

Portal de léxico hispánico: una herramienta para el estudio del léxico

Gloria Clavería
Marta Prat Sabater
Joan Torruella
Cristina Buenafuentes de la Mata
Margarita Freixas
Carolina Julià Luna
Mar Massanell
Laura Muñoz
Sònia Varela
Universitat Autònoma de Barcelona

The Portal de léxico hispánico (Hispanic vocabulary portal) is the result of the research project "Banco de datos diacrónico e hispánico: morfología léxica, sintaxis, etimología y documentación", carried out in the last few years thanks to funding by the Ministerio de Educación y Ciencia (Ref N°: HUM2005-082149-C02-01) and the Generalitat de Catalunya (Ref N°: SGR2005-00568). This website brings together scientific information about the vocabulary of the Ibero-Romance languages and their dialectical variations. The objective of this site is two fold: on the one hand, it aims to provide a tool for all Internet users to obtain both diachronic and synchronic information about Hispanic vocabulary, and on the other hand, it aims to provide a useful tool for the work on the Nuevo Diccionario Histórico of the Real Academia Española. The Portal de léxico hispánico fundamentally brings together information of bibliographic, lexicologic and documentary nature relating them to a wide range of data bases. It contains data from the digital version of the Diccionario crítico etimológico castellano e hispánico by J. Coromines and J. A. Pascual (Madrid: Gredos, 1980-1991) together with data from other sources. The consultation interface of the portal allows bibliographic, diachronic, diatopic, etymologic, graphic, phonetic-phonological, morphosyntactic and semantic information to be searched for specific words, as well as their documentation in ancient and modern texts. In the communication, detailed information will be available about the origin, aims, characteristics and present state of the Portal de léxico hispánico. There will also be a demonstration of the consultation interface available on the Internet.

1. Presentación

El *Portal de léxico hispánico* es una plataforma de Internet de acceso abierto que pretende constituirse en un centro de información sobre el léxico hispánico. Se encuentra en estos momentos en proceso de elaboración en el Seminario de Filología e Informática del Departamento de Filología Española de la Universitat Autònoma de Barcelona y ha sido desarrollado gracias a la financiación obtenida en distintos proyectos de investigación¹.

El objetivo básico del *Portal de léxico hispánico* consiste en reunir, mediante herramientas informáticas, información científica sobre el léxico de las lenguas románicas de la Península Ibérica (español, catalán, gallego y portugués) y sus respectivas variedades; se recogen, además, datos léxicos de carácter tanto sincrónico como diacrónico. Tiene actualmente una doble finalidad: (a) proporcionar a cualquier usuario de la Red una herramienta para la obtención de información sobre el léxico hispánico y (b) contribuir, tanto con la plataforma

¹ Banco de datos diacrónico e hispánico, proyecto concedido por el Ministerio de Educación y Ciencia: (n° de referencia HUM2005-082149-C02-01) y Ajuts per potenciar i donar suport als grups de recerca de la Generalitat de Catalunya (n° de referencia SGR2005-00568).

como con el material que en ella se alberga, a la elaboración del *Nuevo diccionario histórico* de la Real Academia Española.

Por su estructura, el *Portal* consta de dos interfaces: la de introducción de datos y la de consulta. La interfaz de introducción de datos está siendo utilizada en estos momentos por dos equipos de investigación distintos que trabajan de forma coordinada; uno es el equipo de investigación del Seminario de Filología e Infomática de la Universitat Autònoma de Barcelona y el otro es el equipo de Bibliografía de la Real Academia Española. Esta interfaz contiene tres bases de datos relacionadas: *Bibliografía*, *Léxico* y *Documentaciones*; cada una de ellas recoge información de carácter diferente según se expone a continuación (epígrafe 2). La interfaz de consulta se encuentra disponible en acceso abierto en la dirección de Internet <http://portaldelexico.es>, y, desde ella y como se describirá más adelante (epígrafe 3), se pueden realizar búsquedas tanto simples como complejas sobre los datos almacenados.

2. Introducción de datos

2.1. Bibliografía

El apartado de *Bibliografía* del *Portal de léxico hispánico* pretende reunir estudios y monografías que tienen como objetivo fundamental el estudio científico del léxico del español. En él, se han incluido todos los trabajos de Germà Colón, buena parte de ellos recogidos en el *Índex del lèxic estudiat en l'obra de Germà Colón*². Se están introduciendo referencias bibliográficas de lingüistas y filólogos que han dedicado gran parte de sus estudios al léxico — como Malkiel (1988-1989) o Morreale (2006)—; compendios bibliográficos — como el de Dworkin y Gago-Jover (2004)— y una selección de estudios en obras de carácter colectivo (por ejemplo, las actas de los diferentes congresos de historia de la lengua española y de lingüística y filología románicas; obras de carácter misceláneo y libros de homenaje a lingüistas y filólogos).

El criterio principal que se ha adoptado para la inclusión de un trabajo o artículo en la base de datos ha sido el siguiente: el estudio debe contener información léxica o morfo-léxica esencialmente de carácter diacrónico. Además, se han introducido también aquellos trabajos que, sin ser exclusivamente investigaciones de carácter léxico, aportan información sobre las características evolutivas de palabras concretas.

2.2. Léxico

La base de datos de *Léxico* recoge las palabras estudiadas en diferentes fuentes:

Los datos procedentes del DECH, con más de 65.000 palabras (lemas, derivados y compuestos, y voces citadas en el interior del apartado crítico del diccionario).

Los datos procedentes del *Índex del lèxic estudiat en l'obra de Germà Colón*. Se han incluido en el *Portal* las formas pertenecientes a las siguientes variedades lingüísticas: español (11016); catalán (20251); gallego (23); portugués (1149); aragonés (639); asturiano (19); leonés (2) e iberorománico (3). Se trata, por tanto, de 33102 registros.

Las voces analizadas en los trabajos que progresivamente van incluyéndose en la base de datos de *Bibliografía* (véase 2.1.).

La información que se reúne en los diferentes registros es de dos tipos: de contenido lingüístico-filológico y de localización. Para el primero de ellos, se recogen los datos de las palabras que se analizan en las diferentes fuentes vaciadas: *Lema* (representado formalmente según los criterios de lematización del DRAE), *Forma* (antigua o moderna; de una lengua concreta o de alguna de sus variedades), *Categoría gramatical*, *Significado* (si se proporciona en la fuente consultada), *Lengua* a la que pertenece (español, catalán, gallego, portugués y sus respectivas variedades) e *Información* (diacrónica; diastrática; diatópica; documental; etimológica; gráfica; fonético-fonológica; morfológica; sintáctica; y semántica). Para el segundo tipo de

² Para conocer el contenido exacto de este índice y los objetivos por los que se ha confeccionado, vid. Clavería y Buenafuentes (2005), y Clavería y Prat (en prensa).

información, la localización, se recogen las referencias bibliográficas con indicación de la página o las páginas correspondientes en las que se encuentra la forma léxica.

2.3. Documentaciones

El apartado *Documentaciones* reúne la información documental de los términos incluidos en el *Portal de Léxico Hispánico*. Esta base de datos contiene las documentaciones procedentes del DECH. Se está introduciendo, además, la información que proviene de la revisión de todas las documentaciones en las que esta obra remite a una edición del *Diccionario* de la Academia, ya que las dataciones, en estos casos, son un tanto inexactas (Clavería 1999, y Clavería-Morales 1999-2000). También, para completar la documentación de palabras que en el DECH no tienen este tipo de información, se indica su primera aparición en los diccionarios académicos.

En la base de datos se registran los *lemas* compilados en la sección de *Léxico* (según la forma del DRAE), la forma gráfica de la variante documentada (campo *Forma*) y, en el caso de que se encuentre en el interior de una unidad pluriverbal, el sintagma o locución donde aparece (*Forma compleja*). Se identifica la fecha de composición de los textos en los que se documenta cada forma léxica (*Fecha*) y la fecha de la copia manuscrita o del impreso que la contiene (*Fecha secundaria*). También se indica el *Autor* del texto (apellidos, nombre y seudónimo) y el *Título* con el que suele conocerse la obra, según la forma regularizada de la versión electrónica del DECH. Se incluye además un apartado con los *Datos bibliográficos* que se conocen de los textos consultados: el título que aparece en el documento, seguido, en el caso de los impresos, del número de edición, lugar de publicación, impresor, imprenta o editorial y, para los manuscritos, del lugar de la copia y del nombre y procedencia del copista. Finalmente, en la *Localización* se identifica el volumen, parte, capítulo, página o verso del texto en el que se ha documentado la forma léxica y en el campo *Texto* se transcribe el fragmento que la contiene. En el caso de que la documentación se haya obtenido a través de la consulta de una fuente secundaria (como ocurre en las documentaciones extraídas del DECH), se ofrecen los datos de ésta en el apartado *Referencia indirecta*.

3. Consulta de datos

La interfaz de consulta del *Portal de léxico hispánico* permite recuperar informaciones de los tres bancos de datos que actualmente configuran el *Portal*: *Banco de léxico*, *Banco de lexicografía* y *Banco de documentaciones* (Figura 1). En la actualidad, es posible acceder a la consulta y realizar búsquedas en los tres bancos de datos, a partir de cuya relación es posible pasar de uno a otro.

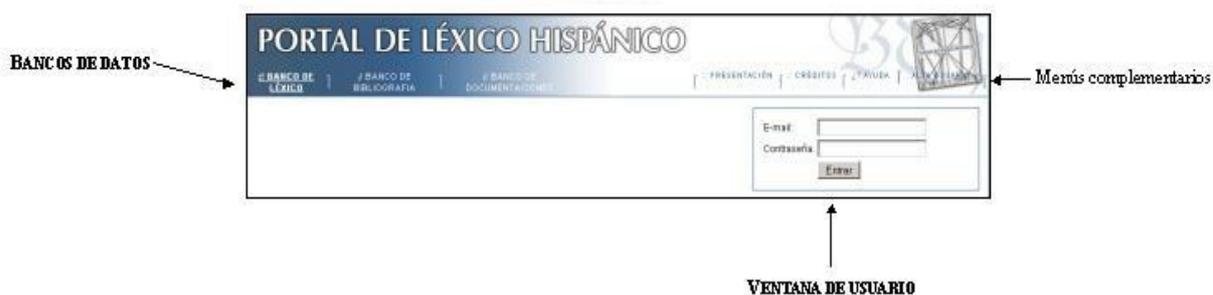


Figura 1

3.1. Banco de léxico

La interfaz de consulta del *Banco de léxico* (Figura 2) se ha diseñado siguiendo los modelos de otros programas elaborados en el *Seminario de Filología e Informática* de la Universitat

Autònoma de Barcelona³, de manera que el *Portal* se organiza en dos partes: a la izquierda, el *Panel de Consulta* y la *Ventana de índice de resultados* y, a la derecha, la *Ventana de resultados*.



Figura 2

En el *Banco de léxico*, las búsquedas pueden ser de dos tipos: simples o compuestas.

3.1.1. Búsquedas simples

Las búsquedas simples se realizan a partir de los parámetros *Palabra* (en cualquiera de sus posibilidades: *Lema*, *Forma* o *Forma compleja*), *Lengua*, *Información* y *Bibliografía* (en cualquiera de sus posibilidades *Artículo/Capítulo*, *Autor* u *Obra*).

La búsqueda por lemas se puede llevar a cabo escribiendo el lema en la *Caja de consulta* o bien seleccionándolo de la lista de lemas que aparece en la *Ventana de índice de resultados* según se active alguna de las letras de la parte inferior de la *Caja de consulta*. Los lemas aparecen clasificados por letras y ordenados alfabéticamente. Cuando se escribe el lema en la *Caja de consulta*, en la *Ventana de índice de resultados* van apareciendo los lemas de modo incrementado. Es posible usar los comodines (# para un solo carácter y \$ para diversos caracteres) para obtener la lista de lemas que “empiezan por” o “terminan por” un carácter o secuencia determinados, o que en su interior poseen una determinada combinación de caracteres.

Las búsquedas por *Forma* y por *Forma compleja* funcionan del mismo modo que la interrogación por *Lema*, pero no dan como resultado una serie de lemas, sino una lista de formas o formas complejas.

Las búsquedas por *Lengua* e *Información*, así como las búsquedas por *Artículo/Capítulo*, *Autor* y *Obra*, se realizan a partir de una ventana dividida en dos partes (Figura 3): a la izquierda aparece el listado de todos los elementos que pueden seleccionarse y a la derecha van apareciendo uno tras otro los elementos seleccionados. De este modo, se pueden buscar, por ejemplo, todas las referencias bibliográficas en las que se estudian palabras pertenecientes a una variedad determinada (gallego y portugués, por ejemplo), o bien todas las palabras de las

³ Vid. Clavería, G.; Muñoz, L.; Prat, M.; Torruella, J. (2005).

que la referencia bibliográfica aporta un tipo de información (morfológica o semántica, por ejemplo), o bien todas las voces estudiadas por un autor o que se hallan en una obra concreta, etc.

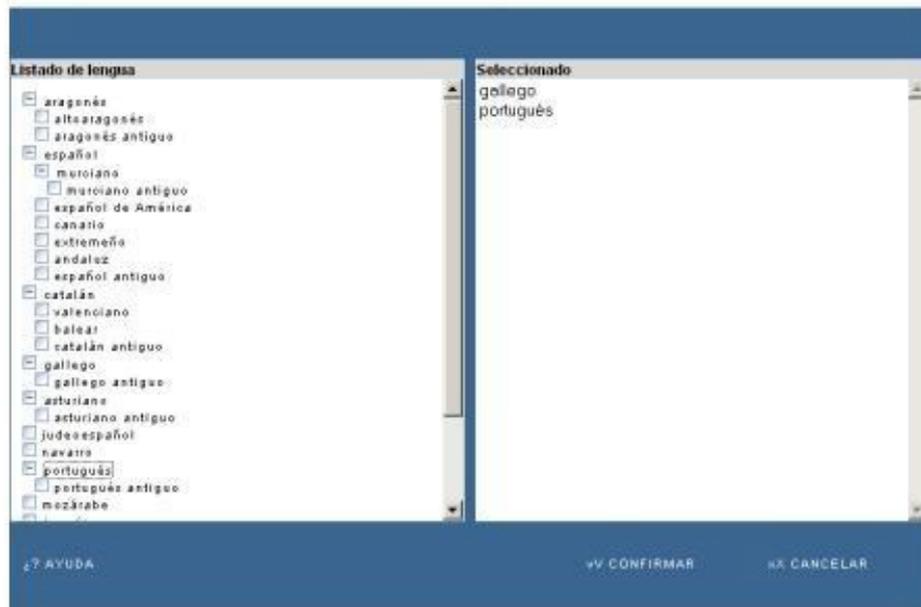


Figura 3

3.1.2. Búsquedas compuestas

Las búsquedas compuestas (*Búsqueda múltiple*) permiten cruzar las informaciones de los distintos parámetros de las búsquedas simples. Por ejemplo, se pueden obtener todas las palabras del aragonés que tienen información sintáctica o todas las obras de un autor determinado que, desde el punto de vista fonético-fonológico, tratan de una palabra o forma compleja concreta, propia de una variedad lingüística como el leonés.

3.1.3. Visualización de los resultados

La disposición de los resultados en la pantalla es siempre la misma: a la izquierda, la *Ventana de índice de resultados*, con la lista de palabras que cumplen las condiciones especificadas en la búsqueda, y, a la derecha, la *Ventana de resultados*, con la ficha o fichas bibliográficas de la palabra seleccionada por el usuario en el *Índice de resultados*.

Las referencias bibliográficas de un mismo lema se presentan agrupadas en bloques, especificados en la barra superior de la *Ventana de resultados* (Figura 4).

Cuando se selecciona un bloque, en la parte superior de la *Ventana de resultados* figuran el *Lema*, la *Categoría* y la *Lengua* común a todas las fichas del bloque y, debajo, aparecen las diferentes fichas léxico-bibliográficas en las que se indica el *Título*, el *Autor* y las *Páginas* de la referencia bibliográfica (o *Sub voce* si se trata del DECH), y los campos *Forma*, *Forma compleja*, *Significado* e *Información*. Al hacer clic encima de cada título, se abre una ventana superpuesta con la ficha bibliográfica completa. Dicha ficha puede imprimirse, como se observa en la Figura 4:

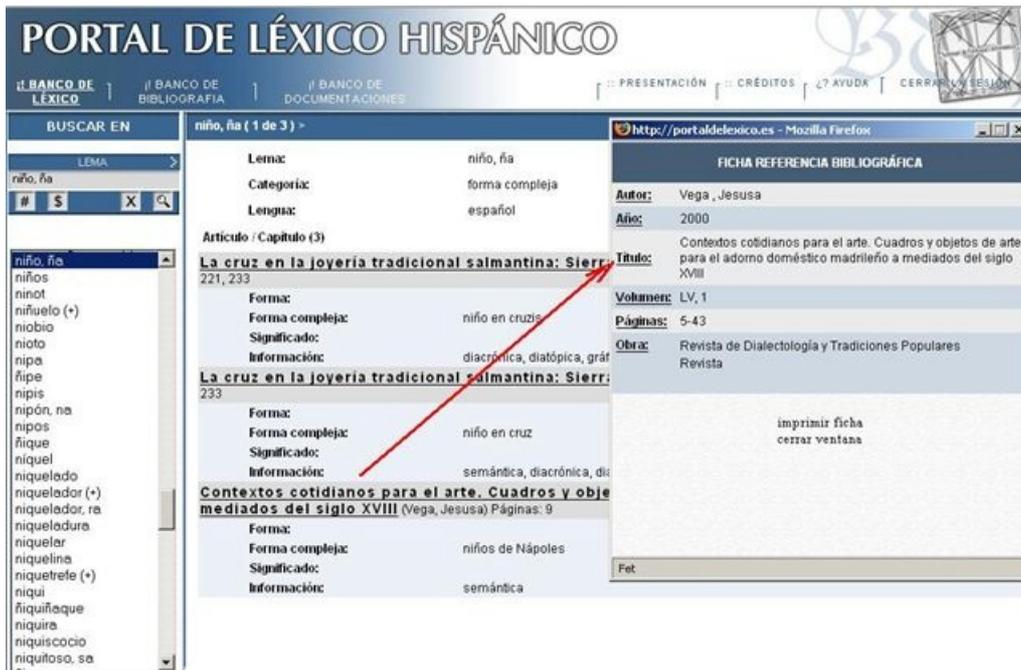


Figura 4

3.2. Banco de bibliografía

El *Banco de bibliografía* (Figura 5) permite acceder a la lista de todas las referencias bibliográficas que se han incluido en el *Portal*. Su interrogación se realiza a través de un *Panel de consulta* que ofrece las opciones *Artículo/Capítulo*, *Autor* y *Obra*. Se puede seleccionar cualquiera de estas opciones con el fin de recuperar la referencia bibliográfica completa, y, si se desea, también se puede realizar un salto hipertextual al *Banco de léxico* para obtener la información sobre las palabras que se estudian en cada una de ellas.



Figura 5

Junto al *Panel de consulta*, aparece la información actualizada del número de *Artículos/Capítulos*, *Autores* y *Obras*, que se hallan en el *Banco de bibliografía* hasta la fecha de la consulta. Tras seleccionar alguno de estos apartados, se obtiene la lista completa de sus fichas.

En la ventana de resultados, al final de cada una de las fichas aparece el título de la obra a la que pertenece. Haciendo clic sobre él, se abre una ventana en la que se ofrecen todos los datos bibliográficos y el índice de todos los artículos de la obra (Figura 6).

PORTAL DE LÉXICO HISPÁNICO

/ BANCO DE LÉXICO | / BANCO DE BIBLIOGRAFÍA | / BANCO DE DOCUMENTACIONES | PRESENTACIÓN | CRÉDITOS | AYUDA | CERRAR

Artículo / Capítulo: 2304 Autores: 1096 Obras: 729

Artículo / Capítulo:
 Autor:
 Obra:

Tes philies tade dora: Miscelánea léxica en memoria de Conchita Serrano
 Formato: Libro

Artículo / Capítulo

Léxico agropecuario de la Villa y Tierra de Madrid (1464-1481) según los Libros de acuerdos del Concejo de la Villa
 Fernández Montes, Melide
 Año: 1999 Págs. 317-328

Los nombres "hábaros" del guesiticismo
 Barcala Muñoz, Andrés
 Año: 1999 Págs. 11-34

Lexicología: contribución al léxico de las instituciones
 Boned Colera, Pilar
 Año: 1999 Págs. 55-60

Sobre el significado léxico de *úsopha* y derivados
 Crespo Oúenes, Emilio
 Año: 1999 Págs. 61-66

Cuestiones de léxico retórico en el tratado de Juan de Guzmán
 Albuquerque García, Luis
 Año: 1999 Págs. 297-304

Los latinismos en el Arte Cisoria de Enrique de Villena
 Calero, Francisco
 Año: 1999 Págs. 305-308

Los nombres de la lechuga. Herencia y superstición
 García Mouton, Pilar
 Año: 1999 Págs. 329-330

Etimología y equivalencia del término "retórica" en los tratados de elocuencia del siglo XVI español
 Luján Alienza, Angel L.
 Año: 1999 Págs. 345-352

Notas de lexicografía diacrónica sobre el término *esperpento*
 Menéndez Orrubia, Carmen
 Año: 1999 Págs. 361-360

Sobre la etimología de los teónimos: el caso del dios galaico-lusitano Bandi : Bandus
 Pedrero Sancho, Rosa María
 Año: 1999 Págs. 417-426

Figura 6

Desde cualquiera de los títulos de los artículos o capítulos de cada una de las referencias almacenadas en el *Banco de léxico*, se puede obtener la lista de palabras estudiadas en su interior.

3.3. Documentaciones

En el apartado *Documentaciones*, como en el *Banco de Léxico*, el usuario puede escoger si quiere consultar el DECH, otras fuentes bibliográficas o todas al mismo tiempo. Las consultas se pueden realizar a partir de tres parámetros principales: *Palabra*, *Lengua* y *Documentaciones*. En los dos primeros, las opciones de búsqueda son las mismas que las que se ofrecen en el *Banco de Léxico*. En *Documentaciones*, se encuentran las opciones *Fecha*, *Fecha secundaria*, *Autor*, *Título* y *Referencia indirecta*. Está prevista la incorporación de búsquedas por *Datos bibliográficos*, que permitirá recuperar las documentaciones en los distintos testimonios manuscritos e impresos de una obra, y por *Localización*, con el fin de reunir las documentaciones concentradas en un fragmento determinado de una obra.

En este apartado también se encuentra la opción de *Búsqueda múltiple* para hacer consultas que crucen informaciones de los tres parámetros principales. Los resultados de las búsquedas se presentan en el *Panel de índice de resultados* y, haciendo clic en una palabra de las encontradas, se muestran todas las documentaciones en la *Ventana de resultados*. Éstas aparecen en forma de tabla con los campos *Fecha*, *Autor*, *Título*, *Documento* y *Referencia indirecta*, y se pueden ordenar por cualquiera de ellos. Por defecto, las documentaciones están ordenadas por el campo fecha, es decir, la primera documentación de la palabra aparece en primer lugar y las siguientes en orden cronológico.

A partir de cada una de las documentaciones, con un simple clic en el botón correspondiente, se puede visualizar, en una ventana sobrepuesta, el texto en el que aparece la documentación (en los casos en que éste se haya incluido) o la ficha entera de la documentación (*Lema, Forma, Forma compleja, Autor, Título, Documento y Referencia indirecta*).

4. Conclusión

El *Portal de léxico hispánico* pretende ser una fuente de información bibliográfica y lexicológica accesible a través de Internet para todos aquellos lingüistas, investigadores y filólogos que deseen obtener información de variado tipo sobre el léxico del español, del catalán, del gallego y del portugués, y de sus respectivas variedades. La cantidad y calidad de las informaciones interrelacionadas que recoge el *Portal* permiten presentarlo como una eficaz herramienta de trabajo que irá creciendo progresivamente hasta convertirse en un gran banco de datos que se constituirá en un valioso apoyo para el estudio de las lenguas románicas de la Península Ibérica.

Referencias bibliográficas

- Clavería, G. (1999). "La documentación en el diccionario etimológico". En Bleuca, J. M.; Clavería, G.; Sánchez, C.; Torruella, J. (eds.). *Filología e informática. Nuevas tecnologías en los estudios filológicos*. Lleida: Editorial Milenio-Seminario de Filología e Informática del Departamento de Filología Española de la Universitat Autònoma de Barcelona. 259-280.
- Clavería, G.; Buenafuentes, C. (eds.). (2005). *Germà Colon: les llengües romàniques juntes i contrastades*. Bellaterra: Servei de Publicacions de la Universitat Autònoma de Barcelona.
- Clavería, G.; Morales, C. (1999-2000). "El DCECH y los diccionarios de la Academia". *Revista de Lexicografia* 6. 37-55.
- Clavería, G.; Muñoz, L.; Prat, M.; Torruella, J. (2005). "Índex del lèxic estudiat en l'obra de Germà Colon" [cd-rom]. En Clavería, G.; Buenafuentes, C. (eds.). *Germà Colon: les llengües romàniques juntes i contrastades*. Bellaterra: Servei de Publicacions de la Universitat Autònoma de Barcelona.
- Clavería, G.; Prat, M. (en prensa). "Concordances lèxiques de l'obra de Germà Colon". *Quaderns de Filologia*.
- [DECH]. Coromines, J.; Pascual, J. A. (1980-1991). *Diccionario crítico etimológico castellano e hispánico*. Madrid: Gredos.
- [DRAE]. Real Academia Española. (2001). *Diccionario de la lengua española*. 22ª ed. Madrid: Espasa Calpe.
- Dworkin, S.; Gago-Jover, F. (2004). *Lexical Studies of Medieval Spanish Texts* [en línea]. <http://college.holycross.edu/lacoronica/lsmst/> [Fecha de consulta: 26 de marzo de 2008].
- Malkiel, Y. (1988-1989). *A Tentative Autobiography. Romance Philology (Special Issue)*. University of California and Los Angeles: University of California Press.
- Morreale, M. (2006). *Estudios escogidos de lengua y literatura española*. Madrid: Gredos.
- [NTLE]. Real Academia Española (2000). *Nuevo Tesoro Lexicográfico de la Lengua Española*. [en línea]. Madrid: Espasa. <http://www.rae.es>.