

CORPUS DIGITALIZADOS Y PALABRAS GRAMATICALES

MAR GARACHANA (*Universitat de Barcelona*)
ESTHER ARTIGAS (*Universitat de Barcelona*)

RESUMEN:

El objetivo de este trabajo es reflexionar acerca de las ventajas e inconvenientes del empleo de los corpus informatizados. El caso que presentamos está vinculado a un proyecto de I+D sobre la gramaticalización de perífrasis verbales (GRAPEVERBA). Para llevar a cabo este estudio, hemos extraído las ocurrencias de los dos corpus académicos, *CORDE* y *CREA*. La falta de una lematización y de un etiquetado en ambos corpus nos ha planteado un problema de difícil solución, puesto que el número de ejemplos obtenido resulta excesivamente elevado. Otro problema tiene que ver con las ediciones textuales de las obras vertidas en los corpus de la Academia, de manera especial en el *CORDE*. Con cierta frecuencia, estas ediciones no son contemporáneas de los manuscritos originales, lo que compromete seriamente las conclusiones que se extraen acerca de la gramaticalización de algunas perífrasis verbales, por ejemplo de *tener + (a/de) + infinitivo*.

PALABRAS CLAVE: lingüística de corpus, gramaticalización, perífrasis verbales.

ABSTRACT:

The aim of this paper is to address the advantages and disadvantages of working with electronic corpora to study the grammaticalization of verbal constructions in the history of Spanish. The data on which we base our study have been extracted from the corpora compiled by the Real Academia Española, i.e., *corde* and *crea*. These corpora lack lemmatization and tagging. Hence, when these corpora are consulted for a given expression, concordance searches often yield an unwieldy amount of irrelevant occurrences. Another difficulty we have encountered when working with these corpora, especially *corde*, has to do with the fact that, all too often, the editions that are used therein are not contemporaneous with the original manuscripts. This seriously compromises the conclusions reached by studies that use these corpora. We illustrate the problem with an analysis of the first documentations of the verbal periphrasis *tener + (de/a) + infinitive*.

KEY WORDS: Corpus Linguistics, Grammaticalization, Verbal Periphrases.

INTRODUCCIÓN¹

Los corpus textuales se han convertido en una herramienta básica para el estudio de la Sintaxis histórica². En pocos minutos, a veces en pocos segundos, disponemos de un listado de ocurrencias que en el pasado hubieran supuesto años de laboriosa búsqueda. En este sentido, son instrumentos imprescindibles para el investigador. Sin embargo, el empleo de estos corpus suscita críticas o «quejas» entre sus usuarios, debido,

¹ Este trabajo se inscribe en el marco del proyecto de investigación *Gramática de las perífrasis verbales del español. Historia, Pragmática y Discurso* (FFI2008-00948/FILO).

² Para el español podemos citar, además de los corpus académicos *CREA* y *CORDE*, el *Corpus del Español* de Mark Davies, el *CODEA* —que se integrará en la red *CHARTA*—, el corpus *Biblia Medieval* (Enrique-Arias 2008), las bases de datos de *Parnaseo* y, recientemente, se puede consultar *on line* la *Biblioteca Digital de Textos del Español Antiguo* del Hispanic Seminary of Medieval Studies (Gago 2011). También hay que mencionar el *ADMYTE*, el *ARTHUS* y el *Corpus del nuevo diccionario histórico (CNDH)*, publicado recientemente. Asimismo, han aparecido en formato digital las ediciones de algunos textos, como, por ejemplo, las cartas de emigrantes a Indias editadas por Fernández Alcaide (2009).

fundamentalmente, a un desajuste entre los objetivos perseguidos por los creadores del corpus —en muchas ocasiones simplemente se trataba de tener reunidos en soporte digital el mayor número de documentos posibles, a fin de aumentar el acervo textual de una lengua³— y los objetivos de los usuarios de dichos corpus. Los desencuentros —y los flechazos— entre corpus y usuarios, las aplicaciones de los corpus digitalizados y las ventajas que ofrecen a las investigaciones lingüísticas hace tiempo que dejaron de ser un tema de conversación en los descansos de los congresos para convertirse en materia de análisis merecedora de libros, simposios, congresos, y de una rama de la Lingüística, a saber, la *Lingüística de corpus*.

Este trabajo es el resultado de una reflexión surgida en el seno del proyecto de investigación GRAPEVERBA, que se ocupa del estudio de los procesos de gramaticalización de las perífrasis verbales del español. Como punto de partida, el equipo decidió extraer las ocurrencias en las que basar el estudio de los dos corpus de la Real Academia (*CORDE* y *CREA*); las páginas que siguen reflejan algunos de los problemas a los que hemos tenido que hacer frente desde nuestra posición de usuarios.

Este artículo queda estructurado en tres apartados, además de la presente introducción. El primero de ellos gira en torno a la experiencia del investigador que utiliza los corpus académicos a fin de llevar a cabo un estudio sobre sintaxis histórica centrado en el análisis de palabras gramaticales. Se trata de sopesar, por un lado, los inconvenientes del empleo de un corpus informatizado y, por otro lado, los beneficios que conlleva el recurso a este. En segundo lugar, nos vamos a centrar en un caso concreto de evolución gramatical cuyos resultados quedan de alguna manera distorsionados debido al empleo del *CORDE*. Más concretamente, vamos a tratar acerca de ciertas cuestiones teóricas y prácticas surgidas al hilo del estudio de la gramaticalización de la perífrasis verbal *tener + (de/a) + infinitivo*. Por último, en el apartado 3, se exponen las conclusiones derivadas de este trabajo.

1. *CORDE* - *CREA* Y PALABRAS GRAMATICALES

La decisión de tomar el *CORDE* y el *CREA* como corpus base para el proyecto *GRAPEVERBA* respondía tanto al deseo de consultar el mayor número posible de fuentes como a la esperanza de obtener un número también elevado de ocurrencias que permitiese analizar de manera detallada el avance del cambio gramatical que conduce a la creación de las perífrasis verbales. Sin embargo, casi desde el principio nos topamos con dificultades no desdeñables: unas derivadas del propio diseño del *CORDE*; y otras, de las necesidades específicas del proyecto.

³ Se sigue así la máxima de Sinclair (1991), «bigger is always better», pero se pierde de vista que el tamaño del corpus, necesariamente, tiene que venir condicionado por las características y por los objetivos perseguidos por los investigadores. El corpus debe ser un utensilio al servicio del investigador, no un fin en sí mismo.

1.1. *CORDE* - *CREA*, etiquetado y lematización

En primer lugar, un trabajo sobre perífrasis verbales requiere que los motores de búsqueda del corpus puedan rastrear palabras gramaticales. En efecto, la pieza que determina la búsqueda es el verbo auxiliar, que muy a menudo se define por su polisemia entre verbo pleno y verbo gramatical. En la mayoría de las perífrasis verbales, la pieza que sirve como verbo auxiliar ha mantenido en la lengua el estrato lingüístico que llevó a la gramaticalización⁴. En otras palabras, en la lengua muy a menudo se conservan dos capas lingüísticas, la más antigua, la del verbo pleno, con significado conceptual, y la más moderna, la del verbo que ha experimentado el proceso de gramaticalización. El principal inconveniente del *CORDE* y del *CREA* en este aspecto es que, al no estar etiquetados, no permiten la búsqueda de palabras gramaticales, ni discriminan, en consecuencia, los casos de heterosemia⁵.

Somos conscientes de que este es un requisito de muy difícil cumplimiento, cualquiera que sea el corpus, ya que muy a menudo los términos que nos interesan se hallan en vías de gramaticalización, y los límites entre los valores antiguos y los modernos —los del verbo pleno y los del auxiliar— son ya no solo imprecisos, sino incluso imposibles de trazar. La dificultad para distinguir entre formas gramaticales y formas léxicas se observa incluso cuando se trata de un corpus sincrónico. Pensemos, por ejemplo, en el verbo *querer* en estructuras como *quiero saber la verdad*, ¿se trata de un verbo pleno o de un auxiliar? Por otra parte, no podemos ignorar los *contextos puente* (*bridging context*, vid. Heine 2002), esto es, los contextos que facilitan el paso del significado léxico al gramaticalizado, donde se generan piezas mixtas que comparten propiedades tanto del elemento léxico como del gramatical.

En este primer aspecto que acabamos de describir someramente, la inadecuación del *CORDE* y del *CREA*, o de cualquier otro corpus, a nuestros particulares intereses se revela insalvable. De hecho, estamos ante un problema circular: por una parte, necesitaríamos corpus etiquetados para avanzar en el conocimiento de la sintaxis histórica, pero, por otra, el etiquetado de un corpus histórico precisa de trabajos previos, como el nuestro, que permitan establecer el número y el tipo de valores con los que hay que trabajar.

A la vista de lo anterior, hemos tenido que rebajar nuestro nivel de exigencia y renunciar a la búsqueda de palabras gramaticales. Lo deseable habría sido, entonces, poder contar con una mínima categorización básica. No se trataría ya de que un verbo apareciera delimitado en sus valores como pleno o auxiliar, sino sencillamente, de que apareciera categorizado como verbo. Tal vez ello no solucionaría el problema de los casos en los que la categoría de llegada ha abandonado el ámbito verbal, como en el caso de *vamos* u *oye*, entre otros —con valores de marcador discursivo—. Y quizá tampoco

⁴ Es lo que se conoce como *layering* o estratificación, como Hopper (1991) denominó a esta característica del cambio gramatical.

⁵ Vid. Rojo (en prensa) para una exposición de los criterios que guiaron la elaboración del *CORDE*.

resolvería la dificultad en las expresiones fraseológicas con valor evidencial del tipo de *vete tú a saber*. Pero facilitaría, sin duda, la limpieza de los datos obtenidos para la mayoría de construcciones.

Tomemos como ejemplo la perífrasis *cuidar* + (preposición) + infinitivo. Estamos, en principio, ante un caso sencillo, porque *cuidar* no es un verbo del que vayamos a presumir una altísima frecuencia; y, por otra parte, al tratarse de un verbo de raíz regular, es posible simplificar la búsqueda, ya que el *CORDE*, efectivamente, permite en estos casos truncar la forma mediante asterisco. El total de ocurrencias que obtenemos a partir de *cuid**, con las correspondientes variantes ortográficas, *coid**, *cuyd**, *coyd**, es, como se muestra en la figura 1, de 50.839. Podemos tratar de averiguar cuántas de ellas corresponden a *cuidadoso* (adjetivo), *cuidadosamente* (adverbio) o *cuidado* (sustantivo), a sabiendas de que en esta última forma estamos incluyendo participios y formas compuestas, afortunadamente poco productivas en la creación de perífrasis. Si eliminamos las ocurrencias no verbales, la cifra se sitúa en 30.867 (*vid.* las filas sombreadas de la figura 1). Es decir, de los resultados que nos permite obtener la búsqueda truncada, previsiblemente más del 60% tendrá que ser eliminado manualmente, ello sin olvidar las ocurrencias de *cuydo* sustantivo, que, al coincidir con la primera persona del singular el presente de indicativo y con la tercera del indefinido, no son susceptibles de ningún tipo de filtrado previo, ya que, en este caso, se trata de formas absolutamente imprescindibles en el estudio de la formación de esta perífrasis.

<i>cuid*</i>	40.302	
<i>coid*</i>	102	
<i>cuyd*</i>	10.367	
<i>coyd*</i>	68	50.839
<i>cuidado/a*</i>	30.757	30.867
<i>coidado/a*</i>	29	
<i>cuydado/a*</i>	6.057	
<i>coydado/a*</i>	24	

Figura 1. Formas obtenidas en el *CORDE* para el verbo *cuidar*.

En el caso de *cuidar*, las búsquedas pueden agilizarse por medio del empleo de un comodín; sin embargo, en muchos otros casos, esto no es posible. En consecuencia, disponer de una mínima lematización es básico⁶. En nuestro caso, el hecho de que el *CORDE* no esté lematizado nos obliga a llevar a cabo una búsqueda a partir de formas verbales concretas. El vaciado de un corpus no lematizado resulta así una tarea tediosa, puesto que es preciso buscar las seis personas gramaticales de cada tiempo verbal, tanto en modo indicativo como en subjuntivo, sin olvidar las formas no personales del verbo.

⁶ Pese a su carácter rudimentario, el *ADMYTE* sí tenía una lematización que facilitaba la búsqueda de las formas verbales. También está lematizado el *ARTHUS* o el *Corpus del español*, citados en la nota 2. En el ámbito internacional, podemos citar corpus también lematizados—y con potentes motores de búsqueda— como el *Helsinki Corpus of English Texts*; el *Sketch Engine*; o el *COCA —The Corpus of Contemporary American English—*, sobre el inglés americano.

Eso supone una búsqueda de 113 formas diferentes, como puede comprobarse en la figura 2.

Presente de indicativo	Pretérito indefinido	Pret. imperfecto de indicativo	Pret. perfecto de indicativo	Pret. pluscuamp. de indicativo	Pretérito anterior
6 personas gramaticales	6 personas gramaticales	6 personas gramaticales	6 personas gramaticales	6 personas gramaticales	6 personas gramaticales
Futuro simple de indicativo	Futuro comp. de indicativo	Condicional simple	Condicional compuesto	Imperativo	Infinitivo Gerundio Participio
6 personas gramaticales	6 personas gramaticales	6 personas gramaticales	6 personas gramaticales	2 personas gramaticales	3 formas
Presente de subjuntivo	Pret. imperfecto de subjuntivo	Pret. perfecto de subjuntivo	Pret. pluscuamp. de indicativo	Futuro simple de subjuntivo	Futuro comp. de subjuntivo
6 personas gramaticales	6 personas gramaticales x doble forma	6 personas gramaticales	6 personas gramaticales x doble forma	6 personas gramaticales	6 personas gramaticales

Figura 2. Parámetros de búsqueda en los estudios sobre el sistema verbal.

Cabe añadir que, curiosamente, el *CORDE* discrimina en el uso de mayúsculas y minúsculas. Ello quiere decir que, si se opta por la búsqueda por formas, habrá que doblar cualquier cifra que se obtenga. Si, finalmente, al número ya particularmente elevado de ocurrencias que exige el estudio de una perífrasis verbal, le añadimos las variantes ortográficas que con frecuencia se encuentran en los textos de épocas pasadas (vacilaciones entre <v>, y <u> para los fonemas /b/-/β/, entre <i>, <j> e <y> para el fonema /i/, presencia o ausencia de acento, etc.), se entenderá fácilmente la necesidad que planteamos de disponer de una lematización y de un etiquetado que arrojen todas las posibles variantes del verbo simplemente introduciendo el infinitivo como lema.

1.2. *CORDE* - *CREA* y filtrado

Al sistema de búsqueda pormenorizada a que acabamos de referirnos, vienen a sumarse, en tercer lugar, los problemas del filtrado que hay que llevar a cabo, necesariamente, ante la imposibilidad de trabajar con el elevado número de ocurrencias que, como se verá, se obtienen en casi todos los casos. El corpus académico no ofrece grandes posibilidades. Si el número de ocurrencias permite el filtrado, las opciones son solo «filtrar por casos» o «por documentos»—con un número de ocurrencias elevado, solo es posible filtrar por documentos—.

REAL ACADEMIA

Resultado de la consulta al banco de datos
Cómo citar el CORPUS

Consulta:	aula, en todos los medios, en CORDE
Resultado:	3 casos en 3 documentos.

Ver estadística

Filtros: Casos Documentos

Ratio: 10

Mantener documentos (Solo para filtro sobre casos).

Filtrar

OBTENCIÓN DE EJEMPLOS

Recuperar

Concordancias: Normal

Clasificación: [] []

Agrupación: [] Marcas: []

Nueva consulta: [CREA](#) [CORDE](#) [Nómina de autores y obras](#) [Ayuda](#).

Figura 3. Ventana desde la que realizar el filtrado en el CORDE.

Pero el filtrado automático es «a ciegas»: no es posible decidir qué obras o qué siglos nos interesan más. Por otro lado, el filtrado por documentos deja completamente al margen el estudio de las tradiciones discursivas, que cada vez se están imponiendo más como piezas claves en la explicación del cambio sintáctico (*vid.*, por ejemplo, los trabajos contenidos en Kabatek 2008). Para contemplar ambas variables, documentos y tradiciones discursivas, tenemos que ignorar el filtrado automatizado e introducir otro parámetro de búsqueda desde «obtención de ejemplos». Sin embargo, este procedimiento ralentiza considerablemente la búsqueda de ocurrencias, ya que estas deben ser obtenidas manualmente. Por otra parte, este procedimiento de recuperación de ejemplos no funciona con números muy elevados de ocurrencias.

REAL ACADEMIA

Resultado de la consulta al banco de datos
Cómo citar el CORPUS

Consulta:	aula, en todos los medios, en CORDE
Resultado:	19683 casos en 939 documentos.

Ver estadística

Filtros: Documentos

Ratio: 10

Mantener documentos (Solo para filtro sobre casos).

Filtrar

OBTENCIÓN DE EJEMPLOS

Recuperar

Concordancias: Documentos Normal

Clasificación: [] []

Agrupación: [] Marcas: []

Nueva consulta: [CREA](#) [CORDE](#) [Nómina de autores y obras](#) [Ayuda](#).

Figura 4. Ventana desde la que realizar el filtrado desde la opción de «obtención de ejemplos».

Vistas las dificultades, podemos optar por prescindir de la obra y de la tradición discursiva como variables para nuestro estudio. Pero puede suceder que, incluso así, el elevado número de resultados de una búsqueda haga imposible filtrar de forma automatizada. En estos casos, es preciso fragmentar las búsquedas, lo cual supone multiplicar el centenar de formas por tantas búsquedas como tengamos que realizar, limitándolas a periodos cronológicos más cortos.

Veamos lo que esto representa en el estudio de las construcciones perifrásticas en que interviene *haber*. Tratándose de un verbo irregular, no existe prácticamente posibilidad de truncar la búsqueda a partir del empleo de comodines, y hay que localizar todas las formas del paradigma. Si nuestro estudio fuera del español actual, esto supondría la consulta del centenar de formas a que hemos aludido antes. Pero, en un estudio de sintaxis histórica como el nuestro, donde han de considerarse necesariamente las variantes con y sin <h>, y con , con <v> y con <u>, acentuadas o no, escritas con <i>, <j> o, incluso con <y>, la situación con la que nos encontramos es la siguiente:

HABER	FORMAS CONSULTADAS	FORMAS ENCONTRADAS	FORMAS NO ENCONTRADAS	OCURRENCIAS
INDICATIVO				
presente	26	26	0	737.230
pretérito imperfecto	72	66	6	377.262
pretérito indefinido	72	66	6	67.478
pretérito perfecto	156	64	92	4.766
pretérito pluscuamperfecto	432	54	378	1.320
pretérito anterior	432	13	419	22
futuro simple	72	70	2	42.553
futuro compuesto	432	20	412	74
condicional simple	132	92	40	18.184
condicional compuesto	792	12	780	39
TOTAL INDICATIVO	2.618	483	2.135	1.248.928

SUBJUNTIVO				
presente	14	13	1	80.126
pretérito imperfecto	192	142	50	66.547
pretérito perfecto	84	29	55	487
pretérito pluscuamperfecto	1.152	43	1.109	413
futuro simple	96	75	21	29.407
futuro compuesto	576	25	551	69
TOTAL SUBJUNTIVO	2.114	327	1.787	177.049
TOTAL CONJUNTO	4.732	810	3.922	1.425.977

Figura 5. Búsquedas en *CORDE*, y resultados obtenidos, para el estudio de la perífrasis *haber + (de/a) + infinitivo*.

Es decir, con independencia de los resultados obtenidos, hemos tenido que realizar 4.732 operaciones de búsqueda, de las cuales 3.922 han resultado infructuosas.

Por otra parte, las 810 que sí arrojan resultados, nos dan un total de 1.425.977 entradas. En este punto, nos resultaría utilísimo poder proceder como mínimo a un filtrado que, además, no fuera aleatorio, sino que permitiera controlar las obras, los siglos o, cuando menos, la tradición discursiva que más nos pueden interesar en cada momento. Pero, como hemos dicho, el *CORDE* solo da opción al filtrado por casos o documentos, y ello únicamente cuando no se supera un número determinado (pero variable) de entradas. Si decidimos añadir los documentos y tradiciones discursivas como variables, entonces la búsqueda es aún más lenta, ya que, como se ha señalado, desde «obtención de ejemplos», no es posible hacer una búsqueda conjunta en todas las obras en las que aparece el término objeto de análisis: es preciso trabajar por páginas, tal y como se muestra en la figura 6. Pero, además, ya hemos señalado que esta opción no siempre es factible, pues a partir de un número de ocurrencias no funciona.

The screenshot shows the 'OBTENCIÓN DE EJEMPLOS' (Obtaining Examples) section of the CORDE interface. At the top, a search bar shows the query 'aula, en todos los medios, en CORDE' and the result '2000 casos en 2000 documentos.' Below this, there are controls for document selection, including a 'Recuperar' button, a 'Documentos' dropdown menu set to 'Normal', and a 'Selección' dropdown menu set to '|x|'. There are also input fields for 'Clasificación', 'Agrupación', and 'Marcas'. Below the controls, a navigation bar indicates 'Pantalla: 1 de 40. Siguiente 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40'. The main content is a table with the following data:

	Casos	Año	Autor	Obra	País	Tema	Publicación
<input checked="" type="checkbox"/>	1	1935	Anónimo	Educación del lactante [Vida femenina. La revista de la mujer inteligente, 15 de febrero de 1935]	ARGENTINA	18.Normas de comportamiento	Vida femenina (Buenos Aires), 1935
<input type="checkbox"/>	1	1935	Anónimo	La vivienda rural [Vida femenina. La revista de la mujer inteligente, 15 de febrero de 1935]	ARGENTINA	18.Problemática social	Vida femenina (Buenos Aires), 1935

Figura 6. Selección de obras desde la opción de «obtención de ejemplos».

En consecuencia, esta indiscriminación de los datos resultantes nos aboca a un filtrado manual que no puede pasar únicamente por una reducción porcentual uniforme, ya que, de aplicarla, estaríamos prescindiendo de todas aquellas formas que presentan un número bajo (y no tan bajo) de ocurrencias. Puede servir de ejemplo de lo que decimos la siguiente tabla, en donde se hallan contenidos los datos del presente de *haber*⁷.

⁷ Hemos excluido de la consulta algunas formas del presente de <haber>, ya que distorsionarían sin duda los datos resultantes y las conclusiones que de ellos pudieran extraerse. Con respecto a las variantes sin <h>, hemos renunciado, para la primera persona del singular, a la búsqueda de <e> y <é>, dado que ambas variantes se solapan en los textos medievales con la conjunción copulativa; en cuanto a la forma <ei>, también de primera persona, la práctica totalidad de ocurrencias que arroja el *CORDE* se documenta en textos latinos y corresponde, en consecuencia, al dativo singular del anafórico de dicha lengua IS, EA, ID. De la segunda persona del singular hemos eliminado los resultados de <as>, por su coincidencia con el sustantivo y, de la tercera, los de <a> y <á>, por su concurrencia con la preposición. Finalmente, tampoco hemos incluido las formas de la segunda persona del plural <aues> y <aves>, cuya consideración introduciría un mayor número de animales alados que de formas verbales.

PRESENTE				
<i>he</i>	109.381	7,72765%	79	7,96%
<i>has</i>	30.210	2,13430%	22	2,22%
<i>ha</i>	343.616	24,27610%	247	24,90%
<i>hemos</i>	33.152	2,34215%	24	2,42%
<i>emos</i>	1.753	0,12385%	1	0,10%
<i>habemos</i>	4.077	0,28804%	3	0,30%
<i>abemos</i>	239	0,01689%	0	0,00%
<i>havemos</i>	575	0,04062%	0	0,00%
<i>avemos</i>	7.080	0,50019%	5	0,50%
<i>hauemos</i>	2.343	0,16553%	2	0,20%
<i>auemos</i>	4.695	0,33170%	3	0,30%
<i>habedes</i>	77	0,00544%	0	0,00%
<i>havedes</i>	60	0,00424%	0	0,00%
<i>hauedes</i>	300	0,02120%	0	0,00%
<i>abedes</i>	82	0,00579%	0	0,00%
<i>avedes</i>	1.871	0,13218%	1	0,10%
<i>auedes</i>	1.532	0,10823%	1	0,10%
<i>hedes</i>	305	0,02155%	0	0,00%
<i>edes</i>	214	0,01512%	0	0,00%
<i>habéis</i>	10.633	0,75121%	8	0,81%
<i>havéis</i>	442	0,03123%	0	0,00%
<i>hauéis</i>	29	0,00205%	0	0,00%
<i>abéis</i>	38	0,00269%	0	0,00%
<i>avéis</i>	2.468	0,17436%	2	0,20%
<i>auéis</i>	75	0,00530%	0	0,00%
<i>han</i>	149.815	10,58427%	108	10,89%
<i>an</i>	21.641	1,52891%	16	1,61%
TOTAL	726.703	51,34078%	522	

Figura 7. Filtrado de las formas del presente del verbo *haber*.

Este tiempo representa el 51,34% del total de entradas de *haber*. Arroja un total de 726.703 ocurrencias, distribuidas, en la proporción que se indica en la columna sombreada en verde, entre las 27 formas que han sido objeto de consulta y que han dado resultados superiores a 1. Si, para *haber*, queremos obtener un corpus total, por ejemplo, de 1.000 entradas, las 726.703 del presente deberán reducirse a 522, con lo cual, de aplicar una reducción proporcional directa, perderemos, como puede observarse en la columna sombreada en azul, formas con más de 440 ocurrencias (el caso de *havéis*). Es más, en otros tiempos, el futuro simple de indicativo sin ir más lejos, un filtro de ese tipo supondría la práctica desaparición de todas sus formas, tal y como queda reflejado en la figura 8.

FUTURO SIMPLE				
<i>abré</i>	96	0,00673%	0	0,00%
<i>avré</i>	324	0,02272%	0	0,00%
<i>auré</i>	57	0,00400%	0	0,00%
<i>habré</i>	990	0,06943%	1	0,10%

<i>havré</i>	47	0,00330%	0	0,00%
<i>hauré</i>	2	0,00014%	0	0,00%
<i>abre</i>	5.286	0,37069%	4	0,40%
<i>avre</i>	87	0,00610%	0	0,00%
<i>aure</i>	150	0,01052%	0	0,00%
<i>habre</i>	33	0,00231%	0	0,00%
<i>havre</i>	5	0,00035%	0	0,00%
<i>haure</i>	21	0,00147%	0	0,00%
<i>abrás</i>	97	0,00680%	0	0,00%
<i>avrás</i>	522	0,03661%	0	0,00%
<i>aurás</i>	33	0,00231%	0	0,00%
<i>habrás</i>	850	0,05961%	1	0,10%
<i>havrás</i>	53	0,00372%	0	0,00%
<i>haurás</i>	4	0,00028%	0	0,00%
<i>abras</i>	278	0,01950%	0	0,00%
<i>avras</i>	332	0,02328%	0	0,00%
<i>auras</i>	651	0,04565%	0	0,00%
<i>habras</i>	14	0,00098%	0	0,00%
<i>hauras</i>	26	0,00182%	0	0,00%
<i>abrá</i>	644	0,04516%	0	0,00%
<i>avrá</i>	2.336	0,16382%	2	0,20%
<i>aurá</i>	547	0,03836%	0	0,00%
<i>habrá</i>	14.065	0,98634%	10	1,00%
<i>havrá</i>	641	0,04495%	0	0,00%
<i>haurá</i>	34	0,00238%	0	0,00%
<i>abra</i>	1.834	0,12861%	1	0,10%
<i>avra</i>	735	0,05154%	1	0,10%
<i>aura</i>	3.005	0,21073%	2	0,20%
<i>abra</i>	466	0,03268%	0	0,00%
<i>havra</i>	46	0,00323%	0	0,00%
<i>haura</i>	383	0,02686%	0	0,00%
<i>abremos</i>	100	0,00701%	0	0,00%
<i>avremos</i>	289	0,02027%	0	0,00%
<i>auremos</i>	151	0,01059%	0	0,00%
<i>habremos</i>	729	0,05112%	1	0,10%
<i>havremos</i>	30	0,00210%	0	0,00%
<i>hauremos</i>	35	0,00245%	0	0,00%
<i>abredes</i>	55	0,00386%	0	0,00%
<i>avredes</i>	157	0,01101%	0	0,00%
<i>auredes</i>	147	0,01031%	0	0,00%
<i>habredes</i>	9	0,00063%	0	0,00%
<i>havredes</i>	3	0,00021%	0	0,00%
<i>hauredes</i>	10	0,00070%	0	0,00%
<i>abréis</i>	31	0,00217%	0	0,00%
<i>avréis</i>	127	0,00891%	0	0,00%
<i>habréis</i>	578	0,04053%	0	0,00%
<i>havréis</i>	22	0,00154%	0	0,00%
<i>hauréis</i>	4	0,00028%	0	0,00%
<i>abreís</i>	4	0,00028%	0	0,00%
<i>avreís</i>	16	0,00112%	0	0,00%
<i>aureís</i>	59	0,00414%	0	0,00%
<i>habreís</i>	54	0,00379%	0	0,00%
<i>havreís</i>	2	0,00014%	0	0,00%
<i>haureís</i>	9	0,00063%	0	0,00%
<i>abran</i>	616	0,04320%	0	0,00%

<i>avran</i>	272	0,01908%	0	0,00%
<i>auran</i>	524	0,03675%	0	0,00%
<i>habran</i>	57	0,00400%	0	0,00%
<i>havran</i>	16	0,00112%	0	0,00%
<i>hauran</i>	222	0,01557%	0	0,00%
<i>abrán</i>	141	0,00989%	0	0,00%
<i>avrán</i>	637	0,04467%	0	0,00%
<i>aurán</i>	64	0,00449%	0	0,00%
<i>habrán</i>	2.546	0,17854%	2	0,20%
<i>havrán</i>	135	0,00947%	0	0,00%
<i>haurán</i>	8	0,00056%	0	0,00%

Figura 8. Filtrado del futuro simple de indicativo del verbo *haber*

En conclusión: el hecho de que el *CORDE* no permita otro filtrado que el de casos o documentos, obliga indefectiblemente al investigador a una criba manual que, en último extremo, puede ocasionar una notable distorsión de los datos objeto de análisis y de las conclusiones lingüísticas que vayan a extraerse a partir de ellos. El objetivo de obtener resultados más fiables a partir de la consulta del corpus se ve así seriamente menoscabado.

1.3. *CORDE* - *CREA*, obtención de resultados

Ante las dificultades que señalamos, hemos ensayado, para *haber*, un sistema de búsqueda controlada o, dicho de otro modo, hemos limitado la búsqueda de formas a algunas de las obras más representativas de los distintos estadios históricos y de las distintas tradiciones discursivas del español (*vid.* el apéndice, al final de este trabajo). Este procedimiento cuenta con la ventaja de que es el investigador el que, desde el principio, controla el filtrado, basándose en los distintos presupuestos teóricos que son aplicables a la historia de la lengua española. Por otra parte, el número de ocurrencias que se obtienen al final se ve reducido notablemente, resultando, en consecuencia, algo más manejable: frente al 1.425.977 de la búsqueda indiscriminada, se obtienen, por este procedimiento, un total de 66.437 casos, con la distribución que podemos observar en la tabla siguiente:

Indicativo	obras	oral	total
presente	23.277	39	23.316
pretérito imperfecto	20.746	709	21.455
pretérito indefinido	10.177	314	10.491
pretérito perfecto	155	168	323
pretérito pluscuamperfecto	135	20	155
pretérito anterior	3	0	3
futuro simple	943	370	1.313
futuro compuesto	0	6	6
condicional simple	992	141	1.133
condicional compuesto	0	2	2
	56.428	1.769	58.197
subjuntivo			
presente	2.273	512	2.785
pretérito imperfecto	4.160	614	4.774
pretérito perfecto	7	19	26
pretérito pluscuamperfecto	22	12	34
futuro simple	618	3	621
futuro compuesto	0	0	0
	7.080	1.160	8.240
totales			
indicativo + subjuntivo	63.508	2.929	66.437

Figura 9. Búsquedas de *haber* en *CORDE*, y resultados obtenidos, aplicando el filtrado por obras.

Una última ventaja radica en el hecho de que el conjunto de formas obtenidas — algo más manejable, como decíamos, pero sin duda todavía elevado— puede ser sometido en última instancia a una reducción porcentual discriminada. Es posible, en efecto, en función de las necesidades del objeto de estudio, reducir las ocurrencias por obras o por periodos históricos de forma, ahora sí, totalmente controlada.

Ahora bien, pese a que el procedimiento del filtrado previo por obras se revela más conveniente para nuestro estudio que cualquiera de las opciones de filtrado que ofrece el *CORDE*, en su aplicación hemos topado con algunos obstáculos derivados todavía de ciertas limitaciones e imperfecciones del corpus académico. El primero de ellos estriba en el hecho de que, cuando la búsqueda de una forma arroja más de 1.000 casos en una sola obra, el *CORDE* no permite recuperar, automáticamente, concordancias. A falta de otra herramienta, en esas ocasiones nos hemos visto forzados a buscar la forma requerida seguida sucesivamente de todas y cada una de las letras del alfabeto —o de la combinación incluso de dos letras— hasta obtener fracciones inferiores a 1.000 (*vid.* figura 10). Solo así hemos podido recuperar las concordancias, no del total, pero sí de un alto porcentaje de los casos presumiblemente existentes. Por lo demás, este procedimiento de búsqueda, ciertamente pedestre, solo es efectivo hasta el siglo XIX y parte del XX, porque, inexplicablemente, no es posible aplicarlo en el *CREA*.

		Hasta finales siglo xv			Siglo xix				Siglo xx
		<i>Crónica del rey don Rodrigo</i>	<i>Istoria de las bienandanzas e fortunas</i>		<i>La Regenta</i>	<i>Fortunata y Jacinta</i>	<i>El Bandolerismo. Estudio social y memorias históricas</i>	<i>Los aires difíciles</i>	
		<i>avía</i>	<i>avía</i>	<i>ovo</i>	<i>había</i>	<i>había</i>	<i>ha</i>	<i>había</i>	
<i>a</i>					143				
	<i>ab</i>	0	0	0		6	1	0	
	<i>ac</i>	15	9	5		13	4	18	
	<i>al</i>	6	18	16		15	2	6	
	<i>am</i>	1	1	0		0	1	0	
	<i>an</i>	0	4	1		2	0	3	
	<i>as</i>	2	1	1		3	3	6	
	<i>at</i>	0	1	0		5	1	5	
	<i>av</i>	1	15	1		0	1	4	
<i>b</i>		12	4	17	15	11	4	6	
<i>c</i>									
	<i>ca</i>	10	8	3	36	23	13	30	
	<i>ce</i>	1	0	0	8	5	0	2	
	<i>ci</i>	5	0	0	1	0	2	2	
	<i>co</i>	28	28	65	93	41	49	100	
	<i>cr</i>	1	16	0	20	8	8	18	
	<i>cu</i>	8	2	0	4	3	3	11	
<i>d</i>		226		191					
	<i>da</i>		20		26	28	19	33	
	<i>de</i>		138		144	137	135	181	
	<i>di</i>		18		84	37	65	48	
	<i>do</i>		7		11	8	0	2	
	<i>du</i>		1		4	0	1	2	
<i>e</i>		95	106	278	151	123	64	165	
<i>f</i>		60	117	304	22	7	7	10	
<i>g</i>		33	34	38	13	7	4	15	
<i>h</i>		10	5	2	121	67	62	72	
<i>i</i>		5	5	0	50	33	25	63	
<i>j</i>		3	6	0	9	5	1	6	
<i>l</i>		57	20	11	67	40	34	78	
<i>m</i>		103	84	162	101	74	30	70	
<i>n</i>		54	16	24	29	24	4	10	
<i>o</i>		21	35	67	62	38	28	81	
<i>p</i>				48	202	110			
	<i>pa</i>	18	16				47	17	
	<i>pe</i>	23	13				22	43	
	<i>pi</i>	1	1				1	1	
	<i>pl</i>	3	1				0	0	
	<i>po</i>	12	17				45	27	
	<i>pr</i>	2	13				45	99	

	pu	6	11			7	14	
q		39	29	10	114	50	17	32
r		16	21	8	77	29	29	76
s		30	62	27	222	111	184	199
t		46	44	18	81	62	45	114
u		20	27	46	31	30	0	20
y		8	9	0	2	2	0	13
z		0	0	0	1	0	0	1

Recuperados	981	983	1.343	1.944	1.157	1.013	1.703	
Totales	1.135	1.100	1.438	2.093	1.307	1.084	1.799	2.177
%	86,43%	89,36%	93,39%	92,88%	88,52%	93,45%	94,66%	

Figura 10. Recuperación de formas con más de 1.000 ocurrencias en una sola obra.

Un segundo inconveniente importante es el de las obras mismas y, más concretamente, el de las ediciones textuales referenciadas en el *CORDE*. En relación con este aspecto pueden reseñarse varias cuestiones:

- Es necesario que la búsqueda corresponda exactamente al título introducido en el *CORDE*. No es posible, por poner un ejemplo evidente, buscar *El Quijote* sino *El ingenioso hidalgo don Quijote de la Mancha* porque, de lo contrario, la búsqueda se realiza indefectiblemente en todas las obras que contienen *El Quijote* en el título (el de Avellaneda, sin ir más lejos).
- Hemos localizado también duplicados de una misma obra, pero con alguna diferencia en los datos. Por ejemplo, para el siglo xx, la *Carta de Salinas* [Cartas. En «Correspondencia (1923-1951)»] es reseñada dos veces con un total de palabras distinto, pese a que, como puede verse en la figura 11, aparentemente, se trata de la misma obra y de la misma edición:

Criterio de clasificación	Fecha de clasificación	Autor	Título	Edición empleada	País	Clasificación temática	Número de palabras
Fecha de escritura	1923	Salinas Pedro	<i>Carta de Salinas</i> . [Cartas. En «Correspondencia (1923-1951)»]	Andrés Soria Olmedo, Barcelona: Tusquets, 1992	España (Madrid)	HISTORIA Y DOCUMENTOS: Cartas y relaciones	885
Fecha de escritura	1923	Salinas Pedro	<i>Carta de Salinas</i> . [Cartas. En «Correspondencia (1923-1951)»]	Andrés Soria Olmedo, Barcelona: Tusquets, 1992	España (Madrid)	HISTORIA Y DOCUMENTOS: Cartas y relaciones	873

Figura 11. Datos correspondientes a la *Carta a Salinas* (*CORDE*).

- Problemas de cronología y de autoría. Por ejemplo, en el siglo xviii, existen varios documentos de distinto autor, y con distintas fechas, titulados todos ellos *Carta*

a *Rodríguez Campomanes* [Epistolario de Pedro Rodríguez Campomanes 1774-1777].

- Asignaciones controvertidas de algunas obras a determinadas tradiciones discursivas. Este es un problema general del corpus académico, de manera que a veces la tipología textual establecida no se ajusta a la norma sobre géneros discursivos. Por poner solo un par de ejemplos, podemos traer a colación el *Testamento del pintor Felices de Cáceres* que, pese a ser un testamento, y por lo tanto, un documento notarial, es consignado dentro de los «Documentos sobre pintores y pintura del siglo XVI». El segundo ejemplo lo constituye un texto de Lope de Vega, concretamente, un croquis de dos obras teatrales (*Esquema de dos piezas teatrales de Lope Félix de Vega Carpio*), que es incluido dentro de los textos notariales⁸.
- Inclusión de ediciones no enteramente fiables. A esta cuestión dedicamos el próximo apartado.

2. CUESTIONES TEÓRICAS. LAS EDICIONES TEXTUALES Y EL EMPLEO DEL *CORDE*

Las dificultades señaladas en el apartado anterior no son el único problema que plantea el estudio de palabras gramaticales a partir del *CORDE*. Otra cuestión no baladí es la que apuntábamos al final de dicho apartado relativa a las ediciones textuales que se toman como referencia. A nadie se le escapa que, en función de las ediciones que se manejen, la explicación de ciertos procesos evolutivos puede verse alterada. En general, el problema fundamental tiene que ver con la datación de los fenómenos objeto de estudio, pero se puede llegar, incluso, a distorsionar las hipótesis explicativas de un cambio gramatical en particular.

No es nuevo que el *CORDE* emplea ediciones textuales que no siempre se ajustan a la fecha para la que son ilustrativas. De manera especial para la etapa medieval no es infrecuente que se tomen ediciones que no coinciden con el manuscrito original ni con la copia más próxima en el tiempo al texto original. De esta manera, se incumple con un requisito básico para fechar los fenómenos históricos con un mínimo de precisión; el de la representatividad de las fuentes. Como se lee en Fernández-Ordóñez (2006: 1795):

⁸ Estas atribuciones pueden ser una consecuencia inmediata del motivo para el que se creó el *CORDE*, a saber, ser la base de un futuro diccionario histórico. Pese a que el *CORDE* no explicita los criterios que sigue para elaborar la tipología textual, el propósito inicial habría favorecido una clasificación de los textos basada en criterios temáticos, no textuales.

Para que la historia de la lengua avance sobre sólidas bases debemos, pues, tratar de garantizar la representatividad de nuestras fuentes, procurando que sean originales, o si no, contemporáneas del momento de composición.

La «contemporaneidad» la cifra Fernández-Ordóñez en un lapso temporal no superior al medio siglo entre el texto original y el texto copiado:

El problema está en determinar qué entendemos por contemporáneas. Atendiendo a las constataciones que hoy nos proporciona la sociolingüística sobre el período mínimo de difusión de los cambios estudiados en tiempo aparente, me atrevería a proponer a copias que no hayan rebasado la frontera del medio siglo entre la composición de la obra y la transcripción del código. Dentro de la historiografía, no son pocas las obras que cumplen esa condición (Fernández-Ordóñez 2006: 1788-1789).

Sin embargo, muchas de las ediciones volcadas en el *CORDE* corresponden a una copia que puede distar considerablemente de la fecha de composición del texto inicial. En nuestro estudio de las perífrasis verbales, la consulta de ediciones textuales que ni se basan en textos originales, ni cumplen con el criterio de contemporaneidad con el original da lugar con frecuencia a resultados algo distorsionados. Así, por ejemplo, los datos del *CORDE* relativos a la aparición de la perífrasis *tener + (de/a) + infinitivo* ratifican los obtenidos años antes por Alicia Yllera (1980: 111-112) en su monografía sobre la evolución de las perífrasis verbales. Concretamente, la consulta del *CORDE* arroja los resultados expuestos en la figura 11 acerca de la documentación de la perífrasis *tener + de/a + infinitivo* en la etapa medieval. De acuerdo con estos datos, la perífrasis surgiría en el siglo XIII e incrementaría levemente su presencia en la centuria siguiente.

	<i>Tener + de + infinitivo</i>	<i>Tener + a + infinitivo</i>
Siglo XIII	85,7% (12/14)	14,3% (2/14)
Siglo XIV	89,2% (33/37)	10,8% (4/37)

Figura 12. Documentación de *tener + de/a + infinitivo* en la Edad Media, según datos extraídos del *CORDE*.

Sin embargo, tanto el corpus académico como la obra de Yllera (1980) se basan en unas ediciones textuales que no cumplen con las exigencias trazadas por Fernández-Ordóñez (2006), a saber, no se trata de fuentes representativas, pues no son ni originales ni contemporáneas al texto original. En nuestro caso, el lapso temporal entre original y copia es de uno o dos siglos, según los casos. Los textos del XIII en los que se documenta la perífrasis en el corpus académico son los siguientes:

- *La General Estoria. Parte dos y parte cinco.*
- *La Estoria de Espanna que fizo el muy noble rey don Alfonso, fijo del rey don Fernando et de la reyna donna Beatriz.*
- *La Estoria de España, II.*
- *El Espéculo.*

- *Los Castigos e Documentos para bien vivir.*
- *El Calila e Dimna.*

Todas estas obras plantean problemas ecdóticos no soslayables, bien conocidos por los investigadores. Pese a ello, en el *CORDE* en ningún caso se especifica que se toman ediciones que, aunque se dan como representativas de textos del siglo XIII, proceden en realidad de copias posteriores. Así se puede ver si se consulta la nómina de autores y obras, extraída del corpus académico y que reproducimos en la figura 12.

Criterio de clasificación	Fecha de clasificación	Autor	Título	Edición empleada	País	Clasificación temática	Número de palabras
Fecha de escritura	c1275	Alfonso x	<i>General Estoria. Segunda parte</i>	Pedro Sánchez-Prieto Borja, Alcalá de Henares: Universidad de Alcalá, 2003	España (Toledo)	HISTORIA Y DOCUMENTOS: Historiografía	249.564
Fecha de escritura	a1284	Alfonso x	<i>General Estoria. Quinta parte</i>	Pedro Sánchez-Prieto, Alcalá de Henares: Universidad de Alcalá, 2002	España (Toledo)	HISTORIA Y DOCUMENTOS: Historiografía	197.268
Fecha de escritura	c1270	Alfonso x	<i>Estoria de Espanna que fizo el muy noble rey don Alfonso, fijo del rey don Fernando et de la reyna donna Beatriz</i>	Pedro Sánchez-Prieto, Alcalá de Henares: Universidad de Alcalá de Henares, 2002	España (Toledo)	HISTORIA Y DOCUMENTOS: Historiografía	259.440
Fecha de escritura	1270-1284	Alfonso x	<i>Estoria de España, II</i>	Lloyd A. Kasten; John J. Nitti, Madison: Hispanic Seminary of Medieval Studies, 1995	España	HISTORIA Y DOCUMENTOS: Historiografía	354.531
Fecha de escritura	a1260	Anónimo	<i>Espéculo de Alfonso X. BNM 10123</i>	Pedro Sánchez-Prieto Borja, Alcalá de Henares: Universidad de Alcalá de Henares, 2004	España	DERECHO: Tratados jurídicos	204.677
Primera edición	1293	Anónimo	<i>Castigos e documentos para bien vivir ordenados por el rey Sancho IV</i>	Agapito Rey, Indiana: Indiana University Publications, 1952	España	PROSA DIDÁCTICA: Literatura sapiencial	79.181
Fecha de escritura	1293	Anónimo	<i>Castigos. BNM ms. 6559</i>	William Plamer y Carig Frazier, Madison: Hispanic Seminary of Medieval Studies, 1992	España	PROSA DIDÁCTICA: Literatura sapiencial	156.180
Fecha de escritura	1251	Anónimo	<i>Calila e Dimna</i>	Juan Manuel Cacho Bleuca; María Jesús Lacarra, Madrid: Castalia, 1993	España	PROSA NARRATIVA BREVE: Tradicional	75.235

Figura 13. Nómina de obras en las que se documenta la perífrasis *tener + (de/a) + infinitivo* en el *CORDE*.

En el caso de la *General Estoria*, de acuerdo con Fernández-Ordóñez, de esta obra solo pueden tomarse en consideración la primera y cuarta partes.

Los dos volúmenes a que me refiero son uno de la primera parte, el ms. A (ms. 816, Biblioteca Nacional, Madrid, s. XIII), editado en 1930 por Solalinde y en 2001 por Sánchez-Prieto, y otro, más lujoso, de la Cuarta parte, el ms. U, fechado en 1280 (ms. Urb. Lat. 539, Biblioteca Apostolica Vaticana, Roma), del que contamos con una transcripción desde 1978 y con ediciones críticas parciales, por el momento (Fernández-Ordóñez 2006: 1785).

Significativamente, en la primera y cuarta partes de la *General Estoria*, que en el *CORDE* se basan en la edición crítica de Sánchez-Prieto, no se recoge ejemplo alguno de *tener + (de/a) + infinitivo*. Los ejemplos proceden de la segunda y quinta partes. Y precisamente de la segunda parte señala Fernández-Ordóñez que:

A la vista de estas joyas originales del *scriptorium* alfonsí, no se entiende bien que se persista en emplear como fuente de datos lingüísticos la segunda parte de la *General estoria*, cuyos testimonios más antiguos datan ya del siglo XIV y cuya única edición existente hasta el momento toma como pauta, además, dos manuscritos diferentes, de muy desigual valor textual (Fernández-Ordóñez 2006: 1785).

La *Estoria de España*, editada en tres ocasiones por Menéndez Pidal con el nombre de *Primera Crónica General* (Menéndez Pidal 1906, 1955, 1977), tampoco está exenta de problemas. Pidal se basó para su edición en dos manuscritos escurialenses. Uno de ellos —el E1 (ms. Y-I-2, Biblioteca del Monasterio de San Lorenzo de El Escorial)— es un códice original del *scriptorium* alfonsí; el otro⁹, en cambio, resulta más problemático, ya que, como probó Diego Catalán (1962: 19-203, 1997: 29-32), no es un original del XIII, sino que se elaboró bajo el reinado de Alfonso XI¹⁰. En consecuencia, como se prueba en Fernández-Ordóñez (2006) y en Rodríguez Molina (2010), de la edición de Menéndez Pidal solo podemos tomar el primer volumen y los 50 primeros capítulos del segundo volumen (caps. 1-565 y 566-616, hasta la p. 350a18) como testimonios representativos del siglo XIII:

Según la edición de Menéndez Pidal, únicamente el primer volumen y los primeros 50 capítulos del segundo pertenecen a la *Estoria de España*, en su versión primitiva (Caps. 1-565 y 566-616, hasta la pág. 350a18). El resto del texto editado por don Ramón en base al testimonio del códice E2 constituye un conglomerado de materiales cronísticos elaborados en distintas épocas y con materiales diversos, pero no es ya la *Estoria de España* tal y como la concibió Alfonso X (Rodríguez Molina 2010: 806).

Pese a ello, como podemos comprobar en la figura 12, toda la *Estoria de España* se da como documento representativo de dicha centuria, identificándose así la *Primera Crónica General* con la *Estoria de España*, tal y como denunciaba Rodríguez Molina (2010).

⁹ Se trata del manuscrito E2 (ms. X-I-4, Biblioteca del Monasterio de San Lorenzo de El Escorial).

¹⁰ El tema ha sido tratado por Fernández-Ordóñez (2006) y Rodríguez Molina (2010).

Los datos no resultan mucho más halagüeños si nos basamos en el *Espéculo*, ya que la edición que se toma en el *CORDE* se basa en el manuscrito BNM 10123, que es una copia de finales del siglo XIV (vid. Craddock 1998). Otro tanto puede señalarse para los *Castigos y documentos*, cuya versión más antigua data del siglo XV, de acuerdo con Sánchez (2005):

En cuanto al soporte manuscrito, parece el más antiguo, datable, a tenor de sus miniaturas, hacia 1430 (Sánchez Sánchez 2005: 54).

En cuanto al *Calila e Dimna*, ya Yllera subrayó la poca fiabilidad de los datos obtenidos de una obra cuyos manuscritos datan de finales del siglo XIV o del siglo XV. Concretamente, estos manuscritos son el A (ms. h-III-9, Biblioteca del Monasterio de San Lorenzo (El Escorial, Madrid) —códice de finales del siglo XIV o principios del XV— y el B (ms. x-III-4, Biblioteca del Monasterio de San Lorenzo (El Escorial, Madrid) —códice del siglo XV, terminado de copiar en 1487—.

Por otra parte, hay que señalar otras dificultades derivadas del hecho de que el *CORDE* mezcle ediciones críticas (por ejemplo, la *GE2*, el *Calila* o la *GE5*) con transcripciones de manuscritos únicos (es el caso del *Espéculo*) o que contienen varias obras (es el caso de la *Estoria de España II*). Como consecuencia de ello, los datos que arroja, pongamos por caso la *G2*, ofrecen diferencias significativas en función de la sección o del manuscrito de base que se tome en consideración (la primera parte de la *GE2*, basada en K, es de principios del XIV, pero las secciones posteriores, basadas en N, son de un siglo posterior). La *Estoria de España II* ofrece una situación similar, pues transcribe el códice escurialense x-I-4, que contiene textos de diferentes épocas y copiados por diferentes manos. En casos como estos, al problema de la distancia cronológica entre original y copia se le suma el de la mezcla indiscriminada de textos, manuscritos y ediciones sin un criterio claro.

Partiendo de estos presupuestos de crítica textual, si prescindimos de las ediciones no fiables, los resultados que obtenemos para nuestro estudio acerca del surgimiento de *tener + de + infinitivo* como perífrasis verbal, resultan bien diferentes a los apuntados en la figura 11. En efecto, como se ha señalado, en las partes de la *General Estoria* que cumplen con los requisitos marcados por la crítica textual, a saber la primera y la cuarta, no se documenta la forma perifrástica que nos ocupa. Y si se desestiman las ediciones de la *Estoria de España II*, del *Espéculo*, de los *Castigos* y del *Calila e Dimna*, únicamente queda un ejemplo de la perífrasis, procedente de la primera parte de la *Estoria de España*:

- (1) Mas porque nos fizimos aqui remembrance de los longobardos —de los que dexamos a contar ca no uuiamos aun por el tiempo en que *tenemos de dezir* desta estoria en que somos, et por esto no lo quisimos meter en obliido (Alfonso X, *Estoria de España, I*, apud Menéndez Pidal (1977: 242).

En consecuencia, la presencia de la perífrasis en el siglo XIII, se convierte en un breve apunte, cuando menos en la lengua escrita, de modo que el surgimiento de la perífrasis parece concentrarse en el siglo siguiente y, más claramente, en el siglo XV (vid. Garachana

y Rosemeyer en prensa). Estos datos advierten claramente de la importancia de atender a ediciones basadas en manuscritos próximos a la fecha de edición del texto. No hay que olvidar que los copistas medievales intervienen en su creación tímidamente, alterando la grafía y la fonética, o profundamente, llegando a crear una nueva obra (Fernández-Ordóñez 2006: 1783). Esto resulta particularmente manifiesto si se coteja la edición de la *Primera Partida* de Arias Bonet, basada en un manuscrito del XIII, el Add. 20.787 del *British Museum*, con la edición de las *Siete Partidas* de Sánchez-Prieto, basada en el manuscrito BNM 1766, y fundamentada en una copia del 1491¹¹. La edición que sigue el manuscrito del *British* presenta la perífrasis *aver + de + infinitivo* allí donde la edición que parte del manuscrito BNM 1766 ofrece *tener + de + infinitivo* (vid. figura 13). Más aún, en la edición del XV hay añadidos en los que se introduce un *tener + de + infinitivo* que da cuenta de la cada vez más sólida extensión de esta construcción verbal como perífrasis de obligación (vid. Garachana y Rosemeyer en prensa).

Edición <i>British</i>	Edición BNM 1766
Ley xxviii ^a : En cuántas maneras puede seer la sentencia no derecha e qué pena deue auer el prelado que la diesse. E por ende touo por bien Sancta Iglesia de mostrar en cuántas maneras es la sentencia non derecha, por que aquellos que la an a dar se sopiessen guardar de la dar en manera que no fuesse contra derecho.	Ley .xx. que en quantas maneras se da la sentencia descomunión iniustamente: & que pena deue auer el prelado que la pone. E porende touo por bien santa iglesia demostrar en quantas maneras es la sentençia non derecha porque aquellos que la dan: o la tienen de dar se sepan guardar della y son tres.
Ley vi ^a De quales omnes manda el derecho de las leyes antiguas sacar de la iglesia Yerroz muy grandes fazen los onbres a las uegadas sin los que dize en la ley ante desta, por que han de fuyr a las yglesias e por esso manda el derecho de las leyes antiguas que los saquen ende sin calonna ninguna assí cuemo a los traydores connoçudos e los que matan a otri a tuerto e los adúlteros e los que fuerçan las uírgenes & e los que han adar a los emperadores & a los reyes sus tributos o sus pechos e otrosí los que les an a dar cuenta dellos.	Ley quinta. quales onbres manda el derecho de las leyes antiguas sacar de la yglesia. yerroz muy grandes fazen los onbres a las vezes sin los que dizen en la ley ante desta porque han de foyr alas yglesias temiendo de pena: & por esto mando el derecho de las leyes antiguas que los saquen dellas sin calupnia ninguna asy commo los traydores connoçidos: & los que matan a otro atuerto & los adulteradores & los que fuerçan virgines & los que tienen de dar cuenta a los enperadores & a los reyes de sus tributos o de sus pechos.
E esto es por que él a dar cuenta a Dios della.	& esto es porque el tiene de dar cuenta a dios dello.

Figura 14. Cotejo de las ediciones del siglo XIII y del XV de la *Primera Partida*.

La comparación de estas dos ediciones pone en tela de juicio la validez de las manifestaciones de *tener + (de/a) + infinitivo* en el siglo XIII obtenidas a partir del *CORDE*, exceptuando, por supuesto, el ejemplo de la *Estoria de España*. Otro argumento que incidiría en la misma dirección sería la ausencia de la perífrasis en las tres biblias

¹¹ Para un estudio minucioso de los cambios que experimentan los manuscritos del XIII copiados o impresos a finales del XV - principios del XVI, vid. el trabajo de Harris-Northall (1996) sobre la *Gran Conquista de Ultramar*.

medievales que el corpus *Biblia Medieval* ofrece para el siglo XIII. Ante esta constatación, se precisa un estudio de la evolución de la perífrasis verbal *tener + de + infinitivo*, al margen del corpus académico, basado en ediciones textuales que permitan una datación más certera de los fenómenos tratados (así se hace en Garachana y Rosemeyer en prensa). Como hemos podido comprobar, las ediciones del siglo XV —muy probablemente también las del XIV— introducen modificaciones en los textos que remitirían a una práctica extendida en la lengua de la época, a saber, el empleo de *tener + (de/a) + infinitivo* como perífrasis de obligación. El copista habría modificado la lengua del texto original acercándola de este modo a su variedad lingüística contemporánea.

En consecuencia, tan discutible puede ser introducir las obras arriba mencionadas como representativas del siglo XIII como fecharlas a partir del año de su copia. Sin embargo, resultaría muy útil introducir en los corpus alguna advertencia para que el investigador tome conciencia del lapso de tiempo que media entre la composición de la obra y la copia conservada para la investigación. Esta indicación podría ofrecer nuevas perspectivas de estudio, puesto que podría permitir, por ejemplo, determinar el grado de asentamiento de un cambio gramatical. Así, en el caso de la perífrasis *tener + de + infinitivo*, si en una copia del XV la construcción perifrástica con *tener* es más frecuente que en obras conservadas en manuscritos del XIII, esto da fe de la pujanza de la nueva construcción, que puede llegar a desplazar a la que aparece en el texto que se está copiando.

Así pues, el caso aquí descrito vuelve a poner sobre la mesa la cuestión de cómo deben manejarse en el estudio de la sintaxis histórica las ediciones basadas en copias de manuscritos cuando entre el original y la copia dista un lapso temporal amplio —o cuando disponemos de más de una versión de un mismo texto antiguo—. Más allá de la crítica, se impone buscar una solución.

3. CONCLUSIÓN

El manejo de los corpus académicos para el estudio diacrónico debería ofrecer al investigador ventajas sustanciales respecto del vaciado manual de textos. Sin embargo, cuando se trata de los corpus de la Real Academia Española, las propias características de estos hacen que dichas supuestas ventajas queden en un muy segundo plano. La falta de etiquetado y de lematización de estos corpus puede obligar a realizar una muy laboriosa búsqueda manual: al no discriminarse automáticamente la categoría gramatical de la forma buscada y al no ofrecerse una estructuración en lemas, los ejemplos obtenidos pueden rebasar ampliamente las posibilidades de un análisis pormenorizado, de manera particular si la pieza de búsqueda es una palabra gramatical. Más todavía, aunque no puede negarse que la consulta del *CORDE* y del *CREA* permite disponer de un importante número de ocurrencias distribuidas en etapas cronológicas y tradiciones discursivas diversas, queda abierta la cuestión de si los resultados obtenidos no hubiesen sido los mismos —en lo relativo al tiempo invertido— llevando a cabo un vaciado de ediciones no digitalizadas.

Otra cuestión que cabe notar tiene que ver con la selección de textos en el *CORDE*, que motiva que algunos de sus aciertos pasen desapercibidos. Así, no es inusual que de una misma obra se introduzca más de una edición textual. Y esta duplicidad arroja en ocasiones repeticiones enojosas en los resultados obtenidos. Pero, además, hemos comprobado que las ediciones incluidas en el corpus antiguo de la RAE no siempre se ajustan a los criterios de fiabilidad que exige la investigación lingüística, de manera que en no pocos casos nos vemos obligados a desestimar ocurrencias —y obras— a fin de que los resultados de los estudios no se vean distorsionados.

Ya para concluir, querríamos subrayar que buena parte del problema al que nos hemos enfrentado parte del concepto mismo de *corpus*. Mientras estos sigan siendo entendidos, como propone la RAE y como proponía Sinclair (1991), como conjuntos lo más extensos y ordenados posibles de datos o textos, estaremos lejos de disponer de una base de documentos útiles para el investigador. Esto es, mientras se siga primando la cantidad por encima de la importancia de la lematización y del etiquetado, mientras el número de textos incluido en el corpus sea más importante que una selección de las ediciones que dé cabida a un número equilibrado y variado de tradiciones discursivas y de variedades diatópicas y diacrónicas, estaremos lejos de disponer de herramientas realmente útiles para las investigaciones lingüísticas.

REFERENCIAS BIBLIOGRÁFICAS

- ADMYTE = MARCOS MARÍN, Francisco *et al.*: *Archivo Digital de Manuscritos y Textos Españoles*. <http://www.admyte.com> [Consulta: 28/03/2012].
- ARTHUS = vv. AA.: *Archivo de textos hispánicos de la Universidad de Santiago de Compostela*. <http://adesse.uvigo.es/data/corpus.php> [Consulta: 28/03/2012].
- CANET, José L.: *Bases de datos Parnaseo*. <http://parnaseo.uv.es/Bases.htm> [Consulta: 28/03/2012].
- CATALÁN, Diego (1962): *De Alfonso X al conde de Barcelos. Cuatro estudios sobre el nacimiento de la historiografía romance en Castilla y Portugal*. Madrid: Gredos.
- CATALÁN, Diego (1997): *De la silva textual al taller historiográfico alfonsí. Códices, crónicas, versiones y cuadernos de trabajo*. Madrid: Fundación Ramón Menéndez Pidal, Universidad Autónoma de Madrid.
- CHARTA = SÁNCHEZ-PRÍETO BORJA, Pedro (coord.): *Corpus hispánico y americano en la red: textos antiguos*. <http://www.charta.es> [Consulta: 28/03/2012].
- CNDH = REAL ACADEMIA ESPAÑOLA (en pruebas): *Corpus del nuevo diccionario histórico*. <http://web.frl.es/CNDHE/>.
- COCA = DAVIES, Mark: *Corpus of Contemporary American English*. <http://corpus.byu.edu/coca/x.asp?w=1093&h=614> [Consulta: 28/03/2012].
- CODEA = SÁNCHEZ-PRÍETO BORJA, Pedro (coord.): *Corpus de Documentos Españoles Anteriores a 1700*. <http://demos.bitext.com/codea> [Consulta: marzo de 2012].
- CORDE = REAL ACADEMIA ESPAÑOLA: Banco de datos (*CORDE*) [en línea]. *Corpus diacrónico del español*. <http://www.rae.es> [Consulta: marzo de 2012].
- CRADDOCK, Jerry R. (1998): «El texto del *Espéculo*», *Initium: Revista Catalana d'Istòria del Dret*, 3, pp. 221-274. <http://escholarship.org/uc/item/4sw0w6mw> [Consulta: 28/03/2012].
- CREA = REAL ACADEMIA ESPAÑOLA: Banco de datos (*CREA*) [en línea]. *Corpus de referencia del español actual*. <http://www.rae.es> [Consulta: marzo de 2012].
- DAVIES, Mark: *Corpus del Español*. <http://www.corpusdelespanol.org/> [Consulta: 28/03/2012].

- ENRIQUE-ARIAS, Andrés (ed.) (2008): *Biblia Medieval*. <http://www.bibliamedieval.es> [Consulta: marzo de 2012].
- FERNÁNDEZ ALCAIDE, Marta (2009): *Cartas de particulares en Indias del siglo xvi. Edición y estudio discursivo*. Madrid/Frankfurt: Iberoamericana/Vervuert.
- FERNÁNDEZ-ORDÓÑEZ, Inés (2006): «La historiografía medieval como fuente de datos. Tradiciones consolidadas y rupturas necesarias», en José Luis Girón Alconchel y José Jesús de Bustos Tovar (coords.), *Actas del VI Congreso Internacional de Historia de la Lengua Española (Madrid, 29 de septiembre - 3 de octubre de 2003)*, vol. II. Madrid: Arco Libros, pp. 1779-1808.
- GAGO JOVER, Francisco (ed.) (2011): *Digital Library of Old Spanish Texts/Biblioteca digital de Textos del Español Antiguo*. Hispanic seminary of medieval studies. <http://www.hispanicseminary.org/t&c/ac/index-es.htm> [Consulta: 28/03/2012].
- GARACHANA, Mar y Malte ROSEMEYER (en prensa): «Rutinas léxicas en el cambio gramatical. El caso de las perífrasis deónticas e iterativas», *Revista de Historia de la Lengua Española*.
- HARRIS-NORTHALL, Ray (1996): «Printed books and linguistic standardization in Spain: the 1503 *Gran conquista de Ultramar*», *Romance Philology*, 50, 2, pp. 123-146.
- HEINE, Bernd (2002): «On the role of context in grammaticalization», en Ilse Wischer y Gabriele Diewald (eds.), *New reflections on grammaticalization*. Amsterdam: John Benjamins (Typological Studies in Language, 49), pp. 83-101.
- HOPPER, Paul J. (1991): «On some principles of grammaticization», en Elizabeth Closs Traugott y Bernd Heine (eds.), *Approaches to Grammaticalization*, vol. I. Amsterdam: John Benjamins (Typological Studies in Language, 19), pp. 17-36.
- KABATEK, Johannes (2008): *Sintaxis histórica del español y cambio lingüístico: Nuevas perspectivas desde las tradiciones discursivas*. Madrid/Frankfurt: Iberoamericana/Vervuert (Lingüística iberoamericana, 31).
- KILGARRIFF, Adam y Pavel RYCHLÝ (dirs.): *Sketch engine*. <http://the.sketchengine.co.uk/auth/corpora/> [Consulta: 28/03/2012].
- KYTÖ, Merja (comp.): *The Helsinki Corpus of English Texts*. <http://icame.uib.no/hc/> [Consulta: 28/03/2012].
- MENÉNDEZ PIDAL, Ramón (ed.) (1906): *Primera crónica general. Estoria de España que mandó componer Alfonso el Sabio y se continuaba bajo Sancho IV en 1289*. Madrid: Bailly-Baillière e Hijos.
- MENÉNDEZ PIDAL, Ramón (ed.) (1955): *Primera crónica general. Estoria de España que mandó componer alfonso el sabio y se continuaba bajo Sancho IV en 1289*, con la colaboración de A. G. Solalinde, M. Muñoz Cortés y J. Gómez Pérez). Madrid: Seminario Menéndez Pidal/Gredos.
- MENÉNDEZ PIDAL, Ramón (ed.) (1977): *Primera crónica general. Estoria de España que mandó componer Alfonso el Sabio y se continuaba bajo Sancho IV en 1289*. Madrid: Seminario Menéndez Pidal/Gredos.
- RODRÍGUEZ MOLINA, Javier (2010): *La gramaticalización de los tiempos compuestos en español antiguo: Cinco cambios diacrónicos*. Madrid: Universidad Autónoma de Madrid. Tesis doctoral dirigida por Inés Fernández-Ordóñez y Leonardo Gómez Torrego.
- ROJO, GUILLERMO (en prensa): «El papel de los corpus en el estudio de la historia del español», en *Actas del VIII Congreso Internacional de Historia de la Lengua Española (Santiago de Compostela, 14-18 de septiembre de 2009)*. http://gramatica.usc.es/~grojo/En_prensa/Corpus_diacronicos.pdf [Consulta: 28/03/2012].

- SÁNCHEZ SÁNCHEZ, Manuel Ambrosio (2005): «Notas sobre los *Castigos* atribuidos a Sancho IV», en Javier San José Lera (coord.), *Praenstans labore Victor: homenaje al profesor Víctor García de la Concha*. Salamanca: Ediciones de la Universidad de Salamanca, pp. 51-70.
- SINCLAIR, John M. (1991): *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- YLLERA, Alicia (1980): *Sintaxis histórica del verbo español*. Zaragoza: Departamento de Filología Francesa, Universidad de Zaragoza.

APÉNDICE

Obras consideradas en el estudio de la perífrasis *haber + (de/a) + infinitivo*:

Obra	Año	Autor	Título	Tipología
1.	1141-1235	Anónimo	<i>Fuero de Madrid</i>	Narrativo
2.	1180	Anónimo	<i>Auto de los Reyes Magos</i>	Narrativo
3.	1194-1211	Anónimo	<i>Liber Regum</i>	Narrativo
4.	1200	Almerich	<i>La fazienda de Ultra Mar</i>	Narrativo
5.	1205-1250	Anónimo	<i>Razón de amor</i>	Narrativo
6.	1230	Berceo, Gonzalo de	<i>Vida de San Millán de la Cogolla</i>	Narrativo
7.	1236	Berceo, Gonzalo de	<i>Vida de Santo Domingo de Silos</i>	Narrativo
8.	1246-1252	Berceo, Gonzalo de	<i>Los Milagros de Nuestra Señora</i>	Narrativo
9.	1270-1284	Alfonso x	<i>Estoria de España</i>	Historiográfico
10.	1275-1284	Alfonso x	<i>General Estoria. Primera parte</i>	Historiográfico
11.	1275-1284	Alfonso x	<i>General Estoria. Cuarta parte</i>	Historiográfico
12.	1293	Anónimo	<i>Gran Conquista de Ultramar</i>	Historiográfico
13.	1340	Anónimo	<i>Crónica de Sancho IV</i>	Historiográfico
14.	1350	Anónimo	<i>Sumas de la historia troyana</i>	Historiográfico
15.	1400	López de Ayala, Pero	<i>Crónica del rey don Pedro</i>	Historiográfico
16.	1400	López de Ayala, Pero	<i>Traducción de las Décadas de Tito Livio</i>	Historiográfico
17.	1425-1430	Anónimo	<i>Crónica del moro Rasis</i>	Historiográfico
18.	ca.1430	Corral, Pedro de	<i>Crónica del rey don Rodrigo</i>	Historiográfico
19.	1431-1436	Díaz de Games, Gutierre	<i>El Victorial</i>	Historiográfico
20.	1449	Martínez de Toledo, Alfonso	<i>Atalaya corónicas</i>	Historiográfico
21.	1450	Ramírez de Guzmán, Vasco	<i>Guerra de Jugurtha de Caio Salustio Crispo</i>	Historiográfico
22.	1455	Pérez de Guzman, Fernán	<i>Mar de Historias</i>	Historiográfico
23.	1467-1475	Escavias, Pedro de	<i>Repertorio de príncipes de España</i>	Historiográfico
24.	1471	Anónimo	<i>Crónica incompleta de los Reyes Católicos</i>	Historiográfico
25.	1473	García de Salazar, Lope	<i>Istoria de las bienandanzas e fortunas</i>	Historiográfico
26.	1480-1486	Pulgar, Hernando de	<i>Claros varones de Castilla</i>	Historiográfico
27.	1481	Anónimo	<i>Crónica de Enrique IV de Castilla 1454-1474</i>	Historiográfico
28.	1482	Pulgar, Hernando del	<i>Crónica de los Reyes Católicos</i>	Historiográfico
29.	1483		<i>Crónica abreviada de España*</i>	Historiográfico
30.	1487	Rodríguez de Almela, Diego	<i>Compilación de las batallas campales</i>	Historiográfico
31.	1487	Valera, Diego de	<i>Crónica de los Reyes Católicos</i>	Historiográfico
32.	1491	Anónimo	<i>Traducción de la historia del noble Vespasiano</i>	Historiográfico
33.	1491	Anónimo	<i>Relación circunstanciada de lo</i>	Historiográfico

			<i>acaecido en la prisión del Rey Chico en Granada</i>	
34.	1492	Enríquez del Castillo, Diego	<i>Crónica de Enrique IV</i>	Historiográfico
35.	1492	Maldonado, Alonso	<i>Hechos del Maestre de Alcántara don Alonso de Monroy</i>	Historiográfico
36.	1499	García de Santa María, Gonzalo	<i>Traducción de la Corónica de Aragón de Fray Gauberto Fabricio de Vagad</i>	Historiográfico
37.	1500	Anónimo	<i>Sancho de Paternoy niega haber sido cómplice en el asesinato de un inquisidor.</i> [Documentos lingüísticos medievales]	Legal
38.	1500	Anónimo	<i>De las melecinas.</i> Bibl. Universitaria Salamanca, ms. 1743	Técnico
39.	1500	Anónimo	<i>Tratado médico (Tratado de la patología general).</i> Madrid, BN I51	Técnico
40.	1504	Anónimo	<i>Sentencia.</i> [Colección diplomática de Santo Toribio de Liébana] (año 1504) 727	Legal
41.	1504	Anónimo	<i>Revisión de sentencia.</i> [Colección diplomática de Santo Toribio de Liébana] 492	Legal
42.	1519-1547	Albo, Francisco	<i>Diario o derrotero del viaje de Magallanes</i> (Texto autobiográfico)	Historiográfico
43.	1528	Delicado, Francisco	<i>La Lozana Andaluza</i>	Narrativo
44.	1535-1536	Valdés, Juan de	<i>Diálogo de la lengua</i>	Argumentativo
45.	1539	Aldana, Bernardo de	<i>Carta de Bernardo Villela de Aldana a Pedro Barrantes Maldonado, siendo recién casado en Alburquerque.</i> [Cartas y documentos relativos a Pedro Barrantes]	Epistolar
46.	1562-1566	Santa Teresa de Jesús	<i>Libro de la vida</i>	Historiográfico
47.	1575	Santa Teresa de Jesús	Sin título. [Cartas relativas a Fray Luis de Granada]	Epistolar
48.	1582	Anónimo	<i>Testamento de Catalina Manrique.</i> [Documentos procedentes del archivo provincial de Toledo (1580-1664)]	Legal
49.	1599	Alemán, Mateo	<i>Primera parte de Guzmán de Alfarache</i>	Narrativo
50.	1600	Anónimo	<i>Pragmáticas relativas a indumentaria española del siglo XVII</i>	Legal
51.	1601	Anónimo	<i>Partida de defunción de Pedro Bocangelino.</i> [Documentos de la casa Bocangelina (1562-1710)]	Legal
52.	1601	Anónimo	<i>Estatutos del colegio de Santo</i>	Legal

			<i>Tomás de la Universidad de Salamanca</i>	
53.	1604	Alemán. Mateo	<i>Segunda parte del Guzmán de Alfarache</i>	Narrativo
54.	1605	Cervantes Saavedra, Miguel de	<i>El ingenioso hidalgo don Quijote de la Mancha</i>	Narrativo
55.	1609	Alemán, Mateo	<i>Ortografía Castellana</i>	Técnico
56.	1611	Anónimo	<i>Relación de los funerales por la reina doña Margarita de Austria. [Actos públicos en Madrid]</i>	Historiográfico
57.	1612	Anónimo	<i>Concierto del matrimonio entre la infanta Ana Mauricio de Austria y el rey Ludovico de Francia. [Actos públicos en Madrid]</i>	Historiográfico
58.	1618-1633	Vega Carpio, Lope de	<i>Epistolario</i>	Epistolar
59.	1620	Quevedo y Villegas, Francisco de	<i>Entremés de Bárbara</i>	Dramático
60.	ca.1620	Vega Carpio, Lope de	<i>El poeta. Entremés</i>	Dramático
61.	1621	Almansa y Mendoza, Andrés de	<i>Carta que escribió un señor de esta Corte a un amigo. [Actos públicos en Madrid]</i>	Epistolar
62.	1630-1633	Quevedo y Villegas, Francisco de	<i>La cuna y la sepultura. Para el conocimiento propio y desengaño de las cosas ajenas</i>	Argumentativo
63.	1640	Gracián, Baltasar	<i>El político don Fernando el Católico</i>	Argumentativo
64.	1675	Zaragoza, Jose	<i>Fábrica y uso de varios instrumentos matemáticos</i>	Técnico
65.	1700	Anónimo	<i>De las propiedades del perro perdiguero</i>	Técnico
66.	1706	Anónimo	<i>Para que se observe y guarde la cédula de 28 de junio de 1696 para el buen cobro de la Real Hacienda. [Cedulario americano]</i>	Legal
67.	1710-ca.1784	Góngora, Diego Ignacio de	<i>Historia del Colegio Mayor de San Tomás de Sevilla</i>	Historiográfico
68.	1729	Luzán, Ignacio de	<i>Arte de hablar, o sea, Retórica de las conversaciones</i>	Argumentativo
69.	1738	Anónimo	<i>El rey manda que la prohibición que por el capítulo cuarto del despacho de 21 de enero de 1735 se estableció para que los comerciantes de la Nueva España no envíen caudales a España de pura negociación, se suspenda en la conformidad que se expresa. [Cedulario</i>	Legal

			americano]	
70.	1740	Anónimo	<i>Se establece una Compañía en San Cristóbal de la Habana dedicada a la conducción de tabacos, azúcar, corambre y otros frutos de la isla de Cuba.</i> [Cedulario americano]	Legal
71.	1758	Isla, José Francisco de	<i>Historia del famoso predicador Fray Gerundio de Campazas alias Zotes</i>	Narrativo
72.	ca.1771	Cadalso, José	<i>Noches lúgubres</i>	Narrativo
73.	1774-1777		<i>Carta a Rodríguez Campomanes.</i> [Epistolario de Pedro Rodríguez Campomanes]	Epistolar
74.	1785	Trigueros, Cándido María	<i>El precipitado (comedia sentimental)</i>	Dramático
75.	1787	Fornier, Juan Pablo	<i>Discursos filosóficos sobre el hombre</i>	Argumentativo
76.	1792	Fernández de Moratín, Leandro	<i>La comedia nueva</i>	Dramático
77.	1794	Fernández de Moratín, Leandro	<i>Cartas de 1794.</i> [Epistolario]	Epistolar
78.	1802	Anónimo	<i>Capitulaciones matrimoniales otorgadas entre los señores D. Bernardo Rodríguez de Toro y Ascanio, D.^a María Teresa Rodríguez de Toro y Alaiza, su hija, y D. Pedro Rodríguez de Toro, su tío, como apoderado del Sr. D. Simón De Bolívar y Palacios.</i> [Documentos sobre la vida privada española]	Legal
79.	1805	Fernández de Moratín, Leandro	<i>El sí de las niñas</i>	Dramático
80.	1814	Anónimo	<i>Real orden mandando que las Aduanas del cordón del Ebro continúen como lo estaban en abril de mil ochocientos ocho.</i> [Colección de cédulas de las provincias vascongadas (1722-1825)]	Legal
81.	1832	Larra, Mariano José de	<i>El castellano viejo.</i> [Figaro. Colección de artículos dramáticos, literarios y de costumbres] Blanco-White Ensayos	Argumentativo
82.	1864	Bécquer, Gustavo Adolfo	<i>Desde mi celda</i>	Epistolar
83.	1868	Botella y de Hornos, Federico de	<i>Descripción geológica-minera de las provincias de Murcia y Albacete</i>	Técnico
84.	1870-1901	Calleja y Sanchez, Julián	<i>Compendio de anatomía descriptiva y de embriología humanas, II</i>	Técnico

85.	1876-1880	Zugasti y Sáenz, Julian	<i>El Bandolerismo. Estudio social y memorias históricas</i>	Historiográfico
86.	1877	Conde de Toreno	<i>Prólogo. [Cartas de Indias]</i>	Epistolar
87.	1884	Castelar, Emilio	<i>Historia del año 1883</i>	Historiográfico
88.	1884-1885	Clarín (Leopoldo Alas)	<i>La Regenta</i>	Narrativo
89.	1885-1887	Pérez Galdós, Benito	<i>Fortunata y Jacinta</i>	Narrativo
90.	1895	Dicenta, Joaquín	<i>Juan José. Drama en tres actos y en prosa</i>	Dramático
91.	1930	Unamuno, Miguel de	<i>Epistolario inédito</i>	Epistolar
92.	1930	Ortega y Gasset, José	<i>Misión de la Universidad</i>	Argumentativo
93.	1940-1956	Arias-Paz, Manuel	<i>Manual de Automóviles</i>	Técnico
94.	1941	Salinas, Pedro	<i>Carta de Salinas. [Cartas. En «Correspondencia (1923-1951)»]</i>	Epistolar
95.	1941	Azorín (José Martínez Ruiz)	<i>Madrid</i>	Historia
96.	1945	García y Bellido, Antonio	<i>España y los españoles hace dos mil años según la Geografía de Strábon</i>	Historia
97.	1946	Anónimo	Decreto de 8 de febrero de 1946 por el que se aprueba la nueva redacción oficial de la Ley Hipotecaria	Legal
98.	1947	Anónimo	Reglamento hipotecario de 1947	Legal
99.	1956	Sánchez Ferlosio, Rafael	<i>El Jarama</i>	Narrativo
100.	1962	Ballarín Marcial, Alberto	<i>La agricultura española en sus aspectos jurídicos</i>	Legal
101.	1976	Alberti, Rafael	<i>Noche de guerra en el museo del Prado</i>	Dramático
102.	1978	Marsé, Juan	<i>La muchacha de las bragas de oro</i>	Narrativo
103.	1982	Savater, Fernando	<i>Invitación a la ética</i>	Argumentativo
104.	1985	Alonso de Santos, José Luis	<i>Bajarse al moro</i>	Dramático
105.	1986	Delibes, Miguel	<i>La hoja roja</i>	Narrativo
106.	1990	Calle Capilla, Ramiro A.	<i>Yoga, una medicina natural</i>	Técnico
107.	2002	Grandes, Almudena	<i>Los aires difíciles</i>	Narrativo
108.	<1900	AA. VV.	OO. VV.	Oral