

Construir un diccionario de derivación del español en el siglo XXI. La arquitectura de la información al servicio de la lexicografía

María Teresa Díaz García
Inmaculada Mas Álvarez
Universidade de Santiago de Compostela

In the age of communication, new technological advances are made everyday in the field of information and linguistics. Deriv@ is a system that makes the most of linguistic data. All data managed by this application are connected with derivational morphology, since derived words constitute the starting point from which relationships with other linguistic elements are established. Deriv@ is a linguistic database with two models of representation: one for all Spanish words created by means of derivation; the other for the corresponding Latin words. We offer a grammatically-analysed corpus, both synchronically and diachronically, to make customised queries according to the user's interests, and also to create a dictionary of derived words. Deriv@ allows and facilitates non-restricted access to all the information available in the two databases. This model is valid for any form derived from Spanish or from other Romance languages. It offers users the possibility of using it real time, thus letting them interact with the system and contribute to its improvement.

1. Introducción

La vinculación entre tecnología y economía en el tiempo empleado suele presidir la planificación de nuestras actividades cotidianas más elementales. Esa conexión resulta determinante en otros ámbitos de mayor relevancia o de mayor especificidad, como el diagnóstico precoz de enfermedades, la formación idónea de individuos, la actividad investigadora en general y la investigación lingüística en particular.

Relacionada con la investigación lingüística está nuestra experiencia como usuarios de diversas fuentes de información: diccionarios en papel, diccionarios electrónicos, *corpora*, etc. Dicha experiencia revela que la tarea de recopilación de datos en los diccionarios de formato tradicional puede no solo resultar monótona, sino también exigir una gran inversión de tiempo. Sin duda, los diccionarios con formato electrónico suponen una agilización de las búsquedas, pues presentan accesos por diferentes puntos, posibilidad de efectuar saltos hipertextuales, etc.; sin embargo, la información facilitada por dicha vía ni puede ser completa —en el sentido de que un mismo diccionario no suele ofrecer información léxica, etimológica, morfológica y sintáctica acerca de la palabra consultada—, ni coincide necesariamente con la deseada por el usuario —en el sentido de que no es él quien establece los criterios de relevancia. Supongamos, a modo de ejemplo, que nos propusiéramos estudiar los procesos de interfijación en español y, en consecuencia, necesitásemos recopilar exhaustivamente las palabras con interfijos; el único modo de llevar a cabo esta tarea en diccionarios de cualquier formato (electrónico o de papel) consistiría en revisar sus entradas una a una, desde la A hasta la Z, lo cual convertiría el proyecto en cuestión en una tarea penosa y lenta.

Todo ello invita a redefinir el diccionario y cualquier obra de información léxica, no solo desde el punto de vista editorial, sino también desde la perspectiva de los usuarios, sobre todo en lo que se refiere al formato electrónico. En esta dirección, es necesario sustituir la idea de los *corpora* y los diccionarios organizados como documentos de contenido fijo, por la idea de la exploración de funciones y de la subsiguiente organización de un producto en permanente proceso de construcción y actualización. Nos referimos, naturalmente, a la denominada “arquitectura de la información”.

Sin que ello suponga una desatención de los medios de información tradicionales, como los diccionarios y otros materiales lexicográficos ya publicados, es indudable que la *lexicografía de corpus* está hoy plenamente constituida como disciplina independiente con las características que se subrayan en la siguiente cita (Pérez Hernández, Moreno Ortiz y Faber Benítez 1999: 206):

La lexicografía de corpus (siguiendo los postulados de la lingüística de corpus) enfatiza la necesidad de utilizar evidencia lingüística proveniente de textos reales, en lugar de confiar únicamente en las fuentes tradicionales de conocimiento léxico: diccionarios y materiales existentes y el conocimiento e intuición del lingüista o lexicógrafo. [...] La lexicografía de corpus, por tanto, se basa, en mayor o en menor medida según el enfoque adoptado, en emplear recursos textuales para obtener la información léxica que incluirá en sus diccionarios y lexicones, haciéndolos mejores y más completos.

2. El tratamiento de la derivación en la lexicografía tradicional

El trabajo que presentamos se mueve precisamente en esta última dirección. Se trata de un primer paso y de la mínima parte de un proyecto de mayor envergadura que puede resultar de gran utilidad en diferentes áreas. Las disciplinas con las que presenta conexiones son múltiples —morfología derivativa, lexicología, etimología, etc.—, por lo que puede resultar útil tanto a fonólogos como a morfólogos, lexicólogos, investigadores de sintaxis e incluso a estudiantes y estudiosos de la lengua española.

En lo que se refiere al tratamiento lexicográfico de las palabras derivadas, el punto de vista tradicional supone que el diccionario no tiene que reflejar los resultados de un mecanismo regular, similar por tanto de palabra a palabra, pues ya sólo desde el punto de vista práctico sería absurdo pretender registrar todas las formaciones posibles. Hay acuerdo, como cabría esperar, en considerar excluidos los eductos de la derivación apreciativa. Esta situación se extiende en parte a la derivación heterogénea. Hay sin embargo dos propiedades distintivas de los procesos heterogéneos —frente a los procesos flexivos y a la derivación homogénea— que entran en juego: la irregularidad y la productividad. En realidad, ambos aspectos están conectados, puesto que cuanto más regular sea un proceso derivativo, más disponible estará para nuevas formaciones y mayor es la posibilidad de que los derivados que resulten de él no queden registrados en el diccionario. Con este criterio se cruza el normativo, que obliga a registrar palabras regulares en la forma pero anómalas en cuanto al significado.

La postura de lexicógrafos y morfólogos no ha sido uniforme en este punto. Para unos, lo fundamental es que el usuario del diccionario encuentre lo que busca en un solo intento, por lo que lo recomendable es incluir en la macroestructura todas las palabras derivadas (Sinclair 1987: 110). Para otros, sin embargo, ya que las palabras derivadas tienen un significado deducible del de sus elementos componentes, es suficiente con que la macroestructura recoja las palabras simples y, si se desea, las derivadas se incluyan en familias léxicas (Corbin y Corbin 1989).

Sea como sea, los criterios que subyacen a la labor lexicográfica son siempre parciales y muy heterogéneos de manera que se da prioridad a objetivos bien diferentes, lo cual condiciona en buena medida el tipo y número de derivados que se registren: la comodidad del usuario, la función descifradora de mensajes, la frecuencia de uso de los derivados, las formas derivadas documentadas, la tarea de producción de mensajes, la economía de espacio en la macroestructura, etc.

El objetivo general y a largo plazo de este proyecto consiste en la creación de un *corpus* analizado gramaticalmente, desde los puntos de vista sincrónico y diacrónico, de todas las palabras derivadas del español y sus correspondientes latinas y ponerlo a disposición de los interesados —mediante DERIV@, una aplicación en entorno web— para realizar consultas personalizables según las necesidades del usuario. Este corpus constituirá el punto de partida para la construcción del *Diccionario de Palabras Derivadas del Español* (DIPADE).

3. Características de DIPADE

Desde nuestro punto de vista es fundamental que el diccionario incluya entre sus lemas aquellos derivados que presentan algún tipo de irregularidad, ya sea en la forma, ya en el significado. Por otra parte, sería conveniente que recogiera los derivados regulares documentados en la lengua. Nuestra propuesta de diccionario no es restrictiva en este sentido y da prioridad a la documentación. Solo quedarán excluidas de DIPADE aquellas palabras no derivadas que además no entran en procesos derivativos como bases.

Por otra parte, con vistas a la elaboración de un diccionario integral, cuya consulta permita acceder a la mayor cantidad posible de información lingüística disponible, nuestro proyecto pretende tomar en consideración diversos aspectos presentados hasta ahora de manera parcial, dispersa o heterogénea:

⇒ Otorgar un mismo peso a sincronía y diacronía, además de considerar los procesos de derivación secundaria. Las dos perspectivas se complementan en el estudio de la derivación por varias razones (Díaz García 2006a: § 2.1.):

- Gran parte de las formas derivadas llegan a nuestros días así formadas desde el latín, siguiendo por tanto las reglas de derivación latina.
- El análisis sincrónico muchas veces no es suficiente para explicar la formación y entender el significado de una palabra; es el caso de voces derivadas del latín que contienen una base no castellana, como *concatenar*, *inducir*, *leguminoso*.
- Existen algunas formaciones lingüísticas, que, según el punto de vista adoptado, pueden ser interpretadas como producto sintáctico o como resultado de un proceso de formación de palabras. Sería el caso de los adverbios en *-mente*.
- Existe la posibilidad de encontrar en la lengua variantes de un mismo afijo que no pueden ser estudiadas en distribución complementaria por causas diacrónicas.

⇒ Incluir toda la familia léxica de una raíz en la misma entrada. En la Figura 1 se muestra un ejemplo de la serie de derivación en la que se inserta el lema *ridiculous*:

Series de derivación		
		Clase
<u>1</u>	ridículo	Nombre (adjetivo y sustantivo) (u.t.c.s.)
<u>3</u>	ridícula	Adjetivo
<u>4</u>	ridículamente	Adverbio
<u>5</u>	ridiculez	Sustantivo
<u>6</u>	ridiculizar	Verbo
<u>7</u>	ridiculous	Adjetivo
<u>8</u>	ridiculosa	Adjetivo
<u>9</u>	reír	Verbo

Figura 1. Series de derivación

⇒ Atender a las variaciones gráficas de la palabra derivada y de su base a lo largo de la historia de la lengua, con sus correspondientes ejemplos¹. En las Figuras 2 y 3, que presentamos a continuación, ejemplificamos la variación gráfica del lema *codicioso* (Figura 2), así como la documentación de una de las variantes, *cudicioso* (Figura 3).

¹ Para la recopilación de las variantes gráficas hemos creado una herramienta generadora de variantes gráficas (VARIGRAF) a partir de diversos *corpora* textuales: ADMYTR, CORDE, CREA, DAVIES.

Variación gráfica		
Derivada	Año	Ejemplo
cobdiçioso	1240-1250	Aguisa a los unos; que sean cobdiçiosos . Faz los otros yrados; los otros enuidiosos.
cobdiçiosso	1400-1421	E este dezia por quanto el era cobdiçiosso de dinero.
cobdizioso	1502	La gente de la mar es cobdiziosa de dyneros y de volver á su casa.
codicioso	1236-1246	Los omnes codiciosos del haver monedado que por ganar riquezas non dubdan fer pecado.
codiçioso	1237	Otrosý non cuple levar a la guerra en la tu merçed gentes nin conpañas ricas nin codiçiosas.
codiciosso	1574	Codiciosso de ver lo que en el coffre venia, anduvole mirando.
codiçiosso	1330-1343	Amos por una dueña estavan codiçiosso , / eran muy bien apuestos e verás quán fermosos.
codizioso	1610	En busca de cauillos que perdidos, andauan codiziosos de hallarlos.
cubditioso	1250	Non por celo de iusticia mas por cubditiosa malitia.
cudicioso	1454-1457	Ciertamente no abrá aquí assaz para los romanos, ca como ellos son muy cudiciosos destas joyas y ornamentos.
cudiçioso	1406	Todos los chacataes se ajuntan al que aquel tesoro oujere ca son cudiçiosos .
cudiciosso	1608	E buen varon es seguir la virtud sola, y de cudiciosso los premios de ella.
cudiçiosso	1580-1600	Si el montero, de muy nueuo o demassiadamente cudiçiosso , asomare a descubrir de vna vez toda la querencia, podra ser visto.
cudizioso	1557	Viene á tomar esta resoluzion con los de Corintio, que pues que eran cudiziosos , deseosos, i ansiosos de alcanzar dones.
cudiziosso	1608	Esto acompañado con lagrimas es moneda que aun Dios (si puede) es cudiziosso en nosotros.

Figura 2. Variación gráfica

Documentación de grafía y ejemplo	
Derivada	cudiçiosso
Autor	Barahona de Soto, Luis
Título	Diálogos de la montería
País	España
Año	1580-1600
	Si el montero, de muy nueuo o demassiadamente cudiçiosso, asomare a descubrir de vna vez toda la querencia, podra ser visto.

Figura 3. Documentación de grafía y ejemplo

⇒ Ofrecer definiciones originales redactadas a partir del *corpus* de ejemplos.

⇒ Incluir ejemplos documentados en todas y cada una de las acepciones y en todas y cada una de las variantes gráficas del lema. Para la documentación de las acepciones, hemos utilizado diferentes *corpora* en línea (véase la nota 1), aunque también disponemos de un *corpus* de elaboración propia que nos ha permitido extraer no pocos ejemplos. Es el caso del que se muestra en la Figura 4:

Documentación ejemplo acepción	
Orden	1
Autor	Delibes, Miguel
Título	El hereje
País	España
Año	1998
Donde	corpus propio
	"El corcel" era un muchacho alto, empeinoso, con las extremidades desproporcionadas, levemente escorado del lado izquierdo.}

Figura 4. Documentación de ejemplo de cada acepción

⇒ Incorporar información sintáctica sobre combinatoria preposicional y verbal. La primera nos parece relevante, pues al estar interesadas en la trayectoria derivativa debemos tener en cuenta las herencias en las preposiciones regidas, por ejemplo en los sustantivos deverbales (Mas Álvarez, I. 1999: 353, 2005: Cap. 3). Respecto a la segunda, en la parte ya elaborada sobre derivados en *—oso*, por ejemplo, se incluye la combinatoria con *ser* o *estar*, un dato que nos permite postular la adscripción de determinados adjetivos a una u otra clase gramatical o léxica—. Véase el ejemplo seleccionado en la Figura 5 a propósito del lema *ansioso*:

Raíz:	ansi-
Clase	Adjetivo
Origen	latino
Es sinónimo de:	
Transcripción Fonética:	[ansjóso]
Corte Silábico:	an-sio-so
Proceso de formación	culta
Uso	--
Predicativo con	Ser y estar
Combinatoria preposicional	por y de
Proceso evolutivo	sufijada de simple
Evolución:	ansioso < ANXIOSUS < ANXIA < ANGERE < ango (gr.)

Figura 5. Información gramatical

⇒ Incluir también, allí donde tenga sentido, información sobre combinatoria léxica, con especial atención a las colocaciones.

4. Una base de datos de palabras derivadas del español

Nuestra propuesta de diccionario de derivación toma forma en la aplicación en entorno web DERIV@. Se trata de una base de datos lingüísticos con dos modelos de representación, uno para las palabras castellanas formadas por el procedimiento de la derivación y otro para las correspondientes latinas. El objetivo de DERIV@ es la consecución de una estructura general que contenga todas las palabras derivadas del castellano con información gramatical seleccionada, para que el *corpus* así constituido y analizado permita la creación de un diccionario de palabras derivadas. DERIV@ permite y facilita el acceso sin restricciones a la información contenida en las dos bases de datos. El modelo resulta válido para cualquier forma derivada del español (o de otras lenguas románicas), y —en su versión para entorno

web— deja abierta la posibilidad de acceso en tiempo real a cualquier usuario, que podrá interactuar con el sistema y contribuir a su perfeccionamiento².

Como ya hemos mencionado, el estudio gramatical de las palabras derivadas se realiza desde los puntos de vista sincrónico y diacrónico. El sincrónico recoge los afijos que existen en la actualidad y su combinatoria, analizándolos y analizando las palabras a las que se adjuntan, describiendo las reglas de esa combinación, aportando ejemplos documentados en diferentes contextos y observando las tendencias que podrían determinar los cambios en el léxico de años venideros. El diacrónico se ocupa de aquellos afijos y reglas que han sido heredados o incorporados en un momento dado de la historia de la lengua, teniendo en cuenta su procedencia, los cambios sufridos en el significante o en el significado, el mantenimiento o la variación de los índices de productividad, etc.

Integrarán DERIV@ todas las palabras castellanas y latinas formadas por el procedimiento de la derivación junto con sus bases. Están excluidas, sin embargo, como ya hemos mencionado, aquellas palabras no derivadas que no constituyen base para la creación de nuevos elementos léxicos. La organización de los datos lingüísticos contenidos en DERIV@, junto con las modalidades de acceso a ellos, hacen posible su consulta como diccionario de uso, diccionario etimológico, diccionario de régimen, instrumento de desambiguación morfológica, etc. La representación de las entidades y sus interrelaciones se efectúa mediante una base de datos relacional, estructurada de tal forma que pueda ser compartida por múltiples y diversos usuarios.

El punto de partida de este proyecto ha sido el estudio de los mecanismos morfológicos involucrados en la formación de palabras por derivación en español y la interacción de los dos elementos implicados en dicho proceso: base y afijo. El tema no carece de interés, puesto que nuestra lengua dispone de un considerable número de afijos y también porque en los últimos años se ha potenciado la vitalidad de no pocas creaciones léxicas. Algunos afijos han experimentado un descenso de su productividad, mientras que otros la han incrementado, entendida ésta en términos tanto cuantitativos como cualitativos (términos referidos al número de derivados que el sufijo puede crear y a la regularidad con la que se crean los eductos derivados).

El elemento básico de gestión es la ficha, en su calidad de representación gráfica de toda la información obtenida sobre una palabra. Hemos procurado que en la documentación de la base de datos se puedan ver con claridad tanto los diagramas de relación como la forma en que fueron establecidas las jerarquías entre entidades. La información que se ofrece de cada elemento léxico es la que sigue:

• Para la *palabra derivada*: *Raíz, Familia léxica, Clase, Origen, Sinónima de, Realización fonética, Corte silábico, Proceso de formación, Uso, Predicativo con, Combinatoria preposicional, Proceso de evolución, Patrón de relación con la base, Valor semántico, Etimología*. Si se trata de una palabra derivada latina: *Clase de palabra, Definición, Uso, Ejemplo, Documentación, Patrón de relación con la base, Definición*. Y de cada acepción: *Uso, Ejemplo, Documentación, Acepción que selecciona de la base, Variación gráfica a lo largo de la historia, Ejemplo, Documentación*.

• Para la *palabra base*: *Clase, Proceso de formación, Valor semántico, Etimología*. Y si es palabra latina: *Clase de palabra, Definición, Ejemplo, Documentación, Declinación, Género, Conjugación, Definición*. Y de cada acepción: *Uso, Ejemplo, Documentación, Variación gráfica a lo largo de la historia, Ejemplo, Documentación*. Derivados de la palabra derivada: *Clase, Definición, Ejemplo, Documentación, Observaciones*.

Somos conscientes de que la realización del *Diccionario de Palabras Derivadas del Español* (DIPADE) es tarea a largo plazo, sin embargo, —y gracias a que esta parcela de la lengua es fácilmente fragmentable—, se podrá ir completando en sucesivas etapas, trabajando con *subcorpora*, y tomando como punto de partida la derivación nominal. De esta manera, la

² Para conocer con más detalle las características de DERIV@, véase Díaz García y Mas Álvarez (2008). Es posible consultar esta aplicación en línea en el sitio <http://www.derivar.com>.

información podrá compartirse a medida que se dé por concluida la parte de análisis y descripción de cada uno de los afijos y palabras que los portan.

En la Figura 6 ofrecemos la primera pantalla del artículo del lema *ambicioso*, tal y como se presenta en la versión en línea de la aplicación. A la izquierda está el menú de exploración con las opciones disponibles.

The screenshot shows the Deriv@ web interface. On the left is a vertical navigation menu with options: Presentación, BD -OSO, Diccionario A-Z, Buscar en un texto, Búsq. Avanzada, Estadísticas, BD -OSUS, Diccionario A-Z, Estadísticas, BD Bibliografía, Búsq. Avanzada, Enlaces, CORDE, CREA, DRAE, and DAVIS. Below the menu is a search box labeled 'Buscador' with a 'Buscar' button. The main content area displays the entry for 'ambicioso' with its phonetic transcription [ambiθjóso] { am-bi-cio-so }. The entry includes sections for MORF. (morphology), SINT. (syntax), SEM. (semantics), and LEX. (lexicon). It also features a table of acceptations (Acep.) with definitions and examples, and a section for the family of words (Fám. léx.).

ambicioso
[ambiθjóso] { am-bi-cio-so }

MORF. adj. y sust. con flexión de género y número; **Evol:** ambicioso < AMBITIOSUS < AMBITIO < AMBIRE < AMB+IRE; **Raíz:** ambici-.

SINT. pvo. con 'ser'.

SEM. adj. calificativo. **Rel. con base:** posesión → que está lleno de (x), que tiene (x).

LEX.

Acep.	Definición / Ejemplo
①	Se aplica a alguien que tiene deseo desmesurado de alguna cosa. [ambición1] » Conchabado con un grupito de políticos tan ambiciosos y trepadores como él, había sido el artífice de la campaña. » Rojo, Alfonso; <i>Matar para vivir</i> , España, 2002.
②	Se dice de algo que entrafña o muestra ambición. [ambición1] » Barroca, expresionista, comprometida, ambiciosa, <i>El tambor de hojalata</i> es, también, la novela de una ciudad. » Vargas Llosa, Mario; <i>La verdad de las mentiras</i> , Perú, 2002.

Fám. léx.: ·ambición (sust.) ·ambicionar (vbo.) ·ambiciosa (adj.) ·ambiciosamente (adv.) ·ambicioso (adj.)

DIACR. Palabra culta formada a partir de AMBITIOSUS; **Origen** latino.

Derivada.	Año	Variación gráfica (CORDE)
ambicioso	1293	<i>E deuisse mucho los Reys guardar dellas por que non sean Reprehendidas ansi como lo son los ambiciosos que desean mas las honrras de quanto deuen.</i>

Figura 6. Presentación web

Como se puede comprobar, además de una presentación general del proyecto que permite al usuario primerizo familiarizarse con las características de la aplicación, se ofrecen diversas posibilidades de acceso a la información en cada uno de los bloques, el de las palabras derivadas del español y el de las palabras latinas³. Junto a la posibilidad de búsqueda directa del lema que nos interesa, tenemos la opción del diccionario en orden alfabético o bien el recurso de búsqueda a partir de la introducción de un texto concreto —hasta un máximo de 5000 caracteres—, en el que la aplicación resaltaré y seleccionará las palabras disponibles en DERIV@. En cuanto a la búsqueda avanzada, permite realizar consultas mediante la utilización de filtros sobre distintos campos de la ficha, en una selección dinámica combinada: el proceso se realiza en dos pasos, cada uno en una pantalla diferente, en el primero construimos el perfil de consulta seleccionando los criterios de búsqueda y en el segundo asignamos los valores concretos que nos interesan. El apartado de “estadísticas”, por último, contiene los resultados de

³ En la actualidad están disponibles como lemas con información completa los adjetivos denominales del español en -oso y los latinos en -OSUS.

algunas consultas estadísticas realizadas con la aplicación sobre la base de datos, lo cual nos permite obtener información general sobre el *corpus*⁴.

5. Para terminar

El valor de la información sobre los procesos de formación de palabras está ligado a la forma en que aquella se estructura, categoriza y organiza para presentarla al usuario interesado en ella. Si no se puede acceder a toda la información contenida en un sistema, o bien el acceso a ella es lento, carecerá de utilidad. Nuestra postura aboga por, con el fin de aprovechar la interactividad que ofrecen las redes electrónicas y su potencial pedagógico, revisar la información vinculada a cada elemento integrante del sistema y crear una interfaz para la integración de esos datos en un entorno multimedia. Adecuar los contenidos y, por supuesto, definir su utilidad en los procesos de consulta interactivos ha sido la tarea principal. De esta forma, la investigación sigue un doble camino que permite trabajar simultáneamente en dos campos: de una parte el del diseño de un modelo válido que permita analizar y describir un *corpus*, y por otra, el de su explotación, junto con la creación de la interfaz que sirve de puente entre el producto lingüístico y el tecnológico. A nuestro modo de ver, este interés, a la vez gramatical y tecnológico, es lo que justifica verdaderamente el acercamiento informático al material léxico.

No existe ninguna obra lexicográfica —que conozcamos— que contenga información gramatical completa acerca de una palabra; hasta ahora era necesaria la consulta de varios diccionarios diferentes si queríamos obtener información gramatical exhaustiva acerca de la sincronía y diacronía de un elemento lingüístico. La dispersión de los datos, así como el inconveniente del manejo de grandes volúmenes nos hizo concebir este proyecto de concentración de la información y de acceso rápido y fácil a ella. El objetivo general y a largo plazo de este proyecto se resume, pues, en la recogida y análisis de un *corpus* de todas las palabras derivadas del español y sus correspondientes latinas, con amplia información gramatical acerca de ellas, que será presentado en diferentes formatos para su consulta.

⁴ Las consultas disponibles en este momento son las siguientes: continuación del derivado latino, evolución de la palabra derivada, semántica de la palabra derivada, predicación con ser o estar, uso del derivado, sinónimo de la palabra derivada, clase del derivado, combinatoria preposicional, relación con la base, clase de la base, procedencia de los ejemplos y origen de la palabra derivada.

Referencias bibliográficas

- [CORDE]. *Corpus Diacrónico del Español* [en línea]. Madrid: Real Academia Española. <http://www.rae.es>.
- Corbin, D.; Corbin, P. (1989). "Sélection et description des dérivés et composés dans le dictionnaire monolingue". En Haussmann, F. J. *et al.* (eds.). *Wörterbücher. An International Encyclopedia of Lexicography*. Berlín: Walter de Gruyter. Vol. 1-2. 937-945.
- [CREA]. *Corpus de Referencia del Español Actual* [en línea]. Madrid: Real Academia Española. <http://www.rae.es>.
- Davies, M. (2003). *Corpus del español* [en línea]. <http://www.corpusdelespanol.org>.
- Díaz García, M. T. (2006a). *Un modelo de diseño, análisis y explotación de un corpus. Sincronía y diacronía del sufijo castellano -oso*. [cd-rom]. Santiago de Compostela: Publicacións e intercambio científico da Universidade de Santiago de Compostela.
- Díaz García, M. T. (2006b). *DERIV@. Base de datos de palabras derivadas del español* [en línea]. <http://www.derivar.com>.
- Díaz García, M. T.; Mas Álvarez, I. (2008). "DERIV@: A Linguistic Database for Spanish Word Formation". En Eichinger, L. M.; Meliss, M.; Domínguez Vázquez, M. J. (eds.). *Wortbildung heute. Tendenzen und Kontraste in der deutschen Gegenwartssprache*. Tübingen: Gunter Narr Verlag. 91-104.
- Marcos Marín, F.; Faulhaber, C. B.; Gómez Moreno, Á. (1999). *ADMYTE II: Archivo digital de Manuscritos y Textos Españoles*. [cd-rom]. Madrid: Micronet.
- Mas Álvarez, I. (1999). "El tratamiento lexicográfico de los sustantivos deverbales". En Samper Padilla, J. A.; Troya Déniz, M. (coords.). *Actas del XI Congreso Internacional de la Asociación de Lingüística y Filología de la América Latina. Las Palmas de Gran Canaria, del 22 al 27 de julio de 1996*. Las Palmas de Gran Canaria: Publicaciones Universidad Las Palmas de Gran Canaria. Vol. 1. 347-358.
- Mas Álvarez, I. (2005). *Construcciones sintácticas de sustantivos deverbales: los complementos adnominales con de*. [cd-rom]. Santiago de Compostela: Publicacións e intercambio científico da Universidade de Santiago de Compostela.
- Pérez Hernández, C.; Moreno Ortiz, A.; Faber Benítez, P. (1999). "Lexicografía computacional y lexicografía de corpus". *Revista española de lingüística aplicada* 1. 175-214.
- Sinclair, J. (1987). "Grammar in the Dictionary". En Sinclair, J. (ed.). *Looking up. An account of the COBUILD Project in lexical computing*. London: Collins. 104-115.
- Pharies, D. (2002). *Diccionario etimológico de los sufijos españoles*. Madrid: Gredos.