

LA EXPLOTACIÓN DE LA *BASE DE DATOS SINTÁCTICOS DEL ESPAÑOL ACTUAL (BDS)**

GUILLERMO ROJO
Santiago de Compostela

en De Kock, Josse (ed.) *Lingüística con corpus. Catorce aplicaciones sobre el español (= Gramática española. Enseñanza e investigación I.7)*, Univ. de Salamanca, 2001, págs. 255-286.

INTRODUCCIÓN

Resulta obvio que en el estado actual de nuestra disciplina, y mucho más en una publicación como esta, carece por completo de sentido dedicar algunos párrafos a demostrar la utilidad y necesidad de los corpus textuales en la investigación lingüística. Es cierto que tanto el corpus de Lovaina¹ como quienes lo diseñaron, construyeron y estudiaron² fueron durante muchos años un elemento bastante extraño y aislado en el conjunto de la investigación gramatical del español, pero también lo es que el panorama ha cambiado considerablemente en los últimos años. En efecto, hoy disponemos ya, o estamos a punto de hacerlo, de la posibilidad de consultar varios corpus de gran tamaño, organizados con finalidades diferentes, en los que podemos obtener con bastante facilidad los materiales que podamos necesitar para nuestros estudios léxicos o gramaticales, tanto de orientación diacrónica como sincrónica.

* Quiero dejar aquí constancia de mi renovado agradecimiento a Victoria Vázquez Rozas y Carmen Silva Corvalán, cuyas lecturas inteligentes y cuidadosas han librado al texto de este capítulo de no pocos errores de los más diversos tipos. Los que queden siguen siendo, por supuesto, de mi exclusiva responsabilidad.

1. Publicado en versión impresa y electrónica con índices y comentarios en *Gramática española. Enseñanza e investigación*, Salamanca, Ediciones Universidad de Salamanca, t. III.1, J. De Kock, R. Verdonk, C. Gómez Molina, *19 textos*, 1ª reimpresión, 1996, 276 págs. + 1 disquette, t. III.2, J. De Kock, C. Gómez Molina y N. Delbecque, *20 textos*, 1992, 276 págs., t. IV.1, J. De Kock, *Índice alfabético, alfabético inverso y de frecuencia de 19 textos*, 1991, 499 págs., t. IV.2, J. De Kock, *Índice alfabético, alfabético inverso y de frecuencia de 20 textos*, 1992, 526 págs., t. V.1, J. De Kock, *Concordancia alfabética de 19 textos*, 1990, 1.964 págs., sólo consultable bajo forma de listado, t. V.2, J. De Kock, *Concordancia alfabética de 20 textos*, 1990, 1.927 págs., sólo consultable bajo forma de listado.

2. Véanse, *idem*, I, *Apuntes metodológicos*, cinco tomos hasta la fecha, y II, *Gramática*, seis tomos.

Es igualmente cierto, sin embargo, que un corpus textual abre nuevas perspectivas para un gramático cuando, por cualquier procedimiento, se ha añadido al texto información gramatical y léxica, esto es, cuando se ha lematizado y etiquetado morfosintácticamente. En ese momento es posible pasar de las consultas simples, basadas en la pura forma ortográfica de las palabras implicadas en nuestra investigación, a búsquedas que proporcionen directamente, por ejemplo, todos aquellos casos presentes en el corpus en que un verbo cualquiera vaya seguido a corta distancia de una conjunción y de otro verbo cualquiera en modo subjuntivo. Aquí se maneja y se requiere información que no está explícita en el texto, sino que ha sido añadida e implica la adición directa o indirecta de conocimiento gramatical al texto «crudo».

Pero un texto etiquetado morfosintácticamente y lematizado es solo un paso más en el camino. Desde un planteamiento general (quiero decir, sin alusión directa a lo que sucede en la investigación del español), estamos hoy en la fase de incorporar a los corpus el resultado del análisis sintáctico, semántico y pragmático. Un corpus analizado sintácticamente nos permite ya localizar todos aquellos casos en los que un verbo cualquiera va acompañado de un complemento directo que se realiza bajo la forma de una cláusula de infinitivo o bien las frases nominales en las que hay un modificador en forma de frase preposicional. Cabe pues, con independencia de cómo se haya introducido, obtener automáticamente datos relevantes para el estudio de lo que sucede en el interior de las frases, las cláusulas y las oraciones o de algún subtipo de una de estas unidades. Algo parecido, aplicado a los ámbitos que les son propios, supone la adición a un texto de los análisis semántico y pragmático.

2. LA BASE DE DATOS SINTÁCTICOS DEL ESPAÑOL ACTUAL (BDS)

2.1. Justificación del proyecto

Cuando los procesos son total o básicamente automáticos, parece claro que los que acabo de señalar corresponden a estadios diferentes, organizados de modo que se puede acceder a cada uno de ellos solamente después de haber llevado a cabo las tareas correspondientes a los previos. Reconocer la presencia de una frase nominal que funciona como sujeto en el interior de una cláusula solo es posible si previamente hemos identificado correctamente la clase de palabras, el lema y los valores que tienen las categorías gramaticales aplicables en cada caso de todos los elementos de la cláusula o, al menos, la mayor parte de ellos. No ocurre así forzosamente, en cambio, cuando el análisis sintáctico, semántico o pragmático se realiza de forma manual y se incorpora al texto de modo que el ordenador pueda posteriormente interpretarlo adecuadamente. Por ejemplo, podríamos proceder al análisis sintáctico de todas las cláusulas de un texto para únicamente identificar y delimitar sus elementos funcionales o añadir a la función el tipo y subtipo de unidad a que pertenecen (frase nominal, cláusula de infinitivo, etc.), pero sin necesidad de entrar en el análisis de cada uno de esos elementos y, por tanto, sin dar la clase de palabras a que pertenecen las que forman

esa cláusula o, cuando menos, sin hacer explícita esa información, que puede haber sido manejada por el lingüista en su proceso de análisis.

El grupo de lingüistas que concibió y desarrolló la *Base de datos sintácticos del español actual* (BDS en adelante) estaba y está todavía interesado fundamentalmente en la estructura y los constituyentes funcionales de la cláusula (entendida como la unidad gramatical que se organiza en torno a un elemento que desempeña la función de predicado). Para progresar en la línea en que habíamos estado trabajando ya desde varios años antes, necesitábamos disponer de datos fiables y cuantitativamente representativos acerca de las construcciones en que puede entrar un verbo, la subcategorización sintáctica y semántica de cada uno de sus argumentos, los verbos que pueden aparecer en diferentes construcciones sintácticas y cuestiones similares. Y el camino natural para conseguirlos no parecía ser el de construir primero una morfología formal, luego un sistema de desambiguación, etc., hasta lograr, varios años después, los primeros ejemplos de secuencias reales analizadas con un porcentaje de aciertos presumiblemente bajos.

La verdad es que en aquel momento (1987) no disponíamos de los conocimientos necesarios para construir un sistema de este tipo, pero también es cierto que no optamos por una vía distinta simplemente por nuestras carencias, sino porque vimos con toda claridad que había otro camino que podía resultar mucho más efectivo y más rentable para nosotros y la comunidad científica en general en un plazo de tiempo mucho más breve: analizar manualmente las cláusulas de un corpus textual relativamente amplio y almacenar el resultado de ese análisis en una base de datos. Al tiempo, este procedimiento hacía posible incorporar a nuestro análisis todos aquellos rasgos que estimáramos que podían resultar pertinentes o simplemente interesantes para la comprensión de lo que sucede dentro de los diferentes tipos de cláusula en español. Después de varios años de trabajo³, obtuvimos una base de datos que contiene el análisis de las aproximadamente ciento sesenta mil cláusulas que integran el conjunto de los treinta y cuatro textos (algo más de un millón cuatrocientas cincuenta mil formas) que componen la parte contemporánea del *Archivo de Textos Hispánicos de la Universidad de Santiago* (ARTHUS). En el momento de escribir este trabajo

3. Financiado en su primera fase (1988 a 1991) por la *Dirección Xeral de Ordenación universitaria e Política científica* de la *Consellería de Educación* de la *Xunta de Galicia* (referencia XUGA 82710088) y por la Dirección General de Investigación científica y técnica del Ministerio de Educación y Ciencia entre 1991 y 1994 (ref. PB90 0376). A lo largo de los diez años transcurridos desde su comienzo, han sido muchas las personas que, en diferentes grados y con distintas tareas, han colaborado en él. En el momento de escribir este trabajo (diciembre de 1998) siguen vinculados a la BDS los siguientes miembros del equipo inicial, todos ellos adscritos a la Universidad de Santiago de Compostela salvo indicación en contrario: Francisco García Gondar, José María García-Miguel (Univ. de Vigo), Belén López Meirama, María del Carmen Losada Aldrey, Inmaculada Mas Álvarez, María José Rodríguez Espiñeira, Guillermo Rojo y Victoria Vázquez Rozas. Además, colaboran activamente en la corrección final y la aplicación de los resultados a diferentes fenómenos gramaticales Conchita Álvarez Lebrede, Pilar Alvariño Alvariño, Cristina Blanco Canosa, Fernando Castro Paredes, Susana Comesaña Iglesias (Univ. de Vigo), Adelaida Gil Martínez, Fátima Gayoso Gómez, Eva Muñiz Álvarez, Marta Rebollo Lemus, María Paula Santalla del Río, Susana Sotelo Docío y David Vázquez Martínez.

(septiembre de 1998), la BDS está ya en su fase final de corrección y confiamos en que saldrá finalmente a la luz hacia mediados de 1999.

2.2. Características generales de la BDS

Así pues, la *BDS* es una base de datos sintácticos cada uno de cuyos registros contiene informaciones correspondientes a los cinco grandes bloques siguientes⁴:

a) Datos referentes al verbo que actúa como predicado y su localización en el texto impreso (texto, página y línea, con, además, columna y número de noticia en el caso de los textos periodísticos).

b) Datos referentes a la cláusula como conjunto: tipo de cláusula, función que desempeña, voz, modalidad, polaridad, forma verbal empleada, forma verbal de la que depende (si es pertinente), orden de los elementos, etc. (vid. tabla 1 del apéndice).

c) Datos referentes a cada una de las funciones sintácticas que pueden existir en la cláusula. Hemos considerado sujeto, complemento directo, complemento indirecto, otros complementos argumentales (se subdividen en suplementos, complementos adverbiales, complementos modales y otros complementos preposicionales), agentes y predicativos (de sujeto complemento directo y otros) (vid. tabla 2 del apéndice). Además de la indicación de su existencia y, cuando es necesario, su carácter, la ficha refleja el tipo de elemento que realiza la función (vid. tabla 3 del apéndice).

d) Observaciones: casos de predicativos no argumentales, ejemplos especialmente interesantes de cláusulas complejas, ausencia de marcas, dequeísmo, etc.

e) Informaciones obtenidas automáticamente a partir del contenido de los campos correspondientes a los bloques anteriores y que son manejadas por los diferentes programas que gestionan la base de datos para hacer estadísticas, realizar las búsquedas solicitadas, etc.

Es sobradamente conocida la existencia de importantes discrepancias acerca de cómo analizar sintácticamente determinadas construcciones o en torno al carácter de ciertas funciones sintácticas. En nuestra aproximación al estudio de la cláusula y del régimen verbal, el dato primordial es, por supuesto, la identificación de los elementos funcionales que aparecen combinados en cada caso con el predicado, pero no es menos importante conocer el tipo y subtipo a que pertenecen las unidades que desempeñan cada una de las funciones. No solo, pues, la indicación de que un cierto verbo se construye con un suplemento, sino también que ese suplemento va introducido por una determinada preposición y que es una cláusula de infinitivo. Como en cualquier proceso de análisis, la construcción de la BDS ha supuesto la necesidad de ir tomando de forma congruente

4. Me limito aquí a exponer los rasgos que considero necesarios para entender lo que viene a continuación. Para más detalles puede consultarse Rojo (1995) y también la información contenida en <http://www.usc.es/~sintx>.

decisiones sobre puntos que afectan a los más variados aspectos. Entre otros muchos, y simplemente para dar una idea del camino seguido, cabe señalar aquí, por ejemplo, los siguientes:

a) Establecimiento del inventario de funciones clausales de carácter argumental. Por citar únicamente el caso más conflictivo, en la BDS se diferencia entre suplemento, complemento adverbial y complemento modal. De todas formas, la ficha se ha organizado de tal modo que, en caso de que llegara a interesar hacerlo así, la asignación de esquemas (que se hace mediante un programa que lee los datos contenidos en la ficha) podría fundir estos tres tipos en una clase única o bien seleccionar uno de ellos y no atender a los otros dos.

b) Carácter argumental o no argumental con respecto a diferentes verbos de elementos con características semejantes. Para continuar en la misma línea de ejemplificación, el locativo de *residir en algún lugar* parece argumental, figura en la valencia del verbo, mientras que no sucede eso con el locativo de *leer un libro en la biblioteca*, que no está en la valencia del verbo y, por tanto, no es argumental. En consecuencia, la ficha correspondiente a un ejemplo del primer tipo debe dejar constancia de la existencia del locativo y atribuirle una función argumental, mientras que no debe ocurrir eso en el segundo caso.

c) Identificación de la función realizada. ¿Cuál es la función de la segunda frase nominal en *la conferencia duró dos horas o el saco pesa ochenta kilos?* ¿Desempeñan la misma función los elementos nominales en *acercarse a Juan* (*acercársele o acercarse a él*) y *acercarse a la pared* (*acercarse a ella*, pero no **acercársele*)?

Sobre cientos de puntos conflictivos del estilo de los que acabo de señalar fue necesario discutir y, tras acordar la solución que parecía más adecuada, incorporarla a un *Manual de fichado* que se consideraba de cumplimiento obligatorio, con independencia de las opiniones personales sobre cada uno de sus puntos. Es el texto que, en su redacción final, explicará a quienes manejen los datos derivados de la BDS cómo se ha realizado el proceso de fichado y, por consiguiente, cómo pueden esperar encontrar los datos de su interés.

Para evitar la dispersión que habría supuesto ir haciendo los análisis siguiendo la secuencia lineal de cada uno de los textos, preparamos una versión electrónica de los textos seleccionados y elaboramos índices de las formas contenidas en cada uno de ellos. De este modo, todos los casos del mismo verbo en cada texto eran fichados uno tras otro por la misma persona, con lo que ello supone para captar las fuentes de dificultad y, en caso necesario, corregir el análisis realizado un rato antes. Naturalmente, los procesos posteriores de revisión han permitido ya trabajar con todos los ejemplos de cada verbo contenidos en el conjunto del corpus.

Por último, un amplio conjunto de programas de manejo de las fichas individuales permite reelaborar su contenido, agrupar los datos y crear ficheros derivados que organizan y resumen la información del general. La reelaboración y refinamiento de su contenido consiste en añadir a cada ficha, a partir de lo que figura en los campos específicos, la indicación del esquema sintáctico que presenta el

verbo en esa cláusula concreta (sujeto - predicado - complemento directo, por ejemplo) y el subesquema (sujeto animado - predicado - complemento directo en forma de cláusula de infinitivo, por ejemplo). Esto último tiene, en mi opinión, gran importancia porque la ficha básica contiene en los campos destinados a reflejar la clase de los elementos que desempeñan las distintas funciones cincuenta y cuatro posibilidades distintas. Queda indicado, por tanto, si la unidad que funciona como sujeto es un pronombre personal de primera persona, un demostrativo, una frase nominal de carácter inanimado, un relativo, etc., lo cual proporciona informaciones de gran interés para ciertas investigaciones. Para el estudio de las construcciones verbales, en cambio, todas esas diferencias pueden ser reducidas, de modo que algunos programas, sin alterar estos datos básicos, añaden la información correspondiente al esquema y el subesquema propio de la cláusula. Dado que se hace mediante programación, es siempre posible cambiar la agrupación de los datos, con lo que los subesquemas resultantes serán, en función de necesidades diversas, más o menos detallados.

A partir de esa reelaboración de los datos básicos, se crean unos ficheros derivados que contienen, respectivamente:

- a) Un registro para cada verbo diferente (por ejemplo, *ABANDONAR*).
- b) Un registro para cada verbo en cada esquema sintáctico en que aparece (por ejemplo, *ABANDONAR* sujeto - complemento directo; *ABANDONAR* sujeto - complemento directo - complemento adverbial, etc.).
- c) Un registro para cada verbo en cada subesquema sintáctico en que ha sido registrado (por ejemplo, *ABANDONAR* sujeto animado - complemento directo animado, *ABANDONAR* sujeto animado - complemento directo inanimado, etc.).

En consecuencia, como todo ello está situado en una base de datos relacional, resulta muy sencillo obtener todos los datos contenidos en la *BDS* acerca de, por ejemplo:

- a) Esquemas sintácticos en que aparece un verbo determinado en el corpus.
- b) Subesquemas sintácticos en que ha sido localizado un verbo determinado.
- c) Verbos que han sido registrados en un cierto esquema o subesquema sintáctico.
- d) Casos de un verbo en un cierto esquema o subesquema con el texto correspondiente.
- e) Verbos que aparecen en el corpus en dos esquemas o subesquemas sintácticos determinados.

En todos los casos aparecen las indicaciones estadísticas pertinentes, de modo que quien hace la consulta puede conocer, por ejemplo, cuántas veces ha aparecido un verbo en el corpus, cuántos de esos casos corresponden a cada esquema y subesquema que presenta (y el porcentaje que suponen con respecto al total del verbo o al total del esquema, respectivamente).

2.3. Corpus, bases de datos sintácticos, *treebanks* y la BDS

La rapidísima descripción del proceso seguido en su construcción y el contenido de la BDS deja patente el carácter original que posee este recurso. En efecto, es evidente que la BDS no es un corpus, sino una base de datos que contiene los resultados del análisis, realizado manualmente, de los aspectos sintácticos fundamentales de todas y cada una de las cláusulas existentes en el corpus constituido por los textos que componen la parte contemporánea de ARTHUS⁵. El almacenamiento de esos análisis individuales en una base de datos electrónica libera a sus responsables de una enorme cantidad de trabajo mecánico, permite la explotación de su contenido por cualquiera de los elementos incorporados a la ficha, automatiza los recuentos y, como acabo de indicar, permite agrupar los datos básicos en la forma que resulte adecuada a cada finalidad.

En tanto que contiene el análisis sintáctico de las aproximadamente 160.000 cláusulas contenidas en los textos examinados (sin ir más allá de los aspectos generales y las características de sus constituyentes funcionales), la BDS es una *base de datos lingüística*, pero no es un corpus anotado sintácticamente ni, por supuesto, un *treebank*. Resulta, me parece, demasiado restringida la definición de Van Halteren, para quien «*linguistic database is therefore synonymous with corpus* [...] or, perhaps more exactly, synonymous with *annotated corpus*, a term which describes a corpus in which specific types of information have been added (or merely made explicit) by some form of annotation» (Van Halteren: 1997, 9). La información contenida en la BDS o en cualquier otro recurso de este tipo es, en los aspectos considerados en cada caso, similar a la que se encuentra en un corpus anotado sintácticamente, por lo que no parece adecuado eliminar un recurso como el que aquí se presenta de la clase de las bases de datos lingüísticos. Sin embargo, las diferencias con un corpus anotado sintácticamente o el *treebank* derivado de él son también claras. La BDS no contiene el texto con los rúbricos sintácticos habituales que, manipulados informáticamente, permiten la posterior recuperación de la información que interesa a quien hace la consulta. La BDS consiste en la codificación del análisis realizado manualmente y la conexión con el texto para permitir la visualización de los ejemplos se hace a través de la referencia (texto, página, línea, más columna y noticia en los textos de prensa) de cada uno.

La organización de la BDS, centrada en torno a la cláusula y al elemento que en ella desempeña la función de predicado, proporciona una enorme facilidad

5. Aunque nosotros mismos hemos empleado otras denominaciones en ciertos casos, la perspectiva correcta es, me parece, considerar que ARTHUS es, como su nombre indica, un *archivo* de textos en el que hemos ido acumulando (y lo seguimos haciendo) textos de muy diferentes épocas y tipos. En su parte contemporánea, ARTHUS contiene 34 textos con un total de 1.450.000 ocurrencias en los que intentamos incluir representantes de todas las clases (narrativa, teatro, ensayo, prensa y lengua oral), tanto de España como de América. Es, pues, un corpus en tanto que pretende constituir una muestra del español tal y como se nos presenta entre 1980 y 1990. No sirve, sin duda, dado su corto tamaño, para estudios léxicos, pero sí resulta adecuado para el estudio del conjunto de fenómenos para el que fue diseñado.

de manejo en aspectos que en un *treebank* resultarían bastante difíciles de obtener en caso de que alguien quisiera estudiarlos. Obtener, por ejemplo, la lista de todos los verbos que llevan un elemento argumental (sea como suplemento o como suplemento adverbial) introducido por una determinada preposición, con indicación de la frecuencia con que cada uno de ellos aparece en el corpus y el porcentaje que supone con respecto al verbo es una consulta ya preparada en la BDS. Saber qué verbos pueden llevar un complemento directo en forma de cláusula de *que* + subjuntivo es también algo inmediato. No se trata, por supuesto, de buscar supuestas superioridades de un tipo de recurso sobre el otro, sino de mostrar que la organización y estructura de la BDS es la que resulta adecuada al objetivo con que fue creada.

De otra parte, los textos analizados para la BDS no están anotados morfológicamente ni lematizados. El análisis se ha hecho tomando la cláusula como unidad central (que es lo adecuado cuando se pretende estudiar el régimen verbal) e identificando los elementos funcionales que la componen y la clase a que pertenecen (con una tipología muy rica, como he señalado anteriormente), pero sin proceder al análisis interno de cada uno de ellos. Por tanto, el modelo más cercano es el *skeleton-parsing* al que se han referido Leech y otros⁶, esto es, el análisis de la cláusula en sus grandes constituyentes funcionales.

Como todas las aproximaciones, la adoptada para construir la BDS tiene ventajas e inconvenientes. A las ventajas ya señaladas, que tienen sentido si se piensa que su ámbito es el estudio de las estructuras clausales y el régimen verbal, puede añadirse el hecho de que la concepción, muy amplia, de los fenómenos de que se deja constancia en la ficha, permite ahora obtener datos del mayor interés sobre aspectos que estaban muy alejados de nuestras preocupaciones iniciales. Por no poner más que un ejemplo, la mayor parte de los datos sobre el orden de los elementos funcionales de las cláusulas monoactanciales en español que aparecen en López Meirama (1997) vienen de la BDS. De modo parecido, datos sobre frecuencia de los verbos, las formas verbales, las perífrasis verbales, la correlación temporal, las formas verbales empleadas en cierto tipo de cláusulas que funcionan en el interior de determinadas oraciones bipolares, etc., han sido todos ellos obtenidos del interior de la BDS.

Las desventajas con respecto a otras aproximaciones aparecen, naturalmente, en aquellos aspectos desvinculados de la estructura predicado - argumentos y en los que, aun teniendo relevancia para ello, no fueron vistos adecuadamente en el momento inicial. No hay en la BDS nada que permita estudiar las frases nominales y las adjetivas, ni siquiera en aquellos casos en que un sustantivo derivado de un verbo rige a otro elemento. Aunque es posible obtener, por ejemplo, datos acerca de las cláusulas que funcionan como condicionante o condicionado de una oración condicional (y, por tanto, estudiar lo que sucede en su interior), la BDS no da forma de unir automáticamente los dos constituyentes de la condicional para estudiarla globalmente. Puesto que no se ha fichado la presencia de

6. Para una presentación rápida de sus conexiones con los *treebanks* puede verse, por ejemplo, Leech and Eves (1997).

elementos no-argumentales, no hay datos sobre los circunstanciales, lo cual tiene un coste muy alto en aquellos casos en los que hay dudas acerca del carácter argumental o no argumental de, por ejemplo, un locativo con un determinado verbo o grupo de verbos.

En definitiva, el camino adoptado en la construcción de la BDS es adecuado (el más adecuado a mi modo de ver) para el objetivo con que fue construida. El mayor problema de la BDS no viene de que no sea un *treebank*, sino de que, en su forma actual, resulta de un análisis manual. Como es bien sabido, cualquier texto, cualquier corpus, muestra siempre unos pocos casos de elementos muy frecuentes y muchos casos de elementos poco frecuentes. En las 160.000 cláusulas fichadas hemos obtenido datos de unos 3.550 verbos. 748 de ellos (el 21%) aparecen una vez; 1.679 verbos (el 47%) aparecen documentados menos de cinco veces. Es evidente que continuar analizando manualmente no es la solución, puesto que supone un enorme esfuerzo para mantener una constitución equilibrada de la BDS. El camino es, evidentemente, construir una gramática formal que practique automáticamente el análisis de los grandes componentes de la cláusula sobre grandes cantidades de textos (esto es, una especie de *skeleton-parsing*) y almacene el resultado en una estructura semejante a la de la actual BDS. Esta es la dirección en que estamos trabajando actualmente⁷.

3. LA EXPLOTACIÓN DE LA BDS

3.1. Generalidades

De lo expuesto en el apartado anterior se deduce que, al lado de los inconvenientes derivados del hecho de proceder de análisis manuales y de no contener textos anotados sintácticamente, la BDS posee también algunas ventajas, entre las cuales destaca la enorme riqueza de la información que contiene acerca de los fenómenos y elementos vinculados a la cláusula. Dado que los datos correspondientes al análisis de cada cláusula han sido registrados en un formato bastante elemental, resulta posible casi siempre buscar la posible relación entre dos fenómenos que, de entrada, podrían parecer desconectados. Por ejemplo, estudiar la posición con respecto al predicado que tiene el sujeto de una cláusula cuando esta función es desempeñada por el pronombre personal de primera persona del singular según el número de argumentos de predicado, la modalidad de la cláusula u otros factores que puedan ser relevantes resulta posible y bastante sencillo si se conoce la estructura de la BDS y los rudimentos del lenguaje de consulta de una base de datos relacional. En este apartado me ocuparé de mostrar, a grandes rasgos, algunos de los tipos de datos que pueden ser obtenidos de la BDS.

7. Se trata del proyecto *Análisis automático del régimen verbal en español* financiado por la Dirección Xeral de Universidades e Investigación de la Xunta de Galicia (XUGA 20402B97) durante los años 1997, 1998 y 1999. El proyecto integra el trabajo de tres equipos radicados en las tres universidades gallegas.

3.2. La frecuencia de los verbos

Es bien conocida la escasez de estudios estadísticos sobre el español, tanto en el léxico como en la gramática. Quien desee saber algo acerca de la frecuencia de los verbos en español se ve hasta ahora forzado a consultar el diccionario de Juilland y Chang Rodríguez (1964). Se trata, sin duda, de una obra excelente, pero concebida y publicada hace ya casi medio siglo, con lo que eso significa en un terreno como el de los corpus lingüísticos, donde tantas cosas han ocurrido en los últimos diez o quince años. En el lado positivo se cuenta el hecho de que las formas están lematizadas, de modo que podemos disponer de datos referidos a las frecuencias de los elementos léxicos, no únicamente de las formas, como sucede, en cambio, en la obra, mucho más reciente, de Alameda y Cuetos (1996). También en el platillo de los rasgos favorables cae el hecho de que se indique la clase gramatical de las palabras, con lo que es posible, aunque a base de realizar un trabajo manual escasamente gratificante, hacer estadísticas sobre clases de palabras, verbos, por ejemplo⁸, y la resolución de las homografías, tanto en la adscripción léxica como en los casos de sincretismo gramatical (primeras y terceras personas de los copretéritos, por ejemplo).

Entre los inconvenientes, cabe señalar los pertenecientes a dos ámbitos distintos. De un lado, todo lo relacionado con la composición del corpus empleado y la selección de lemas. Se trata de un corpus de 500.000 formas, constituido por fragmentos de textos escritos publicados en España entre 1920 y 1940, del cual se han seleccionado solo los 5.024 lemas más usados de acuerdo con ciertos criterios de distribución (cf. Juilland-Chang Rodríguez: 1964, xx). De otro, el propio sistema empleado en la adscripción de formas, que atribuye a un determinado verbo (*haber* o *ir*, por ejemplo), tanto sus usos auxiliares como los independientes o los auxiliados, con lo que ello significa de distorsión para el caso de los verbos con frecuencia de uso en función de auxiliar y de virtual imposibilidad de obtener datos para la frecuencia de las perífrasis verbales⁹.

En la BDS, en cambio, cuando se registra una perífrasis, la ficha correspondiente a la cláusula se asigna al verbo auxiliado (que es el que impone el esquema argumental de la cláusula), pero se indica también la perífrasis utilizada y la forma verbal del auxiliar. De esa forma, los recuentos resultan mucho más adecuados, aunque, por supuesto, no impide que pueda haber discrepancias acerca del carácter perifrástico o no perifrástico de una construcción o de un caso determinado de una construcción.

8. Véanse, por ejemplo, los trabajos de Patterson y Urrutibéity (1975) o Corbella (1987). Estos trabajos están hechos a base de, en muchos casos, detectar y contar manualmente los datos correspondientes a las distintas formas flexionadas, lo cual explica la existencia de algunos desajustes de menor importancia.

9. La finura de la adscripción realizada por Juilland y Chang Rodríguez (1964), sin embargo, llega bastante lejos en algunos casos. Así, en el del verbo *haber*, por ejemplo, la lista correspondiente marca los usos como auxiliar y los diferencia de los otros. Lamentablemente, aunque de forma perfectamente comprensible, no se hace lo mismo en todos los demás casos de uso de los verbos como formas auxiliares de perífrasis.

En el cuadro 1 situó, en orden decreciente, los veinte verbos más frecuentes en español según los datos de la BDS¹⁰ e incorporo, para permitir la comparación, el lugar que ocupan (también según su frecuencia, como es natural) en el corpus de Juilland y Chang Rodríguez (1964) y el corpus LexEsp¹¹. Es fácil observar las discrepancias entre las tres listas y suponer su causa. *Haber* e *ir* en la BDS están mucho más abajo que en las otras dos listas precisamente porque no se ha tenido en cuenta más que sus usos como verbos plenos¹². En sentido contrario, creo que se puede pensar que *decir* ocupa en la BDS una posición más alta que en las otras listas como consecuencia del peso, quizá excesivo, que los textos narrativos tienen en ella¹³.

CUADRO 1

Los veinte verbos más frecuentes en tres corpus distintos

Puesto en la BDS	Verbo	Porcentaje en la BDS	Puesto en LexEsp	Puesto en Juilland-Chang
1	ser	12,50	1	1
2	decir	4,40	7	6
3	estar	3,70	3	3
4	tener	3,00	4	4
5	hacer	2,80	5	7
6	querer	2,70	12	11
7	ver	2,00	9	9
8	dar	1,60	10	10
9	saber	1,60	11	12
10	haber	1,50	2	2
11	ir	1,10	8	5
12	creer	0,97	40	16
13	pasar	0,96	15	13
14	parecer	0,91	14	17
15	hablar	0,81	22	19
16	dejar	0,80	17	20
17	pensar	0,74	23	28
18	poner	0,73	18	23
19	llevar	0,73	19	21
20	quedar	0,71	20	25

10. La frecuencia (y, por tanto, los porcentajes) correspondientes a los verbos *ser*, *decir*, *estar*, *querer* y *haber* que figuran realmente en la BDS son considerablemente menores que los que aparecen en este cuadro. La razón está en que, en un cierto momento del proceso de fichado, se estimó que, siempre con referencia exclusiva a su uso como verbos plenos, estos elementos combinaban una frecuencia muy alta con una variedad de esquemas sintácticos relativamente escasa. Se decidió por ello fichar únicamente uno de cada cinco casos de estos verbos, seleccionados por el orden de aparición en nuestras listas de las formas correspondientes a cada uno. Los porcentajes que figuran en este cuadro corresponden a la frecuencia que presentarían en la BDS en caso de que se hubieran fichado todas sus apariciones. La técnica utilizada aquí con los casos totales del verbo no se puede aplicar, como es lógico, a las formas, esquemas o subesquemas sintácticos de cada uno de ellos.

Como era de esperar, el análisis de las frecuencias de uso devuelve la impresión habitual de acumulación de empleos en unos pocos verbos. Los diez primeros verbos de la BDS suponen un uso conjunto del 39,90%. Los veinte primeros suman el 47,76%, lo cual significa que alrededor de una de cada dos apariciones de un verbo en los textos españoles procede de alguno de los veinte que figuran en el cuadro 1 (recuérdese, para interpretar correctamente este dato, que están excluidos los usos como verbo auxiliar, de modo que *te voy a decir* y *te estoy diciendo* cuentan como casos de *decir*, no de *ir* ni de *estar*).

3.3. La frecuencia de los esquemas sintácticos

Entrando ya en lo que constituye el terreno específico para el que se ha construido la BDS, el primer aspecto de interés general es, por supuesto, el de la frecuencia de los esquemas sintácticos clausales. Como es habitual, llamamos esquema sintáctico de una cláusula a la serie de elementos funcionales que se dan en su interior, pero añadiendo, y esto es ya menos habitual, la consideración de la construcción: activa, pronominal o pasiva perifrástica. Así pues, en la BDS los esquemas sintácticos identificados y manejados son del tipo: construcción activa formada por sujeto, predicado y complemento directo, construcción pronominal formada por sujeto, predicado y suplemento, construcción pasiva perifrástica formada por sujeto, predicado y agente, etc.

Las funciones clausales consideradas han sido: predicado, sujeto, complemento directo, complemento indirecto, suplemento, complemento adverbial, complemento modal, complemento preposicional, complemento agente y complemento predicativo (de sujeto o de complemento directo). Los suplementos, adverbiales, modales y otros preposicionales pueden aparecer hasta dos veces en la misma cláusula. Se considera construcción pronominal, de modo sin duda discutible, toda construcción en que aparezca un clítico concordado con el sujeto, por lo que se funden en un tipo (el de las construcciones pronominales) las cláusulas que, en un análisis que atendiera a estos factores, serían consideradas reflexivas, recíprocas, impersonales, pasivas reflejas o medias. Dadas las dificultades previsibles para

11. LEXEsp es un corpus del español actual constituido por fragmentos de textos tomados de noticias de diferentes temas, editoriales, ensayos, novelas, etc. que contiene unos cinco millones y medio de formas. Es el resultado del proyecto LEXEsp APC 96-0125, financiado por la CAICYT, dirigido por Nuria Sebastián Gallés. Los datos que aparecen aquí proceden de un recuento realizado por mí sobre el resultado de la lematización automática realizada en junio de 1998 por el grupo de la Universidad de Barcelona dirigido por María Antònia Martí. Agradezco a la Dra. Martí y a todo el equipo que participa en el proyecto la generosa cesión del corpus lematizado y anotado morfosintácticamente. Para más datos sobre las características del corpus y el proceso de anotación y desambiguación llevado a cabo, vid. Atserias et. al. (1998) y Carmona et al. (1998). Para los textos que componen el corpus, consúltese la dirección <http://www.uniovi.es/UniOvi/Apartados/Departamento/Psicología/metodos/soft/corpus>.

12. En Juilland y Chang (1964), *haber* aparece con 6.378 casos, de los que únicamente 1.303 son no auxiliares. De tener en cuenta solo estos casos, pasaría al puesto octavo de la lista, bastante próximo al que ocupa en la BDS. Algo parecido sucede en el corpus LEXEsp.

13. La posición ocupada por *creer* en LEXEsp es, sin duda, un error de adscripción de formas a lemas y, por tanto, no debe ser tenido en cuenta.

homogeneizar criterios de clasificación en un terreno tan resbaladizo, preferimos comenzar agrupando todos estos casos en una categoría única y dejar para una revisión unitaria la tarea de añadir la otra caracterización¹⁴.

Como la BDS está todavía en fase final de corrección, no es posible dar datos seguros del número de esquemas registrados, pero sí puede indicarse que hemos documentado 113 esquemas sintácticos (en el sentido que esta expresión tiene en la BDS, esto es, incluyendo la construcción como uno de los factores que intervienen) con frecuencia igual o superior a cinco. Una estimación prudente puede situarnos alrededor de los 150 esquemas registrados. Siempre, claro está, con el inventario de funciones y los criterios de identificación que nosotros hemos empleado.

No es este un tema suficientemente estudiado en español, de modo que puede resultar útil detenerse brevemente en el análisis de algunos de los datos, aunque sin olvidar nunca que estoy manejando una versión todavía provisional. En el cuadro 2 aparece la lista de los esquemas sintácticos más frecuentes en la BDS, que incluye todos los que presentan porcentajes de uso superiores al 1%, con indicación de su frecuencia, el porcentaje que supone sobre el total de casos fichados¹⁵, el número de verbos en los que han sido documentados y el porcentaje que esa cifra supone sobre el total de los verbos registrados en la BDS (3.554).

14. Esa tarea se está realizando ya en el momento de escribir este trabajo (diciembre de 1998). Afortunadamente, el diseño de la BDS es tan flexible que permite mantener los dos análisis. La complejidad que presenta este hecho se verá con claridad si se tiene en cuenta que no se trata de una simple subcategorización, sino de una clasificación de base diferente, que en muchos casos supone cambios en el número y tipo de los elementos funcionales de la cláusula en cuestión.

15. En su estado actual (diciembre de 1998), 161.662 casos. No coincide, claro, con el que se deduce de la tabla que contiene la frecuencia de los verbos puesto que en aquella, como ya he indicado, se ha recalculado la frecuencia de cinco verbos.

CUADRO 2

Esquemas sintácticos con un porcentaje de aparición en la BDS superior al 1%.

Datos provisionales ordenados según la frecuencia del esquema sintáctico.

En cursiva, los esquemas en los que la discrepancia entre las dos perspectivas reflejadas en el cuadro parece más fuerte

Construcción	Esquema funcional	Frecuencia del esquema sintáctico	Porcentaje del esquema sobre el total de ejemplos	Número de verbos que registran el esquema	Porcentaje sobre el total de verbos
activa	Sujeto-Predicado-CDirecto	64.638	40,20	2.434	70,50
activa	Sujeto-Predicado	16.819	10,50	1.168	33,80
<i>activa</i>	<i>Sujeto-Predicado-Predicativo</i>				
	<i>de Sujeto</i>	10.190	6,34	106	3,07
pronominal	Sujeto-Predicado	9.588	5,96	1.352	39,20
activa	Sujeto-Predicado-CDirecto- CIndirecto	8.987	5,59	593	17,20
activa	Sujeto-Predicado-CAdverbial	7.057	4,39	305	8,83
activa	Sujeto-Predicado-Suplemento	5.121	3,19	435	12,60
<i>pronominal</i>	<i>Sujeto-Predicado-Suplemento</i>	4.906	3,05	599	17,30
activa	Sujeto-Predicado-CIndirecto	4.467	2,78	267	7,73
activa	Sujeto-Predicado-CDirecto- Predicativo de CDirecto	3.700	2,30	114	3,30
pronominal	Sujeto-Predicado-CAdverbial	3.128	1,95	313	9,06
activa	Sujeto-Predicado-CDirecto- CAdverbial	3.032	1,89	273	7,91
<i>pronominal</i>	<i>Sujeto-Predicado-CDirecto</i>	3.013	1,87	426	12,30
pronominal	Sujeto-Predicado-Predicativo de Sujeto	2.777	1,73	136	3,94
<i>activa</i>	<i>Sujeto-Predicado-CDirecto- Suplemento</i>	1.995	1,24	367	10,60

Aunque no constituya una sorpresa inesperada, resulta de interés resaltar de nuevo la concentración de la mayor parte de los ejemplos en unos pocos esquemas sintácticos. El más frecuente, a gran distancia de los demás, es el formado por el esquema transitivo biargumental en construcción activa, que supone ya el 40% de los casos. La suma de los cinco esquemas más frecuentes nos lleva al 68,59%, lo cual significa que siete de cada diez cláusulas presentan uno de estos cinco esquemas. Por fin, los quince esquemas que figuran en el cuadro arrojan un porcentaje global del 92,98% del total de los esquemas, con lo que resulta evidente el escasísimo grado de empleo de los 125 ó 130 esquemas restantes.

En el cuadro 2 figuran dos tipos de datos igualmente interesantes: la frecuencia del esquema y el número de verbos en los que ese esquema ha sido registrado. El primero de ellos toma en consideración la frecuencia del verbo y

la frecuencia del esquema. El segundo contempla únicamente el número de verbos en los que ha sido registrado. En algunos casos, las dos perspectivas caminan en paralelo, como sucede con el primero de los esquemas: el esquema activo transitivo ha sido documentado en el 70% de los verbos que aparecen en la BDS. En otros, en cambio, las dos perspectivas proporcionan resultados divergentes. He marcado en cursiva los esquemas en los que la discrepancia entre ambas consideraciones es más fuerte. El caso más llamativo es, sin duda, la construcción activa con predicativo de sujeto. Está situado en tercer lugar si atendemos a la frecuencia del esquema (el 6,34% del total de los ejemplos¹⁶), pero se da únicamente en poco más del 3% de los verbos. Naturalmente, tiene que tratarse de verbos de alta frecuencia con un alto grado de especialización de uso por este esquema. Un análisis un poco más detenido de los datos numéricos muestra una realidad más extremada de lo que podríamos imaginar inicialmente: hay únicamente seis verbos (*andar, estar, parecer, permanecer, resultar y ser*) con frecuencia general superior a 25 en los que este esquema suponga un porcentaje igual o superior al 25% de los empleos registrados del verbo¹⁶, pero esos seis verbos suponen el 86,91% de todos los casos registrados de este esquema¹⁷. Si utilizáramos como criterio de ordenación el número de verbos en que aparece el esquema, la construcción activa formada por sujeto, predicado y predicativo del sujeto dejaría el tercer puesto y pasaría al último lugar de la tabla.

Desde otro punto de vista, solo cuatro de los quince esquemas más frecuentes son triargumentales y, de nuevo según lo previsto, el más frecuente de ellos es el que combina los complementos directo e indirecto que, además, aparece en un número bastante alto de verbos (casi seiscientos). Naturalmente, una buena parte de ellos no son típicamente triargumentales, sino transitivos biargumentales a los que se puede añadir un complemento indirecto (del tipo *alguien lava la cara a alguien*). Menos esperable es, me parece, el alto número de verbos con un esquema biargumental en el que el segundo complemento es un indirecto (*a alguien le gusta algo*).

Es evidente que datos como los reflejados en el cuadro 2 presentan solo una parte de la verdad en tanto que ocultan en sus agrupaciones la diferencia que existe entre los verbos que admiten un esquema de forma ocasional y aquellos que tienen ese esquema como su organización argumental preferida. Como hemos visto hace un momento para el esquema con complemento predicativo, la BDS permite, por supuesto, estudiar esas diferencias y situarlas adecuadamente. Para seguir con el ejemplo anterior, entre los 593 verbos que aceptan el esquema Sujeto - Predicado - Complemento directo - Complemento indirecto en

16. Son unos límites inferiores establecidos aquí de forma puramente tentativa. El de la frecuencia supone que se exige simplemente más del 0,0155% de los ejemplos, es decir, que el verbo aparezca una o más veces cada 6 500 cláusulas fichadas, que es bastante poco. Con los datos que tenemos en este momento, la media de esquemas por verbo es de 3,7, de modo que exigir el 25% de los casos equivale a tomar un verbo de distribución media y pedir que los esquemas que presenta se distribuyan de forma equivalente.

17. El hecho de que dos de estos verbos (*estar y ser*) se hayan visto afectados por el recorte de casos fichados (cf. supra, nota 8) parece de importancia para lo que aquí se discute.

construcción activa hay únicamente 47 con frecuencia superior a 25 en los que este esquema aparezca en más del 25% de los casos registrados y solo 14 en los que el esquema supere el 50% de los usos del verbo. Entre los 267 verbos registrados en construcción activa con el esquema Sujeto - Predicado - Complemento indirecto solo hay 7 (*apetecer, doler, encantar, extrañar, gustar, importare interesar*) con frecuencia general superior a 25 en los que este esquema suponga más del 50% de los casos fichados.

3.4. Los esquemas sintácticos de los verbos

Como ya he señalado, la estructura de la BDS se basa en la existencia de un registro que contiene los datos resultantes del análisis de cada una de las cláusulas presentes en el corpus y la actuación de una serie de programas que, a partir de estos datos primarios, van organizando la información de modo progresivamente elevado. Los datos que hemos comentado en los apartados anteriores suponen una perspectiva de conjunto que debe ser ahora complementada con aspectos más concretos.

Los elementos básicos son, pues, los verbos y los esquemas sintácticos (con todos los detalles necesarios) en que han sido registrados. Al almacenar y agrupar esos elementos en bases de datos derivadas, la posibilidad del doble enfoque es evidente. De un lado, tenemos la opción de estudiar qué esquemas han sido documentados y cuáles son los verbos que los utilizan, con indicación de sus frecuencias absolutas y relativas. De otro, podemos tomar cada uno de los verbos presentes en el corpus y estudiar los esquemas en que aparece, también con las frecuencias absoluta y relativa en cada uno de ellos.

Ya hemos visto en el apartado anterior algún ejemplo de hasta dónde puede llegar el análisis de la BDS en lo referente a los verbos documentados en un esquema. Lo más elemental es, por supuesto, obtener la lista de verbos que han sido registrados en un determinado esquema con la indicación de la frecuencia del esquema y lo que supone con respecto a los usos totales del verbo. Es lo que muestra, en una lista muy parcial que se limita a los veinte primeros casos, el cuadro 3 para el esquema Sujeto - Predicado - Complemento directo - Complemento indirecto. Es fácil detectar, de un simple vistazo, los diferentes pesos que este esquema posee en los verbos que figuran en la relación. El que tiene mayor número de casos, el esquema *alguien abre algo a alguien* no significa más que el 6,5% de los ejemplos de este verbo, que es básicamente biargumental (*alguien abre algo*). Suponen mucho más para el verbo los casos registrados en *adeudar* (*alguien adeuda algo a alguien*), que presenta en este esquema once de los doce ejemplos fichados en la BDS o los veintidós de *aconsejar* (*alguien aconseja algo a alguien*), que constituyen el 40% del total de las apariciones de este verbo.

CUADRO 3

Lista parcial de verbos documentados en la BDS con el esquema SPDI, con indicación de su frecuencia y el porcentaje sobre el total del verbo

Verbo	Frec.	% s/verbo
ABOLLAR	1	100,00
ABONAR	5	35,71
ABRASAR	2	20,00
ABRIR	43	6,51
ABROCHAR	3	20,00
ACARICIAR	19	19,79
ACARREAR	3	30,00
ACEPTAR	1	0,40
ACERAR	1	100,00
ACERCAR	2	0,50
ACERTAR	1	2,33
ACLARAR	18	23,08
ACONSEJAR	22	40,00
ACRECER	1	100,00
ACHACAR	1	12,50
ADELANTAR	4	6,56
ADEUDAR	11	91,67
ADIVINAR	2	2,50
ADJUDICAR	4	26,67
ADJUNTAR	1	33,33

Un esquema sintáctico, tal como se considera en la BDS, es una entidad bien caracterizada, que puede resultar demasiado específica para cierto tipo de ras-treos. Si se está interesado en, por ejemplo, los verbos que admiten un predica-tivo de complemento directo, es evidente que hay que aceptar la copresencia del complemento directo en los esquemas relevantes para la búsqueda corres-pondiente, pero puede resultar totalmente indiferente la presencia o ausencia de un complemento indirecto, un suplemento o, en muchos casos, el que la cons-trucción sea activa o pronominal. La BDS permite hacer búsquedas que se orga-nicen por una o más funciones con independencia del resto de los elementos concurrentes en los esquemas, lo cual proporciona gran libertad y soltura en las interrogaciones a la base.

En el cuadro 4 puede verse la lista parcial que responde a la interrogación sobre los verbos que han sido registrados en construcción activa y con un suplemento. La lista, parcial y fragmentada para que muestre casos suficientemente distintos entre sí, presenta diferentes combinaciones del suplemento con o sin complemento directo o indirecto, pero siempre en construcción activa. Naturalmente, los verbos aparecen tantas veces como esquemas en los que se encuentra la función buscada que poseen (véase, por ejemplo, el caso de *acostumbrar*). Como en el caso anterior,

se marca siempre la frecuencia absoluta y relativa, con lo que es fácil ver la importancia de la función (y el esquema en que aparece) en el conjunto de los casos de un verbo. *Abogar*, *acceder* o *acertar* son verbos básicamente biargumentales con clara tendencia a la combinación con un suplemento.

CUADRO 4

Lista (parcial) de verbos que han sido localizados en la BDS en esquemas que contienen un suplemento en construcción activa

Verbo	Esquema	Frec.	% s/verbo
ABANDONAR	SD SP	2	1,02
ABARCAR	S SP	1	2,70
ABDICAR	S SP	1	100,00
ABJURAR	S SP	1	100,00
ABOGAR	S SP	8	88,89
ABRUMAR	SD SP	2	13,33
ABSOLVER	SD SP	2	20,00
ABUNDAR	S SP	4	20,00
ABUSAR	S SP	9	75,00
ACABAR	S SP	46	20,00
ACCEDER	S SP	18	47,37
ACEPTAR	SD SP	1	0,40
ACERTAR	S SP	25	58,14
ACHACAR	SD SP	2	25,00
ACOMPAÑAR	SD SP	3	1,33
ACOMPASAR	SD SP	1	100,00
ACORDAR	S SP	1	0,36
ACOSTUMBRAR	S SP	5	8,93
ACOSTUMBRAR	SD SP	1	1,79

Con un planteamiento similar, resulta de gran interés el estudio de los verbos que han sido registrados en dos esquemas sintácticos determinados. En efecto, el estudio de las regularidades sintáctico-semánticas presentes en los verbos que muestran variaciones en su significado en paralelo con los cambios de esquema sintáctico se ve enormemente beneficiado con esta posibilidad de búsqueda. Por poner un ejemplo ilustrativo, el cuadro 5 muestra algunos de los verbos que han sido registrados en el esquema Sujeto - Predicado - Complemento directo en construcción transitiva y el esquema Sujeto - Predicado en construcción pronominal, esto es, el tipo de correspondencia *alguien abre la puerta / la puerta se abre* (que, por supuesto, no se da exactamente de este modo en todos los verbos que figuran en la relación). Como siempre, la indicación de las frecuencias da idea del peso de los esquemas en el conjunto del verbo.

CUADRO 5

Listá (parcial) de verbos que han sido documentados en la BDS en los esquemas sujeto - predicado - comp. directo en construcción activa y sujeto - predicado en construcción pronominal

ESQUEMAS:	Act. SD		CPm. S	
		% S/Verbo		% S/Verbo
ABANDONAR	171	86,80	3	1,52
ABANICAR	1	50,00	1	50,00
ABATIR	2	33,33	2	33,33
ABLANDAR	4	66,67	2	33,33
ABOLIR	6	75,00	1	12,50
ABORDAR	36	87,80	3	7,32
ABRASAR	4	40,00	1	10,00
ABRAZAR	80	69,57	17	14,78
ABRIGAR	10	76,92	3	23,08
ABRIR	372	56,28	104	15,73
ABROCHAR	6	40,00	1	6,67
ABURRIR	5	11,36	22	50,00
ACABAR	30	13,04	65	28,26
ACALLAR	7	70,00	3	30,00
ACARICIAR	61	63,54	1	1,04

La perspectiva complementaria consiste en considerar los datos tomando el verbo, no el esquema, como elemento orientador. Eso es lo que aparece en el cuadro 6 para el verbo *abandonar*. También aquí es fácil ver que *abandonar*, que tiene en la BDS un número elevado de ejemplos, muestra un esquema muy claramente dominante entre los diez que hemos podido documentar: el activo transitivo del tipo *alguien abandona a alguien/algo* (vid. infra, § 3.5. para la diferencia entre las dos posibilidades). En efecto, este esquema aparece en el 86,8% de todos los casos de este verbo, con lo que los nueve esquemas restantes tienen que repartirse el exiguo 13,2% que queda disponible.

CUADRO 6

Esquemas registrados en la BDS para el verbo abandonar,
con indicación de su frecuencia y porcentaje sobre el verbo

ABANDONAR		[197 casos; 10 esquemas]	
Act.	S	2	1,02 %
Act.	SD	171	86,80 %
Act.	SD AD	5	2,54 %
Act.	SD SP	2	1,02 %
CPrn.	S	3	1,52 %
CPrn.	S SP	8	4,06 %
Pas.	S	2	1,02 %
Pas.	S A	2	1,02 %
Pas.	S AD	1	0,51 %
Pas.	S SP	1	0,51 %

Una distribución distinta es la que muestra, con el mismo número de casos, el verbo *asomar*, que, aunque presenta también un esquema claramente más utilizado que los demás, deja ver una distribución bastante mejor repartida entre las construcciones en que ha sido registrado.

CUADRO 7

Esquemas registrados en la BDS para el verbo asomar,
con indicación de su frecuencia y porcentaje sobre el verbo

ASOMAR		[197 casos; 6 esquemas]	
Act.	S	45	22,84 %
Act.	S AD	10	5,08 %
Act.	SD	26	13,20 %
Act.	SD AD	1	0,51 %
CPrn.	S	37	18,78 %
CPrn.	S AD	78	39,59 %

La lista de los verbos contenidos en la BDS con los esquemas sintácticos en que ha sido documentado cada uno de ellos y la indicación de sus frecuencias constituye el equivalente del *Diccionario de construcciones verbales* a que me he referido en algunas ocasiones (cf., por ejemplo, Rojo: 1992 y 1994). Incluso en esta forma tan esquemática, constituye un recurso de utilidad inestimable, que hemos aplicado ya con excelentes resultados para la complementación automática de información sintáctica de los verbos en el lexicon asociado a una gramática formal (cf. Álvarez, Alvarino, Santalla y Sotelo: 1998).

3.5. Los subesquemas sintácticos

Pero la BDS posee información mucho más detallada sobre cada uno de los verbos que figuran en ella. Como ya he indicado, la ficha básica contiene la clase de unidad a que pertenece cada uno de los elementos funcionales existentes en una cláusula, tomada de un inventario de 53 posibilidades distintas (vid. tabla 3 del apéndice). Naturalmente, para el estudio del régimen de los verbos tal grado de detalle produce una incómoda fragmentación de la información, de modo que, manteniendo la indicación específica en la ficha básica, el subesquema se atribuye mediante la agrupación de las posibilidades iniciales en grandes bloques (frase nominal de carácter animado, frase nominal de carácter inanimado, cláusula introducida por *que* + indicativo, cláusula de infinitivo, etc.). El subesquema, pues, consiste en la adición al esquema de las características semántico-sintácticas que se consideran habitualmente relevantes para cada uno de los elementos funcionales que componen la cláusula.

Es posible, por tanto, obtener la lista de verbos que han sido localizados en construcción transitiva con el esquema Sujeto - Complemento directo - Complemento indirecto (vid. cuadro 3), pero añadiendo ahora la indicación de que el sujeto y el complemento directo deben ser animados, mientras que la función de complemento directo debe ser realizada por una cláusula completiva con la conjunción *que* y en modo indicativo. Los resultados son del estilo de los que aparecen en el cuadro 8.

CUADRO 8

Relación (parcial) de verbos presentes en la BDS que admiten el esquema SPDI en construcción activa, con sujeto e indirecto animados y directo en forma de cláusula con *que* y modo indicativo

SUBESQUEMA:	Act. San Dcqñlan	
	F %	S/Esq.
ACLARAR	5	27,78
ACONSEJAR	1	4,55
ADELANTAR	1	25,00
ADVERTIR	16	51,61
ANUNCIAR	9	56,25
ASEGURAR	44	61,11
AVISAR	3	37,50

Como en los casos anteriores, se ve inmediatamente la importancia relativa de este subesquema para cada uno de los verbos, pero ahora, sobre todo, aparece la posibilidad de contrastar la lista de verbos que admiten un complemento directo en forma de cláusula de indicativo con la de los que han sido registrados con un complemento directo en subjuntivo, que es lo que muestra el cuadro 9.

CUADRO 9

Relación (parcial) de verbos presentes en la BDS que admiten el esquema SPDI en construcción activa, con sujeto e indirecto animados y directo en forma de cláusula con *que* y modo subjuntivo

SUBESQUEMA:	Act. San DcqsIan	
	F	% S/Esq.
ACONSEJAR	8	36,36
ADVERTIR	2	6,45
AGRADECER	7	14,00
ASEGURAR	3	4,23
AVISAR	1	12,50

Es fácil observar que solo *agradecer* aparece únicamente en la lista de los verbos que pueden tener completiva en subjuntivo. Dado que las listas reales, que trabajan sobre la totalidad de los ejemplos no suelen ser tan reducidas, la BDS proporciona también la posibilidad de obtener la relación de verbos que pueden aparecer en dos subesquemas determinados, que es lo que aparece en el cuadro 10 para los dos que acabamos de tomar en consideración.

CUADRO 10

Relación (parcial) de verbos registrados en la BDS que admiten el esquema SPDI en construcción activa, con sujeto e indirecto animados y directo en forma de cláusula con *que* y modo indicativo o subjuntivo

SUBESQUEMAS:	Act. San DcqiIan		Act. San DcqsIan	
	F	% S/Esq.	F	% S/Esq.
ACONSEJAR	1	4,55	8	36,36
ADVERTIR	16	51,61	2	6,45
ASEGURAR	44	61,11	3	4,23
AVISAR	3	37,50	1	12,50

Una lista de este tipo muestra no solo la existencia de las dos posibilidades, sino también, en la medida en que la BDS es representativa del español actual, la importancia relativa de cada una de ellas, puesto que se nos ofrece el porcentaje de lo que cada subesquema supone con relación al esquema (en este caso, Sujeto - Predicado - Complemento directo - Complemento indirecto). Se ve de este modo que el indicativo en la completiva es la opción claramente dominante en *advertir* y *asegurar*, mientras que tiene muy poco peso en *aconsejar*¹⁸.

18. El único ejemplo de este tipo registrado en la BDS procede de la obra de Carmen Martín Gaité *Los usos amorosos de la posguerra española* (Barcelona, Anagrama, 8ª ed., 1988):

Por otro lado, la BDS contiene los datos generales de cada una de las cláusulas examinadas (la polaridad, por ejemplo), de modo que es relativamente simple analizar automáticamente alguno de los factores que puedan causar esta diferencia en el modo de la cláusula dependiente.

Desde la otra perspectiva, como hemos visto ya para los esquemas, la BDS proporciona todos los subesquemas en que ha sido registrado cada uno de los verbos que contiene. El cuadro 11 muestra la salida correspondiente al verbo *abandonar*, cuya organización en esquemas sintácticos hemos visto ya en el cuadro 6.

CUADRO 11

Relación de esquemas y subesquemas del verbo *abandonar*

ABANDONAR		[197 casos; 10 esquemas; 18 subesquemas]
Act.	S	(F = 2; 1,02% s/verbo)
San (F = 2; 100,00% s/esq.)		
Act.	SD	(F = 171; 86,80% s/verbo)
San Dan (F = 32; 18,71% s/esq.)		
San Din (F = 122; 71,35% s/esq.)		
Sin Dan (F = 11; 6,43% s/esq.)		
Sin Din (F = 6; 3,51% s/esq.)		
Act.	SD AD	(F = 5; 2,54% s/verbo)
San Dan ADin (<i>en</i>) (F = 1; 20,00% s/esq.)		
San Din ADin (<i>en</i>) (F = 3; 60,00% s/esq.)		
San Din ADin (<i>sobre</i>) (F = 1; 20,00% s/esq.)		
Act.	SD SP	(F = 2; 1,02% s/verbo)
San Dan SPin (<i>a</i>) (F = 1; 50,00% s/esq.)		
San Din SPin (<i>a</i>) (F = 1; 50,00% s/esq.)		
CPm.	S	(F = 3; 1,52% s/verbo)
San (F = 1; 33,33% s/esq.)		
Sin (F = 2; 66,67% s/esq.)		
CPm.	S	SP (F = 8; 4,06% s/verbo)
San SPcf (<i>a</i>) (F = 1; 12,50% s/esq.)		
San SPin (<i>a</i>) (F = 7; 87,50% s/esq.)		

«Recuerdo haber escuchado a cierto profesor de Formación Política [...] aconsejarnos en uno de sus discursos que si nos decían algún piro por la calle, *no debíamos limitarnos* a callar o a apretar el paso con apuro, porque eso era anticuado» (p. 65).

Nótese que la aparición del indicativo en el condicionado de la oración incrustada es posible por la presencia del verbo modal: *nos aconsejó que no debíamos limitarnos, nos aconsejó que nos limitáramos, *nos aconsejó que nos limitábamos*. No es fácil detectar la existencia de estas posibilidades si no existe el propósito explícito de analizar todos los ejemplos presentes en un corpus textual.

Pas.	S	(F = 2; 1.02% s/verbo)
Sin (F = 2; 100,00% s/esq.)		
Pas.	S A	(F = 2; 1,02% s/verbo)
San Aan (F = 2; 100,00% s/esq.)		
Pas.	S AD	(F = 1; 0,51% s/verbo)
San ADin (<i>en</i>) (F = 1; 100,00% s/esq.)		
Pas.	S SP	(F = 1; 0,51% s/verbo)
Sin SPin (<i>a</i>) (F = 1; 100,00% s/esq.)		

Además de los datos generales, en esta presentación figuran los esquemas en que aparece cada verbo, con su frecuencia y el porcentaje que suponen con respecto a la totalidad del verbo y, en cada uno de ellos, los diferentes subesquemas registrados, también con indicación de su frecuencia y el porcentaje que suponen con respecto al esquema¹⁹.

En un cuadro como el anterior, que sale directamente de los datos contenidos en la BDS, encontramos toda la información referente a las combinaciones sintácticas en que entra el verbo *abandonar* y, por tanto, estamos ante lo que puede considerarse ya una entrada bastante detallada de un auténtico diccionario de construcciones verbales o un diccionario sintáctico del estilo de los elaborados para el portugués por Busse (1994) o Luft (1995). En efecto, además de los datos generales, el cuadro 11 contiene información sobre los esquemas y subesquemas en que entra *abandonar*, diferenciando en las funciones que lo permiten entre los casos en que esa función es desempeñada por un elemento nominal animado, un nominal inanimado, una cláusula de infinitivo, etc. Puede verse también, en los complementos preposicionales o adverbiales, la preposición que los rige o introduce, de modo que sabemos que hay una construcción del tipo *alguien abandona algo*, otra del tipo *alguien abandona algo en/sobre algún lugar*, una tercera como *alguien se abandona a algo*, etc.

La lista de los esquemas y subesquemas en que ha sido registrado un verbo, que en muchas ocasiones resulta difícil de asociar con construcciones reales, resulta mucho más cómoda y atractiva si añadimos uno o más ejemplos representativos de cada una de las construcciones, al estilo de lo que figura en obras como las de Busse (1994) o Luft (1995). Para que tenga sentido en una base de datos que deriva de un corpus, los ejemplos han de ser reales (lo que no equivale forzosamente a identidad textual) y, por tanto, requiere el examen de los ejemplos registrados en cada subesquema, para poder incorporar ejemplos realmente representativos de lo que es el verbo en cada una de las construcciones en que aparece. Todo eso es lo que figura en el cuadro 12 para este mismo verbo.

19. Las abreviaturas utilizadas son fácilmente interpretables. *SP* es suplemento, *AD* complemento adverbial y *A* significa complemento agente. Las indicaciones de los subesquemas se refieren al carácter animado (*an*) o inanimado (*in*) de los elementos; *cf* quiere decir cláusula de infinitivo.

La diferencia más clara entre el cuadro 12 y la entrada de un diccionario sintáctico del estilo de los mencionados deriva directamente del hecho de que la BDS se ha construido mediante el análisis exhaustivo (salvo en el caso de ciertos verbos en los que se tomaba uno de cada cinco ejemplos) de todas las cláusulas contenidas en un corpus. Es, naturalmente, la información acerca de la frecuencia de cada uno de los esquemas y subesquemas que han sido documentados para cada verbo. Ciertamente, el trabajar exclusivamente a partir de lo que se encuentra en un corpus tiene también ciertas desventajas, especialmente en los verbos de baja frecuencia, en los que la aparición de un cierto esquema en el único ejemplo que figura en el corpus no puede ser considerada como realmente representativa del comportamiento de ese verbo. No obstante, a partir de un cierto grado de frecuencia los inconvenientes se diluyen y aparecen en cambio las indudables ventajas de poder observar los usos del verbo en el sentido más estricto del término. La BDS permite aquilatar al máximo las características construccionales de los verbos más frecuentes del español contemporáneo²⁰.

CUADRO 12

Esquemas y subesquemas del verbo *abandonar*

ABANDONAR		[197 casos; 10 esquemas; 18 subesquemas]
Act.	S	(F = 2; 1,02% s/verbo)
San		(F = 2; 100,00% s/esq.)
<i>El boxeador abandonó al tercer asalto.</i>		
Act.	SD	(F = 171; 86,80% s/verbo)
San Dan		(F = 32; 18,71% s/esq.)
<i>El padre abandonó a sus hijos.</i>		
San Din		(F = 122; 71,35% s/esq.)
<i>Julián abandonó los estudios muy pronto.</i>		
<i>Los viajeros abandonaron la ciudad al día siguiente.</i>		
Sin Dan		(F = 11; 6,43% s/esq.)
<i>Los espasmos no han abandonado al paciente.</i>		
Sin Din		(F = 6; 3,51% s/esq.)
<i>La nube abandonó la montaña.</i>		
Act.	SD AD	(F = 5; 2,54% s/verbo)
San Dan ADin (en)		(F = 1; 20,00% s/esq.)
<i>La mujer abandonó al niño en un descampado.</i>		
San Din ADin (en)		(F = 3; 60,00% s/esq.)
<i>El niño abandona los juguetes en cualquier sitio.</i>		
San Din ADin (sobre)		(F = 1; 20,00% s/esq.)
<i>Néstor abandonó suavemente el libro sobre la mesa.</i>		

20. El ejemplo ilustrativo del subesquema que consta de sujeto y suplemento en forma de cláusula de infinitivo en construcción pronominal (*no le resulta fácil abandonarse a ser uno más*) constituye, sin duda, un uso muy poco común, que probablemente necesita más atención de la que se

Act.	SD SP	(F = 2; 1,02% s/verbo)
San Dan SPin (a)		(F = 1; 50,00% s/esq.)
<i>Me abandonaste a la duda sobre tu identidad.</i>		
San Din SPin (a)		(F = 1; 50,00% s/esq.)
<i>Los fugitivos abandonaron sus bienes a la furia de los enemigos.</i>		
Cprn.	S	(F = 3; 1,52% s/verbo)
San		(F = 1; 33,33% s/esq.)
<i>[La envolvió en caricias y] Flora se abandonó con un primer llanto.</i>		
Sin		(F = 2; 66,66% s/esq.)
<i>[Aumentó el cultivo de la coca y] los demás cultivos se abandonaron.</i>		
CPrn.	S SP	(F = 8; 4,06% s/verbo)
San SPcf (a)		(F = 1; 12,50% s/esq.)
<i>No le resulta fácil abandonarse a ser uno más.</i>		
San SPin (a)		(F = 7; 87,50% s/esq.)
<i>La mujer se abandonaba a un fatalismo absurdo.</i>		
Pasiva	S	(F = 2; 1,02% s/verbo)
Sin		(F = 2; 100,00% s/esq.)
<i>Los paquetes confiscados fueron abandonados en el puerto.</i>		
Pasiva	S A	(F = 2; 1,02% s/verbo)
San Aan		(F = 2; 100,00% s/esq.)
<i>A veces, los enfermos son abandonados por sus familias.</i>		
Pasiva	S AD	(F = 1; 0,51% s/verbo)
Sin ADin (en)		(F = 1; 100,00% s/esq.)
<i>La víctima fue abandonada en un descampado.</i>		
Pasiva	S SP	(F = 1; 0,51% s/verbo)
Sin SPin (a)		(F = 1; 100,00% s/esq.)
<i>El edificio fue abandonado a su suerte.</i>		

le puede prestar cuando solo se trata de ilustrar una aparición entre las dieciocho posibilidades construccionales registradas en el verbo *abandonar*. El ejemplo que figura en el cuadro está inspirado en el único que aparece en la BDS, procedente de la obra de Rafael Sánchez Ferlosio *La homilía del ratón* (Madrid, Ed. El País, 1986):

«Le sabe a poco (a tan poco, digamos, como al Barça le sabría no ser más que un equipo) ser español tan sólo en un sentido tan inerte, tan llano y contingente como el que la convención del uso más común y cotidiano parece suponer. O tal vez no le inspira confianza alguna *abandonarse a serlo* con arreglo a la libre decisión del predicado, dejando su españolez a merced de la incontrollable circunstancia de que el asémico ser copulativo del gramático relegue enteramente al solo predicado la responsabilidad de decidir, por su propia y exclusiva virtualidad significativa, el más fuerte o más débil compromiso de pregnancy conforme al cual el sujeto ha de sentirse alcanzado y afectado por la predicación». (p. 45).

Se refiere a lo implicado por la expresión *rabiosamente español* y la longitud de la cita, necesaria para la comprensión correcta del texto y el valor que en él presenta el verbo, ilustra las dificultades que plantea la inclusión de ejemplos *reales*, que, naturalmente, no pueden consistir sistemáticamente en citas textuales.

4. LA INFORMACIÓN SEMÁNTICA

La enorme riqueza de información que contiene la BDS presenta, sin embargo, una ausencia clara: la información semántica. Es claro que la visión real del comportamiento sintáctico de un verbo aparece cuando somos capaces de poner en relación los cambios en el esquema o subesquema sintáctico utilizado con las diferencias en el significado. Por poner referentes claros, es el salto que va desde un diccionario de construcciones verbales como el de Busse (1994) a un diccionario como el de Cuervo o el de Borba (1991), en el que se establece la relación entre las diferentes acepciones que tiene un verbo y los esquemas y subesquemas sintácticos capaces de expresarlos.

Está fuera de duda que el trabajo de añadir la acepción correspondiente a cada ejemplo no podía ser realizado al tiempo de hacer el análisis sintáctico. Por una parte, habría supuesto una fuerte distorsión del proceso de fichado, al obligar a quien lo llevaba a cabo a enfrentarse con fenómenos muy distintos, aunque evidentemente conectados. Por otra, habría retrasado considerablemente el período de análisis, ya muy largo en la parte estrictamente sintáctica. Decidimos, por tanto, dejar el estudio del significado para un momento posterior, un momento en que ya se pudiera trabajar con todos los ejemplos fichados para cada verbo, que parecía una condición imprescindible para la distribución adecuada de los valores de cada ejemplo. La estructura de la base de datos diseñada era, por otro lado, suficientemente flexible como para poder añadir nuevos campos a la tabla central en cualquier momento.

En este momento, cuando la revisión de los análisis sintácticos está a punto de finalizar, todo indica que la estrategia decidida fue plenamente acertada. Hemos acometido la tarea de asignar a cada ejemplo la acepción que le corresponde, pero no ha sido posible ir más allá de unos cuantos verbos porque el trabajo que supone es inmenso y necesita como mínimo tanto tiempo como la parte sintáctica.

No es este el lugar de desarrollar esta cuestión, de forma que me limitaré a indicar que, para decirlo en pocas palabras, llevar a cabo ese trabajo implica forzosamente reorganizar por completo la entrada del diccionario. La solución simple de tomar un determinado diccionario aceptándolo como algo que nos es dado y que no vamos a discutir ni cuestionar e ir asignando a cada ejemplo la acepción que le corresponde en el diccionario seleccionado es inviable. Los diccionarios de español tienen características muy distintas y parten de presupuestos diferentes, pero, salvo el de Cuervo, tienen en común una notable falta de atención a los aspectos sintácticos que, sin embargo, muestran a poco que se investigue fuertes correlaciones con los semánticos. Para no citar más que un caso claro, las tan habituales indicaciones *úsase también como pronominal* o *úsase más como pronominal* o similares para un verbo caracterizado inicialmente como transitivo o la caracterización *transitivo y pronominal* que se emplea en otros diccionarios puede ocultar diferencias de significado tan fuertes como las que se dan entre *dormir* y *dormirse* o *enojar* y *enojarse* (cfr. Ahumada: 1987, 171-178; Rojo: 1997, 40-41). Las distorsiones que estos desajustes producen en un diccionario convencional se harían insoportables en un diccionario sintáctico, de modo que

no es posible utilizar alguno de los existentes como marco estructurador para el significado y, en consecuencia, es necesario elaborar la entrada por completo. Todo ello supone, de otra parte, el problema de cómo representar adecuadamente la información sintáctica en un diccionario, tema que se ha discutido en varias ocasiones en los últimos años y en el que, lógicamente, tampoco podemos entrar aquí²¹.

La información que ya contiene la BDS, el modo en que está organizada y la posibilidad de utilizar de nuevo procedimientos informáticos en el reanálisis de los registros trazan una vía ciertamente interesante para la elaboración de un diccionario sintáctico completo. Hemos preparado ya algunos verbos para comprobar las posibilidades de este acercamiento y hemos podido comprobar que los resultados son realmente interesantes. La primera parte del trabajo consiste en identificar y organizar las acepciones y subacepciones del verbo. Se pasa luego a analizar cada uno de los ejemplos de ese verbo para incorporarle la acepción que le corresponde. Es un proceso que lleva habitualmente a reconsiderar la organización inicial y, en consecuencia, a rehacer en ciertos puntos la organización de las acepciones. Una vez terminada la asignación de acepciones a cada uno de los ejemplos registrados, un nuevo conjunto de programas se encarga de hacer los recuentos necesarios, de modo que, al final, podemos disponer de la frecuencia de cada una de las acepciones, la distribución de esquemas y subesquemas sintácticos por acepciones, con la frecuencia absoluta y relativa de cada uno, y la distribución de acepciones por esquemas y subesquemas, de nuevo con las frecuencias de cada uno.

Las ventajas que supondrá disponer de una entrada de este tipo son evidentes y proceden por un lado del hecho de que la BDS es un recurso computacional y por otro de que deriva del análisis exhaustivo del corpus. Del primer factor obtenemos la posibilidad de presentar la doble organización de la información, imposible de conseguir por los procedimientos tradicionales. Gracias a ello, podemos ver con total claridad la vinculación existente entre las acepciones y los subesquemas sintácticos en cualquiera de las dos direcciones. Del segundo rasgo viene la posibilidad de disponer de estadísticas fiables, no ya para los esquemas y subesquemas, sino también para las acepciones y, por supuesto, para cada una de las asociaciones que estos dos caracteres establecen entre sí.

BIBLIOGRAFÍA

- Ahumada Lara, Ignacio (1987): «La definición verbal y régimen lexemático: su indicación formal en la lexicografía hispánica», en *Amistad a lo largo. Estudios en memoria de Julio Fernández Sevilla y Nicolás Marín López*, Universidad de Granada, 1987, 13-25.
 Alameda, José Ramón y Fernando Cuétos: *Diccionario de frecuencias de las unidades lingüísticas del castellano*, Univ. de Oviedo, 1995, 2 vols.

21. En la Lexicografía española se ha tratado con cierta detención el tema del llamado *contorno* de la definición. Vid., entre otros, Seco (1979), Porto (1988), Ahumada (1989), Porto (1997), Rojo (1997).

- Álvarez Lebreo, Conchita, Pilar Alvarino, Adelaida Gil, Teresa Romero, Paula Santalla y Susana Sotelo (1998): «Avalon, una gramática formal basada en corpus», en *Procesamiento del Lenguaje Natural*, 23, 1998, 132-139.
- Atserias, J., J. Carmona, I. Castellón et alii (1998): «Morphosyntactic Analysis and Parsing of Unrestricted Spanish Text», en *Proceedings of the 1st International Conference on Language Resources and Evaluation (LREC'98)*, Granada. El texto puede conseguirse también en <http://www.ub.es/ling/labcat.htm>.
- Borba, Francisco da Silva (coord.) (1991): *Dicionário gramatical de verbos do Português Contemporâneo do Brasil*, São Paulo, Unesp, 1991².
- Busse, Winfried (coord.) (1994): *Dicionário sintático de verbos portugueses*, Coimbra, Almedina, 1994.
- Carmona, J. Cervell, L. Márquez et alii (1998): «An Environment for Morphosyntactic Processing of Unrestricted Spanish Text», en *Proceedings of the 1st International Conference on Language Resources and Evaluation (LREC'98)*, Granada. El texto puede conseguirse también en <http://www.ub.es/ling/labcat.htm>.
- Corbella, Dolores (1987): «Algunos datos-estadísticos del paradigma verbal español», en *In Memoriam Inmaculada Corrales*, Univ. de La Laguna, 2 vols., I, 1987, 145-159.
- Garside, Roger, G. Leech and A. McEnery (eds.) (1997): *Corpus Annotation. Linguistic Information from Computer Text Corpora*, Londres, Longman, 1997.
- Juilland, Alphonse and E. Chang-Rodríguez (1964): *Frecuency Dictionary of Spanish Words*, La Haya, Mouton, 1964.
- Leech, Geoffrey and Elisabeth Eyes (1997): «Syntactic Annotation: Treebanks», en Garside, Leech and McEnery (1997).
- López Meirama, Belén (1997): *La posición del sujeto en la cláusula monoactancial en español*, Universidade de Santiago de Compostela, 1997.
- Luft, Celso Pedro (1995): *Dicionário prático de regência verbal*, São Paulo, Ática, 1995.
- Patterson, William and Héctor Urrutibéheity: *The Lexical Structure of Spanish*, La Haya, Mouton, 1975.
- Porto Dapena, José Álvaro (1988): «Notas lexicográficas: La información sintáctica en los diccionarios comunes», *Linguística española actual*, 10, 1988, 133-151.
- (1997): «Algunas observaciones sobre el contorno de la definición lexicográfica», en Almeida, Manuel y Josefa Dorta (eds.): *Contribuciones al estudio de la Lingüística hispánica*, Homenaje al Profesor Ramón Trujillo, 2 tomos, Montesinos, Tenerife, 1995, I, 211-226.
- Rojo, Guillermo (1992): «El futuro *Diccionario de construcciones verbales del español actual*», en Martín Vide, Carlos (ed.): *Actas del VIII Congreso de Lenguajes naturales y lenguajes formales*, Univ. de Barcelona, 1992, 41-50.
- (1994): «Problemas lingüísticos e informáticos en los diccionarios de construcción y régimen», en *Actas del Congreso de la Lengua Española* (Sevilla, 7 al 10 de octubre de 1992), Madrid, Instituto Cervantes, 1994, 307-315.
- (1995): «La base de datos sintácticos del español actual», *Español actual*, 59, 1993, 15-20 (pero 1995).
- Rojo, Guillermo (1997): «Gramática y Diccionario», en González Calvo, José Manuel y Jesús Terrón González (eds.): *IV Jornadas de Metodología y Didáctica de la Lengua española*, Univ. de Cáceres, 1997, 33-48.
- Seco, Manuel (1979): «El 'contorno' en la definición lexicográfica», en *Homenaje a Samuel Gili Gaya (in memoriam)*, Barcelona, 1979, 183-191. Cito por su reed. con el título «El contorno en la definición» como cap. 2 de Seco, M.: *Estudios de Lexicografía española*, Madrid, Paraninfo, 1987, 35-45.
- Van Halteren, Hans (1997): *Excursion into Syntactic Databases*, Amsterdam, Rodopi, 1997.

APÉNDICES

TABLA 1

BDS: Datos generales de la cláusula

-
- > Verbo
 - > Referencia
 - > Tipo de cláusula
 - ◆ 11: Independiente
 - ◆ 12: Coordinada
 - ◆ 21: Cláusula de *que*
 - ◆ 22: Cláusula de infinitivo
 - ◆ 23: Otras completivas
 - ◆ 31: Cláusula relativa
 - ◆ 32: Nominalización de cláusula relativa
 - ◆ 4: Cláusula adverbial
 - ◆ 5: Cláusula de gerundio
 - ◆ 6: Cláusula de participio
 - ◆ 7: Cláusula constituyente de oración bipolar
 - ◆ 81: Cláusula no completiva precedida de otras conjunciones
 - ◆ 82: Cláusula con verbo en forma personal precedida de locución
 - ◆ 83: Cláusula de infinitivo precedida de locución
 - ◆ 9: Otras cláusulas
 - > Función de la cláusula
 - > Voz
 - > Polaridad
 - > Modalidad
 - > Perífrasis verbales
 - > Forma verbal
 - > Forma verbal de la cláusula dominante
 - > Número y persona
 - > Predicado complejo
 - > Número de argumentos
 - > Orden de elementos
-

TABLA 2

BDS: Datos de elementos funcionales

➤ SUJETO

- ◆ Tipo de sujeto
- ◆ Tipo de unidad
- ◆ Animación
- ◆ Determinación
- ◆ Número

➤ COMPLEMENTO DIRECTO

- ◆ Marca
- ◆ Tipo de unidad
- ◆ Animación
- ◆ Determinación
- ◆ Número
- ◆ Clítico
 - *me*
 - *nos*
 - *te*
 - *lo*
 - *la*
 - *le* (referente masculino)
 - *le* (referente femenino)
 - *le* (referente plural)
 - *lo* (tratamiento de usted)
 - ...
- ◆ [...]

➤ COMPLEMENTOS PREPOSICIONALES Y ADVERBIALES REGIDOS

- ◆ Tipo de complemento
 - Suplemento
 - Modal
 - Adverbial
 - Otros
 - ◆ Tipo de unidad
 - ◆ Animación
 - ◆ Determinación
 - ◆ Número
-

TABLA 3

BDS: Tipos y subtipos de unidades

10	Frase nominal
11	Demostrativos
12	Indefinidos y numerales
13	Cláusula de relativo nominalizada
14	Nominalización de adjetivo, posesivo, fprep, etc.
15	Posesivos
21	Pronombre personal de 1. ^a persona
22	Pronombre personal de 2. ^a persona
23	Pronombre personal de 3. ^a persona
...	
28	Pronombre <i>lo</i> reproductor de predicativos
...	
311	Relativo <i>que</i>
312	Relativo <i>quien</i>
...	
411	Cláusula de <i>que</i> + indicativo
412	Cláusula de <i>que</i> + subjuntivo
...	
51	Cláusula de infinitivo con sujeto recuperable del sujeto de la cláusula dominante
52	Cláusula de infinitivo con sujeto recuperable del complemento directo de la cláusula dominante
...	
611	Cláusula con <i>si</i> + indicativo
612	Cláusula con <i>si</i> + subjuntivo
...	
81	Cláusula relativa
91	Frase adjetiva
91	Frase adverbial
...	

