1599	<i>Orden para la cura y preservación</i>	Madrid: Nacional R/1879
1600-1600	<i>Libro de la enfermedad de las bubas</i>	Madrid: Nacional R/3612
1604-1604	<i>Defensa de las criaturas de tierna</i>	Madrid: Nacional R/51863

*En busca de lo hablado en lo escrito en los corpus diacrónicos del español: una comparativa con los corpus anglosajones*

	edad	
1651-1651	<i>Trágico suceso</i>	Madrid: Nacional MSS/62350
1674-1674	<i>Tratado de raras y peregrinas yervas</i>	Madrid: Nacional R/15538
<b>Textos navarro-aragoneses</b>		
<b>Fecha</b>	<b>Título</b>	<b>Biblioteca</b>
1260-1280	<i>Vidal Mayor</i>	Los Angeles: Getty Museum, Ludwig XIV 6
1301-1400	<i>Fuero de la Novenera</i>	Salamanca: Universitaria, 2652
1301-1400	<i>Libro de Apolonio</i>	Escorial: Monasterio K.III.4
1301-1400	<i>Santa María Egipcíaca</i>	Escorial: Monasterio K.III.4
1301-1400	<i>Tres Reyes d'Orient</i>	Escorial: Monasterio, K.III.4
1330-1330	<i>Fuero general de Navarra</i>	Madrid: Nacional MSS/248
1330-1400	<i>Fuero de Navarra (versión C)</i>	Salamanca: Universitaria, 1947
1340-1341	<i>Fuero de Navarra (versión A)</i>	Madrid: Nacional MSS/17653
1340-1360	<i>Fueros de Aragón</i>	Madrid: Nacional MSS/458
1357-1390	<i>Viajes de Juan de Mandevilla</i>	Escorial: Monasterio, M.III.7
1375-1400	<i>Vidas semblantes I</i>	Paris: Nationale, Esp. 70
1375-1400	<i>Vidas semblantes II</i>	Paris: Nationale, Esp. 71
1375-1400	<i>Vidas semblantes III</i>	Paris: Nationale, Esp. 72
1376-1396	<i>Crónica de los conquistadores I</i>	Madrid: Nacional MSS/2211
1376-1396	<i>Crónica de los conquistadores II</i>	Madrid: Nacional MSS/10134 2
1376-1396	<i>Breviarium ab urbe condita</i>	Paris: Arsenal, 8324
1376-1396	<i>Grant crónica de Espanya III</i>	Madrid: Nacional MSS/10134
1377-1396	<i>Flor de las ystorias de Orient</i>	Escorial: Monasterio, Z.I.2
1377-1393	<i>Crónica de Morea</i>	Madrid: Nacional MSS/10131
1377-1396	<i>Historia contra los paganos</i>	Valencia Universidad Pontificia
1377-1396	<i>Libro de actoridades ~ Ram de flors</i>	Escorial: Monasterio, Z.I.2
1377-1396	<i>Secreto Secretorum</i>	Escorial: Monasterio, Z.I.2
1380-1396	<i>Historia Troyana</i>	Madrid: Nacional MSS/10801
1384-1396	<i>Guerra del Peloponeso</i>	Madrid: Nacional MSS/10801
1385-1385	<i>Grant crónica de Espanya I</i>	Madrid: Nacional MSS/10133
1385-1385	<i>Crónica del tudense</i>	Stockholm: Kungliga, D 1272a
1385-1385	<i>Crónica de San Isidoro</i>	Stockholm: Kungliga, D 1272a
1385-1390	<i>Libro de Palladio</i>	Madrid: Nacional MSS/10211
1385-1396	<i>Historia de los godos de San Isidoro</i>	Stockholm: Kungliga, D 1272a
1390-1410	<i>Cuento de Tristán de Leonis</i>	Roma: Vaticana, 6428
1393-1393	<i>Crónica de los emperadores</i>	Escorial: Monasterio, 10131
1396-1396	<i>Libro de Marco Polo</i>	Escorial: Monasterio, Z.I.2
1400-1410	<i>Crónica general de España</i>	Escorial: Monasterio, X.II.22
1425-1450	<i>Livres dou trésor</i>	Gerona: Catedral, 20-a-5
1490-1490	<i>Traslación del doctor Chatón</i>	Escorial: Monasterio, 32-V-19 (30)
1500-1500	<i>Stratagemata</i>	Madrid: Nacional MSS/10198
<b>Textos legales españoles</b>		

<b>Fecha</b>	<b>Título</b>	<b>Biblioteca</b>
1223-1223	<i>Fuero Viejo de Alcalá</i>	Alcalá de Henares, AMA [H] F.V.A
1250-1350	<i>Fuero de Alba de Tormes</i>	Alba de Tormes: Archivo Municipal
1256-1265	<i>Libro de las leyes</i>	London: British Library Add. 20787
1260-1300	<i>Fuero Juzgo</i>	New York: Hispanic Society, B2567
1260-1280	<i>Vidal Mayor</i>	Los Angeles: Getty Museum, Ludwig XIV 6
1276-1300	<i>Fuero de Béjar</i>	Béjar: Archivo Municipal
1289-1289	<i>Fuero de Avilés</i>	Avilés: Archivo Municipal
1289-1289	<i>Fuero de Zamora</i>	Madrid: Real Academia Española, 293
1290-1310	<i>Fuero de Ledesma</i>	Ledesma: Archivo Municipal
1295-1295	<i>Fuero de Oviedo</i>	Oviedo: Archivo Municipal, t. 1 n. 1
1296-1300	<i>Fuero de Burgos</i>	Philadelphia: Free Library: Lewis E.-245
1300-1371	<i>Espéculo</i>	Madrid: Nacional, MSS/10123
1301-1400	<i>Fueros de Castiella</i>	Madrid: Nacional, MSS/431
1301-1400	<i>Leyes del estilo</i>	Escorial: Monasterio, Z.II.8
1301-1400	<i>Leyes del estilo</i>	Escorial: Monasterio, Z.III.11
1301-1400	<i>Leyes del estilo</i>	Madrid: Nacional, MSS/5764
1301-1400	<i>Fuero Real</i>	Escorial: Monasterio, Z.III.16
1301-1400	<i>Fuero de la Novenera</i>	Salamanca: Universitaria, 2652
1301-1400	<i>Fuero de Salamanca</i>	Salamanca: Archivo Municipal
1330-1330	<i>Fuero general de Navarra</i>	Madrid: Nacional, MSS/248
1330-1400	<i>Fuero de Navarra (versión C)</i>	Salamanca: Universitaria, 1947
1340-1341	<i>Fuero de Navarra (versión A)</i>	Madrid: Nacional, MSS/17653
1340-1360	<i>Fueros de Aragón</i>	Madrid: Nacional, MSS/458
1348-1351	<i>Ordenamiento de Alcalá</i>	Madrid: Nacional, RES/9
1350-1450	<i>Fuero de Briviesca</i>	Madrid: Nacional, MSS/9199
1351-1352	<i>Ordenamiento de Alcalá</i>	Madrid: Nacional, VITR/15/7
1401-1500	<i>Libro de las tahurerías</i>	Salamanca: Universitaria, 1962
1401-1500	<i>Libro de las tahurerías</i>	Salamanca: Universitaria, 1862
1400-1410	<i>Leyes del estilo</i>	Escorial: Monasterio, Z.II.14
1401-1500	<i>Libro de las tahurerías</i>	Escorial: Monasterio, Z.II.4
1401-1500	<i>Libro de las tahurerías</i>	Escorial: Monasterio, Z.II.5
1401-1600	<i>Libro de las tahurerías</i>	Escorial: Monasterio, Z.II.6
1400-1410	<i>Libro de las tahurerías</i>	Escorial: Monasterio, Z.II.14
1404-1500	<i>Libro de las tahurerías</i>	Escorial: Monasterio, Z.I.8
1404-1500	<i>Libro de las tahurerías</i>	Escorial: Monasterio, Z.I.9
1410-1410	<i>Libro de las tahurerías</i>	New York: Hispanic Society, B2193
1412-1500	<i>Libro de las tahurerías</i>	Escorial: Monasterio, Z.I.6
1425-1432	<i>Libro de las tahurerías</i>	Madrid: Nacional, MSS/23
1450-1550	<i>Libro de las tahurerías</i>	Madrid: Palacio, II-1963
1483-1483	<i>Libro de las tahurerías</i>	Madrid: Nacional, MSS/691 (2)
1484-1484	<i>Ordenanzas Reales</i>	Madrid: Nacional, INC/1338
1490-1500	<i>Libro de las tahurerías</i>	Valladolid: Santa Cruz, 139

1491-1491	<i>Siete Partidas</i>	New York: Hispanic Society
1497-1497	<i>Leyes del estilo</i>	San Marino: Huntington, 9566.6
1500-1500	<i>Leyes del estilo</i>	Oxford: Bodleian, Inc. d. S6. 1500.1
1501-1600	<i>Libro de las tahurerías</i>	Salamanca: Universitaria, 2056
1501-1600	<i>Libro de las tahurerías</i>	Madrid: Nacional, MSS/5784.
1505-1505	<i>Leyes de Toro</i>	Valladolid: Chancillería, PERGAMINOS, CAJA,1,1
1525-1525	<i>Libro de las tahurerías</i>	Madrid: Fundación Lázaro Galdiano, 439
<b>Textos bíblicos españoles</b>		
<b>Fecha</b>	<b>Título</b>	<b>Biblioteca</b>
1200-1210	<i>La Fazienda de Ultramar</i>	Salamanca: Biblioteca Universitaria 1997
1250-1250	<i>Biblia romanceada E8</i>	Escorial: Monasterio Y.i.6
1250-1250	<i>Biblia romanceada E8</i>	Escorial: Monasterio I.i.8
1250-1250	<i>Biblia romanceada E6</i>	Escorial: Monasterio I.i.6
1270-1280	<i>General Estoria</i>	Madrid: Biblioteca Nacional MSS/816
		Madrid: Biblioteca Nacional MSS/10237
		Évora: Biblioteca Pública CXXV/2-3
		Escorial: Monasterio O-I-11
		Biblioteca Vaticana, Urb. Lat. 539
1420-1430	<i>Biblia romanceada E3</i>	Escorial: Monasterio I.i.3
1420-1430	<i>Biblia de Ajuda</i>	Lisboa: Biblioteca da Ajuda, ms. 52-XIII-1
1420-1420	<i>Biblia romanceada E19</i>	Escorial: Monasterio J.ii.19
1400-1430	<i>Biblia romanceada E7</i>	Escorial: Monasterio I.i.7
1400-1430	<i>Biblia romanceada E5</i>	Escorial: Monasterio I.i.5
1400-1429	<i>Biblia de Évora</i>	Evora: Biblioteca Pública CXXIV/1-2
1400-1430	<i>Biblia romanceada E4</i>	Escorial: Monasterio I.i.4
1400-1430	<i>Biblia romanceada de Oxford</i>	Oxford: Bodleian, Canon. Ital. 177
1400-1459	<i>Biblia romanceada de la Biblioteca Nacional</i>	Madrid: Biblioteca Nacional MSS/10288
1400-1459	<i>Biblia romanceada de la Real Academia de la Historia</i>	Madrid: Real Academia de la Historia
1422-1430	<i>Biblia romanceada de Mosé Arragel</i>	Madrid: Biblioteca de la Casa Ducal de Alba
1450-1450	<i>Libro de Ester</i>	Salamanca: Biblioteca Universitaria 2015
1398-1407	<i>Libro de Job</i>	Madrid: Biblioteca Nacional MSS/10138
1400-1459	<i>Evangelios y Epístolas paulinas</i>	Madrid: Biblioteca Nacional MSS/9556
<b>Textos poéticos españoles</b>		
<b>Fecha</b>	<b>Título</b>	<b>Biblioteca</b>
1301-1350	<i>Poema de Mío Cid</i>	Madrid: Biblioteca Nacional VITR/7/17
1301-1400	<i>Libro de Apolonio</i>	Escorial: Monasterio K.III.4
1301-1400	<i>Vida de Santa María Egipcíaca</i>	Escorial: Monasterio K.III.4
1301-1400	<i>Libre dels tres reys d'Orient</i>	Escorial: Monasterio K.III.4
1301-1400	<i>Libro de Alexandre</i>	Madrid: Biblioteca Nacional VITR/5/10
1325-1325	<i>Obras de Gonzalo de Berceo</i>	Madrid: Real Academia Española 4
1330-1400	<i>Libro de buen amor (T)</i>	Madrid: Biblioteca Nacional VITR/6/1

1389-1389	<i>Libro de buen amor (G)</i>	Madrid: Real Academia Española 19
1401-1500	<i>Cancionero castellano de París (Dutton PN9)</i>	París: Nationale Esp. 231
1421-1440	<i>Libro de buen amor (S)</i>	Salamanca: Biblioteca Histórica de la Universidad 2663
1441-1460	<i>Mocedades de Rodrigo</i>	París: Nationale Esp. 12
1444-1500	<i>Cancionero castellano de París (Dutton PN5)</i>	París: Nationale Esp. 227
1445-1450	<i>Cancionero de Baena (Dutton PN1)</i>	París: Nationale Esp. 37
1451-1500	<i>Cancionero castellano de París (Dutton PN2)</i>	París: Nationale Esp. 216
1456-1480	<i>Proverbios morales de Semtob</i>	Madrid: Real Academia Española: M-RAE, RM-73
1458-1500	<i>Cancionero de París (Dutton PN10)</i>	París: Nationale Esp. 233
1460-1460	<i>Cancionero de Salvá (Dutton PN13)</i>	París: Nationale Esp. 510
1461-1480	<i>Poema de Fernán González</i>	Escorial: Monasterio b.IV.21 (2)
1465-1479	<i>Danza de la muerte</i>	Escorial: Monasterio b.IV.21 (1)
1470-1470	<i>Cancionero de París (Dutton PN12)</i>	París: Nationale Esp. 313
1485-1485	<i>Coplas y glosas de Mingo Revulgo</i>	Londres: British Library, IA.53204
1489-1489	<i>Coronación</i>	Nueva York: Hispanic Society of America
1490-1490	<i>Traslación del doctor Chatón</i>	Escorial: Monasterio, 32-V-19 (3o)
1491-1500	<i>Cancionero castellano y catalán de París (Dutton PN4)</i>	París: Nationale Esp. 226
1493-1493	<i>Cirugía rimada</i>	Madrid: Nacional MSS/2153
1493-1493	<i>El catón en latín y en romance</i>	Madrid: Nacional INC/401
1496-1496	<i>Cancionero de las obras de Juan del Encina</i>	Madrid: Real Academia Española I-8
1501-1510	<i>Romance «Rey que no hace justicia»</i>	Madrid: Palacio II/1520 (2)
1512-1546	<i>Cancionero de la Trivulziana</i>	Milano: Biblioteca Trivulziana Ms. 940
1521-1521	<i>Castigos &amp; enxemplos de catom</i>	Oporto: Biblioteca Pública, Y'3-37
1533-1533	<i>Castigos y exemplos de Caton: nuevamente impresos</i>	Viena: Nationalbibliothek, 32-T-28
1542-1542	<i>Castigos y exemplos de Caton: Nueuamente impresso</i>	Londres: British Library, C.39.f.28
1543-1543	<i>Castigos y exemplos de Caton: Nueuamente impresso</i>	Madrid: Biblioteca Nacional, R/3145
1543-1543	<i>Castigos y exemplos de Caton</i>	Madrid: Biblioteca del Marqués de Morbecq
1563-1563	<i>Exemplos de Caton.</i>	Madrid: Biblioteca Nacional, R/12736
1586-1586	<i>Exemplos de Caton</i>	Nueva York: Hispanic Society of America of America
1609-1609	<i>Exemplos de Caton</i>	Cambridge: Houghton Library, SC6 C3375 A609c
<b>Textos tempranos de Celestina</b>		
<b>Fecha</b>	<b>Título</b>	<b>Impresor</b>
1499-1499	<i>Comedia de Calisto y Melibea</i>	Burgos: Fradrique de Basilea
1500-1500	<i>Comedia de Calisto y Melibea</i>	Toledo: Pedro Hagenbach

*En busca de lo hablado en lo escrito en los corpus diacrónicos del español: una comparativa con los corpus anglosajones*

1501-1501	<i>Comedia de Calisto y Melibea</i>	Sevilla: Estanislao Polono
1501-1510	<i>Comedia de Calisto y Melibea</i>	Fragmento manuscrito del Acto I
1507-1507	<i>Tragicomedia de Calisto y Melibea</i>	Zaragoza: Jorge Coci
1510-1510	<i>Tragicomedia de Calisto y Melibea</i>	Toledo: Pedro Hagenbach
1511-1511	<i>Tragicomedia de Calisto y Melibea</i>	Sevilla: Jacobo Cromberger
1513-1515	<i>Tragicomedia de Calisto y Melibea</i>	Sevilla: Jacobo Cromberger
1514-1514	<i>Tragicomedia de Calisto y Melibea</i>	Valencia: Juan Joffre
1518-1520	<i>Tragicomedia de Calisto y Melibea</i>	Sevilla: Jacobo Cromberger
1516-1516	<i>Tragicomedia de Calisto y Melibea</i>	Roma: Marcelo Silber
1518-1518	<i>Tragicomedia de Calisto y Melibea</i>	Valencia: Juan Joffre
1520-1520	<i>Tragicomedia de Calisto y Melibea</i>	Roma: Antonio de Salamanca
1523-1523	<i>Tragicomedia de Calisto y Melibea</i>	[Sevilla?] [Venecia?]: [s.n.] [Juan Batista Pedrezano]
1525-1525	<i>Tragicomedia de Calisto y Melibea</i>	Barcelona: Carles Amorós
1525-1525	<i>Tragicomedia de Calisto y Melibea</i>	Sevilla: Jacobo & Juan Cromberger
1526-1526	<i>Tragicomedia de Calisto y Melibea</i>	Toledo: Casa de Remon de Petras
1526-1526	<i>Tragicomedia de Calisto y Melibea</i>	Valladolid: Nicolas Tierry
1528-1528	<i>Tragicomedia de Calisto y Melibea</i>	Sevilla: Jacobo & Juan Cromberger
1529-1529	<i>Tragicomedia de Calisto y Melibea</i>	Valencia: Juan Viñao
1530-1540	<i>Tragicomedia de Calisto y Melibea</i>	Medina del Campo: Pedro Tovans

**Apéndice II: Relación detallada de los textos incluidos en el Archivo Digital de Manuscritos y Textos Españoles**

<b>Autor</b>	<b>Obra</b>
<b>Libros</b>	
Diego de Valera.	<i>Ceremonial de príncipes.</i>
Diego de Valera.	<i>Tratado de las armas (Madrid Nacional ms. 1341).</i>
Diego de Valera.	<i>Tratado de las armas (Madrid Nacional R-2302).</i>
Diego de Valera.	<i>Tratado de las armas (Roma Casanatense 1098).</i>
Diego de Valera.	<i>Tratado en defensa de virtuosas mujeres.</i>
Fernando Mejía.	<i>Nobiliario vero.</i>
Frontino.	<i>Strategematon; Arte de Caballeria.</i>
Juan Manuel.	<i>Libro de las armas.</i>
Juan Manuel.	<i>Libro de los estados.</i>
Juan Manuel.	<i>Libro del caballero y del escudero.</i>
Juan Manuel.	<i>Libro infinito.</i>
Leomarte.	<i>Sumas de la historia troyana.</i>
<b>Textos científicos</b>	
Alfonso Chirino.	<i>Espejo de medicina.</i>
Alfonso Chirino.	<i>Menor daño de medicina.</i>
Alfonso de Toledo.	<i>Invencionario.</i>
Alfonso X.	<i>Astronomía.</i>
Alfonso X.	<i>Cánones de Albateni.</i>
Alfonso X.	<i>Formas e imágenes.</i>
Alfonso X.	<i>Lapidario.</i>
Alfonso X.	<i>Libro de ajedrez, dados y tablas.</i>
Alfonso X.	<i>Libro de las cruces.</i>
Alfonso X.	<i>Libro de las tafurerías.</i>
Alfonso X.	<i>Picatrix.</i>
Alfonso X.	<i>Rabizag-Libro del cuadrante señero.</i>
Alfonso X.	<i>Tablas de Zarquiel.</i>
Alfonso XI.	<i>Libro de la montería.</i>
Ali aben Ragel; Alfonso X.	<i>Judizios de las estrellas.</i>
Anónimo, atribuido al Doctor Gómez de Salamanca.	<i>Propiedades del romero.</i>
Anónimo.	<i>Dancus rex; Libro de cetrería.</i>
Anónimo.	<i>De las medecinas.</i>
Anónimo.	<i>Libro de los azores; también llamado de los Gavilanes.</i>
Anónimo.	<i>Tesoro de los remedios.</i>
Anónimo.	<i>Tratado de la música.</i>
Anónimo.	<i>Tratado del alboraique.</i>
Anónimo; atribuido a Daneos y Galaciano.	<i>Tratado de cetrería.</i>
Antonio de Nebrija.	<i>Tabla de la diversidad de los días y horas.</i>

*En busca de lo hablado en lo escrito en los corpus diacrónicos del español: una comparativa con los corpus anglosajones*

Bernardino de Montana de Monserrat.	<i>Libro de la anatomía del hombre.</i>
Bernardo de Gordonio.	<i>De ingeniis curandorum morborum; Tratado de las tablas de los diez ingenios de curar enfermedades agudas.</i>
Bernardo de Gordonio.	<i>De regimine acutorum morborum; Regimiento de las agudas.</i>
Bernardo de Gordonio.	<i>Lilio de medicina.</i>
Bernardo de Gordonio; Anónimo tr.	<i>Prognostica; Las pronósticas.</i>
Bernardo de Gordonio; Anónimo tr.	<i>Tratado de los niños con el regimiento del ama.</i>
Diego de Cobos	<i>Cirugía rimada.</i>
Domingo Marcos Durán.	<i>Glosa sobre Lux bella.</i>
Domingo Marcos Durán.	<i>Lux bella.</i>
Enrique de Aragón.	<i>Arte cisoria.</i>
Estéfano de Sevilla.	<i>Visita y consejo de medicos.</i>
Evangelista.	<i>Libro de cetrería.</i>
Fernando Álvarez.	<i>Regimiento contra la peste.</i>
Fernando de Valdés.	<i>Tratado de la utilidad de la sangría en las viruelas y otras enfermedades de los muchachos.</i>
Francisco López de Villalobos.	<i>Sumario de la medicina.</i>
Gerardus Falconarius; Anónimo tr.	<i>Tratado de las enfermedades de las aves de caza.</i>
Gilberto.	<i>Recetas.</i>
Guillelmus Falconarius; Anónimo tr.	<i>Libro de los halcones.</i>
Isaac Israeli.	<i>Tratado de las fiebres.</i>
Jehuda ha-Levi.	<i>Libro del Cuzari.</i>
Juan de Ketham.	<i>Fasciculus medicinae; Compendio de la salud humana.</i>
Juan Enríquez.	<i>Secretos de la medicina.</i>
Juan Manuel.	<i>Libro de la caza.</i>
Julián Gutiérrez de Toledo.	<i>Cura de la piedra.</i>
Lanfranco de Milano.	<i>Cirugía mayor (Madrid Nacional ms. 2147.)</i>
Lanfranco de Milano.	<i>Cirugía mayor (Madrid Nacional ms. 2165).</i>
Lope de Barrientos.	<i>Tratado de la divinanza (Madrid Nacional ms. 6401).</i>
Lope de Barrientos.	<i>Tratado de la divinanza (Escorial Monasterio h-III-13).</i>
Luis del Mercado.	<i>Instituciones para el aprovechamiento y examen de los algebristas.</i>
Macer.	<i>Macer herbolario.</i>
Manuel Díez de Calatayud; Martín Martínez de Ampíes tr.	<i>Libro de albeitería.</i>
Moamyn.	<i>Libro de las animalias que cazan.</i>
Moisés ben Samuel de Roquemaure, también atribuido a Juan de Aviñón.	<i>Sevillana medicina.</i>
Mondino dei Luzzi; Anónimo tr.	<i>Tratado de la fisonomía en breve suma contenida.</i>
Paladio; tr. Anónimo.	<i>De re rustica; Tratado de Agricultura.</i>
Pedro Arias de Benavides.	<i>Secretos de cirugía.</i>

Pero López de Ayala.	<i>Libro de la caza de las aves.</i>
Pietro de Tossignano; atribuido a Vasco de Taranta.	<i>De epidemia et peste; Tratado de la peste.</i>
Saladino da Ascoli; Alonso Rodríguez de Tudela tr.	<i>Compendio de los boticarios.</i>
Teodorico Borgognoni.	<i>Libro de los caballos.</i>
<b>Crónicas y biografías</b>	
Alfonso Martínez de Toledo.	<i>Atalaya de las Crónicas.</i>
Alfonso X.	<i>Estoria de España I.</i>
Alfonso X.	<i>Estoria de España II.</i>
Alfonso X.	<i>General estoria I.</i>
Alfonso X.	<i>General estoria II.</i>
Alfonso X.	<i>General Estoria IV.</i>
Alfonso X.	<i>General Estoria V (Escorial Monasterio I-I-2).</i>
Alfonso X.	<i>General estoria V (Escorial Monasterio R-I-10).</i>
Alfonso X.	<i>General estoria VI.</i>
Anónimo, Pseudo Rodrigo Jiménez de Rada.	<i>Historia de los reyes godos.</i>
Anónimo.	<i>Anales de los reyes de Castilla.</i>
Anónimo.	<i>Crónica de veinte reyes.</i>
Anónimo.	<i>Crónica del Cid.</i>
Anónimo; atribuido a Lope de Barrientos.	<i>Abreviación del Halconero.</i>
Boccaccio.	<i>Teseida.</i>
Boccaccio; Pero López de Ayala tr. Alonso de Cartagena tr.	<i>Caída de principes.</i>
Diego de Valera.	<i>Crónica de España.</i>
Diego Rodríguez de Almela.	<i>Compilación de las batallas campales.</i>
Diego Rodríguez de Almela.	<i>Escritura de cómo y por qué razón no se debe dividir, partir ni enajenar los reinos y señoríos de España.</i>
Diego Rodríguez de Almela.	<i>Letra sobre los matrimonios y casamientos entre los reyes de Castilla y de León de España con los reyes y casa de Francia hechos.</i>
Diego Rodríguez de Almela.	<i>Valerio de las historias escolasticas y de España.</i>
Eutropio; Juan Fernández de Heredia tr.	<i>Breviarium ab urbe condita.</i>
Fernán Sánchez de Valladolid.	<i>Crónica de Alfonso X.</i>
Fernán Sánchez de Valladolid.	<i>Crónica de Sancho IV.</i>
Fernando del Pulgar.	<i>Claros varones de Castilla.</i>
Fernando del Pulgar.	<i>Letras.</i>
Gauberto Fabricio de Vagad.	<i>Crónica de Aragon.</i>
Giovanni Boccaccio.	<i>De las mujeres ilustres en romance (De claris mulieribus).</i>
Guido de Columna.	<i>Historia troyana (Pamplona Arnaldo Guillén de Brocar 1499. Madrid Nacional I-733)</i>
Guido delle Colonne; Juan Fernández de Heredia tr.	<i>Historia Troyana (Madrid Nacional ms. 10801).</i>

*En busca de lo hablado en lo escrito en los corpus diacrónicos del español: una comparativa con los corpus anglosajones*

He'tum Korikos'ci; Juan Fernández de Heredia tr.	<i>Flor de las historias de orient.</i>
Johannes Zonaras; Juan Fernández de Heredia tr.	<i>Crónica de los emperadores.</i>
Juan Fernández de Heredia tr.	<i>Crónica de Morea.</i>
Juan Fernández de Heredia.	<i>Crónica de los conquistadores I.</i>
Juan Fernández de Heredia.	<i>Crónica de los conquistadores II.</i>
Juan Fernández de Heredia.	<i>Grant Crónica de Espanya I.</i>
Juan Fernández de Heredia.	<i>Grant Crónica de Espanya III.</i>
Juan Fernández de Heredia.	<i>Libro de autoridades-Rams de flors.</i>
Juan II.	<i>Carta a Don Alvaro de Luna.</i>
Juan II.	<i>Carta a Enrique IV sobre don Álvaro de Luna.</i>
Juan II.	<i>Carta a la condesa doña Juana Pimentel y al conde don Juan mujer e hijo de don Álvaro de Luna.</i>
Lucas de Tuy.	<i>Crónica del tudense.</i>
Pablo de Santa María.	<i>Suma de las Crónicas de España.</i>
Paulo Orosio; Juan Fernández de Heredia tr.	<i>Historia contra los paganos.</i>
Pedro Afonso de Barcelos; tr Anónimo.	<i>Crónica de 1344 I.</i>
Plutarco; Juan Fernández de Heredia tr.	<i>Vidas paralelas (París Nationale Esp. 70).</i>
Plutarco; Juan Fernández de Heredia tr.	<i>Vidas paralelas (París Nationale Esp. 71).</i>
Plutarco; Juan Fernández de Heredia tr.	<i>Vidas paralelas (París Nationale Esp. 72).</i>
Salustio; Francisco Vidal de Noya tr.; atribuido a Vasco de Guzmán tr.	<i>Guerra de Jugurtha.</i>
Salustio; Francisco Vidal de Noya tr.	<i>Conjuración de Catalina.</i>
Salustio; Vasco Ramírez de Guzmán tr.	<i>Conjuración de Catilina.</i>
Salustio; Vasco Ramírez de Guzmán tr.	<i>Guerra de Yugurta.</i>
San Isidoro; Anónimo tr.	<i>Crónica.</i>
San Isidoro; Anónimo tr.	<i>Historia de los godos.</i>
Séneca; Anónimo tr.	<i>Juego de Claudio emperador.</i>
Tucídides; Juan Fernández de Heredia tr.	<i>Guerra del Peloponeso.</i>
<b>Diccionarios y gramáticas</b>	
Alfonso de Palencia.	<i>Universal vocabulario de latín en romance.</i>
Antonio de Nebrija.	<i>Dictionarium hispano-latinum.</i>
Antonio de Nebrija.	<i>Dictionarium latino-hispanicum.</i>
Antonio de Nebrija.	<i>Gramática castellana.</i>
Antonio de Nebrija.	<i>Introducciones latinae.</i>
Antonio de Nebrija.	<i>Reglas de ortographía en la lengua castellana.</i>
Rodrigo Fernández de Santaella.	<i>Vocabulario eclesiástico.</i>
<b>Enciclopedias</b>	

Alfonso de la Torre.	<i>Visión delectable.</i>
Anónimo.	<i>Libro del conocimiento de todos los reinos (Madrid Nacional ms. 1997).</i>
Anónimo.	<i>Libro del conocimiento de todos los reinos (Madrid Nacional ms. 9055).</i>
Anónimo.	<i>Libro del conocimiento de todos los reinos (Salamanca Biblioteca Universitaria 1890).</i>
Bartholomaeus Glanville; fr Vicente de Burgos tr.	<i>Liber de proprietatibus rerum; Propiedades de las cosas.</i>
Brunetto Latini.	<i>Libro del tesoro.</i>
Francesc Eiximenis.	<i>Libro de las donas.</i>
Juan Manuel.	<i>Crónica abreviada.</i>
San Isidoro.	<i>Semejanza del mundo.</i>
<b>Textos legales</b>	
Alfonso Díaz de Montalvo.	<i>Ordenanzas reales.</i>
Alfonso X.	<i>Espéculo.</i>
Alfonso X.	<i>Fuero Juzgo.</i>
Alfonso X.	<i>Fuero real (Escorial Monasterio Z-III-16).</i>
Alfonso X.	<i>Fuero real (Filadelfia Free Library Rare Books E-MS 245).</i>
Alfonso X.	<i>Siete partidas (Sevilla Meinardo Ungut y Estanislao Polono 10-25-1491. Madrid Nacional I-766).</i>
Alfonso X.	<i>Siete partidas I (Londres British Library Add 20787).</i>
Alfonso XI.	<i>Ordenamiento de Alcalá (Madrid Nacional Res 9).</i>
Alfonso XI.	<i>Ordenamiento de Alcalá (Madrid Nacional Vitrina 15-7).</i>
Anónimo	<i>Leyes del estilo (Escorial Monasterio Z-III-11).</i>
Anónimo.	<i>Fuero General de Navarra Versión A.</i>
Anónimo.	<i>Fuero general de Navarra Versión B (Barcelona Catalunya 1015).</i>
Anónimo.	<i>Fuero General de Navarra Versión B (Dresde Landesbibliothek App 1761).</i>
Anónimo.	<i>Fuero General de Navarra Versión B (Madrid Nacional ms. 800).</i>
Anónimo.	<i>Fuero general de Navarra Versión C (Madrid Nacional ms. 248).</i>
Anónimo.	<i>Fuero general de Navarra Versión C (Salamanca Biblioteca Universitaria 1947).</i>
Anónimo.	<i>Fuero viejo de Alcalá de Henares.</i>
Anónimo.	<i>Fueros de la Novenera.</i>
Anónimo.	<i>Leyes del estilo (Escorial Monasterio Z-II-14).</i>
Anónimo.	<i>Leyes del estilo (Escorial Monasterio Z-II-8).</i>
Anónimo.	<i>Leyes del estilo (Madrid Nacional ms. 5764).</i>
Anónimo.	<i>Leyes del estilo (Salamanca Impresor de la Gramática de Nebrija 1497-02-10 / Transcrito a base de Huntington 9566-6).</i>
Anónimo.	<i>Leyes del estilo (Salamanca Impresor de la Gramática de Nebrija 1500-04-12 / Transcrito a base de Bodleian INC d S6 1500-1).</i>

*En busca de lo hablado en lo escrito en los corpus diacrónicos del español: una comparativa con los corpus anglosajones*

Anónimo.	<i>Tratado jurídico aljamiado.</i>
Blanca de Portugal.	<i>Fuero de Briviesca.</i>
Juana la Loca.	<i>Leyes de Toro.</i>
Reyes Católicos.	<i>Ordenanzas sobre los abogados (1499).</i>
Reyes Católicos.	<i>Capítulos de gobernadores, asistentes y corregidores.</i>
Reyes Católicos.	<i>Cuaderno de las leyes nuevas de la hermandad.</i>
Reyes Católicos.	<i>Leyes hechas por la brevedad y orden de los pleitos (26 de mayo de 1499). (Salamanca Impresor de la Gramática de Nebrija ca. 1500. Madrid Nacional I-1801).</i>
Reyes Católicos.	<i>Leyes hechas por la brevedad y orden de los pleitos (26 de mayo de 1499). (Salamanca Juan de Porras ca. 1511, ff 1v-7v. Madrid Nacional I-1405).</i>
Reyes Católicos.	<i>Ordenanzas de la ciudad de Sevilla (1492-05-30).</i>
Reyes Católicos.	<i>Ordenanzas sobre los escribanos de la ciudad de Sevilla (1492-05-30).</i>
Reyes Católicos.	<i>Ordenanzas sobre los paños.</i>
Seudo-Fernando III.	<i>Libro de los fueros de Castilla.</i>
Vidal de Canellas.	<i>Fueros de Aragón.</i>
<b>Textos poéticos</b>	
¿Pedro de? Cartagena.	<i>¿Qué cosa es amor? ff 18va-19va.</i>
Ambrosio Montesino.	<i>Coplas hechas sobre la Pasión.</i>
Anónimo.	<i>Cancioneiro geral.</i>
Anónimo.	<i>Cancionero castellano de París (París Nationale Esp. 216).</i>
Anónimo.	<i>Cancionero castellano de París (París Nationale Esp. 226).</i>
Anónimo.	<i>Cancionero castellano de París (París Nationale Esp. 227).</i>
Anónimo.	<i>Cancionero castellano de París (París Nationale Esp. 231).</i>
Anónimo.	<i>Cancionero castellano de París (París Nationale Esp. 233).</i>
Anónimo.	<i>Cancionero castellano de París (París Nationale Esp. 313).</i>
Anónimo.	<i>Cancionero de romances.</i>
Anónimo.	<i>Cancionero de Salvá.</i>
Anónimo.	<i>Cantar de Mio Cid.</i>
Anónimo.	<i>Coplas hechas sobre el casamiento de la hija del rey de España.</i>
Anónimo.	<i>Danza de la muerte.</i>
Anónimo.	<i>Glosa al Romance "Rey que no hace justicia."</i>
Anónimo.	<i>Libro de Alexandre.</i>
Anónimo.	<i>Libro de Apolonio.</i>
Anónimo.	<i>Libro dels tres reys d'orient.</i>
Anónimo.	<i>Mocedades de Rodrigo.</i>
Anónimo.	<i>Poema de Fernán González.</i>

Anónimo.	<i>Vida de Santa María Egipcíaca.</i>
Fernán Pérez de Guzmán.	<i>Coplas sobre Cartagena.</i>
Fernando de Yanguas.	<i>El nunc dimittis, ff 19vb-20vb.</i>
Gonzalo de Berceo.	<i>Obras de Gonzalo de Berceo.</i>
Juan Alfonso de Baena comp.	<i>Cancionero de Baena.</i>
Juan de Mena.	<i>Coronación.</i>
Juan del Encina.	<i>A su amiga porque se le escondía en viéndola.</i>
Juan del Encina.	<i>Cancionero.</i>
Juan del Encina.	<i>Coplas del memento homo.</i>
Juan del Encina.	<i>Égloga de Plácida y Victoriano.</i>
Juan del Encina.	<i>Egloga de tres pastores.</i>
Juan del Encina.	<i>Testamento de amores.</i>
Juan del Encina.	<i>Todos servid al Amor.</i>
Juan del Encina.	<i>Villancico nuevo sobre el perdón de Santa María Magdalena.</i>
Juan Ruiz.	<i>Libro de buen amor (Madrid Nacional Vitrina 6-1 Toledo).</i>
Juan Ruiz.	<i>Libro de buen amor (Madrid Real Academia Espanola 19 Gayoso).</i>
Juan Ruiz.	<i>Libro de buen amor (Salamanca Biblioteca Universitaria 2663).</i>
Luis de Góngora yB193:C219 Argote.	<i>Obras.</i>
<b>Prosa literaria</b>	
Alfonso Martínez de Toledo.	<i>Arcipreste de Talavera; Corbacho.</i>
Anónimo.	<i>Cuento de Tristán de Leonís.</i>
Anónimo.	<i>Enrique fi de Oliva.</i>
Anónimo.	<i>Gran conquista de Ultramar.</i>
Anónimo.	<i>Gran Conquista de Ultramar.</i>
Anónimo.	<i>Gran conquista de Ultramar.</i>
Anónimo.	<i>Historia de la doncella Teodor.</i>
Anónimo.	<i>Historia del noble Vespasiano.</i>
Anónimo.	<i>La vida de Lazarillo de Tormes y de sus fortunas y adversidades (Alcalá de Henares Salzado 1554-02-26)</i>
Anónimo.	<i>La vida de Lazarillo de Tormes y de sus fortunas y adversidades (Amberes Martín Nucio 1554).</i>
Anónimo.	<i>La vida de Lazarillo de Tormes y de sus fortunas y adversidades (Burgos Juan de Junta 1554).</i>
Anónimo.	<i>Libro del caballero Zifar.</i>
Anónimo.	<i>Pierres y Magalona.</i>
Anónimo.	<i>Tristán de Leonís.</i>
Diego de San Pedro.	<i>Arnalte y Lucenda.</i>
Diego de San Pedro.	<i>Cárcel de amor.</i>
Enrique de Villena.	<i>Doce trabajos de Hércules.</i>
Fernando de Rojas.	<i>Comedia de Calisto y Melibea (Burgos Fadrique de Basilea 1499. Hispanic Society of America).</i>

Fernando de Rojas.	<i>Comedia de Calisto y Melibea (Madrid Palacio 1520 ff 93v-100v).</i>
Fernando de Rojas.	<i>Tragicomedia de Calisto y Melibea (Valencia Juan Joffre 1514-02-21).</i>
Fernando de Rojas.	<i>Tragicomedia de Calisto y Melibea (Zaragoza Jorge Coci 1507 / Madrid Real Academia de la Historia 2-7-2/3566).</i>
Jean D'Arras.	<i>Historia de la linda Melosina.</i>
John Gower; Robert Payne tr. portugues; Juan de Cuenca tr. castellano.	<i>Confessio Amantis.</i>
Juan de Flores.	<i>Grimalte y Gradissa.</i>
Juan de Flores.	<i>Triunfo de amor (Madrid Nacional ms. 22019).</i>
Juan de Flores.	<i>Triunfo de amor (Sevilla Colombina 5-3-20).</i>
Juan Manuel.	<i>Conde Lucanor.</i>
Juan Rodríguez del Padrón.	<i>Siervo Libre de Amor.</i>
<b>Textos religiosos</b>	
Alfonso de Cartagena San Juan Crisóstomo.	<i>Apología sobre el salmo "Judica me Deus".</i>
Anónimo.	<i>Arte de bien morir-Breve confesionario.</i>
Anónimo.	<i>Biblia Latina.</i>
Anónimo.	<i>Biblia romanceada judio cristiana.</i>
Anónimo.	<i>Sermones contra los judíos y moros.</i>
David; Hermann el Aleman tr.	<i>Salterio.</i>
Jean Gerson.	<i>Meditación del corazón; De meditatione cordis.</i>
Juan Manuel.	<i>Tratado de la Asunción de la Virgen.</i>
Pseudo-Jean Gerson.	<i>Imitacioacute;n de Cristo; Imitatio Christi.</i>
Pseudo-San Bernardo.	<i>Libro Llamado Infancia Salvatoris.</i>
San Basilio; Pero Díaz de Toledo tr.	<i>Tratado de la reformación de la ánima.</i>
San Juan Crisóstomo + Alfonso de Cartagena.	<i>Quod nemo laeditur nisi a seipso + Glosa a San Juan Crisóstomo.</i>
<b>Libros sapienciales</b>	
Alfonso de Cartagena.	<i>Oracional de Fernán Pérez de Guzmán.</i>
Alfonso de Valladolid.	<i>Libro de las tres creencias.</i>
Anónimo.	<i>Castigos y doctrinas que un sabio daba a sus hijas.</i>
Bidpai.	<i>Ejemplario contra los engaños y peligros del mundo.</i>
Bonium	<i>Bocados de oro.</i>
Cicerón; Alfonso de Cartagena tr.	<i>De inventione; Retórica.</i>
Cicerón; Alonso de Cartagena tr.	<i>De officiis; De los oficios.</i>
Esopo.	<i>Esopete ystoriado (Toulouse Paris Clebat 1488. Manchester John Rylands)</i>
Esopo.	<i>Ysopete ystoriado (Zaragoza Juan Hurus 1489. Escorial Monasterio 32-I-13).</i>
Íñigo López de Mendoza.	<i>Bías contra Fortuna.</i>
Maestre Pedro.	<i>Libro del Consejo y de los Consejeros.</i>
Moises ben Maimon.	<i>Moreh Nevukim; Mostrador y enseñador de los turbados.</i>

Pedro Gómez Barroso.	<i>Libro del Consejo y de los Consejeros.</i>
Pierre Bersuire.	<i>Morales de Ovidio.</i>
Platón; Pero Díaz de Toledo tr.	<i>Axioco-Fedrón.</i>
Pseudo-Aristóteles; Juan Fernández de Heredia tr.	<i>Secretum secretorum.</i>
Sancho IV.	<i>Castigos y documentos para bien vivir (Escorial Z-III-4, ff 1ra-95vb).</i>
Sancho IV.	<i>Castigos y documentos para bien vivir (Madrid Nacional 6603).</i>
Sancho IV.	<i>Castigos y documentos para bien vivir (Madrid Nacional ms. 6559, ff 1r-244v).</i>
Sem Tob.	<i>Proverbios Morales.</i>
Seudo-Aristoteles.	<i>Poridat de las poridades.</i>
<b>Libros de viajes</b>	
Bernardo de Breidenbach; Martín Martínez de Ampiés tr.	<i>Peregrinatio in Terram Sanctam; Viaje siquier peregrinación de la tierra sancta.</i>
Jean de Bourgogne.	<i>Viaje de Juan de Mandevilla.</i>
Juan de Mandevilla; Anónimo tr.	<i>Libro de las maravillas del mundo.</i>
Marco Polo; Juan Fernández de Heredia tr.	<i>Libro de Marco Polo.</i>
Marco Polo; Rodrigo Fernández de Santaella tr.	<i>Libro de Marco Polo.</i>
Martín Martínez de Ampiés.	<i>Tratado de Roma.</i>
Ruy González de Clavijo.	<i>Historia del gran Tamorlán.</i>