

Nuevas herramientas para la historia de la Edad Media hispánica: los corpus textuales informatizados

New tools for the history of the Hispanic Middle Ages: The Computerized Textual Corpus

Ana Isabel CARRASCO MANCHADO
Universidad Complutense. Madrid
aicarras@ghis.ucm.es

RESUMEN

El objetivo de esta investigación es fomentar y ensayar nuevos métodos de análisis de las fuentes medievales, mediante el empleo de los Corpus Textuales Informatizados. Hoy en día, la investigación lingüística en España puede considerarse pionera en el desarrollo de corpus textuales informatizados. Sólo el CORDE (Real Academia de la Lengua) permite la búsqueda simultánea de información sobre 12.865 obras y documentos de la época medieval. La historia medieval no puede ignorar el uso de estos recursos para conseguir un mayor apoyo empírico a la interpretación derivada de la formulación de hipótesis de la historia política o social. Nuestro objetivo es informar a la comunidad de medievalistas de los principales corpus textuales informatizados y sus aplicaciones básicas para la investigación histórica.

Palabras clave: Metodología de la Historia. Corpus Textuales Informatizados. CORDE. Corpus del Español. Lexicología. Lexicometría. Discurso político

ABSTRACT

The objective of this research is to encourage and develop new methods of analysis for medieval sources by employing the Computerized Textual Corpus. To the present day, linguistic research in Spain can be considered pioneer in the development of these technologies. The CORDE of the Spanish Royal Academy in itself permits the simultaneous search for information in 12,865 works and documents of the medieval period. Historical investigation cannot ignore the use of these resources to obtain greater empirical support for the interpretations derived from the formulation of hypotheses in political or social history. This article attempts to provide information about the main Computerized Textual Corpus and their basic applications for research.

Keywords: Historical methodology. Computerized Textual Corpus. CORDE. Spanish Corpus. Lexicology. Lexicometry. Political discourse

Sumario: Corpus Textuales informatizados. Admyte. Corde. Corpus del español. Corpus “territoriales”. Corpos de otros ámbitos idiomáticos hispánicos. Corpus latinos. Corpus textuales que integran criterios paleográficos y diplomáticos. Otros corpus. Uso de Corpus Textuales Informatizados entre los historia-dores. Esbozo de un ejemplo de aplicación: la noción de “acuerdo”.

Como en cualquier otra disciplina científica, el trabajo del historiador debe constantemente estar sometido a una labor de reflexión metodológica¹, para que no se paralicen los avances renovadores que, desde hace varias décadas, han experimentado los distintos ámbitos temáticos de la historia en general, y de la historia medieval en particular. Especialmente, desde los años 90 del siglo xx, el imparable avance de las nuevas tecnologías de la información obliga a no dejar de lado la indagación metodológica, con el objetivo de comprobar la eficacia que tienen en el campo de la investigación histórica los recursos que están apareciendo. Son ya muchos los recursos informáticos disponibles para el ámbito científico de las humanidades, pero su perfeccionamiento parece discurrir a una mayor velocidad que la de su aplicación al campo de la historia². Quizás haya llegado el momento de que el historiador medievalista ensaye en sus investigaciones algunos de esos recursos informáticos cuya eficacia ya ha sido probada para otras disciplinas humanas afines a la historia, tales como la filología³. Los ámbitos de la historia social o de la historia política de la Edad Media hispánica pueden verse beneficiados de la aplicación de estos recursos, y también la comprensión general de la época, ya que toda comprensión general pasa por precisar el significado de los testimonios textuales que nos han transmitido parte de ese pasado medieval. Muchos de los recursos disponibles pueden resultar apropiados para profundizar en el sentido profundo de los textos.

En la actualidad existen de forma accesible y gratuita una serie de Corpus textuales informatizados que deben ser conocidos por los medievalistas que trabajen sobre ámbito hispánico. La lectura directa de las fuentes medievales no puede asegurar la comprensión total de un texto que ha sido producido en una civilización distinta a la nuestra, con un sistema de producción del significado también diferente, aunque el texto esté escrito en romance. Por ello, resulta imprescindible aplicar a los textos un filtro que ayude a afinar el significado y la comprensión de los enunciados. Las posibilidades de tratamiento y análisis textual que proporcionan estos corpus para su aplicación a las fuentes empleadas habitualmente en las distintas investigaciones están todavía por ensayar.

Actualmente, la investigación lingüística en España puede considerarse pionera en el desarrollo de corpus textuales informatizados. Sólo el CORDE, de la Real Academia de la Lengua, permite la búsqueda simultánea de información en unas

¹ Este artículo forma parte del proyecto «Nuevos métodos para la historia social y política de la Edad Media hispánica: aplicaciones históricas de los Corpus textuales informatizados» (HAR2010-17860), que dirijo desde el Departamento de Historia Medieval en la Facultad de Geografía e Historia (UCM).

² Sobre las posibilidades de la informática humanística y su situación en el actual esquema de planes de estudios adaptados al EESS en Europa, véase LUCÍA MEGÍAS, José Manuel: “La Informática Humanística: una puerta abierta para los estudios medievales en el siglo XXI”, *Revista de poética medieval*, 20 (2008), pp. 165-167.

³ La lingüística computacional es la rama humanística que mayor avance ha logrado en el campo tecnológico; véase el conjunto de trabajos recogidos en ENRIQUE-ARIAS, Andrés (ed.), *Diacronía de las lenguas iberorrománicas. Nuevas aportaciones desde la lingüística de corpus*, Madrid-Francfort del Meno, 2009.

12.865 obras y documentos del período medieval⁴. Venimos constatando el escaso conocimiento y el prácticamente nulo uso que hacen los historiadores de este importante Corpus y de las posibilidades que puede ofrecer para la investigación histórica. Pensamos que los historiadores de la Edad Media hispánica habituados a trabajar no solo con documentación inédita, sino con muchas fuentes actualmente ya publicadas (crónicas, tratados, obras literarias, colecciones documentales, legislación...), no pueden obviar la utilización de estos recursos que permiten el procesamiento de los datos textuales de forma masiva y simultánea, de manera que es posible ajustar la interpretación de determinados conceptos que solo pueden ser comprendidos si se estudian de una forma sincrónica y diacrónica. La formulación de ciertas hipótesis de historia política o social podría apoyarse sobre bases más empíricas cuando se dispone de esta masa textual en un solo corpus, y de las aplicaciones informáticas adecuadas para realizar un análisis más complejo.

Hasta el siglo xx, buena parte de los proyectos de investigación histórica medieval se ocupaban de la publicación de cartularios o de colecciones documentales. Actualmente, gracias a los progresos informáticos, aunque también se siguen publicando nuevas fuentes inéditas, ahora en formatos electrónicos, parece que la investigación se orienta a conseguir medios que permitan procesar esa misma información histórica ya publicada de manera más eficaz⁵. Pero en el ámbito del medievalismo, aún no se han dado los pasos necesarios para diseñar ese tipo de bancos de datos, a pesar de que el mayor trabajo, el más difícil, ya ha sido, en buena medida, realizado: reunir un extenso volumen de documentación publicada. El *Codiphis. Catálogo de colecciones diplomáticas hispanolusas*⁶ es suficientemente expresivo de la inmensa documentación que podría estar disponible para su incorporación a un corpus informatizado. La idea de que es preciso abrir una etapa nueva –pasar de la publicación de colecciones a la creación de procedimientos informáticos de análisis–, ya fue planteada hace algunos años, en el entorno mismo de la elaboración del *Codiphis*⁷. Pero no parece que se haya avanzado mucho en una labor tan fundamental.

⁴ Un corpus textual es un corpus que incluye obras completas. Los llamados «grandes» son los que incluyen millones de palabras.

⁵ Prueba de ello son proyectos como el *Corpus de Textos Jurídicos Andaluces* (CTJA), el *Diccionario de Textos Concejiles Andaluces* (DITECA), o el *Archivo Informático de Textos Jurídicos Medievales de Andalucía* (ARINTA), desarrollado en el Departamento de Filología española I y Filología románica de la Universidad de Málaga. Pero, como puede verse, son filólogos los que han iniciado esta segunda fase de tratamiento documental.

⁶ GARCÍA DE CORTÁZAR, José Ángel, MUNITA LOINAZ, José Antonio y FORTÚN, Luis Javier: *Codiphis. Catálogo de colecciones diplomáticas hispanolusas de época medieval*, Santander, 1999, 2. Vols.

⁷ «Conviene plantear también la necesidad de aplicar las nuevas tecnologías a la publicación de colecciones documentales y no sólo mediante el empleo del CD-Rom o el DVD como soporte, del que ya existen algunos ejemplos. Sugiero la conveniencia de elaborar bases de datos documentales, accesibles en red, en las que volcar los documentos íntegramente o en forma de catálogo, mediante campañas sucesivas que atiendan criterios geográficos o tipológicos», FORTÚN CIRIZA, Luis Javier: “La edición de fuentes documentales para el estudio de la Edad Media hispana”, en José Antonio MUNITA y José Ramón DÍAZ DE DURANA (ed.), *XXV años de historiografía hispana (1980-2004)*, Universidad del País Vasco, 2007, p. 40.

A la luz de la experiencia de los corpus filológicos, la mejor opción sería realizar un banco de datos con los documentos volcados íntegramente, diseñando un interfaz de búsqueda adecuado. De momento, y hasta que esas iniciativas se concreten en proyectos reales, vamos a dar a conocer los principales corpus textuales filológicos, que ya de por sí, como veremos, proporcionan herramientas muy útiles para el historiador.

La labor de investigación que desarrolla el historiador ha de pasar por el análisis crítico de las fuentes escritas que constituyen su principal fuente de información. Para ello es preciso contar con el concurso de las ciencias y técnicas historiográficas, tales como la paleografía o la diplomática, pero también con el apoyo de aquellas otras disciplinas que se centran en el análisis del lenguaje, especialmente de aquellas que incorporan la perspectiva histórica, como la lingüística histórica (especialmente la semántica diacrónica), o la sociolingüística histórica. Otras ramas de la lingüística también pueden resultar útiles, tales como la lexicografía, la lexicología, o incluso la pragmática⁸. No hay que olvidar tampoco el apoyo de las disciplinas que analizan los testimonios y sus condiciones de elaboración y de transmisión, con criterios literarios, como la filología románica, para el caso de los textos escritos en lenguas romances, que aporta información básica para determinar la autoridad o validez de los testimonios.

En las últimas décadas, la incorporación de técnicas informáticas de tratamiento textual o hipertextual a la investigación en estos campos ha supuesto una verdadera revolución en la investigación, hasta el punto de convertir en incompletas muchas de las investigaciones realizadas con anterioridad al momento en que surge la posibilidad de uso de estas técnicas informáticas (muchos trabajos sobre semántica histórica realizados con anterioridad a la aparición del CORDE están ya siendo revisados)⁹. La posibilidad de disponer de un corpus con información textual masiva, datada y, en algunos casos, también lematizada, permite obtener listas de frecuencias y de concordancias de términos, lo cual permite estudiar un mayor aporte de información ob-

⁸ La pragmática se dedicaba al estudio del habla, pero actualmente estudia también la relación entre la oralidad y la lengua escrita, lo cual permite añadir una aproximación histórica. Desde la perspectiva de la pragmática histórica pueden descubrirse lecturas insospechadas en textos medievales. Así, se ha analizado, por ejemplo, la presencia de una segunda lectura, de naturaleza jurídica, en los Milagros de Nuestra Señora, de Gonzalo de Berceo. A juicio del investigador, la obra está tan plagada de alusiones al derecho romano, que puede leerse también como una obra propagadora del nuevo derecho; KABATEK, Johannes: “Tadiciones discursivas jurídicas y elaboración lingüística en la España medieval”, *Cahiers de linguistique hispanique médiévale*, 27 (2004), p. 260.

⁹ Para ámbito histórico, por ejemplo, el estudio de GUTIÉRREZ NIETO, José Ignacio: “Semántica del término “comunidad” antes de 1520: las asociaciones juramentadas de defensa”, *Hispania*, 37 (1977), pp. 319-367, que sigue siendo muy utilizado en el análisis de la Comunidades, puede ser revisado y, quizá, matizado, a la luz de los resultados que ofrece el CORDE: sólo realizando la búsqueda «comunidad», sin incluir variantes ortográficas (como veremos, uno de los defectos del CORDE es que no está lematizado), entre 1300 y 1520, ofrece un resultado de 355 casos en 121 documentos, que convendría analizar para precisar el significado; y entre 1521 y 1550, 205 casos en 34 documentos, lo cual podría poner en cuestión la idea de que ese término estuviera en desuso en Castilla la Vieja; búsqueda «comunidad», REAL ACADEMIA ESPAÑOLA: Banco de datos (CORDE) [en línea]. Corpus diacrónico del español. <<http://www.rae.es>> [22-1-2011].

jeto de análisis, de forma seriada, así como obtener estadísticas léxicas. El resultado se beneficia así de una perspectiva empírica más fiable que si se hubiera empleado para la misma investigación únicamente el esfuerzo lector directo y una clasificación parcial. Ello ha dado origen en filología, incluso, a una nueva rama que es la de la lingüística de Corpus, centrada, precisamente, en la explotación metodológica para la historia de la lengua de estos nuevos recursos textuales informáticos que son los corpus lingüísticos informatizados. Con la aparición de esta nueva generación de corpus, gracias al avance tecnológico, se han fortalecido las bases empíricas de la filología¹⁰.

Si el uso de corpus textuales diacrónicos en formato informático está revolucionando las disciplinas filológicas mencionadas, es de esperar que su aplicación a la historia general, y a la de la Edad Media en particular, produzca, cuanto menos, unos resultados excelentes, ya que tales recursos pueden permitir la búsqueda de información simultánea en una buena parte de las fuentes escritas editadas que suelen constituir la base de muchas investigaciones que se están llevando a cabo actualmente. Piénsese, por ejemplo, en las investigaciones centradas en las crónicas o en el análisis de la rica tradística de la Baja Edad Media (que pueden manejarse desde el CORDE o desde el Corpus del Español), o en colecciones documentales editadas en los últimos años (como, por ejemplo, volúmenes de la colección Fuentes Documentales medievales del País Vasco recogidos también en el CORDE). Estudios sobre la percepción social de grupos sociales, sobre mentalidades, sobre las instituciones medievales, sobre las tradiciones legislativas o sobre las prácticas políticas, o sobre la cultura política, pueden verse también mejorados con el análisis conceptual que puede realizarse a partir, por ejemplo, del corpus legislativo alfonsí (incluido en ADMYTE, en el Corpus del Español y en el CORDE), o a partir de la colección de Cortes de León y Castilla (igualmente consultable desde el CORDE), por sólo citar unas pocas fuentes.

Con objeto de comprender el valor de las herramientas que proponemos, describiremos, en primer lugar, los principales corpus textuales informatizados y sus características. Veremos también cómo algunos de estos recursos ya han sido tímidamente empleados en el ámbito científico del medievalismo, aunque no siempre de la manera más eficaz, lo cual justifica el seguir indagando en las posibilidades que ofrecen.

CORPUS TEXTUALES INFORMATIZADOS

El desarrollo de corpus textuales informatizados lleva produciéndose, en el ámbito de la filología hispánica, desde hace ya cuarenta años, proporcionando un material textual masivo que apenas tiene comparación con empresas similares en el resto de Europa. Es la razón por la cual es posible pasar ya a la fase de estudio de sus distintas aplicaciones fuera del campo de la filología¹¹. Presentamos brevemente los

¹⁰ Como iniciación al uso de los corpus y a su tipología, resulta útil la síntesis de RAFEL I FONTANALS, Joaquín y SOLER I BOU, Joan: "El procesamiento de corpus. La lingüística empírica", en MARTÍ, María Antonia y ALONSO, Juan Alberto: *Tecnología del lenguaje*, Barcelona, 2003, pp. 41-73.

¹¹ Los propios lingüistas han reivindicado la necesidad de incorporar la historia general a sus investigaciones, pues la historia de la lengua no puede prescindir tampoco de las circunstancias sociales (perfil social del autor o del destinatario) o políticas (circunstancias de la transmisión o conservación)

principales corpus textuales susceptibles de ser aplicados a una renovación metodológica en el ámbito histórico.

ADMYTE

El *Archivo Digital de Manuscritos y Textos Españoles* (ADMYTE) es el proyecto electrónico más antiguo, concebido ya en los años 70 del siglo XX. El equipo internacional fue coordinado por el Hispanic Seminary of Medieval Studies (HSMS), de la Universidad de Wisconsin, Madison (EE. UU)¹². El objetivo era conseguir reunir una biblioteca de obras manuscritas medievales, para lo cual se efectuó una rigurosa selección de manuscritos y se realizaron transcripciones semipaleográficas¹³. Contiene 290 obras medievales íntegras, sumando un total de 54.000 páginas¹⁴. Algunas de estas obras han sido transcritas a partir de diversos manuscritos, lo cual permite estudiar variantes sobre un mismo aspecto. La biblioteca apareció en formato CD-ROM, en varios discos. Pero no sólo ofrecía la posibilidad de acceso a las transcripciones, sino que este recurso permite, además, obtener listas de vocabulario, de concordancias, de frecuencias, proporcionando así instrumentos de medición lexicométrica. La versión on-line ofrece posibilidades nuevas (como acceder a imágenes digitales de los manuscritos) y la consulta directa en la red¹⁵.

Aunque ADMYTE ya integra la obra legislativa completa de Alfonso X el Sabio, se está desarrollando actualmente un proyecto monográfico dedicado a toda su obra en prosa, la *Biblioteca Digital de la Obra en Prosa de Alfonso X*¹⁶. Su utilización resultará muy útil para la clarificación de cuestiones relacionadas con la práctica jurídico-política bajo su reinado.

que determinan la formación de un texto; ENRIQUE-ARIAS, Andrés: "Introducción. Lingüística de Corpus y diacronía de las lenguas iberorrománicas", en Andrés ENRIQUE-ARIAS (ed.), *Diacronía de las lenguas iberorrománicas*, pp. 11-16.

¹² Sobre las posibilidades que ofrecía ADMYTE, véase el capítulo de la obra de MARCOS MARÍN, Francisco: *Informática y humanidades*, Madrid, 1994, p. 180 y ss.

¹³ El proyecto estaba íntimamente ligado al dirigido por Charles B. Faulhaber, destinado a crear una base de datos bio-bibliográfica sobre todos los textos medievales peninsulares: *Bibliografía Española de Textos Antiguos* (BETA), *Bibliografía de Textos Antigos Galegos y Portugueses* (BITAGAP) y *Bibliografía de Textos Antics Catalans, Valencians i Balears* (BITECA), los tres integrados en versión web PhiloBiblon [Volumen 2011, Number 1 (January)]; <http://bancroft.berkeley.edu/philobiblon/> [consultado el 19-3-2011].

¹⁴ Puede verse la presentación de este proyecto del nuevo ADMYTE ON LINE en la propia página web del proyecto: <http://www.admyte.com/presentacion.htm> [consultado el 13-3-2011], así como la relación de las obras integradas en la nueva versión.

¹⁵ <http://www.admyte.com> (consultado el 13-1-2011). La desventaja de ADMYTE es que no es gratuito. No obstante, muchas de las transcripciones que aparecieron publicadas en CD-ROM fueron incorporadas al CORDE o al Corpus del Español, de manera que puede trabajarse con ellas a través de estas otras aplicaciones.

¹⁶ GAGO-JOVER, Francisco: "La Biblioteca Digital de la Obra en Prosa de Alfonso X: pasado, presente y futuro", conferencia al *Coloquio Internacional sobre Corpus Diacrónicos en Lenguas Iberorromances*, Palma de Mallorca, 2007 (cf. ENRIQUE-ARIAS, Andrés: "Introducción", p. 14).

CORDE

El CORDE (*Corpus Diacrónico del Español*), forma parte del Banco de Datos de la Real Academia Española, que fue creado en 1990 junto con el CREA (*Corpus de Referencia del Español Actual*). Ambos corpus son el fundamento de las labores del Instituto de Lexicografía de la Real Academia Española. Es el corpus histórico del español más extenso, formado por más de 300 millones de palabras, escritas desde los inicios de la lengua hasta 1974, procedentes de obras históricas, jurídicas (ordenamientos y códigos legales), tratadística, obras didácticas, literarias de todo tipo y también documentales. Es el sistema de referencia básico que utiliza la Real Academia de la Lengua para documentar el origen de las palabras. De hecho, es hoy en día el principal instrumento para la elaboración del *Nuevo diccionario histórico de la lengua española*. El volumen total de obras y de documentación integrado es muy elevado: para todo el período histórico, hasta 1974, incluye 5.500 títulos de obras y documentos. Para el período medieval, distinguiendo por siglos, contamos con 2.185 obras extensas y documentos desde los inicios de la lengua hasta el año 1200; 3.428 hasta el año 1300; 2.264 hasta el año 1400 y 4.988 hasta el año 1500. Esas 12.865 obras conforman para toda la época medieval un total de en torno a 44 millones de palabras. El objetivo que ha guiado la selección de obras integradas en el CORDE ha sido crear un corpus representativo de la lengua en su desarrollo histórico, de manera que se incluyen textos de todo género y de una amplia distribución territorial¹⁷. Se ha tratado de incluir la mejor edición posible de todos los textos¹⁸.

El CORDE admite búsquedas de palabras o de combinaciones de palabras en períodos cronológicos determinados o en un año concreto. Los casos resultantes, que pueden ordenarse cronológicamente (o por otros criterios), se presentan en diversas pantallas, distribuidos según queramos trabajar con los documentos en los que aparecen, con los párrafos de contexto, con las concordancias o con las estadísticas. Pueden efectuarse las búsquedas en todo el corpus o en un autor u obra concreta, o en un género o tipo de obras determinado¹⁹.

A pesar de la inmensa utilidad del CORDE para el historiador, no es una herramienta muy utilizada. Una razón de la escasa utilización del CORDE por parte de los historiadores, que puede hacerse extensiva al uso de todos los corpus, puede ser el hecho de que cuando los resultados de las búsquedas son fallidos (por problemas técnicos que generan las irritantes pantallas de error), o resultados no muy satisfactorios, el investigador se desalienta y decide no seguir utilizándola. Ciertamente, el CORDE tiene sus limitaciones: es un corpus codificado y etiquetado, pero no está

¹⁷ Véase la descripción del CORDE en SÁNCHEZ, Mercedes y DOMÍNGUEZ, Carlos: “El Banco de datos de la Real Academia española: CREA y CORDE”, *Per Abbat*, 2 (2007), pp. 142-146.

¹⁸ Aunque la autoridad filológica de algunas fuentes debe ser revisada, a juicio de LUCÍA MEGÍAS, José Manuel: “El hipertexto ante el reto de los textos medievales: nuevas reflexiones sobre informática humanística”, en GONZÁLEZ, Aurelio y WALDE MOHENO, Lilian von (eds.), *Temas, motivos y contextos medievales*, México, 2008, pp.9-14 [en eprint.ucm.es/8980, versión de 27-05-2009, consultado el 19-1-2011].

¹⁹ Se accede al interfaz de búsqueda por <http://corpus.rae.es/cordenet.html>. (consultado el 13-2-2011).

lematizado, lo cual implica que las búsquedas deben ser literales, de manera que hay que multiplicar las búsquedas según las variantes ortográficas de los términos (que previamente hay que conocer), o utilizar algunos comodines²⁰. Pero un buen conocimiento de todas sus posibilidades (y también de sus defectos) es la premisa fundamental para el máximo aprovechamiento del recurso. En cualquier caso, condición previa no es únicamente el dominio técnico, sino el conocimiento del funcionamiento de la lengua y de su desarrollo histórico, así como de las condiciones de producción del significado en la Edad Media. Si planteamos la utilización de estos recursos no es para efectuar meras búsquedas de vocabulario, sino para intentar comprender en todos sus matices el significado de ese vocabulario y el de cualquier enunciado medieval. La distinción de Saussure entre significante y significado es la premisa básica. Dicho de otro modo, no son búsquedas terminológicas lo que interesa preguntar al Corpus, sino búsquedas semasiológicas que nos permitan reconstruir los campos semánticos adecuados a nuestra investigación.

CORPUS DEL ESPAÑOL

Hemos aludido a las limitaciones que afectan a las búsquedas en el CORDE (aunque quizá estas limitaciones resulten más graves en el caso de los estudios lingüísticos que en el caso de los estudios históricos). Ya hemos dicho que el CORDE no ha sido lematizado. El CORDE no ha sido anotado en su diseño con lemas, no tiene marcas gramaticales ni semánticas. Por ejemplo, si el término buscado es un infinitivo como «aconsejar», sólo recupera la forma en infinitivo, no en el resto de formas verbales. Para subsanar estos problemas y crear posibilidades de búsquedas más complejas, se creó el Corpus del Español. También es un corpus de libre acceso y fue diseñado por Mark Davis, desde la Brigham Young University en el año 2002, siendo revisado en 2007. El corpus incorpora una compleja tecnología que lo hace más eficaz para las búsquedas lingüísticas, y más rápido²¹. Incorpora textos y transcripciones desde el siglo XIII al XX, con un total de 100.000.000 millones de palabras. A los siglos XIII al XV le corresponden unos 18 millones de palabras. Las obras medievales proceden de las transcripciones de ADMYTE y del HSME, así como de las integradas en la Biblioteca Virtual Miguel de Cervantes²².

Este corpus soluciona los problemas de lematización y de desambiguación (confusiones, por ejemplo, entre sustantivos y verbos en primera persona) del CORDE, de manera que es posible realizar, por ejemplo, búsquedas de cualquier verbo en cualquiera de sus formas gramaticales, con sólo introducir el infinitivo, búsquedas combinadas de cualquier sustantivo más un adjetivo, etc. Su sistema de búsqueda es,

²⁰ Sobre los defectos (y virtudes) del CORDE, MUÑOZ ARMILLO, Laura: “Herramientas para la investigación sobre lingüística diacrónica en la WEB”, en ROMERO, Laura y JULIÀ, Carolina (coord.), *Tendencias actuales en investigación diacrónica de la lengua*, Barcelona, 2009, pp. 104-109, en donde se aportan también algunas indicaciones para subsanar esas limitaciones.

²¹ Al interfaz de búsquedas se accede por la página <http://corpusdelespanol.org> (consultado el 13-1-2011).

²² Una presentación del *Corpus del Español* en DAVIS, Mark: “Un corpus anotado de 100.000.000 palabras del español histórico y moderno”, en *Procesamiento del lenguaje natural*, 29, 2002, pp. 21-27.

por tanto, más rico que el del CORDE, y menos literal. Una novedad importante es que incorpora la posibilidad de recuperar sinónimos de cualquier palabra o de grupos de palabras. También pueden obtenerse listas de calificativos asociados a determinados sustantivos objeto de la búsqueda. Se pueden generar listados de palabras que están colocadas en el entorno de la palabra que buscamos, o de sus sinónimos. Este tipo de búsquedas genera interesantes resultados para analizar campos semánticos²³.

El resultado de cada búsqueda se presenta en forma de listado de frecuencias, pudiendo acceder también al término buscado en su contexto de párrafo. La búsqueda puede limitarse a ciertos siglos, pero no a intervalos de años concretos, lo cual, para el historiador, supone cierta desventaja respecto al CORDE, que sí admite búsqueda por años.

CORPUS “TERRITORIALES”

Hemos llamado corpus “territoriales” a los corpus textuales que están apareciendo en diversos territorios del ámbito idiomático del español peninsular. Son corpus que, como el CORDE, se están desarrollando fundamentalmente para servir al estudio de la lengua y del habla histórica en espacios delimitados históricamente y pertenecientes al área de influencia del romance castellano. La ventaja de contar con estos corpus es la de disponer de fuentes no incorporadas en el CORDE, muchas de las cuales son fuentes documentales procedentes de archivos locales. Para la historia resulta muy útil, ya que se amplía el listado de fuentes municipales que pueden ser analizadas con la metodología aplicada a los corpus. Pero muchos de estos proyectos aún no se han desarrollado suficientemente o están en sus inicios. En cualquier caso, es preciso estar alerta a la aparición de corpus de este tipo y hacer un inventario de los existentes, de los responsables de su diseño y de su estado de actualización, así como de las posibilidades que ofrecen. De momento, destacamos los corpus desarrollados para ámbito andaluz, ya citados, el *Corpus de Textos Jurídicos Andaluces* (CTJA), el *Diccionario de Textos Concejiles Andaluces* (DITECA), y el *Archivo Informático de Textos Jurídicos Medievales de Andalucía* (ARINTA). Los equipos responsables de estos corpus ya han generado resultados para el avance del conocimiento histórico²⁴. Un ejemplo de ello es la tesis doctoral de Pilar López Mora, *Estudio del léxico de las “Ordenanzas del Concejo de Cór doba” (1435)*, realizada sobre la base del corpus y que desvela de forma sistemática nudos de la estructura de organización social de

²³ Partiendo de casos prácticos, puede verse la comparación entre el *Corpus del Español* y el CORDE en DAVIS, Mark: “Creating useful historical corpora: a comparison of CORDE, the *Corpus del Español*, and the *Corpus do português*”, en ENRIQUE-ARIAS, Andrés (ed.), *Diacronía de las lenguas iberorrománicas*, pp. 137-166.

²⁴ La directora del proyecto es Inés CARRASCO CANTOS. Véase una presentación del CTJA en CARRASCO CANTOS, Inés: “El Corpus de Textos Jurídicos Andaluces (léxico de las instituciones municipales)”, en ECHENIQUE ELIZONDO, María Teresa, SÁNCHEZ MÉNDEZ, Juan P.: (coord.), *Actas del V Congreso Internacional de Historia de la Lengua Española, V alencia 31 de ener o -4 de febr ero 2000*, Madrid, 2002, pp. 1.937-1.946; y del proyecto DITECA en CARRASCO CANTOS, Inés: “El Diccionario de Textos Concejiles de Andalucía”, en CAMPOS SOUTO, Mar, COTELO GARCÍA, Rosalía y PÉREZ PASCUAL, José Ignacio: *Historia del léxico español*, A Coruña, 2007, pp. 25-38.

la ciudad de Córdoba (la ciudad y término jurisdiccional, topónimos, grupos sociales, la familia, gobierno municipal, fiscalidad concejil, aspectos de la administración de justicia y de la organización socio-económica de la ciudad)²⁵. No se ofrece una mera explicación del vocabulario, sino que se desgrana la relación de significados vinculados con todos esos aspectos, aportando además el número de frecuencias de los vocablos. Con este análisis, se profundiza en el conocimiento de las ordenanzas municipales de Córdoba, y de la fuente que las ha transmitido, y se ofrece una base empírica sólida para abordar un análisis histórico más complejo.

Otros corpus formados con material de archivo que están desarrollándose actualmente, y que hay que mencionar en este apartado, son el *Corpus Documental del Antiguo Reino de Murcia*, en un estado de desarrollo todavía muy incipiente²⁶, y el *Corpus Diacrónico del español del Reino de Granada* (CORDEREGRA), que incluye obras y documentos desde la creación del Reino de Granada en 1492: declaraciones de testigos en probanzas y pleitos, correspondencia oficial, entre la que se encuentra el epistolario del conde de Tendilla, cartas privadas, informaciones y averiguaciones, memoriales y peticiones de particulares, partes médicos, testamentos, cartas de dote y arras, inventarios de bienes²⁷. Este corpus, por el momento, es para uso interno del equipo investigador, que ya está publicando los primeros resultados de su aplicación²⁸.

Añadamos también como proyecto en curso el corpus de textos medievales castellanos de ámbito aragonés confeccionado por el Grup d'Història i Contacte de Llengües (GHCL), como base del *Diccionario de castellano de la Corona de Aragón*. Incluye manuscritos originales e impresos, actualmente 1.400.000 formas procedentes de documentos jurídicos y administrativos, muchos seleccionados del Archivo de la Corona de Aragón (documentos de protocolo con testimonios aljamiados, registros de cancillería del *sellell secret*, documentación de rentas reales, documentación nobiliar), obras de medicina, tratados morales, obras literarias, entre las que se incluyen la Crónica de Aragón, de Gauberto Fabrizio de Vagad, novelas de ficción, poesía de cancionero... El corpus estará lematizado, de manera que, para utilizar esas fuentes, resultará más ágil su manejo que el de otros corpus, como el CORDE²⁹.

²⁵ Puede consultarse en red, LÓPEZ MORO, Pilar: *Estudio del léxico de las "Ordenanzas del concejo de Córdoba" (1435)*, Málaga, Universidad de Málaga, 2003.

²⁶ ABAD MERINO, Mercedes: "Uso de corpus documental en sociolingüística histórica y retos para su elaboración en el sureste peninsular (siglos XV-XVII)", en ENRIQUE-ARIAS, Andrés (ed.), *Diacronía de las lenguas iberorrománicas*, pp. 251-266.

²⁷ Descripción del proyecto en CALDERÓN CAMPOS, Miguel y GARCÍA GODOY, María Teresa: "El *Corpus Diacrónico del Español del Reino de Granada (CORDEREGRA)*", en ENRIQUE-ARIAS, Andrés (ed.), *Diacronía de las lenguas iberorrománicas*, pp. 229-249.

²⁸ CALDERÓN CAMPOS, Miguel: "Escribanos y notarios en el Corpus del Reino de Granada (CORDEREGRA). Los inventarios de bienes como fuente lexicográfica", conferencia pronunciada en julio de 2010 en el Seminario de Lengua Española, *Fuentes documentales para el Nuevo diccionario histórico*, organizado por la Fundación Duques de Soria.

²⁹ Información sobre este corpus en LLEAL GALCERÁN, Coloma: "El Diccionari del Castellà del segle XV a la Corona D'Aragó del GHCL de la Universitat de Barcelona", *Estudis Romànics*, 30, 2008, pp. 241-247 y ANGLADA ARBOIX, Emilia: "Un banco de datos electrónicos: a propósito de la confección de un

Se encuentra disponible actualmente otro corpus que posee una particularidad respecto a los corpus mencionados en este apartado. Se trata del *Corpus diacrónico vasco* (CODIVAS), también llamado *Corpus On-line de Vasconia*. La particularidad de este corpus es que, aunque ideado por lingüistas (el proyecto está dirigido por Ángeles Líbano Zumalacárregui), su concepción tuvo lugar en el marco del proyecto histórico *De la lucha de bandos a la hidalguía universal. Transformaciones sociales, políticas, económicas y culturales en el País Vasco al final de la Edad Media* (dirigido por José Ramón Díaz de Durana en la Universidad del País Vasco). El corpus ofrece, actualmente, la posibilidad de realizar búsquedas onomásticas y léxicas, con número de frecuencias, en documentación y en textos de los siglos XIV al XVI (la *Colección documental de los Parientes Mayores de Álava, 1332-1512, y de Guipúzcoa, 1319-1438*; la obra *Buenas Andanzas e Fortunas de Lope García de Salazar*)³⁰. Nos encontramos aquí con el resultado de una colaboración interdisciplinaria fructífera entre profesionales de la lengua y de la historia. El corpus permitirá analizar el léxico romance relacionado con la sociedad y los conflictos socio-políticos que tuvieron lugar en el País Vasco entre los siglos XIV y XV, y profundizar en el conocimiento de los linajes, gracias a la información antroponímica y toponímica contenida en el corpus³¹. No es, por tanto, un corpus ideado con criterios exclusivos filológicos, sino pensado para adaptarse a las investigaciones históricas. Lo cual indica que actualmente los historiadores medievalistas españoles están empezando a tomar conciencia de la necesidad de contar con recursos que ayuden a profundizar en el conocimiento del sentido de las fuentes que habitualmente se manejan, apoyándose en el procesamiento informático de los datos textuales.

CORPUS DE OTROS ÁMBITOS IDIOMÁTICOS HISPÁNICOS

Los avances que ha experimentado para el estudio del español la historia de la lengua o de la lexicografía también se constatan para el estudio de otras lenguas de la realidad idiomática del actual estado español, lenguas que se gestaron igualmente como romances en la época medieval. Por tanto, hay que conocer los corpus textuales que los filólogos catalanes o gallegos han creado para el estudio de esas lenguas. Así, es posible abarcar el conjunto de las realidades históricas de los reinos ibéricos medievales, lo cual proporcionará la posibilidad de establecer análisis comparativos o interrelacionados.

diccionario del castellano del siglo XV en la Corona de Aragón”, en Mar CAMPOS SOUTO, Rosalía COTELO GARCÍA, José Ignacio PÉREZ PASCUAL (coords.), *Historia del léxico español*, La Coruña, 2007, pp. 9-24.

³⁰ La base sigue ampliándose, esperando abarcar el período de los siglos XIII-XVI. Presentación e interfaz de búsqueda en <http://www.ehu.es/anlibano/index.html> (consultado el 19-3-2011).

³¹ LÍBANO ZUMALACÁRREGUI, Ángeles y VILLACORTA MACHO, Consuelo: “El corpus digitalizado bajomedieval de Vasconia: realidades y perspectivas de investigación”, en *Comunicación social en el siglo XXI. XII Simposio Internacional de Comunicación Social. Santiago de Cuba, 17-21 enero de 2011*, vol. I, pp. 43-48 (consultable en red: <http://www.santiago.cu/hosting/linguistica/actas.php?Simposios=XII&Actas=1>, fecha de consulta, 17-3-2011).

Para el catalán, se encuentran disponibles en red el *Corpus Textual Informatitzat de la Llengua Catalana* (CTILC)³² y el *Corpus Informatitzat del Català Antic* (CICA)³³. Éste último resulta especialmente interesante, por la relación de fuentes medievales que incorpora: desde juramentos de paz y tregua, textos legislativos importantes, como los Usatges de Barcelona, procesos criminales, capítulos de *greuges*, manuales de *consells*, cartas reales, dietarios de la Generalitat, crónicas como la de Desclot o la de Muntaner, sermones de Vicente Ferrer, otras obras religiosas o morales, tratados médicos, epistolarios, y muchas otras fuentes imprescindibles para el conocimiento de la historia medieval en el ámbito catalán.

Para el gallego contamos con el *Tesouro Medieval Informatizado da Lingua Galega* (TMILG)³⁴. El 80% de las fuentes incluidas en el corpus es prosa notarial, el 20% prosa histórica y el 10%, verso³⁵. Además de obras en gallego de Alfonso X (las *Cantigas de Santa María*, pero también las *Partidas*), encontramos crónicas religiosas, cartularios monásticos, fueros (Santiago, Castelo Rodrigo), tratados de albeitería, documentos notariales...

En estrecha relación con el gallego, contamos también con el *Corpus do Português*, diseñado por Mark DAVIS y por Michael J. Ferreira. El interfaz de búsqueda es igual que el del Corpus del Español, de manera que comparte sus virtudes para el procesamiento de unos cinco millones de palabras pertenecientes a obras de los siglos XIII-XV³⁶.

CORPUS LATINOS

El sistema lingüístico en la Edad Media es predominantemente latino hasta el siglo IX, y latino y romance, desde el siglo XII en adelante. Entre los siglos IX y XII se extiende un período en el que se produce la mezcla del latín y del romance en los testimonios escritos: un latín romanceado y un romance latinizado. Esta situación tan interesante ha llevado a los lingüistas a ocuparse del léxico de los testimonios escritos latinos de los inicios de la Edad Media peninsular, con el objetivo de estudiar el proto-romance. Actualmente se han incorporado también las nuevas tecnologías a estos análisis, creando herramientas y corpus textuales que interesan también a los historiadores. El más avanzado actualmente, y que se encuentra disponible en red, es el *Corpus Documentale Latinum Gallaeciae* (CODOLGA)³⁷. El equipo está dirigido por José E. López Pereira, desde la Universidad de La Coruña. El objetivo del grupo responsable no es sólo incluir documentación medieval, sino todo documento o texto latino relacionado

³² Acceso a través de <http://ctilc.iec.cat/> (consultado el día 14-2-2011).

³³ Acceso a través de <http://webs2002.uab.es/sfi/cica/> (consultado el día 14-2-2011).

³⁴ Acceso a través de <http://ilg.usc.es/tmilg/> (consultado el día 14-10-2010).

³⁵ Véase la descripción de este recurso en PICHEL, Ricardo y VARELA BARREIRO, Xavier: "O corpus do galego medieval *Tesouro Medieval Informatizado da Lingua Galega (TMILG)*", en ENRIQUE-ARIAS, Andrés (ed.), *Diacronía de las lenguas iberorrománicas*, pp. 195-216.

³⁶ Acceso al listado de obras y al interfaz de búsquedas a través de <http://corpusdoportugues.org/x.asp> (consultado el 14-2-2011).

³⁷ El usuario puede registrarse libremente y acceder al interfaz de búsquedas y contenidos a través de <http://corpus.cirp.es/codolga/> (consultado el 14-10-2010).

con Galicia hasta 1500, inédito o editado. Se incluyen, claro está, los principales cartularios monásticos (Samos, Sobrado, Oseira, Celanova, Xubia, Lourenzá, Carboeiro, Caaveiro, Antealtares, San Martiño Pinario y otros), documentación de las catedrales de Santiago, Mondoñedo, Lugo, Tui y Ourense y otros textos literarios o historiográficos (*Crónica de Hidacio*, *Annales Portugalenses Veteres*). En total, en la actualidad hay 11.807 registros de 148 ediciones, más 180 registros de la documentación real anterior al año 1000, que se ha transcrito expresamente para el corpus³⁸. El interfaz de búsqueda recupera resultados de concordancias para palabras concretas (y algunas variantes por medio de comodines) o de grupos de palabras. Se han incluido también criterios paleográficos, de manera que pueden obtenerse interesantes estadísticas partiendo de las búsquedas léxicas (por soporte documental, por otorgante –particular, real, episcopal o pontificio, o sin especificar–, por procedencia y por fecha).

El corpus sigue ampliándose anualmente, y ya está generando resultados (algunos de ellos publicados desde la misma página del proyecto), no sólo referidos al ámbito de la historia de la lengua³⁹, sino también al de los estudios históricos, pues desde los inicios se comprendió el valor de esta herramienta para el mejor conocimiento de la realidad medieval en ámbito gallego⁴⁰. Por ejemplo, gracias a la posibilidad de estudiar la toponimia de forma sistemática y contextual con el CODOLGA, es posible conocer mejor los orígenes de ciertos concejos⁴¹, o realizar estudios de conjunto más amplios sobre el desarrollo y evolución del poblamiento y del territorio⁴².

En la línea de este proyecto se inscriben otros actualmente en proceso de elaboración, y que aún no han sido puestos enteramente a disposición de la comunidad investigadora. No obstante, es importante seguir el desarrollo de estos proyectos, puesto que, una vez publicados y en pleno rendimiento, proporcionarán la posibilidad de consultar nuevos corpus para los distintos ámbitos territoriales ibéricos. Así, los promotores del proyecto vinculado con la redacción del *Glossarium Mediae Latinitatis Cataloniae* han considerado necesario digitalizar los textos de referencia base

³⁸ DÍAZ DE BUSTAMANTE, José Manuel: “Noticia e presentación dunha nova ferramenta de investigación: Corpus Documentale Latinum Gallaeciae (CODOLGA)», *Compostellanum* vol. 49 (2004), pp. 755-762 e ID., «Notas á xestión de bases de datos textuais: a documentación latina medieval do Reino de Galicia», en BOULLÓN AGRELO, Ana Isabel, COUCEIRO PÉREZ, Xosé Luís y FERNÁNDEZ REI, Francisco (ed.): *As tebras alumeadas: estudos fíloxicos ofrecidos en homenaxe a Ramón Lorenzo*, Santiago de Compostela, 2005, pp. 115-126.

³⁹ COUCEIRO, Xosé Luís: “Codolga na investigación do léxico hispánico primitivo”, en Ana Isabel BOULLÓN AGRELO, *As tebras alumeadas: estudos fíloxicos ofrecidos en homenaxe a Ramón Lorenzo*, Santiago de Compostela, 2005, pp. 83-102.

⁴⁰ LÓPEZ ALSINA, Fernando: “La red y las fuentes documentales medievales: el ejemplo de CODOLGA”, en *Pescar o navegar. La Edad Media en la Red*, Zaragoza, 2005, pp. 19-40.

⁴¹ REI, Maxi: “Referencias documentais ó concello de Abegondo durante a Alta Idade Media”, *Anuario brigantino*, N.º. 27 (2004), pp. 179-198.

⁴² Como puede observarse en los trabajos de SÁNCHEZ PARDO, José Carlos: “Una propuesta de análisis geohistórico de las aldeas tradicionales gallegas”, *Cuadernos de Estudios Gallegos*, Vol 54, No 120 (2007), pp. 103-134; ID., «Las iglesias rurales y su papel en la articulación territorial de la Galicia medieval (ss. VI-XII): un caso de estudio», *Melanges de la Casa de Velázquez*, N.º 40, 1 (2010), pp. 149-170, o su tesis doctoral, *Territorio y poblamiento en Galicia entre la antigüedad y la plena Edad Media*, Santiago de Compostela, 2008.

de la elaboración del diccionario latino, un amplio corpus fundamentalmente documental que incluye documentos notariales de todo tipo, textos legales y costumbres locales (como la compilación del *Liber Iudicum* realizada por el jurista barcelonés Bonsom, los *Usatges* o el *Liber feudorum*), pero también textos literarios como las epístolas y poemas del abad Oliva, textos científicos o hagiográficos⁴³. Otro corpus que estará a disposición de la comunidad científica será el *Corpus Documentale Latinum Valentiae* (CODOLVA), integrado en el proyecto dirigido por Juan Francisco MESA SANZ desde la Universidad de Alicante. El proyecto tiene como objetivo estudiar la implantación del latín en el Reino de Valencia desde su conquista, a través del estudio de la cancillería regia, y para ello crearán un corpus documental informatizado que incluirá toda edición de documentos relativos al Reino de Valencia. En la presentación del proyecto se dice expresamente que está previsto facilitar su utilización a los investigadores de los diversos campos de la ciencia histórica y jurídica⁴⁴.

En el ámbito leonés y castellano, se viene desarrollando desde hace años el *Lexicon Latinitatis Medii Aevi Regni Legionis* (LELMAL) por un equipo formado por filólogos e historiadores de la Universidad de León. El objetivo es crear un corpus de textos digitalizados (de documentos y de historiografía) y un diccionario. Actualmente se han digitalizado todos los documentos del reino asturleonés publicados hasta 2001 y textos cronísticos⁴⁵, y se ha publicado el estudio de 3.020 términos latinos⁴⁶. La tercera fase del proyecto se encuentra en preparación, y está orientada a incluir también y analizar el corpus de la documentación y obras latinas del reino castellano. Al término de esa fase, quedará concluido el nuevo corpus y diccionario latino del reino de León y Castilla (LELMACEL). El proyecto cuenta con el patrocinio de la Real Academia de la Historia y la Union Académique Internationale⁴⁷.

CORPUS TEXTUALES QUE INTEGRAN CRITERIOS PALEOGRÁFICOS Y DIPLOMÁTICOS

Entre los proyectos de elaboración de Corpus textuales, hemos encontrado algunos cuyos objetivos son abiertamente multidisciplinares, pues pretenden abordar el análisis del documento en su dimensión material, lingüística e histórica. Así, el corpus

⁴³ GÓMEZ RABAL, Ana: “La realització d’un projecte lexicogràfic: el *Glossarium Mediae Latinitatis Cataloniae*; la seva història, el seu present, les seves perspectives”, *Estudis Romànics*, 30 (2008), pp. 215-221.

⁴⁴ MESA SANZ, Juan Francisco: “Corpus Documentale Latinum Valentiae”, en *II Simposio Internacional Vers una sintaxi històrica del català: metodologia i objectius. 10-12 nov 2004*, Versión electrónica 28/10/2007, <http://hdl.handle.net/10045/2532> (consultado el 14-3-2011).

⁴⁵ A partir de LÓPEZ PEREIRA, José E., DÍAZ BUSTAMANTE, José Manuel, VÁZQUEZ BUJÁN, Manuel Enrique y LAGE COTOS, María Elisa: *Corpus Historiographicum Latinum Hispanicum Saeculorum VIII-XII: Concordantiae*, Hildesheim-Zürich-Nueva York, 1993. Relación completa de la documentación incluida en <http://www3.unileon.es/grupos/lexicon/corpusleon.htm> [consultado el 20-2-2011].

⁴⁶ Recientemente publicados: PÉREZ, Maurilio (dir.), *Lexicon Latinitatis Medii Aevi Regni Legionis (s. VIII-1230) Imperfectum*, Tuhnhout, 2010.

⁴⁷ La página principal del grupo LELMACEL: *Lexicon Latinitatis Medii Aevi Castellae et Legionis (s. VIII-1230)* en <http://www.3unileon.es/grupos/lexicon> (consultado el 21-11-2010).

resultante no ofrece únicamente la transcripción del documento, con la posibilidad de incorporar búsquedas léxicas, sino que añade criterios de presentación y de edición paleográfica del documento, e imágenes facsimilares. Un proyecto centrado en la documentación de archivo inédita es el *Corpus de Documentos Españoles Anteriores a 1700* (CODEA). Se desarrolla desde 1996 y está integrado en los proyectos del Grupo de Investigación de Textos para la Historia del Español (GITHE) de la Universidad de Alcalá de Henares⁴⁸. El objetivo expreso del CODEA es aproximarse al documento histórico desde una perspectiva amplia, abarcando las ciencias y técnicas historiográficas (paleografía, numismática, epigrafía), y la historia general, sin olvidar los aportes de disciplinas relacionadas con la historia de la lengua y con la cultura escrita, pues la finalidad es valorar los diversos factores que influyen en el documento final⁴⁹. De momento, el CODEA incluye 1.500 documentos perfectamente catalogados y clasificados según una amplia tipología (documentos cancillerescos, municipales, eclesiásticos, de particulares) procedentes de fondos de distintos archivos españoles, fechados desde el siglo XI. A ese volumen se incorporarán las transcripciones generadas por el GITHE, entre ellas, las de los documentos medievales de la catedral de Toledo⁵⁰.

OTROS CORPUS

En este apartado hemos incluido algunos corpus que también han sido creados para servir al conocimiento de la lengua (en este caso castellana), pero basándose no en un amplio corpus de tipologías textuales variadas, sino en un sólo tipo de obra o género. Pero no por ello resultan menos útiles. Especialmente importante es poder disponer en red del *Corpus Biblia Medieval*, que permite “acceder en paralelo” a la transcripción de la casi la totalidad de los manuscritos de biblias medievales romanceadas en castellano que se han conservado, así como a la imagen facsímil de los originales. El corpus ha sido diseñado y es mantenido por Mark Davis, director técnico del proyecto. Según la página de presentación del corpus, están incluidos los primeros romanceamientos de la *Biblia* (misc. XIII), la *Biblia prealfonsí* (E8/E6), *La General estoria* (GE), la *Biblia Escorial* I.i.3 (E3), la *Biblia de Ajuda* (Aj), la *Biblia Escorial* I.ii.19 (E19), la *Biblia Escorial* I.i.7 e I.i.5 (E7/E5), la *Biblia Évora* (Ev), la *Biblia Escorial* I.i.4 (E4), la *Biblia de la Biblioteca Nacional de Madrid* (BNM), la de la Real Academia de la Historia (RAH), la *Biblia de Alba* (Alba), la *Miscelánea XV* (Misc. XV), así como la *Vulgata latina*, y la *Biblia hebrea*⁵¹. El proyecto está dirigido por Andrés Enrique-Arias, de la Universitat de les Illes Balears, y ofrece, además

⁴⁸ Desde la página del GITHE puede accederse a las numerosas publicaciones del grupo (<http://www.textoshispanicos.es/> fecha de consulta 21-11-2010).

⁴⁹ Descripción del corpus en SÁNCHEZ-PRieto BORJA, Pedro, PAREDES GARCÍA, Florentino, MARTÍNEZ, Rocío, MIGUEL FRANCO, Ruth, SIMÓN PARRA, María y VICENTE, Irene: “El *Corpus de Documentos Españoles Anteriores a 1700 (CODEA)*”, en Andrés ENRIQUE-ARIAS (ed.), *Diacronía de las lenguas iberorrománicas*, pp. 25-38.

⁵⁰ Información completa en la página del grupo: <http://demos.bitext.com/codea> (consultado el 15-2-2011).

⁵¹ Acceso a todos los contenidos del *Corpus Biblia* a través de <http://www.bibliamedieval.es> (consultado el 15-2-2011).

del corpus, amplia información crítica sobre los códices, bibliografía sobre biblia romanceada, artículos, y acceso a otros recursos similares, como el *Corpus Biblicum Catalanicum*, que igualmente pretende formar un corpus de traducciones al catalán de la Biblia, incluidas las medievales⁵².

El análisis de la Biblia resulta fundamental para comprender buena parte del universo conceptual medieval. Tanto si se transmitía en latín, como si se transmitía en romance, la Biblia proporcionó material inacabable para el imaginario y para la concepción del mundo en la visión medieval. Por ello, es importante poder contar con un recurso como el Corpus Biblia, que permite conocer, a través del léxico, cómo influyó el vocabulario de la Biblia en la forma de actuar y de transformarse la visión medieval de la sociedad.

Otro corpus realizado con una concepción similar al *Corpus Biblia*, pero sobre una fuente completamente distinta es el *Electronic Corpus of 15th Century Castilian Cancionero Manuscripts*, proyecto dirigido por Dorothy Severin desde la Universidad de Liverpool, en colaboración con las de Birmingham y Barcelona. El corpus está realizado sobre la base de la magna obra de Brian Dutton, incluyendo la información anexa de su Catálogo-Índice⁵³, corpus al que se han añadido otras transcripciones de Dorothy Severin y Fiona Macguire. El interfaz de búsqueda permite acceder al texto de los manuscritos y también realizar búsquedas léxicas⁵⁴. La poesía de cancionero ha sido una fuente empleada en los últimos años para analizar numerosas cuestiones relacionadas con las prácticas cortesanas o con los usos culturales nobiliarios, las relaciones sociales, así como un medio de transmisión del discurso político, de manera que este recurso servirá para perfilar hipótesis planteadas sobre la base de esta fuente.

USO DE CORPUS TEXTUALES INFORMATIZADOS ENTRE LOS HISTORIADORES

La lectura directa de las fuentes medievales, aunque resulte imprescindible, no puede constituir por sí sola un método de investigación histórica fiable: es algo obvio que no merecería la pena ser indicado. El esfuerzo de lectura de los textos es siempre limitado. Por ello no hay que desdeñar los procedimientos tecnológicos con los que se pueda ir más allá de la simple lectura y de la realización de fichas. Partiendo de un mayor aporte de información y de medios de rápido procesamiento de esa información es posible efectuar las clasificaciones de los datos, o realizar comparaciones, de una manera más eficaz y completa. Hasta hace muy poco no se disponía de bases de datos que permitieran el manejo simultáneo de la información de tantas fuentes medievales importantes. Los corpus que acabamos de describir permiten realizar esta labor. La consulta simultánea de una masa de textos, la obtención de concordancias

⁵² Acceso al *Corpus Biblicum Catalanicum* a través de <http://www.abcat.org/cbcat/inici.php>, (consultado el 15-2-2011).

⁵³ DUTTON, Brian (ed.), *Cancionero del siglo xv: c. 1360-1520*, Salamanca: Universidad de Salamanca, 1990-1991, 7 vols., y *Catálogo-Índice de la Poesía Cancioneril del siglo xv*, Madison, Hispanic Seminary of Medieval Studies, 1982.

⁵⁴ Acceso a través de <http://cancionerovirtual.liv.ac.uk> (consultado el 15-2-2011).

o de agrupaciones léxicas, que pueden dar lugar a la realización de estadísticas, el estudio de las distintas relaciones de significado que pueden analizarse de forma combinada en la sincronía y en la diacronía..., todo ello puede resultar de gran utilidad si se incorpora como práctica habitual en el trabajo medievalista, pues es importante delimitar en toda investigación los campos semánticos que resulten pertinentes para la comprensión de las fuentes, antes de aplicar conceptos que la mayor parte de las veces resultan anacrónicos. La estructura del sentido de las fuentes se reconstruye mejor partiendo de las propias fuentes y de las relaciones de significados que se forman en el contexto determinado y a lo largo del tiempo: los conceptos historiográficos que se apliquen a las investigaciones resultarán, de ese modo, menos artificiosos o anacrónicos. Intentar que los textos se adapten a nuestros conceptos puede derivar en explicaciones imperfectas y formalistas, alejadas de la perspectiva histórica. Por ello es importante no dejar de utilizar los diccionarios históricos (y no limitarse a uno sólo) y toda la información posible que pueda aclararnos el significado de las palabras y la relación semántica que se establece entre ellas. Actualmente se percibe un renovado interés en el medievalismo hispánico por introducir el análisis del léxico en las investigaciones históricas, de forma metódica y sistemática, tal y como reflejan, por ejemplo, algunos proyectos de historia económica⁵⁵.

Si la historia de la lengua, rama hermana de la filología, pero también de la historia, ha reforzado la base empírica de sus investigaciones empleando los corpus textuales, reduciendo así la perspectiva que reposa únicamente en la intuición, no hay por qué pensar que no pueda ocurrir lo mismo con los estudios históricos generales. Para el caso de la historia medieval puede resultar más útil, si cabe, pues el colectivo medievalista no puede prescindir del estudio de las fuentes incluidas en estos corpus.

Las aplicaciones a la historia de estas herramientas están todavía por explorar. Por ello consideramos necesario realizar el intento. Los corpus textuales que hemos descrito incluyen una tecnología y unas posibilidades de búsqueda, que, utilizadas al máximo, pueden de por sí generar importantes resultados. En otros países parece que se hayan empleado en los estudios históricos los corpus elaborados para el estudio de la historia de la lengua. Pero ello es debido, quizá, al menor desarrollo que ha alcanzado en otros países este tipo de corpus, en comparación con los corpus del español⁵⁶. O quizá se deba a la preferencia por otros métodos de análisis léxico.

Fuera de España, han tenido mayor extensión los métodos de la estadística léxica o lexicometría. Los corpus textuales descritos proporcionan una importante base

⁵⁵ Me estoy refiriendo aquí a proyectos como el que dirigió Eloísa Ramírez Vaquero sobre “Recursos naturales y espacios de montaña. Léxico y prácticas relacionadas con los recursos naturales en las sociedades pirenaicas. Antigüedad-Edad Moderna», entre 2004 y 2006 en la Universidad Pública de Navarra, o el proyecto en curso sobre léxico de la fiscalidad en Occidente: « Les mots de l’impôt dans l’Occident méditerranéen. Glossaire critique de fiscalité médiévale » dirigido conjuntamente por Manuel Sánchez MARTÍNEZ (Institució Milà i Fontanals-CSIC) y Denis Menjot (Université de Lyon). La base con el glosario de fiscalidad está disponible en <http://www.mailxxi.com/fiscalitat2009v1/> (consultado el 10-4-2011).

⁵⁶ Para comprender el potencial de los Corpus españoles, baste la comparación del más representativo, el CORDE, con su equivalente inglés, el llamado *Helsinki Corpus*: el CORDE es sesenta veces más extenso que el corpus inglés; DAVIS, Mark, “Creating useful historical corpora”, *art. cit.*, p. 139.

cuantificable, pero cuentan también con limitaciones, precisamente de orden cuantitativo y estadístico. La práctica de la estadística léxica se ha desarrollado por medio de otras aplicaciones informáticas, *logiciels*, bases de datos con las que procesar la información textual y obtener estadísticas más fiables⁵⁷. El principio es inverso al de la aplicación de los corpus textuales que hemos descrito, pues en este caso, quien hace la investigación histórica debe construir el corpus. De manera que suelen utilizarse corpus de textos menos extensos o incluso el corpus puede formarse partiendo de un sólo texto vertido al ordenador. La aplicación informática utilizada genera índices de frecuencia, listas de las palabras más frecuentes, o de las estructuras sintácticas más frecuentes, listas de colocaciones o de coocurrencias, y otros indicadores léxicos. La variabilidad de frecuencias debe servir para la reconstrucción de los campos léxicos o semánticos que servirán de base para la interpretación histórica⁵⁸.

La lexicometría o la estadística léxica tienen también sus complicaciones y sus imprecisiones, pero si algo han puesto de manifiesto es la necesidad de que el historiador preste atención a las estructuras semánticas de los textos históricos que analizan. Alain GUERREAU, que mostró hace tiempo los límites de la estadística lingüística, pero también su enorme utilidad para el historiador, llamó la atención sobre ello. El objetivo debe ser llegar a la estructura semántica de un texto histórico, desentrañando las redes de asociación, de jerarquía, de oposición y de analogía de significados, así como las variables sociolingüísticas, operación indispensable para la lectura y comprensión de un texto y de la sociedad que lo origina. Sólo así pueden aflorar las múltiples lecturas de una fuente y la relación de significados. Esa estructura semántica está inserta en la estructura social, objeto de estudio del historiador, de modo que no es posible desdeñar ciertos procedimientos cuantitativos léxicos, por muy rudimentarios que sean, para llegar a la comprensión del texto. Al plantear diversos tipos de investigaciones de estas características (análisis de coocurrencias o de segmentos repetidos, examen sistemático de los contextos de una lista de palabras, comparaciones globales sobre listas elaboradas por uno mismo o por otros), afirma Guerreau que la mera determinación de las palabras-clave puede dar lugar a resultados fructíferos⁵⁹.

La tradición de ensayo de análisis factoriales de léxico aplicados al discurso histórico político o social cuenta ya con una larga trayectoria fuera de España, fun-

⁵⁷ Algunas de estas aplicaciones utilizadas hasta ahora en historia, aunque, a buen seguro, ya superadas, en BONIN, Emmanuel y DALLO, Alain: «Hyperbase et Lexico 3, outils lexicométriques pour l'historien», *Histoire et mesure*, XVIII, ¾, 2003, pp. 389-402.

⁵⁸ Un modelo de este tipo de investigación puede verse en el trabajo de GUERREAU, Alain: «Le champ sémantique de l'espace dans la vita de Saint Maieul (Cluny, début du XI^e siècle)», *Journal des Savants* (1997), pp. 363-419. Este historiador basa sus interpretaciones históricas en el análisis semántico de forma habitual (otro trabajo suyo, «Vinea», en Monique GOULLET, y Michel PARISSÉ (ed.), *Les historiens et le latin médiéval, Colloque tenu à la Sorbonne les 9, 10 et 11 septembre 1999*, París, 2001, pp. 67-73.

⁵⁹ Véanse las consideraciones de este autor, así como los trabajos que en su época comenzaban a aplicar la lexicometría a la historia medieval: GUERREAU, Alain: “Pourquoi (et comment) l'historien doit-il compter les mots?”, *Histoire & Mesure* (1989), IV-1/2, pp. 81-105.

damentalmente en Francia⁶⁰, en donde se vinculó con la incorporación del estructuralismo al campo de la historia⁶¹. Y a pesar de las críticas que sostienen que aún no se han abordado con rigor los métodos que permitan esclarecer la lógica interna de sentido de la sociedad medieval⁶², actualmente continúan vigentes ciertos proyectos que contribuyeron a esa renovación metodológica. Uno de los ejes de investigación del Laboratoire de Médiévisique Occidentale de Paris (LAMOP), dirigido por Jean-Philippe Genet, es *Histoire médiévale et informatique*, cuya finalidad es potenciar el uso de las nuevas tecnologías de la información y de sus diversas aplicaciones. Precisamente, desde este eje se promueve la aplicación lexicológica, la lexicología cuantitativa y el empleo de instrumentos de análisis semánticos a partir del estudio de los contextos⁶³. Jean-Philippe Genet reflexionaba hace años sobre la colaboración entre historiadores y lingüistas para el desarrollo de la historia del discurso político, y uno de los problemas que planteaba el director de este proyecto eran los deficientes corpus textuales existentes⁶⁴. Si en España no ha habido un desarrollo similar de la tecnología informática léxica que pueda ser aplicada a la historia medieval⁶⁵, sin embargo, la gran magnitud de los Corpus textuales disponibles hoy podría compensar, en parte, esa dificultad.

⁶⁰ MATORÉ, Georges: *Le vocabulaire et la société médiévale*, París, 1986 fue uno de los primeros intentos de aplicación de la lexicología al conocimiento de la sociedad medieval, aunque hoy resulta insuficiente. La aplicación de la lexicometría a la historia tiene lugar en los años 80 del siglo XX; las revistas *Histoire & Mesure*, publicada desde 1986 y *Le médiéviste et l'ordinateur. Histoire médiévale, informatique et nouvelles technologies*, publicada desde 1989 (hasta 2007) constituyeron un marco para su desarrollo.

⁶¹ Resume esta trayectoria CEAMANOS LLORENS, Roberto: "Lingüística e historia social. Los comienzos de una aventura interdisciplinar en la historiografía francesa", *Brocar*, 26 (2002), pp. 265-277. Llama la atención que una obra pionera de la relación entre lingüística e historia, como fue la de ROBIN, Régine: *Histoire et linguistique*, París, 1973, no haya tenido traducción en España.

⁶² Alain GUERREAU se muestra muy crítico con los escasos avances de la metodología léxica en los estudios medievales. Según GUERREAU, quedan por elaborar las reglas prácticas del análisis metódico de los campos semánticos: «es la tarea más urgente para los medievalistas»; GUERREAU, Alain: *El futuro de un pasado. La Edad Media en el siglo XXI*, Barcelona, 2002, pp. 138-222.

⁶³ Las líneas de investigación y de trabajo de este equipo de investigación del LAMOP se declaran en la página del grupo <http://lamop.univ-paris1.fr/spip.php?rubrique58> [fecha de consulta 10-2-2011]. Se trata además de un proyecto con finalidad pedagógica, para incorporar a los estudiantes de últimos cursos o de doctorado a la iniciación de la investigación.

⁶⁴ GENET, Jean Philippe : «Le médiéviste, la naissance du discours politique et la statistique lexicale: quelques problèmes», en C. BOURLET et A. DUFOUR (ed), *L'écrit dans la société médiévale. Divers aspects de sa pratique du XI^e au XV^e siècle*, París, 1991, pp. 289-298.

⁶⁵ A pesar de ello, desde muy pronto se constatan iniciativas, pero para otros períodos históricos, Juana Ugarte ideó, junto con otras colaboradoras del Centro de Cálculo Electrónico del CSIC, en 1977, un sistema analítico para el estudio histórico de los procesos discursivos en torno a la palabra «economía». El resultado fue un estudio que planteaba el discurso como un hecho social, puesto que tenía en cuenta el proceso comunicativo, que tiene, en sí mismo, un carácter social por excelencia: UGARTE BLANCO, Juana:, *Discurso. Historia. Informática: la palabra economía en los textos económicos españoles del siglo XVIII*, Oviedo, 1996.

La estadística léxica o la lexicometría es, por tanto, un campo todavía por desarrollar en el ámbito del medievalismo español⁶⁶. En algo más de una década se han introducido en España temas de estudio, en historia cultural o política⁶⁷, vinculados de algún modo con las corrientes historiográficas en boga en los años 90 del siglo xx. Se incorporaron nuevos temas y perspectivas de análisis, como, por ejemplo, los introducidos por la corriente que, por entonces, se proclamaba como «Nueva Historia política», que ha venido dando sus frutos en los últimos años. Pero hay que reconocer que esa renovación temática no ha venido acompañada de la suficiente renovación metodológica⁶⁸.

Los corpus textuales objeto de nuestro análisis apenas han sido utilizados en historia para suplir la ausencia de una metodología lexicométrica, tal vez debido a su reciente desarrollo. Apenas se conocen los potenciales de tales recursos en Historia Medieval, ni en la historia general. En parte, también, porque las innovaciones metodológicas se incorporan con mayor lentitud.

ADMYTE no resulta desconocido en el ámbito histórico. Aunque no con la frecuencia que se debería, su uso se ha ido introduciendo, cuando se trabaja con las fuentes transcritas en ADMYTE, especialmente con obras de carácter legislativo⁶⁹. Este recurso permite trabajar sobre testimonios manuscritos, lo cual resulta más apropiado que manejar ediciones decimonónicas. Al utilizar las ediciones electrónicas de ADMYTE puede decirse que se han introducido mejoras en la investigación, al emplearse unas versiones más fiables de las obras objeto de estudio, que las que han transmitido otras ediciones. Sin embargo, aún no existen trabajos que hayan aprovechado las listas de vocabulario, de frecuencias o las concordancias que también proporciona el Hispanic Seminary of Medieval Studies, a través de sus publicaciones en CD-ROM, de manera que aún no se han explotado todas las posibilidades metodológicas que proporciona este recurso. La totalidad del vocabulario de las obras

⁶⁶ Aunque existe algún trabajo sobre temática hispánica, pero no fraguado en el contexto investigador español, véase el empleo de análisis factoriales por FORONDA, François: «S'emparer du roi. Un rituel d'integration politique dans la Castille trastamara» en François FORONDA, Jean- Philippe GENET y José Manuel NIETO (dirs.), *Coups d'État à la fn du Moyen Âge?*, Madrid, 2005, pp. 213-329.

⁶⁷ NIETO SORIA, José Manuel: "La renovación de la historia política en la investigación medieval: las relaciones de poder", en *Relaciones de poder en la provincia de Cuenca y su evolución*, Cuenca, 1997, pp. 37-74.

⁶⁸ Tampoco parece que se haya avanzado demasiado en otros ámbitos de la historia, a pesar de propuestas realizadas en la década de los 90' del siglo xx, como la que hacía Carlos BARROS, que proponía aplicar la semántica a las Historia de las Mentalidades: BARROS, Carlos: "Historia de la Mentalidades: posibilidades actuales", en José María SÁNCHEZ NISTAL (ed.), *Problemas actuales de la Historia*, Salamanca, 1993, p. 56, n. 37.

⁶⁹ Algunos ejemplos de uso de ADMYTE: han consultado las transcripciones de ADMYTE de obras legislativas como las *Siete Partidas*, MORÍN, Alejandro: "Matar a la adúltera: el homicidio legítimo en la legislación castellana medieval", *Cahiers de linguistique et de civilisation hispaniques médiévales* 24, 2001, p. 354; el *Espéculo* o el *Ordenamiento de Alcalá*, NIETO SORIA, José M.: "Tiempos y lugares de la «realza sagrada» en la Castilla de los siglos XII al XV", *Cahiers de linguistique et de civilisation hispaniques médiévales*, 15 (2003), pp. 272-273; las *Leyes del Estilo*, OLIVA HERRER, R. Hipólito: "Espacios de comunicación en el mundo rural a fines de la Edad Media: la escritura como contrapeso del poder", *Medievalismo*, 16 (2006), p. 99.

de Alfonso X y de una parte de la documentación real ha sido ya clasificada por los filólogos, pero esos estudios aún no han sido incorporados a los estudios históricos⁷⁰.

Sólo por el hecho de constituir uno de los instrumentos más fiables para determinar el significado y el uso de determinada palabra en determinada época, o cuándo comienza a ser usada en romance, el CORDE resultaría una herramienta indispensable para la Historia Medieval. El CORDE proporciona una base sólida para discusiones conceptuales o terminológicas que resultan muy pertinentes⁷¹. Sin embargo, no es ésta la única aplicación en historia. No se encuentran demasiadas referencias al empleo del CORDE en publicaciones del ámbito del medievalismo. Y cuando es declarado su uso, el aprovechamiento parece limitado. Aunque a veces es empleado como un apoyo muy claro para la comprensión del significado⁷², no debería sustituir las consultas en un diccionario histórico⁷³, pues se podría incurrir en inexactitudes considerables o en ambigüedad. Otras veces se emplea simplemente para localizar una cita o referencia en un texto, o para consultar de forma rápida una fuente⁷⁴. Puede utilizarse también como apoyo a la transcripción de una fuente⁷⁵.

⁷⁰ KASTEN, Lloyd, NITTI, John y JONXIS-HENKEMANS, Wilhelmina (eds.), *The Electronic Texts and Concordances of the Prose Works of Alfonso X el Sabio*, Madison: HSMS, 1996.

⁷¹ Como las que hace Luis FERNÁNDEZ GALLARDO en su reseña de la obra de «Alain Dumerger, *Croisades et croisés au Moyen Âge*», *Atalaya*, nº 11, abril 2009 [en ligne], mis en ligne le 20 avril 2009. URL: <http://atalaya.revues.org/index444.html>. Consulté le 3 décembre 2010, a propósito de la introducción en el romance castellano del siglo XIII de los términos «cruzada» y «cruzados» (n. 3), o de la crítica a la interpretación que hace Dumerger de la idea de cruzada aplicada a la Reconquista. Luis FERNÁNDEZ GALLARDO precisa, usando el CORDE, que el término «Reconquista», tal y como fue considerado por la historiografía, fue acuñado a finales del siglo XVIII (n. 5). Este historiador incorpora habitualmente el uso del CORDE en sus últimos trabajos.

⁷² Un ejemplo es el empleo de David NOGALES RINCÓN, para aclarar el uso de dos términos en el texto que edita, «sciencia» y «orador» (por conocidas que parezcan resultar ciertas palabras, ninguna tiene un significado evidente); NOGALES RINCÓN, David, «En torno a la sabiduría en el cuatrocientos castellano: el prólogo a la traducción castellana del *De regno ad regem Cypri* de Tomas de Aquino dirigida a Fernando el Católico», *Memorabilia*, 12 (2009-2010), pp. 137-138.

⁷³ Sobre el escaso y desigual uso que hacen los historiadores medievalistas de los diccionarios históricos también habría que realizar algunas indicaciones. No parece que sean muy conocidos diccionarios como el *Diccionario del español medieval* de Bodo MÜLER, Heidelberg, 1987- (véase una reseña de este diccionario por Elena BAJO PÉREZ, en *Romanische Forschungen*, 118/4 (2006), pp. 486-491), o el *Nuevo Tesoro lexicográfico del español (siglos XIV-1726)*, de Lidio NIETO y Manuel ALVAR, Madrid, 2007. Tampoco parece muy utilizado el *Diccionario de la prosa castellana del rey Alfonso X*, de Lloyd A. KASTEN y John J. Nitti, en 3 vols, Nueva York, HSMS, 2003.

⁷⁴ Por ejemplo, NIETO SORIA, José Manuel: *Medievo constitucional*, Madrid, 2007, pp. 69 y 213 maneja una obra de Feijoo, citando la versión publicada en CORDE. En un trabajo reciente, este autor utiliza además los datos estadísticos que aporta el CORDE, como apoyo a su argumentación en torno al concepto de «España» en el siglo XV: Id., «Conceptos de España en la época de los Reyes Católicos», *Norba. Revista de Historia*, vol 19 (2006), pp. 105-123, pero en su búsqueda no introduce todas las variantes léxicas del término «España» que se usaban en las fuentes medievales, de manera que la estadística resulta incompleta.

⁷⁵ David NOGALES RINCÓN, para desarrollar la abreviatura de «reg» en un texto, como «regno» y no como «regimiento», tal y como habían preferido otros editores; NOGALES RINCÓN, David, «En torno a la sabiduría», p. 327, n. 139.

Hay, por tanto, cierto conocimiento del CORDE en los trabajos de historia medieval, pero todavía no se observa un empleo intensivo de este recurso. A veces, incluso, no se presta mucha atención a la forma de citarlo. No basta con usarlo, hay que decir cómo se ha usado. En cierta investigación, para precisar las variantes de significado de un concepto aludido se cita el CORDE, pero se deja la responsabilidad de la consulta (y comprobación) a quien lee el estudio. El autor no declara el tipo de búsqueda que ha realizado, ni en qué período cronológico, ni los datos sobre el número de casos, ni tampoco aporta una interpretación pertinente sobre la base de las concordancias o del contexto resultado de su búsqueda⁷⁶. En resumen, el CORDE, o bien se toma por un diccionario histórico, o se emplea con un uso equivalente al que se haría en la propia fuente escrita en papel o en una versión corriente de la obra digitalizada. Estos usos pueden resultar útiles, pero no introducen especiales novedades metodológicas⁷⁷.

ESBOZO DE UN EJEMPLO DE APLICACIÓN: LA NOCIÓN DE “ACUERDO”

Como hemos visto, son escasísimos los trabajos que en el ámbito de la historia medieval incorporan métodos de estudio lexicológico o aproximaciones lexicométricas, métodos que en otros ámbitos se han revelado como apropiados para no dejar al albur de la mera intuición el avance de determinadas tesis que explican la dinámica política medieval o las relaciones sociales. Tampoco ha sido muy fructífero el uso del CORDE⁷⁸. Es preciso mencionar aquí, por tanto, dos trabajos recientes que han puesto de manifiesto la utilidad de recursos como el CORDE. Me estoy refiriendo a un trabajo propio, «Léxico político en el Seguro de Tordesillas»⁷⁹ y al trabajo de François Foronda, sobre «el miedo al rey», ambos aparecidos en 2007. Especialmen-

⁷⁶ Así procede, por ejemplo, Jesús Rodríguez Velasco, cuando, para precisar el concepto de *ordo* dice que «Por lo común, al *ordo*, se le suele llamar también *estado*; en español medieval es más común hallar el sustantivo *estado* que *orden*», y en nota remite al CORDE y al *Corpus del Español*, pero sin detallar la búsqueda realizada, ni en qué período; sin explicar los resultados aportados, ni el número de casos, ni la frecuencia por años, de manera que es el lector el que debe comprobar si efectivamente es “más común” *estado* que *orden*. Véase RODRÍGUEZ VELASCO, Jesús: *Ciudadanía, soberanía monárquica y caballería; poética del orden de caballería*, Madrid, 2009, p. 11, n. 2.

⁷⁷ No es el caso, quizá, de otros corpus disponibles, como el CODOLGA (me remito a la descripción hecha de este corpus más arriba).

⁷⁸ El uso que hace del CORDE Óscar Villarroel, en un artículo reciente sobre el análisis léxico de las fórmulas de tratamiento del obispo Pedro de Castilla, descendiente del rey Pedro I, no parece relevante, pues tan sólo se recogen las menciones al prelado en tres crónicas incluidas en el corpus (y no parece que, de haber manejado las versiones de esas mismas crónicas en papel, la interpretación hubiera sido distinta); VILLARROEL GONZÁLEZ, Óscar: “El reflejo léxico de las posturas políticas: vaivenes políticos de un prelado de sangre regia”, *Espacio. Tiempo y Forma. Serie III. Historia Medieval*, 20 (2007), pp. 268-269.

⁷⁹ CARRASCO MANCHADO, Ana Isabel: “Léxico político en el Seguro de Tordesillas: conflicto, pactos y autoridad real”, en François FORONDA y Ana Isabel CARRASCO, *Du contrat d’alliance au contrat politique. Cultures et sociétés politiques dans la péninsule Ibérique à la fin du Moyen Âge*, Toulouse, 2007, pp. 85-138.

te, el primero de los trabajos se planteó como un primer ensayo sobre los potenciales metodológicos del CORDE para abordar la cuestión del lenguaje político, o del uso político del lenguaje, en la Castilla bajomedieval.

En el artículo de François Foronda, se intentó una primera aproximación al campo léxico del miedo, partiendo de un corpus cronístico escrito entre 1340 y 1380, centrado básicamente en Fernán Sánchez de Valladolid y Pedro López de Ayala. En las fuentes analizadas, y a través del léxico, se percibe la presencia de una «comunidad emocional» en torno al miedo, de la que participaban los ricos hombres y la familia real, vinculada con la «política del espanto» que parece ponerse en funcionamiento por esas fechas. Foronda toma como indicio de la alta presencia de esta forma de metus/miedo, los casos recogidos a partir del CORDE. El término «miedo» es buscado de forma combinada en conexión con la palabra «rey», y con otros vocablos que, por su cercanía y relación en el corpus analizado, Foronda sitúa en el campo léxico del miedo: «sosiego», «recelo», «miedo», «temor», «sospecha», «espanto», y algún otro como «imaginación», que reveló matices nuevos⁸⁰. En este artículo, el CORDE proporcionaba cierta base cuantitativa, por muy aproximativa que fuera, a un objeto de estudio que puede resultar un tanto inaprensible, como es el de las emociones políticas, pero que está interesando actualmente a los investigadores⁸¹. Un tema como el del “miedo político” obliga a estudiar de forma metódica el léxico que nos revela dicha emoción.

Una utilización más sistemática del CORDE se intentó en el trabajo sobre *El seguro de Tordesillas*, pieza historiográfica peculiar, atribuida al conde de Haro Pedro Fernández de Velasco y redactada en fechas posteriores a 1440. La obra recoge, mezclando narración y documentación, el proceso negociador que se llevó a cabo en Tordesillas, en 1439, bajo arbitraje del conde de Haro, siguiendo el procedimiento de un «seguro». La única edición de esta obra, que es la que se ha incluido en el CORDE, está hecha a partir de una obra impresa en 1611 y transcrita por Pedro Mantuano, secretario de Juan Fernández de Velasco. Esto hacía imprescindible la consulta complementaria del Ms. 9224 de la BNE, pero, haciendo el debido cotejo, no se resta utilidad a los resultados lexicométricos que pudieran efectuarse con la edición incluida en el CORDE. Los índices de frecuencia de aparición de una palabra en el texto (el número de casos que proporciona el programa) constituyeron un fundamento para la interpretación del texto y de la realidad política que refleja, complementando así las interpretaciones que surgieron de la lectura comprensiva de la obra y del análisis de la bibliografía y de otras fuentes manuscritas con las que se contrastó el texto.

La primera lectura de la obra generó una clasificación léxica inicial, de manera que se distinguieron varios bloques temáticos que circunscribían los problemas po-

⁸⁰ Véase la tabla aproximativa que elabora el autor con el número de casos obtenidos, y la interpretación de la misma en FORONDA, François: « El miedo al rey », *e-Spania* [En ligne], 4 | 2007, mis en ligne le 14 janvier 2010, Consulté le 05 octobre 2010. URL : <http://e-spania.revues.org/2273> 04 janvier 2011. URL : <http://e-spania.revues.org/2273>.

⁸¹ BOQUET, Damien y NAGY, Piroška : *Le sujet des émotions au Moyen âge*, Paris, 2009; y los trabajos de ROSENWEIN, Barbara H. : « Histoire de l'émotion: méthodes et approches », *Cahiers de civilisation médiévale*, vol. 49, 2006, pp. 33-48, e ID., *Emotional communities in the early Middle Ages*, Ithaca, 2006.

líticos presentes en un texto que era claro reflejo del momento crítico y convulso en el que se escribió. De ese modo, delimitando campos léxicos (palabras asociadas) y efectuando una jerarquía de palabras, a partir de las más frecuentes (o de las asociaciones de palabras más frecuentes), se podía determinar con mayor seguridad, que el de la simple lectura, el grado de importancia que se le daba en la obra a determinada cuestión, la jerarquía de valores políticos y los problemas fundamentales explícitos o implícitos en la negociación. Aparecían de ese modo las palabras clave que remitían a problemáticas clave de la obra. Tales palabras se agrupaban según apuntaran al conflicto, a la negociación, al compromiso, al gobierno, a la autoridad o a los valores políticos. La aproximación lexicométrica era muy simple, pero proporcionó una base amplia para la interpretación de los problemas históricos reflejados en el texto. En vez de tratar los campos léxicos de una manera formalista, se procedió para su interpretación a ponerlos en conexión con otras fuentes, crónicas y documentales del momento, y con el contexto histórico. El objetivo inicial era tratar de analizar un hecho histórico específico (la negociación del seguro en Tordesillas, en 1439) desde la práctica política y desde la perspectiva de su plasmación testimonial, viendo cómo se conjugaron las diferentes fuerzas en acción y como justificaban su acción con formulaciones políticas. Asuntos clave del reinado de Juan II se veían reflejados en el texto, en algún caso bajo aspectos completamente novedosos⁸².

A la vista de estos resultados, y dado el avance logrado en España en la elaboración de corpus textuales diacrónicos, tal y como se ha mostrado en las descripciones de los sucesivos corpus, y el potencial que pensamos poseen para el desarrollo de los estudios históricos medievales, es el momento de ensayar diversas aplicaciones a la Historia Medieval hispánica. Para ello hemos constituido un equipo de trabajo multidisciplinar, procedente del ámbito de la historia y de la filología, que se encargará de dar a conocer los principales corpus informáticos y de explorar nuevos métodos aplicados a la historia política o social de la Edad Media hispánica, empleando los Corpus textuales informatizados y el tratamiento lexicológico de las fuentes medievales⁸³. Nuestro objetivo es, por tanto, aprender a utilizar de forma más eficaz los corpus textuales e idear las posibilidades de su aplicación a la historia. Partiendo de la hipótesis de que la consulta masiva de las fuentes medievales recogidas en los corpus, y su tratamiento informático, proporcionan una base cuantitativa y empírica para los análisis históricos, es posible plantearse múltiples líneas de investigación.

Los ámbitos de análisis concretos que se elijan para realizar tales aplicaciones exigirán una metodología específica. El volumen de información que se maneje no será el mismo si se trata, por ejemplo, de analizar la ideología política de don Juan Manuel a través de sus obras, que si se plantea el análisis del discurso político en las cortes de León y Castilla, la visión de la sociedad en sermonarios y tratados doctrina-

⁸² La desnaturalización de todos los actores políticos durante la negociación, realizada al amparo del « poderío real absoluto » del monarca, había pasado completamente desapercibida en la lectura tradicional que habían hecho los historiadores que estudiaron esta obra. Remito, para una mayor explicación a CARRASCO MANCHADO, Ana Isabel: «Léxico político en el *Seguro de Tordesillas*, art. cit.», pp. 125-129.

⁸³ El equipo está compuesto por dos historiadores medievalistas, François Foronda, de la Universidad de París I (Pantéon-Sorbonne), y yo misma; una filóloga románica, Sophie Hirel, de la Universidad de París Este; y una lingüista Elena BAJO PÉREZ, de la Universidad de Salamanca.

les, o si se analiza un concepto o un campo léxico específico y su evolución en todo el corpus medieval recogido, o en un período más corto, o si se limita por ámbitos territoriales o por géneros textuales. Los objetos de estudio concretos exigirán un tratamiento u otro, pero la metodología debe incluir siempre una aproximación léxica, empleando alguno de los corpus descritos, o combinando su uso. Se puede trabajar en una fuente o en varias fuentes integradas en los corpus textuales, en función de las investigaciones concretas que se propongan. Se puede optar también por trabajar sobre un tema en todo el corpus, aunque puede resultar una tarea desmesurada. No obstante, el trabajo sobre un corpus completo puede aportar una primera visión general sobre la cuestión a analizar. Los corpus pueden ayudar a delimitar los campos léxicos o semánticos apropiados al tema a tratar y al período cronológico que se estudie⁸⁴. Los datos procedentes de los corpus, procesados de forma informática, nos pueden proporcionar listas de palabras-clave, tablas de frecuencias y, en la medida de lo posible, estadísticas que apoyen la interpretación. Pero la estadística, siempre problemática y limitada, no puede obviar otras tareas, como combinar los resultados obtenidos del manejo de los corpus con otros datos obtenidos de la lectura y análisis de otras fuentes documentales o literarias, manuscritas o impresas, que no se hallan incluidas en dichos corpus. Utilizar los corpus textuales no implica prescindir de los archivos correspondientes ni de las bibliotecas históricas. La base cuantitativa resultante debe servir de apoyo para la interpretación histórica de las problemáticas analizadas, como fundamento que avale o que invalide las hipótesis de trabajo que previamente hayan sido formuladas según la lectura de la bibliografía y de las interpretaciones vigentes sobre tales problemas. El aporte lexicométrico debe ser debidamente contextualizado e interpretado. Pero hay que tener en cuenta que ciertas investigaciones, como por ejemplo las que analizan el lenguaje, el discurso o las concepciones políticas o ideológicas, no pueden ser llevadas a buen término sin emplear algún tipo de método lexicométrico o lexicológico.

Como puede verse, la investigación resulta compleja e implica el manejo de un volumen de información textual que puede resultar inmenso, tanto si la investigación del tema en cuestión se circunscribe a un período de varios siglos, a un siglo o a un período de años. Pongamos un ejemplo concreto. Si planteáramos una investigación sobre la noción de «acuerdo» en la sociedad medieval hispánica entre los siglos XIII y XV, sólo con una primera aproximación no sistemática a los dos corpus del español (CORDE y Corpus del Español), tendríamos que manejar, de entrada, los siguientes datos. Si buscáramos en el CORDE para el período 1200-1500, los términos *acuerd*/acord**⁸⁵, obtendríamos los siguientes casos:

⁸⁴ Tarea nada fácil, pero, a juicio de Alain GUERREAU, necesaria: «Campo semántico: este es en definitiva el tema central al que todo método debe adaptarse [...] La fase central de estructuración de los campos es una operación propiamente histórica, que requiere la paciencia, la imaginación y el rigor del investigador. Cada campo plantea sus propios problemas, que remiten a estructuras sociales: las parcelas, las posiciones sociales, las virtudes, no pueden estructurarse partiendo de los mismos principios en cualquier tiempo y lugar»; Alain GUERREAU, *El futuro de un pasado*, p. 149.

⁸⁵ REAL ACADEMIA ESPAÑOLA: Banco de datos (CORDE) [en línea]. Corpus diacrónico del español <<http://www.rae.es>> [21-1-2011].

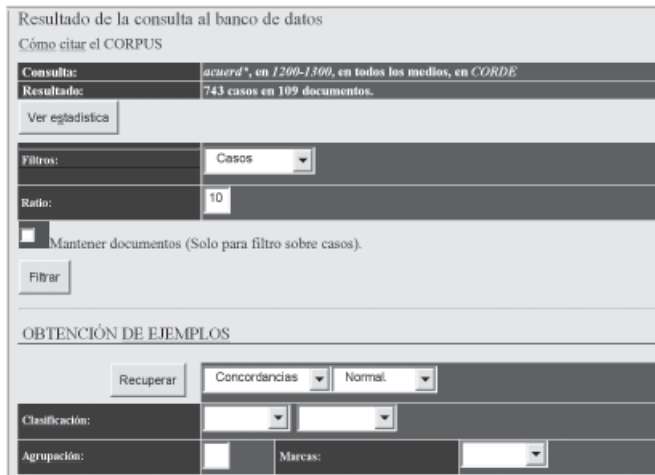


Fig. 1: RAE, BANCO DE DATOS.- CORDE: Pantalla de búsqueda

TABLA 1 RESULTADOS CORDE

acuerd*	casos	documentos	acord*	casos	documentos
1200-1300	743	109	1200-1300	880	124
1300-1400	911	171	1300-1400	1.590	219
1400-1500	2217	420	1400-1500	5.104	846

A continuación habría que revisar en la base de datos cada uno de los casos en su contexto de párrafo, sometiéndolos a una primera criba. Tendríamos también que trabajar con los documentos y con las concordancias, agrupándolas según distintos parámetros (diplomáticos, literarios, históricos, geográficos, sociales, lingüísticos...).

Concordancias (RAE)					
Consulta:		<i>acuerd*, en 1200-1300, en todos los medios, en CORDE</i>			
Resultado:		715 casos en 107 documentos.			
Pantalla: 23 de 29.					
Nº CONCORDANCIA	TÍTULO	PAÍS	TEMA	AÑO	AUTOR PUBLICACIÓN
560	e los fijosdalgo & ante los alcalles para auer su acuerdo con ellos. Et don Diago Lopes fizo'l relación ** a				
1284	Anónimo		Libro de los fueros de Castilla		ESPAÑA
			10.Ordenamientos y códigos leg Bares, Hispanic Seminary of Medieval Studies (Madison), 1993		
561	lamó a los fijosdalgo & a los alcalles et tomó su acuerdo sobre ello: et fallaron que antiguamente que ** a				
1284	Anónimo		Libro de los fueros de Castilla		ESPAÑA
			10.Ordenamientos y códigos leg Bares, Hispanic Seminary of Medieval Studies (Madison), 1993		
562	lla les ponie, por tal de la desponer, ovieron su acuerdo, e enbiaron su carta con su acusacion al obis ** a 1284				
Anónimo			Traducción de las Cantigas de Santa María		ESPAÑA 21.individual
			John E. Keller, Robert W. Linker, Real Academia Española (Madrid), 1974		
563	ere despedido delos mayordomos. [7] * de commo se acuerden de dos en dos. * Depues que el jues & los al **				
1284 - 1295	Anónimo		Fuero de Cuenca		ESPAÑA
			10.Ordenamientos y códigos leg Rafael de Ureña y Smenjaud, Academia de la Historia (Madrid), 1935		
564	mnes non puede seer vencida si fueren todos de un acuerdo. Capitulo xii de los cabdiellos e de quáles ** c				
1285	Anónimo		Libro de los cien capítulos		ESPAÑA
			14.Enciclopedias y compendios Marta Haro Cortés, Vervuert-Iberoamericana (Frankfurt am Main-Madrid), 1998		
565	ores non pueden vencer, e los que son todos de un acuerdo aquellos vencen. Las grandes voces a la ora d ** c				
1285	Anónimo		Libro de los cien capítulos		ESPAÑA
			14.Enciclopedias y compendios Marta Haro Cortés, Vervuert-Iberoamericana (Frankfurt am Main-Madrid), 1998		

Fig. 2: RAE, BANCO DE DATOS.- CORDE: : Pantalla de concordancias

Documentos (RAE)							
Consulta:				<i>acuerd*</i> , en 1200-1300, en todos los medios, en CORDE			
Resultado:				743 casos en 109 documentos.			
Cómo citar el CORPUS				Documentos.			
Pantalla: 1 de 3.		Siguiete		1 2 3			
Casos	Año	Autor	Obra	País	Tema	Publicación	
<input type="checkbox"/>	1	1239	Anónimo	Confirmación de Fernando III de unos acuerdos [Documentos del Archivo Histórico Nacional (al200-a149 ...	ESPAÑA	10.Documentos notariales	Pedro Sánchez-Prieto, Universidad de Alcalá (Madrid), 1999
<input type="checkbox"/>	1	1272	Anónimo	Avenencia entre los concejos de Hayuela y Santurdejo [Documentos del Archivo Histórico Nacional (al2 ...	ESPAÑA	10.Documentos notariales	Pedro Sánchez-Prieto, Universidad de Alcalá (Madrid), 1999
<input type="checkbox"/>	1	1275	Anónimo	Privilegio del infante don Fernando [Documentos del Archivo Histórico Nacional (al200-a1492)]	ESPAÑA	10.Documentos notariales	Pedro Sánchez-Prieto, Universidad de Alcalá (Madrid), 1999
<input type="checkbox"/>	1	a 1296	Anónimo	Fuero de Alcaraz	ESPAÑA	10.Ordenamientos y códigos legales	Jean Roudil, Librairie C. Klincksieck (París), 1968

Fig. 3: RAE, BANCO DE DATOS.- CORDE: Pantalla de documentos [fragmento]

Resultados con estadísticas (RAE)									
Consulta:		acuerd ^a , en 1200-1300, en todos los medios, en CORDE							
Resultado:		743 casos en 109 documentos.							
Cómo citar el CORPUS									
Estadísticas									
Año	%	Casos	País	%	Casos	Tema	%	Casos	
1293	24.85	85	ESPAÑA	100.00	743	19.- Prosa histórica	47.91	356	
1270	23.09	79				10.- Prosa jurídica	23.95	178	
1300	11.11	38				12.- Prosa narrativa	16.55	123	
1251	9.94	34				15.- Prosa científica	4.17	31	
1256	8.77	30				14.- Prosa didáctica	3.90	29	
1254	6.14	21				22.- Verso narrativo	2.69	20	
1257	1.75	6				17.- Prosa religiosa	0.53	4	
1277	1.75	6				16.- Prosa de sociedad	0.13	1	
1290	1.75	6				21.- Verso lírico	0.13	1	
Otros	10.81	37							

Fig. 4 : RAE, BANCO DE DATOS.- CORDE: Pantalla de estadísticas

Pero, lógicamente, las formas de llegar a un acuerdo no se ciñen a aquello que expresamente se denomina «acuerdo». Habría que determinar el campo semántico, comenzando, por ejemplo, por trazar todos los sinónimos que en la época medieval equivalen a acuerdo. Para ello no debemos acudir a un diccionario moderno de sinónimos, sino recurrir al Corpus del Español, para ver los sinónimos que efectivamente se usaban en la época para referirse al hecho de llegar a un acuerdo. Una primera búsqueda de sinónimos de acuerdo nos daría el siguiente resultado distribuido por siglos y número de casos (sólo como sustantivo, sin considerar los sinónimos del verbo acordar):

Corpus del español: sinónimos de *acuerdo*⁸⁶, ordenados según frecuencia por número de casos y siglos:

⁸⁶ DAVIES, Mark. (2002-) Corpus del español (100 millones de palabras, siglo XIII - siglo XX). Disponible en <http://www.corpusdelespanol.org>. [revisado a 9-6-2011].

TABLA 2 RESULTADOS CORPUS DEL ESPAÑOL

6.715.712 palabras [siglo XIII]		2.669.561 palabras [siglo XIV]		8.161.410 palabras [siglo XV]	
PAZ	1661	PAZ	541	PAZ	1254
ENTENDIMIENTO	597	ENTENDIMIENTO	462	ENTENDIMIENTO	881
ACUERDO	505	ACUERDO	248	ACUERDO	303
ENTENDIMIENTOS	37	TRATO	26	TRATO	204
CONCIERTO	17	ENTENDIMIENTOS	21	TRATOS	102
ACTA	11	FACTO	18	CONCIERTO	64
UNIDADES	7	TRATOS	13	ENTENDIMIENTOS	58
CONCIERTOS	7	ACUERDOS	3	CONTRATO	53
ACUERDOS	6	ACTA	3	FACTO	51
FACTO	6	AMÉN	1	CONTRATOS	28
UNIDAD	5	UNIÓN	0	CONFORMIDAD	26
AVENENCIA	5	UNIDADES	0	FACTOS	26
AMÉN	4	UNIDAD	0	UNIDAD	11
FACTOS	3	RESOLUCIONES	0	CONCIERTOS	11
CONTRATOS	3	FACTOS	0	ACTA	11
CONTRATO	2	FEDERACIONES	0	RESOLUCIONES	7
UNIÓN	1	CONTRATOS	0	SOLUCIONES	5
ARREGLO	1	CONTRATO	0	NEGOCIACIÓN	5
FEDERACIONES	1	CONCIERTOS	0	APAAÑO	5
RESOLUCIONES	1	CONCIERTO	0	ACUERDOS	5
TRATO	1	AVENENCIA	0	PACES	4
TRATOS	0	ARREGLO	0	AVENENCIA	4
				AMÉN	3
				INTELIGENCIA	2
				NEGOCIACIONES	2
				UNIDADES	1
				PLEBISCITOS	1
				PLEBISCITO	1
				COMPROMISO	1
				ARMONÍAS	1
				ARMONÍA	1
				ACTAS	1

De nuevo, tendríamos que ver cada uno de estos casos en sus contextos de párrafo, en las fuentes correspondientes recogidas en el Corpus del Español (y someterlos a criba). A continuación habría también que realizar las búsquedas de esos mismos términos y sus variantes morfológicas en el CORDE (puesto que incluye muchos más documentos que el Corpus del Español) y analizar los resultados en sus contextos de párrafo y realizar una nueva clasificación. Todavía puede precisarse más el campo semántico del acuerdo, buscando más sinónimos de cada una de estas palabras, o de las que presentan una mayor frecuencia (consultando de nuevo el Corpus del Español y elaborando una nueva lista). Asimismo, habría que determinar las palabras colocadas en la proximidad de estos términos, en los enunciados en los que aparecen, de ese modo se obtiene otro listado de palabras relacionadas, términos afines que también deben ser estudiados en su contexto de párrafo en las obras correspondientes. También es posible realizar esa operación con el Corpus del Español (y generar listados), pero omitimos el resultado, por ser demasiado extenso.

Este bloque de datos léxicos nos proporciona una primera red de clasificaciones léxicas. Si nuestra investigación abarcara el conjunto hispánico, habría que repetir estas operaciones, en la medida de lo posible, en los corpus textuales disponibles de los distintos ámbitos idiomáticos. Sin olvidar que también habría que considerar el dominio de la lengua latina.

Como puede verse, los corpus textuales informatizados facilitan la labor de identificación de las fuentes y la elaboración de clasificaciones, pero no ahorran la prác-

tica de una labor concienzuda y metódica, descendiendo cada vez más en los niveles de significado, abriendo nuevas relaciones, hasta poder llegar a reconstruir una visión general de la realidad social o política que analicemos. Cuando obtengamos esa visión estructurada, estaremos en disposición de responder mejor a las preguntas que nos formulemos.

En los próximos años esperamos poder ofrecer algunas muestras de aplicación de estos corpus al ámbito de la historia. Si planteamos esta investigación sobre los corpus filológicos es porque desde la historia aún no existen las aplicaciones informáticas apropiadas que faciliten el análisis de la inmensa masa documental ya publicada. Los corpus realizados por lingüistas y filólogos resultan muy útiles pero son todavía insuficientes. Sería deseable poder contar con corpus textuales que incluyeran todas las fuentes medievales publicadas, diseñados, a ser posible, con criterios de análisis histórico. El proceso ya abierto por la filología, y el desarrollo tecnológico aplicado a las humanidades es un buen punto de partida. Pero esa compleja y necesaria labor, para que diera frutos, solo podría ser abordada colectivamente, con la participación de toda la comunidad científica de medievalistas hispánicos, y con el apoyo de las respectivas instituciones de investigación⁸⁷.

⁸⁷ Me remito a las sugerencias de FORTÚN, Luis Javier: «La edición de fuentes documentales, *art. cit.*», p. 40.