

Borrador final

Sobre la frecuencia de verbos y esquemas sintácticos

Guillermo Rojo

Universidade de Santiago de Compostela

Resumen

Haciendo uso de la enorme cantidad de información existente en la *Base de datos sintácticos del español* (BDS), el presente trabajo se propone analizar algunos de los muy diversos aspectos que surgen cuando se establece la relación entre la frecuencia de los esquemas sintácticos y la frecuencia de los verbos que se insertan en ellos. En el último apartado se estudian diversos procedimientos de cuantificación de esta relación y se concluye que las pruebas basadas en la aplicación del *collexeme analysis*, propuesto por Stefanowitsch y Gries, parecen dar los resultados más interesantes.

Palabras clave: BDS, esquema sintáctico, 'collexeme analysis', frecuencias en sintaxis.

Abstract

Using the great amount of information contained in the *Base de datos sintácticos del español* (BDS), the present paper analyzes some of the different aspects involved when the relation between the frequency of syntactic constructions and the frequency of verbs is established. In the last section, different methods for quantifying this relation are studied. The conclusion is that the results based in the collexeme analysis, proposed by Stefanowitsch and Gries, seem to be the most interesting of them all.

Keywords: BDS, syntactic construction, collexeme analysis, frequency in syntax.

1. La frecuencia de los esquemas sintácticos

Hasta no hace mucho tiempo, la inmensa mayoría de los estudios sobre frecuencia de elementos y fenómenos lingüísticos se centraba en el léxico, lo cual ha supuesto la necesidad de llevar a cabo ciertos reajustes teóricos en el momento en que se pretende trabajar sobre las frecuencias de los fenómenos sintácticos. En efecto, aunque es innegable que el léxico presenta un buen número de problemas que es necesario resolver antes de proceder a las operaciones de simple recuento, los elementos sintácticos resultan considerablemente más complicados, de modo que la pura transferencia mecánica de técnicas y conceptos no suele dar buenos resultados. Los esquemas sintácticos son mucho más abstractos, requieren la identificación de los elementos con los que operan (básicamente, las funciones sintácticas) y exigen, además, la delimitación entre elementos valenciales y no valenciales (o argumentales y no argumentales). Todos esos rasgos introducen complicaciones adicionales que, como es lógico, se multiplican si queremos incorporar a nuestros análisis los esquemas semánticos.

En un trabajo reciente (cf. Rojo 2010), he tratado de demostrar que la comprensión de la importancia de la frecuencia en sintaxis requiere el manejo de la distinción entre la frecuencia de inventario y la frecuencia de uso por una parte y la consideración amplia de los conceptos de elemento y clase (de elementos) por otra. La frecuencia de inventario es la que corresponde al número de elementos que compone una clase gramatical: número de verbos que aparecen en un

diccionario o en el leuario extraído de un corpus, número de verbos que pueden presentar un esquema sintáctico determinado, número de construcciones sintácticas que, en una lengua determinada, pueden entrar en la consideración de cláusulas biargumentales, etc. La frecuencia de uso se refiere a la que los elementos presentan en un conjunto determinado de textos: número de casos de sustantivos que se encuentran en un corpus, total de casos que un cierto verbo presenta en un esquema sintáctico determinado, total de casos de un cierto esquema sintáctico, etc. Los ejemplos utilizados pretenden mostrar también el juego entre elementos y clases, de modo que podemos ocuparnos, en pasos sucesivos, de la consideración de la frecuencia de uso de las condicionales irreales de pasado con el esquema *si hubiera tenido, habría dado*, de las oraciones condicionales irreales de pasado, de las oraciones condicionales o, más en general, de las oraciones bipolares (cf. Rojo: 2010).

En el primer trabajo que dediqué a la frecuencia de los esquemas sintácticos (Rojo 2003) apliqué ya, aunque sin emplear todavía estas denominaciones, la distinción entre frecuencia de inventario y frecuencia de uso y traté de mostrar algunas de sus características más señaladas. Por citar únicamente un caso claro, se menciona allí la importante diferencia de configuración que existe entre el esquema activa + sujeto + CD, que aparece en el 39,06% de las cláusulas de la BDS¹ y se documenta en el 70,44% de los verbos registrados, y el esquema activa + sujeto + predicativo de sujeto, que supone el 6,34% de las cláusulas, pero es utilizado únicamente por el 1,83% de los verbos. Esta diferencia, reelaborada y ampliada en Rojo (2010), es válida para la consideración general de la distribución de los esquemas sintácticos en una fase determinada de una lengua (o de un cierto tipo de textos), pero no permite conocer adecuadamente la relación entre verbos concretos y esquemas sintácticos determinados. En línea con lo avanzado en Rojo (2010), me propongo aquí profundizar en el aspecto cuantitativo de esos vínculos. En el apartado 2 me ocupo de las características generales de la distribución de los esquemas sintácticos. En el apartado 3 exploro diferentes modos de poner de manifiesto las relaciones entre los esquemas sintácticos y los verbos que los utilizan en mayor medida.

2. Verbos y número de esquemas

La cara más general de toda esta cuestión es la constituida por la capacidad de los verbos para figurar en diferentes esquemas sintácticos y la consiguiente discrepancia entre los que tienden

¹ La *Base de datos sintácticos del español actual* (BDS), construida en la USC por nuestro grupo de investigación, contiene el análisis de las aproximadamente 160 000 cláusulas contenidas en un corpus de aproximadamente 1,5 millones de formas. Para detalles y recuperación de datos sobre verbos y esquemas sintácticos, cf. <http://www.bds.usc.es>.

a concentrar sus usos y aquellos con los distribuyen en un número alto de esquemas. La BDS, aunque extraída de un corpus de tamaño reducido y, como consecuencia de ello, con ciertas disfuncionalidades en los verbos de frecuencias bajas, contiene, sin embargo, un conjunto de datos de gran interés para el estudio de estas características.

Los 3427 verbos documentados en la BDS arrojan un total de 157 esquemas sintácticos distintos y 10 197 combinaciones de verbos y esquemas sintácticos,² lo cual supone que, por término medio, cada verbo aparece en 2,97 esquemas sintácticos diferentes. Como es de esperar, el análisis de lo que sucede realmente en la BDS produce un panorama bastante heterogéneo. De una parte, esos 157 esquemas sintácticos detectados tienen pesos muy distintos. El 17,2% de ellos presenta una única documentación (y algunos proceden, sin duda, de errores en la codificación utilizada). Según se señala en Rojo (2010), los 34 esquemas que alcanzan porcentajes de aparición superiores al 0,1% suponen en conjunto el 98,36% de todas las cláusulas analizadas, lo cual implica que los 122 esquemas restantes no llegan al 2% de la BDS.³ Por otro lado, las cifras más reducidas de esquemas por verbo están muy condicionadas por la gran cantidad de verbos de frecuencia baja (el 21,67% de los documentados en la BDS aparecen una sola vez). En el otro extremo, los verbos que aparecen en muchos esquemas, entre los que destacan *dar* (48), *pasar* (46), *hacer* (42) y *poner* (40). Conviene, pues, para hacerse una idea más adecuada de lo que sucede aquí añadir al valor de la media de esquemas por verbo (2,97, como hemos visto), el de la desviación estándar (3,29), la varianza (10,79) y, para poder comparaciones cómodas, el coeficiente de variabilidad (1,11). Los datos generales de distribución de número de esquemas son los que aparecen en el cuadro 1.

Número de esquemas	Número de verbos	% sobre el total de esquemas
1	1368	39,80
2	784	22,81
3	412	11,99
4 - 5	427	12,42
6 - 10	359	10,45
11 - 20	83	2,41
21 - 48	14	0,41
Totales	3447	100,29

Cuadro 1

Número y porcentaje de verbos documentados en la BDS en distintos números de esquemas. Fuente: BDS (<http://www.bds.usc.es>). Elaboración propia.

El predominio de los verbos que presentan uno o dos esquemas sintácticos es evidente (más del

² Tal como es entendido el término en la BDS, un esquema sintáctico está constituido por la unión de un valor de voz (activa, media, pasiva pronominal o pasiva perifrástica) y una serie de funciones sintácticas argumentales. Cf. Rojo 2003.

³ Cf. también Rojo 2003: 417 y sigs. Las cifras difieren ligeramente a causa de la utilización de una base de cálculo distinta, pero no hay cambios de interés en los aspectos mencionados aquí.

60% en conjunto), pero no puede olvidarse que una parte considerable de su peso procede de la baja frecuencia de muchos de ellos.

Un panorama bastante diferente aparece si tomamos de la BDS únicamente aquellos verbos que presentan un número de casos que dé garantías de representatividad individual. Para este trabajo, he establecido el corte en 15 o más casos, esto es, una frecuencia normalizada situada en torno a 10 casos por millón de formas, con lo que se seleccionan 1016 verbos.⁴ La media de esquemas por verbo se eleva ahora hasta 6,00, con una desviación estándar de 4,62 y una varianza de 21,32. El coeficiente de variabilidad, en cambio, se reduce a 0,77. Los datos correspondientes a este subconjunto aparecen en el cuadro 2.

Número de esquemas	Número de verbos	% sobre el total de esquemas
1	32	3,15
2	119	11,71
3	127	12,50
4 - 5	291	28,64
6 - 10	350	34,45
11 - 20	83	8,17
21 - 48	14	1,38
Totales	1016	100,00

Cuadro 2

Número y porcentaje de verbos con frecuencia normalizada igual o superior a 10 por millón documentados en la BDS en distintos números de esquemas. Fuente: BDS (<http://www.bds.usc.es>). Elaboración propia

Las diferencias con lo que muestra el cuadro 1 son bastante claras. La eliminación de los verbos con frecuencias bajas produce el aumento de la media de esquemas por verbo y, al tiempo, el aumento de la desviación estándar y la varianza y la esperable disminución del aumento en el coeficiente de variabilidad por el aumento de la media. La gran mayoría de los verbos se agrupa ahora en el centro de la tabla, es decir, presenta entre cuatro y diez esquemas. El gráfico 1 presenta las diferencias entre lo que sucede en la totalidad de la BDS y el subconjunto seleccionado (verbos con frecuencia normalizada igual o superior a 10 apariciones por millón de formas).

⁴ En términos generales, son los verbos correspondientes al subconjunto de los 5000 lemas más frecuentes o con mayores índices de uso. Entre los 5024 lemas con índices más altos del FDSW (Juilland & Chang 1964) hay 954 verbos. Entre los 5000 lemas más frecuentes del corpus *Cumbre* (Almela et alii 2005) aparecen 1074 verbos. Los 1016 con frecuencia normalizada igual o superior a 10 en la BDS se sitúan en el mismo ámbito de uso.

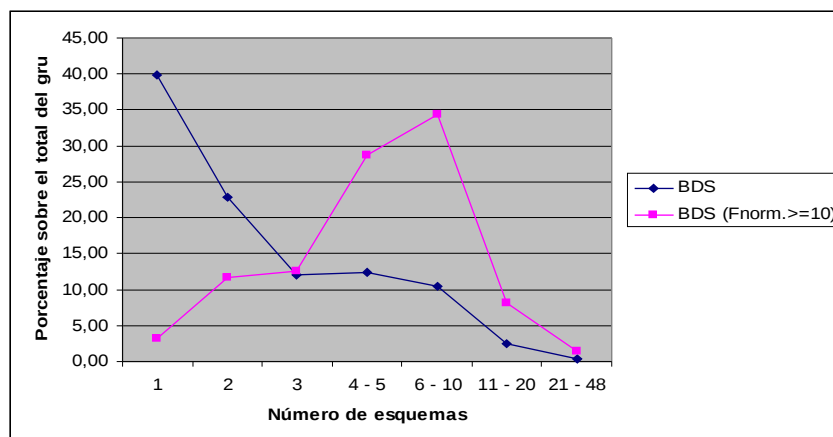


Gráfico 1

Porcentajes de los verbos en la BDS y en el subconjunto de los que tienen frecuencia normalizada igual o superior a 10 por millón según el número de esquemas en que se documentan.

Parece que el aumento en la frecuencia de uso de un verbo debería traducirse, en términos generales, en la ampliación del número de esquemas sintácticos en que podemos documentarlo, tanto por las características del verbo como por los factores puramente estadísticos. Sin que ello suponga negar esa tendencia, los datos obtenidos de la BDS muestran sin embargo la existencia de un número importante de verbos frecuentes con propensión a concentrar sus usos en un número relativamente reducido de esquemas. Como hemos visto, hay 32 verbos (el 3,15% del subconjunto de trabajo) que presenta un único esquema. No son verbos de frecuencias muy altas, pero destacan *consistir*, *carecer* o *residir*, todos ellos con frecuencias normalizadas superiores a 35 casos por millón. Si ampliamos a uno o dos esquemas, encontramos ya un total de 151 verbos (el 14,86%). Y si elevamos el tope superior a 3 aparecen 278 verbos, que suponen el 27,36% del total considerado. Esto supone que, en números redondos, un porcentaje próximo al 30% de los verbos de frecuencias altas concentran sus usos en un máximo de tres esquemas sintácticos.

El dato relevante no parece, de todas formas, el número de esquemas con que se haya podido documentar un verbo en la BDS. En efecto, un verbo que presente dos esquemas puede mostrar los ejemplos distribuidos al 50%, pero también puede utilizar uno de ellos en el 95% o el 98% de los casos. El perfil del verbo será distinto en cada una de esas posibilidades, de modo que resulta de interés tener una estimación del grado en el que los verbos tienden a concentrar sus empleos en pocos esquemas sintácticos. Los datos obtenidos para los verbos de la BDS con frecuencia normalizada igual o superior a 10 casos por millón son los que figuran en el cuadro 3.

Número de verbos que concentran en un esquema un porcentaje de sus usos igual o superior a	Número	Porcentaje	Frecuencia normalizada media (por millón)
100	32	3,14	19,25
95	65	6,39	295,01
90	111	10,92	215,32
75	310	30,51	116,98
50	731	71,94	79,24

Cuadro 3

Número y porcentaje de verbos con frecuencia normalizada mayor o igual a 10 casos por millón que concentran en un único esquema sintáctico la mayoría de sus usos documentados en la BDS. Fuente: BDS (<http://www.bds.usc.es>).
Elaboración propia

Aunque no aparecen cifras espectacularmente altas, creo de interés resaltar que los porcentajes de verbos con alto grado de concentración de usos en un esquema son importantes y que, además, se trata en muchos casos de verbos de frecuencia elevada, como muestran las medias que figuran en la última columna. De acuerdo con el cuadro 3, algo más del 10% de los verbos que estamos estudiando presentan el 90% o más de sus usos en un único esquema, dato que adquiere mayor importancia si observamos que, según la media de las frecuencias normalizadas (215,32 por millón), se trata de verbos con altos índices de aparición. Son datos esperables, pero es interesante poder cuantificar el grado de especialización constructiva de verbos como *haber* (el 99,25% de sus apariciones en la construcción impersonal que en la BDS se caracteriza como activa con un CD), *existir* (97,89% en el esquema activa+sujeto), *gustar* (el 97,44% en el esquema activa + sujeto + complemento indirecto), *ser* (el 95,33% en el esquema activa + sujeto + predicativo de sujeto), *necesitar* (92,01 en el esquema Activa + sujeto + complemento directo) o *querer* (90,38 en este mismo esquema). En términos generales, resulta que algo más del 70% de los verbos pertenecientes al subconjunto que hemos delimitado concentra un mínimo del 50% de sus apariciones en un único esquema sintáctico.

Por lo ya insinuado acerca de las distintas frecuencias de las construcciones, lo esperable es que no sean muchos los esquemas en los que estos verbos concentran sus usos. En efecto, de los 157 detectados en el subconjunto (en realidad, todos los que se dan en la BDS), únicamente en 23 de ellos (el 14,64% del total) aparece algún verbo que tenga en ese esquema un mínimo del 50% de sus usos documentados en la BDS. Si elevamos el porcentaje hasta el 75% para acercarnos a tasas que puedan mostrar esquemas realmente preferidos, el número se reduce a 18 esquemas (el 11,46%

de los documentados). Son los que aparecen en el cuadro 4.⁵

Voz	Esquema	Número de verbos	% s/verbos del subconjunto que documentan ese esquema	% s/verbos que concentran sus usos en un esquema sintáctico en %>=75
Activa	SD	213	24,01	68,70
Activa	S	34	5,63	10,96
Activa	S SP	17	8,01	0,32
Media	S SP	15	7,07	4,83
Activa	S AD	8	6,20	2,58
Activa	SDI	4	1,00	1,29
Activa	S I	4	2,32	1,29
Media	S	3	0,77	0,96
Media	S AD	3	2,75	0,96
Activa	D	1	33,33	0,32
Activa		1	16,66	0,32
Media	S MD	1	12,5	0,32
Media	S PS	1	2,70	0,32
Activa	S PS	1	1,81	0,32
Media	S PR	1	1,63	0,32
Activa	SD PD	1	1,16	0,32
Activa	SD AD	1	0,67	0,32
Activa	SD SP	1	0,54	0,32

Cuadro 4

Número de verbos que presentan un porcentaje de sus usos igual o superior al 75% en los esquemas sintácticos mencionados. Fuente: BDS (<http://www.bds.usc.es>). Elaboración propia.

Aunque es evidente que los datos contenidos en el cuadro no nos permiten ir más allá de una simple descripción cuantitativa, creo que surgen de él algunos aspectos de especial interés para la visión general de las relaciones entre verbos y esquemas sintácticos. En primer lugar, solo 9 de los 18 esquemas que pasan el filtro tienen más de un verbo en estas circunstancias. Los porcentajes de la penúltima columna permiten diferenciar con claridad entre los casos que corresponden a esquemas poco frecuentes o, al menos, poco representados en el subconjunto seleccionado y los que, en cambio, son de esquemas de mayor peso en general o en este subconjunto. En la primera situación se encuentran los esquemas activa + CD, activa + 0 argumentos o media + sujeto + complemento modal. Son esquemas poco frecuentes en los que un solo verbo (*haber*, *llover* y *comportar(se)*, respectivamente) tiene una presencia muy destacada. En la situación opuesta se encuentra, por ejemplo, el verbo *depositar*, que es el único del amplio grupo de verbos

⁵ Las claves utilizadas son: S = sujeto, D=comp. directo, I = comp. indirecto, SP = suplemento, AD = comp. adverbial, MD = comp. modal, PR = otro comp. preposicional argumental, A = agente, PS = predicativo de sujeto, PD = predicativo de comp. directo, PO = predicativo de otros complementos.

documentados en el esquema activa + sujeto + CD + complemento adverbial que concentra en ese esquema el 75% de sus usos; de ahí ese escaso 0,67% que representa con respecto al subgrupo al que pertenece.

En segundo lugar, hay un grupo de cuatro esquemas que se distinguen del resto por su importancia cuantitativa y, como era de esperar, el esquema transitivo es el que lo hace de forma más destacada. En efecto, hay 213 verbos –el 20% de los que forman el subconjunto seleccionado– que concentran en este esquema un porcentaje igual o superior al 75% de sus usos documentados en la BDS. Sin duda es el caso más notable, pero puede apreciarse también la importancia que adquiere la consideración conjunta del esquema activa/media + sujeto + suplemento.

3. Verbos y esquemas sintácticos

Veamos ahora otro aspecto, sin duda el más interesante, de esta cuestión: la relación existente entre un cierto verbo y un esquema sintáctico determinado, entendida en el sentido de la predilección o repugnancia mostrada por un verbo hacia un esquema sintáctico. Me ocuparé en este apartado de comparar las características que presentan algunos de los diferentes modos de estudiar esta conexión.

La primera posibilidad, la más evidente, consiste en el simple recuento del número de casos documentados de cada verbo en un determinado esquema para identificar los que lo presentan en mayor medida y también, al menos en algunos casos, los que no presentan ese esquema en el corpus analizado. El cuadro 5 muestra los datos correspondientes a los 25 verbos que presentan más casos del esquema Activa + S + CD + CI en la BDS.

Rango	Verbo	Frecuencia en el esquema	F. norm. en el esquema (por millón)	% s/total del verbo
1	DECIR	1991	1327,33	23,26
2	DAR	1315	876,67	41,54
3	HACER	545	363,33	9,98
4	PREGUNTAR	352	234,67	39,11
5	CONTAR	339	226,00	49,78
6	PEDIR	272	181,33	50,56
7	PONER	245	163,33	17,65
8	QUITAR	190	126,67	62,91
9	PERMITIR	151	100,67	52,98
10	OFRECER	121	80,67	45,83
11	EXPLICAR	114	76,00	25,73
12	PROPONER	100	66,67	55,56
13	ENSEÑAR	100	66,67	49,50
14	TRAER	97	64,67	25,13
15	DEJAR	96	64,00	6,16
16	RECORDAR	82	54,67	9,90
17	DEVOLVER	80	53,33	64,52
18	ENTREGAR	76	50,67	48,72
19	ABRIR	75	50,00	11,36
20	REGALAR	73	48,67	83,91
21	MOSTRAR	70	46,67	22,29
22	ASEGURAR	69	46,00	28,75
23	ECHAR	66	44,00	16,67
24	PRESTAR	61	40,67	57,01
25	PRESENTAR	60	40,00	16,53

Cuadro 5

Verbos que en la BDS tienen mayor número de casos en el esquema Activa + Sujeto + CD + CI ordenados por frecuencia. Fuente: BDS (<http://www.bds.usc.es>). . Elaboración propia

El análisis, incluso superficial, de las primeras líneas del cuadro deja ver las virtudes y defectos de este enfoque: los primeros puestos están ocupados por verbos de frecuencia muy alta que utilizan este esquema en un porcentaje apreciable de sus apariciones, aunque no sea excesivamente importante. Los dos rasgos positivos (frecuencia alta y porcentaje importante) se dan en *decir* y *dar*, pero *hacer* ocupa el tercer lugar a pesar de que sus usos en ese esquema no llegan al 10%. Algo parecido, aunque en menor medida porque tienen menos apariciones, sucede con *dejar* o *recordar*. En pocas palabras, esta perspectiva prima el número total de documentaciones en el esquema considerado, aunque el verbo en cuestión no lo tenga en los primeros lugares de sus preferencias.

Un panorama bastante diferente aparece, como es lógico, cuando utilizamos el porcentaje sobre los usos del verbo en lugar de la frecuencia general (total o normalizada). En el cuadro 6

aparecen los 25 verbos que tienen porcentajes más altos de sus usos en el esquema Activa + S + CD + CI.

Rango	Verbo	Frecuencia en el esquema	F. norm. en el esquema (por millón)	% s/total del verbo
1	REGALAR	73	48,67	83,91
2	ATRIBUIR	53	35,33	72,60
3	DEVOLVER	80	53,33	64,52
4	QUITAR	190	126,67	62,91
5	AGRADECER	49	32,67	62,82
6	CONCEDER	40	26,67	58,82
7	PRESTAR	61	40,67	57,01
8	COMUNICAR	42	28,00	56,00
9	PROPONER	100	66,67	55,56
10	PERMITIR	151	100,67	52,98
11	PEDIR	272	181,33	50,56
12	CONTAR	339	226,00	49,78
13	ENSEÑAR	100	66,67	49,50
14	ENTREGAR	76	50,67	48,72
15	ENVIAR	44	29,33	47,31
16	COSTAR	55	36,67	47,01
17	OFRECER	121	80,67	45,83
18	IMPEDIR	56	37,33	43,41
19	DAR	1315	876,67	41,54
20	PREGUNTAR	352	234,67	39,11
21	EXPLICAR	114	76,00	25,73
22	TRAER	97	64,67	25,13
23	DECIR	1991	1327,33	23,26
24	PONER	245	163,33	17,65
25	HACER	545	363,33	9,98

Cuadro 6

Verbos que en la BDS tienen mayor número de casos en el esquema Activa + Sujeto + CD + CI ordenados por porcentaje de uso en ese esquema. Fuente: BDS (<http://www.bds.usc.es>). . Elaboración propia.

Los resultados son muy diferentes. Los verbos que ocupan los primeros puestos en la primera ordenación están muy lejos de los puestos de cabeza en la segunda y lo mismo sucede en sentido contrario. Son pocos los verbos que figuran entre los 25 primeros en las dos perspectivas. *Quitar* es la excepción más marcada, puesto que oscila entre los puestos octavo y cuarto. Evidentemente, al trabajar con porcentajes se aplanan las enormes diferencias debidas a la mayor o menor frecuencia y los puestos más elevados son ocupados por verbos no muy frecuentes, pero con un alto porcentaje de utilización del esquema Activa + S + CD + CI. El efecto es claro, a pesar de que aquí estamos tomando en consideración únicamente los verbos que tienen una frecuencia normalizada igual o superior a 10 casos por millón. Lo esperable es que las discrepancias entre los

dos enfoques se reduzcan si se sube la exigencia en la frecuencia normalizada mínima (o bien, en sentido contrario, se introduce un filtro en el porcentaje mínimo del esquema).

Aunque presentan las diferencias que hemos señalado, estos dos enfoques tienen en común el hecho de que reducen su perspectiva a lo que sucede con cada uno de los verbos analizados, sin tener en cuenta lo que se da en la totalidad del corpus y la mayor o menor frecuencia general del esquema. Con los datos generales de la BDS, el esquema Activa + S + CD + CI es el cuarto en frecuencia de uso (el 5,54% de las cláusulas del corpus) y lo utiliza alguna vez el 18,16% de los verbos documentados en la BDS (cf. Rojo 2010: cuadro 10). El porcentaje medio de uso en la BDS (5,54%) se modifica ligeramente en el subconjunto que estamos considerando aquí: 10 204 casos sobre un total de 181 777 cláusulas, esto es, el 5,61%.

Tener en cuenta los datos cuantitativos de cada verbo en general y en un determinado esquema, ponerlo en conexión con lo que sucede con todos los demás verbos presentes en el corpus estudiado e intentar cuantificar esa relación es la característica general de la aproximación diseñada por Sfejanowitsch y Gries, a la que han dado la denominación general de *collostructional analysis*:

Collostructional analysis is the application of (quantitative) collocational analysis within a constructional view of language (hence its name, a blend of *construction* and *collocational analysis*) (Stefanowitsch & Gries 2005: 5).

Se trata, pues, de averiguar el grado de asociación entre un determinado lexema (un verbo en nuestro caso) y una construcción (aquí, un esquema sintáctico). Esta es la primera opción del método general, a la que denominan *collexeme analysis*. Frente a la aproximación clásica (estrictamente léxica) a las co-locaciones,⁶

instead of a node word, we look at a construction (such as the ditransitive, the past tense, the imperative, etc.), and instead a user-defined span, we look at the words occurring in a particular slot provided by that construction (we refer to such words as [potential] *collexemes*) (Stefanowitsch & Gries 2005: 6).

En la práctica, para cada cruce hay que conocer la frecuencia del verbo *V* en el esquema *E*, la frecuencia de *V* en todas las demás construcciones en que aparece, la de los demás verbos del corpus en *E* y la de los demás verbos en las demás construcciones. Se llega así a un cuadro de doble entrada, con cuatro casillas (más los totales de fila y columna, naturalmente). En el caso de *dar* en el esquema Activa + S + CD + CI, los datos de la BDS son los que figuran en la tabla 1.

	En Act.+ S + CD + CI	En los demás esquemas	Totales
--	-------------------------	--------------------------	---------

⁶ Los términos *colocación*, *colocado*, etc., que son los utilizados habitualmente en español, resultan, en mi opinión, poco congruentes con el significado que se les da por el peso del que tienen en la lengua habitual, alejado ya del valor etimológico original. *Co-locación* o *conlocación* y términos conectados resultan mucho más acordes con el significado técnico que poseen.

<i>Dar</i>	1315	1851	3166
Los demás verbos	8889	169 709	178 598
Totales	10204	171 560	181 764

Tabla 1

Distribución de los casos del esquema Activa + S + CD + CI en la BDS. Fuente: BDS (<http://www.bds.usc.es>).
Elaboración propia.

A partir de ahí, mediante los procedimientos habituales, cabe estimar si esa distribución se puede explicar como simple resultado del cruce de las frecuencias del verbo y el esquema, con una desviación atribuible al azar o bien es necesario recurrir a la hipótesis de que existe una correlación especial (positiva o negativa) entre verbo y esquema. La aplicación de esos cálculos a todos los verbos de la BDS permitirá cuantificar la relación existente en cada caso y, por tanto, ordenar los verbos en función de los estadísticos obtenidos.

Mediante las rutinas adecuadas, he extraído los datos de la BDS y he calculado el χ^2 correspondiente a cada verbo. Dado que esta prueba tiene ciertas debilidades bien conocidas, he utilizado también un programa escrito por S. Gries para R (Gries 2007) que lleva a cabo los pesadísimos cálculos necesarios para hallar el índice de Fisher-Yates. El cuadro 7 muestra, como en los casos anteriores, la lista de los 25 primeros verbos en cada uno de los estadísticos utilizados. Por economía de espacio y para facilitar la comparación entre los diferentes cálculos, incorporo los datos reelaborados de los cuadros anteriores. Este cuadro, pues, contiene la relación de todos los verbos que ocupan uno de los 25 primeros puestos en los cuatro cálculos realizados, con indicación del rango que les corresponde en cada uno de ellos.

Verbo	Frecuencia en el esquema	F. norm. en el esquema (por millón)	% s/total del verbo	Rango por frecuencia	Rango por porcentaje sobre el total del verbo	Rango por X ² cuadrado (positivo)	Rango por Fisher-Yates (atracción)
DAR	1315	876,67	41,54	2	41	1	1
DECIR	1991	1327,33	23,26	1	73	2	2
CONTAR	339	226,00	49,78	5	23	3	3
PREGUNTAR	352	234,67	39,11	4	44	5	4
PEDIR	272	181,33	50,56	6	19	4	5
QUITAR	190	126,67	62,91	8	8	6	6
PERMITIR	151	100,67	52,98	9	17	7	7
OFRECER	121	80,67	45,83	10	34	11	8
REGALAR	73	48,67	83,91	20	1	8	9
PROPONER	100	66,67	55,56	12	15	9	10
ENSEÑAR	100	66,67	49,50	13	24	12	11
DEVOLVER	80	53,33	64,52	17	7	10	12
PONER	245	163,33	17,65	7	98	17	13
ENTREGAR	76	50,67	48,72	18	25	14	14
ATRIBUIR	53	35,33	72,60	28	5	13	15
PRESTAR	61	40,67	57,01	24	13	15	16
EXPLICAR	114	76,00	25,73	11	65	23	17
AGRADECER	49	32,67	62,82	29	9	16	18
HACER	545	363,33	9,98	3	159	37	19
COSTAR	55	36,67	47,01	27	31	18	20
TRAER	97	64,67	25,13	14	67	27	21
IMPEDIR	56	37,33	43,41	26	37	22	22
COMUNICAR	42	28,00	56,00	31	14	20	23
CONCEDER	40	26,67	58,82	32	12	19	24
ENVIAR	44	29,33	47,31	30	29	25	25
ASEGURAR	69	46,00	28,75	22	58	31	26
PROPORCIONAR	33	22,00	67,35	33	6	21	27
REPROCHAR	26	17,33	81,25	35	2	24	28
ROGAR	25	16,67	75,76	36	4	26	31
CAUSAR	31	20,67	51,67	34	18	32	34
MOSTRAR	70	46,67	22,29	21	77	42	35
OTORGAR	19	12,67	61,29	37	10	41	42
ASIGNAR	13	8,67	76,47	40	3	44	46
ECHAR	66	44,00	16,67	23	102	59	47
PRESENTAR	60	40,00	16,53	25	103	64	48
FROTAR	14	9,33	50,00	38	20	54	51
RESERVAR	14	9,33	50,00	39	21	55	52
AHORRAR	11	7,33	61,11	42	11	57	59
ARREBATAR	12	8,00	50,00	41	22	63	61
ABRIR	75	50,00	11,36	19	144	93	64
ABROCHAR	8	5,33	53,33	43	16	80	70
RECORDAR	82	54,67	9,90	16	160	115	74
DEJAR	96	64,00	6,16	15	200	552	153

Cuadro 7

Verbos situados en los 25 primeros puestos según alguno de los factores empleados. La ordenación del cuadro es la que corresponde al índice de Fisher-Yates. En negrita, algunas discrepancias especialmente llamativas. Fuente: BDS (<http://www.bds.us.es>). Elaboración propia.

Como era de esperar, los rangos derivados del χ^2 y de la prueba de Fisher-Yates son bastante congruentes. Por supuesto, hay algunas ligeras diferencias en la ordenación resultante en cada prueba, pero, en realidad, el único caso realmente llamativo es el de *hacer*, que se sitúa en el puesto 19 si usamos el índice de Fisher-Yates y queda relegado al 37 si empleamos el χ^2 . Salvo en los casos de *hacer*, que es el más llamativo, *proporcionar*, *reprochar* y *traer* (que no están entre los 25 primeros con la prueba del χ^2 , las discrepancias se reducen a ligeras diferencias en el puesto ocupado por el verbo en el interior de la lista de los 25 que presentan valores más altos en la correlación.

Cualquiera de estos dos índices muestra diferencias bastante fuertes con las otras dos ordenaciones. Muy evidentes con la basada en el porcentaje de empleo del verbo en este esquema. También muy marcados, aunque menos que en el caso anterior, en la establecida sobre la frecuencia de uso del verbo en ese esquema. Véase, por ejemplo, lo que sucede con *dar* y *decir*, situados en los dos primeros puestos en tres de las listas y relegados a los puestos 41 y 73, respectivamente, en la que se basa en el porcentaje de uso. *Hacer*, que ocupa el tercer puesto en la ordenación por frecuencia, aparece en puestos inferiores o muy inferiores en las otras dos.

Las limitaciones de espacio impiden profundizar en el análisis pero creo que lo visto es suficiente para concluir que la línea del **collexeme analysis**, propuesta por Stefanowitsch y Gries, constituye un método adecuado para valorar y cuantificar las relaciones entre verbos y esquemas sintácticos.

Referencias bibliográficas

- Almela Pérez, Ramón, Pascual Cantos, Aquilino Sánchez, Ramón Sarmiento y Moisés Almela (2005): *Frecuencias del español: diccionarios y estudios léxicos y morfológicos*, Madrid: Universitas.
- Gries, Stefan Th. (2007): *Coll.analysis 3.2. A program for R for Windows 2.x*. Descargable de <http://www.linguistics.ucsb.edu/faculty/stgries/teaching/groningen/coll.analysis.r> (comprobado el 27/12/2010).
- Juilland, Alphonse & Eugenio Chang-Rodríguez (1964): *Frequency Dictionary of Spanish Words*. La Haya: Mouton.
- Rojo, Guillermo (2003): “La frecuencia de los esquemas sintácticos clausales en español”, en F. Moreno Fernández *et al.* (coords.): *Lengua, variación y contexto. Estudios dedicados a Humberto López Morales*, Arco/Libros: Madrid, vol. I, 413-424.
- (2010): “Frecuencia de inventario y frecuencia de uso”. En prensa en la RSEL.
- Stefanowitsch, Anatol & Stefan Th. Gries (2005): “Covarying collexemes”, en *Corpus Linguistics and Linguistic Theory*, 1/1, 1-43.