

CHAPTER TWO

LINGÜÍSTICA DE CORPUS Y VALENCIA VERBAL

JOSE MARÍA GARCÍA-MIGUEL

1. Introduction

Las propiedades combinatorias, o *valencia*, de los verbos y otros elementos léxicos constituyen un problema clave para el estudio de las interrelaciones entre léxico y sintaxis y entre sintaxis y semántica. El concepto y el término de valencia proceden de Tesnière (1959), quien establece una analogía entre los verbos y los elementos químicos,

On peut comparer le verbe à une sorte de atome crochu susceptible de exercer son attraction sur un nombre plus ou moins élevé d'actants, selon qu'il comporte un nombre plus ou moins élevé de crochets pour les maintenir dans sa dépendance. Le nombre de crochets que présente un verbe et par conséquent le nombre d'actants qu'il est susceptible de régir, constitue ce que nous appellerons la valence du verbe (Tesnière 1959: 238)

Tanto en la propia obra de Tesnière como en la elaboración del concepto de valencia por parte, sobre todo, de lingüistas alemanes queda claro que la valencia verbal no se limita al número de actantes, sino que también concierne a sus características sintácticas y semánticas, por lo que suele hablarse, siguiendo a Helbig (1992) de valencia cuantitativa o valencia lógica, de valencia sintáctica (propiedades y relaciones sintácticas de cada actante) y de valencia semántica (propiedades y relaciones semánticas de cada actante). Esto convierte la teoría de la valencia en un marco teórico adecuado para describir todo tipo de propiedades sintagmáticas de los actantes, lo cual incluye lo que en otros marcos teóricos se conoce como marcos de subcategorización (Chomsky 1965), marcos predicativos (Dik 1989), estructura argumental, esquemas sintáctico-semánticos y otros. El conjunto de conceptos relacionados con

la valencia ha demostrado su importancia para la lingüística en general; y en especial para la lingüística de corpus, para la lexicografía y para el procesamiento del lenguaje natural. En opinión de Schulte im Walde (2009: 954), en procesamiento del lenguaje la mitad de los fallos de parsing están causados por información valencial inadecuada.

La teoría de la valencia permite sistematizar las superficiales indicaciones sintácticas de los diccionarios “tradicionales”. Por ejemplo, el *Diccionario de la lengua española* de la Real Academia Española (DRAE) limita su información sintáctica a abreviaturas del tipo *tr.* “transitivo”, *intr.* “intransitivo”, *U.t.c.intr.* “úsase también como intransitivo”, y similares. Ciertamente, también se emprendieron monumentales obras como el *Diccionario de construcción y régimen de la lengua castellana* (Cuervo 1994 [1872-]), que no pudo completarse hasta mucho después de iniciado; pero sólo con la aplicación a la lexicografía de la teoría de la valencia, a partir de los años 70, se empieza a sistematizar la información sintáctica en los diccionarios. A partir de la publicación de la obra de G. Helbig y W. Schenkel (1969) *Wörterbuch zur Valenz und Distribution deutscher Verben* y, posteriormente, del *Kleines Valenzlexikon deutscher Verben* de U. Engel y H. Schumacher (1976), se ha ido generalizando progresivamente la confección de numerosos diccionarios de valencias, esto es, diccionarios destinados a ofrecer información sobre las posibilidades combinatorias de los verbos o de otras unidades léxicas como los nombres y adjetivos. En la década de los ochenta y los noventa, se han ido publicando trabajos lexicográficos sobre la valencia, régimen verbal, o construcciones sintácticas de los verbos para lenguas como el francés (Busse y Dubost 1977), el portugués (Borba 1990, Busse 1994, Luft 1995), el inglés (Garrudo 1991), el polaco (Polánski *et al.* 1980-1992), el eslovaco (Nižníková y Sokolová 1998) o el italiano (Blumenthal y Rovere 1998), por mencionar algunas diccionarios monolingües, a lo que habría que añadir proyectos de diccionarios bilingües como el Diccionario contrastivo de valencias verbales español-alemán (Domínguez *et al.* 2008).

Entre las obras recientes más destacadas, habría que citar el *Valenzwörterbuch deutscher Verben* (VALBU) de Schumacher *et al.* (2004) y el *Valency Dictionary of English* (VDE) de Herbst *et al.* (2004). VALBU, desarrollado en Institut für Deutsche Sprache (IDS) de Mannheim, es un diccionario monolingüe de valencia verbales de alemán que contiene 638 lemas, con unas 3000 acepciones verbales, para los que se ofrece además de información valencial también detallada información morfológica, fraseológica y estilística. La versión electrónica on-line E-VALBU (<http://hypermedia2.ids-mannheim.de/evalbu/>) no es una simple copia de la versión en papel sino que presenta una reorganización del

Tabla 1. Muestra (parcial) de una entrada en E-VALBU

GEWINNEN 2	
Strukturbeispiel:	jemand/etwas gewinnt etwas gegen jemanden irgendwie
Im Sinne von:	jemand/etwas geht in etwas gegen jemanden irgendwie als Sieger hervor
Satzbauplan:	K_{sub} , (K_{akk}) , (K_{prp}) , (K_{adv})
Belegungsregeln:	• K_{sub}: NP im Nom/ProP im Nom/GWS • K_{akk}: NP im Akk/ProP im Akk/GWS • K_{prp}: <i>gegen</i> +Akk • K_{adv}: Art und Weise • AdjP /AdvP /<i>durch</i>+Akk /<i>in</i>+Akk /<i>mit</i>+Dat /<i>nach</i>+Dat

El *Valency Dictionary of English* [VDE] (Herbst *et al.* 2004) describe la valencia de 511 verbos, 274 nombres y 544 adjetivos del inglés contemporáneo, con ejemplos tomados del corpus Bank of English, de 320 millones de palabras.

Tabla 2. VDE: Preámbulo y significado de WIN

win verb		Active: ¼	Passive: ¼	General: 0					
I	[N] _A /[by N]				IV	[against N] _P	D2	T5-6	Q
	[V-ing] _A			T1	V	[at N] _P	D3		
	[that-CL] _A			T1	VI	[by N/SCORE]	D4	T2.6	Q
II	[N] _{P-2}		D1	T1-4	Q	[SCORE]	D5	T5	
III	[N] ₁			T1		VII		T4	
	[for N]			T3					

A person or a group of people^I can win

(i) a contest or competition^{II} against someone else^{IV} by a certain score^{VI}, i.e. defeat the other competitors

(ii) something that they want such as support, money or an award^{II}, i.e. get it

Las propiedades verbales codificadas en VDE incluyen el mínimo y máximo de actantes en cláusulas activas y pasivas, un inventario de actantes semánticos (I-VII, ...) descritos en términos de categorías sintácticas (por ejemplo, [N]_A/[by N], para sujeto activo y frase con *by* de la pasiva), y una lista de los esquemas sintácticos en los que se registra cada actante el corpus analizado (T1, D1, ...).

Tanto VALBU como VDE son diccionarios basados en corpus. Estos y otros proyectos similares (más adelante nos detendremos en el proyecto ADESSE) nos animan a reflexionar sobre las aportaciones de la lingüística de corpus a la descripción de la valencia verbal (y viceversa). Hay muchas maneras de hacer lingüística con corpus, pero también algunas constantes que caracterizan esta metodología emergente. En primer lugar, el mismo hecho de trabajar con corpus, esto es con conjuntos de textos naturales en soporte informático que se toman como representativos de una lengua o variedad de lengua, da una perspectiva diferente que si tomamos como datos ejemplos contruidos artificialmente. Pero además los métodos característicos de la lingüística de corpus consisten en el análisis sistemático y exhaustivo de todos los datos, el análisis cuantitativo (frecuencias, probabilidades,...) y prioritariamente en la descripción a partir de patrones de coaparición, lo que otorga especial relevancia al contexto (Gries 2006: 4). En mi opinión, destaca especialmente el hecho de que el manejo de corpus textuales obliga a intentar un examen exhaustivo de todos los casos contenidos en el corpus:

There is no prior selection of data which we are meant to be accounting for and data we have decided to ignore as irrelevant to our theory. This principle of "total accountatibility" for the available observed data is an important strength of CCL[= computer corpus linguistics] (Leech, 1992:112).

La exhaustividad implica una revisión continua de las categorías cualitativas preestablecidas y un movimiento continuo de ida y vuelta de los datos a la teoría y viceversa. Por eso, no hay o no debería haber una contraposición entre lingüísticas teóricos "de sillón" y lingüistas de corpus, sino que "the two kinds of linguists, wherever possible, should exist in the same body" (Fillmore 1992: 35). De la misma manera, es artificial o simplemente un cuestión de perspectiva la contraposición entre "corpus-based linguistics" y "corpus-driven linguistics" (Tognini-Bonelli 2001), según se ponga el énfasis en el camino que va de los categorías descriptivas a los datos del corpus o el que va de los datos del corpus a las categorías que emergen de él. Ciertamente, si el punto de partida es el corpus no analizado y no anotado, la metodología de la lingüística de

corpus se limita en gran medida a fenómenos a los que sólo puede accederse haciendo búsquedas sobre palabras particulares, tales como ejemplos contextualizados (concordancias KWIC), combinaciones frecuentes de palabras (“colocaciones”) y sólo tras un cierto proceso de abstracción patrones como los que constituyen la *Pattern Grammar* de Hunston y Francis (2000).

Sin embargo, la investigación en sintaxis y semántica (y la valencia es un fenómeno sintáctico y semántico) requiere el análisis de propiedades estructurales más abstractas que las que pueden observarse fácilmente de esa manera (Manning 2003: 294). Es decir, resulta necesario ir más allá del texto como cadena de caracteres, porque no siempre es suficiente con buscar en texto plano, ni siquiera basta siempre con corpus anotados morfosintácticamente. Además de corpus con anotación morfosintáctica, necesitamos también corpus con anotación sintáctica, semántica, discursiva, ... lo cual implica categorización (agrupación de casos similares). Pero por otro lado, ningún conjunto de categorías preestablecidas suele valer para analizar exhaustiva e indiscutiblemente un corpus cualquiera. Además, la lingüística de corpus reciente ha mostrado que debe superarse una manera de hacer sintaxis basada en categorías abstractas que pueden ser ocupadas por elementos léxicos: “The evidence now becoming available casts grave doubts on the wisdom of postulating separate domains of lexis and syntax” (Sinclair 1991: 104). La idea de un continuo léxico-sintaxis con un amplio número de expresiones idiomáticas y combinaciones preconstruidas es compartida también por la Gramática de Construcciones (Goldberg 2003). El concepto de valencia se asocia con el número de huecos [“slots”] abiertos por un predicado para ser ocupados por categorías sintácticas y elementos léxicos. Pero también tiene que ver con las restricciones impuestas sobre cada hueco y debe ser compatible, como veremos, con la probabilidad de que los actantes adopten ciertas realizaciones específicas, lo cual debe comprobarse con datos cuantitativos basados en corpus. Además, el concepto de valencia debería ser compatible con el principio de idiomatidad, lo que obliga a definir patrones combinatorios a diferentes niveles de abstracción, y no sólo para lexemas verbales.

En este contexto, han aparecido en los últimos años un número considerable de recursos informatizados directamente relacionados con la valencia y la lingüística de corpus. Además de diccionarios de valencias basados en corpus, como VALBU y VDE, algunos de ellos en versión electrónica como el propio E-VALBU o el léxico valencial del checo VALLEX (Zabokrtsky and Lopatkova 2007), tenemos también otros tipos de diccionarios combinatorios basados en corpus como el DEC (Mel’čuk *et al.* 1984-1999) o, para el español, REDES (Bosque 2004). Pero sobre

todo, tenemos una serie de recursos basados en corpus que incluyen como parte esencial la sistematización de las propiedades sintácticas y semánticas de los verbos y otras unidades léxicas, incluyendo su descripción y clasificación semántica y sus propiedades combinatorias. Entre esos recursos destacaremos para el inglés PropBank (Palmer *et al.* 2005), FrameNet (Fillmore *et al.* 2003. Fillmore 2007) y Pattern Dictionary of English Verbs (Hanks and Pustejovsky 2005), y para el español Spanish FrameNet (Subirats 2009), SenSem (Vázquez *et al.* 2006) y el léxico derivado de AnCora (Taulé *et al.* 2008), además de ADESSE, que comentaremos a continuación. Muchos de estos recursos apenas ofrecen información sobre frecuencia en el corpus de cada patrón sintagmático, a diferencia de lo que ofrece VALEX (Korhonen *et al.* 2006), un lexicon con la frecuencia de los patrones de subcategorización de 6397 verbos del inglés. A cambio, tenemos recursos con información semántica bastante detallada. Su desarrollo demuestra también la necesidad de anotación sintáctica y semántica bien del corpus bien de los recursos léxicos derivados de corpus de las nociones claves de la valencia semántica no pueden deducirse de otras más superficiales, por lo que la anotación sintáctico-semántica de corpus requiere ineludiblemente anotación manual, aunque los recursos léxicos en desarrollo pudieran servir de base para la automatización de anotaciones semánticas ulteriores. Frente al desarrollo de diversos “treebanks” y similares desde los años 90, la primera década del siglo XXI ha visto el desarrollo de varios proyectos que añaden a la anotación sintáctica también información semántica relevante. Además de los proyectos ya mencionados, destaquemos el Prague Dependency Treebank (Hajič 2004) para el checo, y SALSA (Burchardt *et al.* 2006), que incorpora información semántica de FrameNet al treebank Tiger del alemán.

2. ADESSE

En la páginas siguientes hablaremos de ADESSE, “*Base de datos de verbos, Alternancias de Diátesis y Esquemas Sintactico-Semánticos del Español*” (<http://adesse.uvigo.es/>), un proyecto¹ de la Universidad de Vigo cuyo objetivo es disponer de una base de datos con información sintáctica y semántica sobre los verbos y cláusulas de un corpus del español (corpus ARTHUS de 1,5 millones de palabras). ADESSE es una versión, ampliada con información semántica, de BDS, “*Base de datos sintácticos del español*” (<http://www.bds.usc.es/>), un proyecto desarrollado en la Universidad de Santiago de Compostela en los años 90 (Rojo 2001). Una diferencia importante con respecto a los diccionarios de valencias

publicados es que en ADESSE no se analizan sólo los patrones valenciales de un número limitado de verbos, para los que se han seleccionado ejemplos suficientes en el corpus (por ejemplo, 511 verbos en VDE, 638 verbos en VALBU), sino que se desarrolla con intención de exhaustividad: todas las cláusulas del corpus, no una selección suficiente para la descripción de los verbos. Se pretende incluir el análisis de los patrones valenciales de todos los ejemplos de todos los verbos. Dado el tamaño del corpus y la distribución estadística esperable, tendremos muchos ejemplos de unos pocos verbos, y muy pocos ejemplos de la mayoría de los verbos registrados en el corpus.

Tabla 3. ADESSE: Número de verbos y frecuencia en número de cláusulas (no se cuentan verbos auxiliares, anotados junto con el verbo principal)

Verbos (lemas)	N cláusulas
25	>1000
45	> 500
276	> 100
472	> 50
1210	> 10
1672	> 5
3433	> 0 (todos)

El tamaño (y la composición) del corpus son insuficientes para una descripción detallada del léxico y, en menor medida, la gramática del español. Pero ese tamaño reducido es adecuado para un análisis manual exhaustivo de numerosos fenómenos léxicos, sintácticos y semánticos, permitiendo que se convierta en un interesante campo de pruebas que serían irrealizables en corpus más amplios. Las principales propiedades que se han anotado en la base de datos (sumando las anotaciones de BDS y ADESSE) aparecen ejemplificadas en la Tabla 4.

Tabla 4. (Parte de) una ficha en ADESSE: *Los hermanos Vicario les habían contado sus propósitos a más de doce personas* [CRO: 61, 22]

	PREDICADO		
<i>verbo</i>	CONTAR		
<i>significado</i>	I “narrar, relatar”		
<i>clase</i>	Comunicación		
<i>voz</i>	Activa		
<i>Tiempo</i>	Antepretérito Indicativo		
...			
	ARGUMENTOS		
<i>Rol</i>	Emisor	Mensaje	Receptor
<i>Indice_arg</i>	Arg1	Arg2	Arg3
<i>Función sint.</i>	Suj	ODir	OInd
<i>Concordancia</i>	3pl.Suj [-n]		3pl.Dat [<i>les</i>]
<i>Categoría</i>	FN	FN	FN
<i>Núcleo léxico</i>	[<i>hermano</i>]	[<i>propósito</i>]	[<i>persona</i>]
<i>Tipo semántico</i>	Animado	Abstracto	Animado

Las principales dimensiones codificadas en ADESSE comprenden, para cada argumento nuclear: función sintáctica, realización morfo-sintáctica (categoría, preposición, etc.), rol semántico, tipo semántico (animado, inanimado, abstracto, proposicional), núcleo léxico, posición (con respecto al verbo), determinación y número. Además, para para cada verbo y cada cláusula, se ha anotado la voz (activa, media, pasiva), el significado del verbo (acepción), el tipo semántico de proceso, categorías flexivas (tiempo, aspecto, modo), perífrasis verbales, polaridad, modalidad, y algunos otros rasgos morfosintácticos.

En la versión online de la base de datos (<http://adesse.uvigo.es/data>) se pueden obtener informes sobre verbos (acepciones, potencial valencial, realizaciones valenciales y perfil combinatorio), sobre clases semánticas de verbos (roles de participantes, verbos más frecuentes y construcciones más frecuentes de la clase), sobre esquemas sintácticos (verbos y clases más frecuentes) y existe un formulario de búsquedas avanzadas, que permite obtener casi cualquier combinación de verbo, clase y propiedades léxicas, sintácticas y semánticas de argumentos.

En la base de datos hay anotaciones sobre argumentos [actantes], pero no circunstanciales. El criterio básico para determinar qué actantes reciben anotación son la obligatoriedad semántica y la especificidad. Además se han anotado siempre las funciones sintácticas centrales (SUJ, ODIR, OIND), aunque no sean obligatorias. Secundariamente, se han anotado

también muchos argumentos adicionales evocados semánticamente por el predicado, aunque no sean obligatorios.

3. Verbos y esquemas construccionales: Aspectos de la valencia en el corpus

3.1. Valencia semántica

Aunque frecuentemente se ha presentado la valencia como exigencia de determinados constituyentes sintácticos, la idea que subyace a ADESSE es que las bases de la estructura argumental deben buscarse en la semántica y en el discurso, no en propiedades puramente formales de los elementos léxicos. Los verbos y los esquemas sintácticos de una lengua nos proporcionan instrumentos para representar escenas o situaciones del mundo real o imaginado, de modo que al utilizar una palabra (un verbo) o una construcción gramatical (esquema) estamos conceptualizando una situación como realización o ejemplar de cierta clase o tipo. Por ejemplo, los verbos siguientes conceptualizan ciertos tipos de escenas de las que podemos decir que ocurren cuando intervienen los actantes que se indican con cada verbo:

(1) Situaciones y actantes

CASARSE: alguien (A1) SE CASA con alguien (A2)

SENTIR: alguien (A1) SIENTE algo (A2)

REGALAR: alguien (A1) REGALA algo (A2) a alguien (A3)

AVISAR: alguien (A1) AVISA [de] algo (A2) a alguien (A3)

GANAR: alguien (A1) GANA algo (A2) a alguien (A3)

Dando algunos pasos más en el mismo proceso de abstracción, cabe generalizar sobre diferentes tipos de situaciones abstrayendo lo que tienen en común las situaciones definidas por verbos semánticamente relacionados. Por ejemplo, *ganar*, *vencer* y *perder* aunque diferentes semánticamente evocan un mismo tipo de escena en la que compiten un competidor y un antagonista. Este punto de vista semántico-conceptual está en línea con la perspectiva adoptada por FrameNet: “the semantic valence properties of a word are expressed in terms of the kinds of entities that can participate in *frames* of the type evoked by the word. We call these roles *frame elements*” (Fillmore *et al.* 2003: 237). En la Tabla 5 tenemos un ejemplo:

Tabla 5. Elementos del marco conceptual “Finish_Competition” en FrameNet (<http://framenet.icsi.berkeley.edu/>)²

Frame: <u>Finish_Competition</u>	
Core Frame Elements:	
<u>Competition</u>	The event in which the <u>Competitors</u> compete with each other.
<u>Competitors</u>	The individuals who compete against each other.
<u>Competitor</u>	The individual involved in the <u>Competition</u> .
<u>Opponent</u>	The individual that the <u>Competitor</u> has been competing against.
Non-Core:	
<u>Rank</u>	The rank (according to <u>Score</u>) that a <u>Competitor</u> achieves.
<u>Score</u>	A quantity whereby progress is measured during the game, and Rank and Prizes are awarded to the <u>Competitors</u> afterward.
<u>Margin</u>	The amount by which the <u>Competitor's</u> and <u>Opponent's</u> <u>Scores</u> differ.
...	
Lexical Units:	<i>draw.n, lose.v, loss.n, tie.v, victory.n, win.n, win.v, winner.n, ...</i>

ADESSE tiene propósitos diferentes y no incluye la definición de marcos conceptuales similares a los de FrameNet. Pero necesita también una caracterización de tipos de situaciones y de participantes en esas situaciones. En el nivel más bajo de abstracción, cada verbo o cada sentido verbal evoca un tipo particular de situación. En sucesivos procesos de abstracción se agrupan los verbos en clases jerarquizadas que definen tipos de procesos (Albertuz 2007). En la Tabla 6 se muestran los dos niveles más altos de la jerarquía utilizada en ADESSE (similar, pero no igual, a la tipología de procesos de Halliday 1985) junto con el número de verbos que se han adscrito a cada clase en la base de datos. Para cada una se han definido también clases más específicas hasta un total de 62 clases en la versión más reciente.

Tabla 6. Tipos de situaciones: clases primarias de verbos en ADESSE

CLASES primarias		Ejs	VERBOS	CLAUS.
1 MENTAL	11 Sensación	<i>Gustar</i>	576	36.417
	12 Percepción	<i>Ver</i>	(14%)	(23%)
	13 Cognición	<i>Saber</i>		
2 RELACIONAL	21 Atribución	<i>Ser</i>	425	29.816
	22 Posesión	<i>Tener</i>	(11%)	(19%)
3 MATERIAL	31 Espacio	<i>Ir</i>	2302	52.082
	32 Cambio	<i>Abrir</i>	(58%)	(33%)
	33 Otros hechos	<i>Tocar</i>		
	34 Conductas	<i>Reír</i>		
4 VERBAL		<i>Decir</i>	336 (8%)	15.595 (10%)
5 EXISTENCIAL		<i>Haber</i>	200 (5%)	11.568 (7%)
6 MODULACION		<i>Hacer-II</i> (caus.)	161 (4%)	13.444 (9%)
TOTAL			4000	158922

En cuanto a los papeles de los participantes, ADESSE comparte con FrameNet la idea de no utilizar una lista única de roles semánticos (o “papeles temáticos”) de validez general y definirlos en relación con el tipo de situación o marco conceptual. Así se definen roles en diferentes niveles de generalidad (García-Miguel and Vaamonde 2010). Tendremos pues roles específicos de cada VERBO (representan “potencial valencial”):

(2)

ESCRIBIR: <A0:Escritor, A1:Texto, A2:Receptor, A3:Asunto>

ENSEÑAR: <A0: Docente, A1: Discente; A2: Contenido >

GANAR: <A1: Ganador, A2:Perdedor>

VENCER: < A1: Vencedor, A2:Vencido >

Y para cada clase semántica, que agrupa entradas verbales a diferentes niveles de abstracción, se definen roles de participantes específicos de cada clase:

(3)

41 Comunicación <Comunicador, Mensaje, Receptor, Asunto>

13 Cognición <Conocedor, Contenido>

338 Competición <Competidor, Antagonista>

Tanto en la definición de tipos de escena designados por un verbo como en los designados por una clase se utiliza un sistema de numeración correlativa para cada argumento semántico (A0, A1, A2, etc...) que en parte es arbitrario, en parte refleja probabilidad de aparición, y en parte refleja roles semánticos generalizados: por ejemplo, A0 es siempre el Iniciador en procesos causativos.

Los esquemas sintáctico-semánticos o patrones valenciales consisten explícitamente en el emparejamiento de cada función sintáctica con el índice de uno de los argumentos específicos del verbo.

3.2. Actantes obligatorios y opcionales

Definida la *valencia* como el número de actantes que un verbo es susceptible de regir (Tesnière 1959: 238), el propio Tesnière añade que “certains valences peuvent rester inemployées ou libres. C’est le cas lorsqu’on dit *Alfred chante* en face de *Alfred chante une chanson*” (Tesnière 1959: 239). En consonancia con esto, al presentar los diccionarios de valencias lo que Helbig (1992) llama “valencia cuantitativa” o “valencia lógica” suelen ofrecer bien una cuantificación de actantes obligatorios y opcionales, como hace el VDE, o bien una lista de actantes caracterizados formalmente donde se indican entre paréntesis los opcionales, como hace VALBU

[VDE] *win* Active: 1/4

[VALBU] *gewinnen*: NomE, (AkkE), (PräpE), (AdvE)

La distinción entre lo que es posible con un verbo y lo que está sintácticamente realizado, también es utilizada en ADESSE. Por ejemplo, el marco conceptual de *escribir* puede describirse por medio de cuatro roles semánticos: A0-Escritor, A1-Texto, A2-Receptor y A3-Asunto. Aunque ocasionalmente podríamos encontrar explícitos esos cuatro argumentos, como en (a), los esquemas sintácticos de *escribir* suelen seleccionar un subconjunto, destacando ciertos participantes en la escena y dejan los demás no expresados, como en (b) o (c):

(4)

Juan [A0] le escribió una carta [A1] a su madre [A2] sobre su recuerdos de infancia [A3]

Juan [A0] escribió una carta [A1]

Juan [A0] le escribió a su madre [A2]

Sin embargo, la observación del corpus no nos permite una distinción estricta entre actantes obligatorios y opcionales. Los esquemas valenciales de un verbo aparecen en el corpus con diferente frecuencia, desde los más usuales hasta los raros o inesperados. Es decir, la obligatoriedad u opcionalidad de los argumentos de un predicado no es una cuestión de sí o no, sino una gradación; por lo que debemos entender que los actantes obligatorios son aquellos elementos (referenciales) más frecuentemente vinculadas al verbo en los textos.

Tabla 7. Argumentos of Escribir (N cláusulas = 321)

Indice	Etiqueta	Frecuencia	
0	Escritor	300	(93.5 %)
1	Texto	208	(64.8 %)
2	Receptor	85	(26.5 %)
3	Asunto	10	(3.1 %)

Ninguno de los actantes de escribir es estrictamente obligatorio, pero las diferencias en la frecuencia de realización sugieren diferentes grados de implicación por parte del lexema verbal y generan expectativas sobre qué podemos encontrarnos al usar un verbo. El verbo *escribir* parece asociarse más bien con un participante “escritor” y, en menor medida, un “texto”. Ambos, además, parecen evocados incluso cuando no están presentes, pues difícilmente puede concebirse el evento que llamamos *escribir* sin ellos, y en ese sentido son semánticamente obligatorios. Pero los argumentos “receptor” y “asunto” deben considerarse como argumentos opcionales, compatibles con ese verbo pero de ningún modo obligatorios.

Lógicamente, el grado variable de exigencia de los actantes por parte de cada verbo puede diferenciar verbos cuasi-sinónimos. Por ejemplo, tanto *ganar* como *vencer* admiten hasta tres actantes (competidor-ganador, antagonista-perdedor y competición-objetivo), pero mientras *vencer* se combina frecuentemente con los dos primeros (*alguien vence a alguien* y raramente con el tercero, *ganar* exige el primero (*alguien gana*) y admite opcionalmente tanto el segundo (*alguien gana a alguien*) como el tercero (*alguien gana algo*):

Tabla 8. Argumentos of *Ganar y Vencer*

	GANAR-II (N=80)		VENCER (N=86)	
A1 COMPETIDOR (Ganador)	80	(100 %)	86	(97.7 %)
A2 ANTAGONISTA (Perdedor)	23	(28,8 %)	74	(84.1 %)
A3 OBJETIVO (Competición)	28	(35 %)	2	(2.3 %)

Pero las diferencias de frecuencia de los actantes plantean un problema de más largo alcance, pues ponen en cuestión la posibilidad de establecer de manera categórica qué combinaciones sintagmáticas admite un verbo.

Many subcategorization distinctions presented in the linguistics literature as categorical are actually counterexemplified in studies of large corpora of written language use. We can get a much better picture of what is going on by estimating a probability mass function (pmf) over the subcategorization patterns for the verbs in question. We can put a probability function over the kinds of dependents to expect with a verb or class, conditioned on various features. (Manning 2003: 302)

Desde una perspectiva funcional, basada en el uso, los patrones valenciales son generalizaciones sobre uso registrado, donde se hace abstracción de detalles particulares y cada ejemplar se interpreta por analogía con ejemplares previamente percibidos. En este tipo de modelos la frecuencia de uso es un factor crucial para explicar el conocimiento del lenguaje y su adquisición. Y, por supuesto, esto es también válido para la estructura argumental, o valencia:

“Argument structure” needs to be replaced by a greatly enriched probabilistic theory capturing the entire range of combinations of predicates and participants that people have stored as sorted and organized memories of what they have heard and repeated over a lifetime of language use” (Thompson and Hopper 2001: 47)

Es decir, en vez de argumentos obligatorios y opcionales, nuestra experiencia lingüística nos proporciona expectativas sobre la referencia a cierto tipo de participante en las escenas evocadas por el verbo, a partir de la frecuencia de usos previos. Como la obligatoriedad es uno de los principales criterios para la distinción entre actantes y circunstanciales [argumentos y adjuntos], esta distinción también es gradual.

However, such a frequency-based account is not satisfactory, because frequency of occurrence needs to be distinguished from argument status.

Many arguments are optional, while some adjuncts are (almost) compulsory (e.g., a how adjunct for worded as in He worded the proposal very well). (Manning 2003: 305)

Por eso, una cosa es el grado de obligatoriedad y otra qué complementos cuentan para describir la valencia del verbo. En BDS, y ADESSE, se utilizan como criterios adicionales a la exigencia semántica de los participantes (aunque no siempre estén expresados), la especificidad (complementos que no se combinan con cualquier clase de verbos) y en cualquier caso, siempre se anotan los actantes destacados por el hecho de expresarse mediante las funciones sintácticas centrales (SUJ, OD, OI), aunque no estén exigidos por el verbo. Esto último nos permite destacar que la valencia verbal no se limita a los participantes semánticamente obligatorios. Y esto es especialmente importante en una perspectiva construccionista. Encontramos frecuentemente que los verbos aparecen en construcciones que incluyen argumentos adicionales que no se deducen del significado del verbo. Pero incluso si consideramos sólo los actantes más esenciales es necesario considerar los esquemas construccionales como tales y no como simple concatenación de actantes individuales:

With regard to the view of whether valency phenomena are to be described more appropriately in terms of a complement inventory or also in terms of valency patterns, it seems that both views will have to be considered (Herbst 2007: 27)

Veremos a continuación algunos aspectos de la asociación entre verbos y construcciones argumentales (es decir, patrones valenciales).

3.3. Esquemas construccionales

Entendemos por *construcción* cualquier patrón de asociación entre expresión y contenido [“constructions are stored pairings of form and function” (Goldberg 2003: 219). Una construcción argumental (“Argument Structure Construction”) es un tipo de construcción que proporciona los medios básicos de expresión clausal en una lengua.. Los esquemas construccionales argumentales admiten diferentes formatos alternativos o combinados

Categorías sintácticas: **NP tells NP that-clause**

Casos: *Gewinnen*: **Nom Akk gegen+Akk**

Funciones sintácticas: *Enseñar*: **Suj ODir OInd**

En un corpus con anotación morfosintáctica (el tipo más común de anotación gramatical) es relativamente simple encontrar patrones de categorías sintácticas. Por eso, el formato en términos de categorías sintácticas puede guardar cierta similitud con algunas fórmulas de la llamada *Pattern grammar* (Hunston and Francis 2000), una sistematización de las preferencias contextuales inducidas del corpus en una gramática de esquemas o patrones [“patterns”], yendo de los *patterns* a los *items* presentes en él y luego a otros *patterns* en los que esos *items* ocurren. En opinión de los proponentes de esta metodología, el análisis estructural en términos de funciones sintácticas “is a pointless exercise” (Hunston and Francis 2000: 152), que no añade nada. Sin embargo, Teubert (2007) argumenta que no es suficiente con patrones léxico-categoriales y que necesitamos los conceptos básicos de la teoría de la valencia “to link local, lexical grammar to general grammar” (225). A eso debemos añadir que la estructura del inglés y del alemán permiten deducir, muchas veces, la función sintáctica de patrones categoriales o de caso, respectivamente. En español, es imprescindible el recurso a funciones sintácticas debido al orden relativamente libre y a la realización de actantes mediante concordancia o clíticos. (a) tiene el mismo esquema sintáctico que (b): Suj ODir OInd. La diferencia está que en (a) cada actante se expresa mediante una frase, mientras que en (b) OInd y ODir se expresan sólo mediante los clíticos *se* y *lo*, respectivamente, mientras que la frase que funciona como sujeto aparece pospuesta al verbo.

(5)

- a. El padre le enseñaba latín al hijo.
- b. Se lo enseñó su padre.

Otra cosa es que la información puramente categorial o funcional sea suficiente para definir los patrones valenciales admitidos por un verbo. Volveremos más adelante sobre el grado de detalle necesario.

La estrategia básica aplicada en ADESSE para anotar esquemas valenciales parte de la distinción entre potencial valencial y realizaciones valenciales (Ágel 1995). El potencial valencial de un verbo se entiende como el conjunto de argumentos semánticos admitidos por ese verbo, esto es, los participantes inherentes o compatibles con la escena evocada por el verbo. Las realizaciones valenciales son el conjunto de patrones argumentales efectivamente expresados por las estructuras sintácticas. Por ejemplo, hemos visto en la Tabla 8 actantes evocados por los verbos *ganar* y *vencer*, y el diferente grado de obligatoriedad que muestra cada uno de ellos. De ahí se deriva también que de cada verbo se encontrarán el corpus

patrones valenciales alternantes, que escogen entre los actantes posibles y focalizan algunos de ellos sobre otros. En las tablas siguientes se ofrecen, con ejemplos, las realizaciones valenciales registradas con *vencer* y *ganar-II*, y la frecuencia de cada esquema.

Tabla 9. Argumentos y esquemas sintáctico-semánticos del verbo *Vencer* en ADESSE

Potencial valencial del verbo VENCER				
A1 (Vencedor)	A2 (Vencido)	A3 (Competición)		
Realizaciones valenciales			N	Ejemplos
A1:Suj	A2:ODir		65	<i>Los liberales <u>vencieron</u> a la Iglesia</i> [TIE:151]
A1:Suj			12	<i>Habían <u>vencido</u> los buenos.</i> [USOS:013]
A1:Suj		A3:Obl(en)	2	<i>los carlistas <u>vencieron</u> en la batalla de Lácar</i> [TER:019]

Tabla 10. Argumentos y esquemas sintáctico-semánticos del verbo *Ganar-II* en voz activa en ADESSE

Potencial valencial del verbo GANAR-II “superar, vencer”				
A1 (Ganador)	A2 (Perdedor)	A3 (Objetivo)		
Realizaciones valenciales			N	Ejemplos
A1:Suj	A2:ODir		16	<i>Javier Sánchez <u>ganó</u> a Tomás Carbonell</i> [1VOZ:47]
A1:Suj			34	<i>Ustedes <u>ganan</u>, pero no tienen razón</i> [PAS:081]
A1:Suj		A3:ODir	22	<i>Guus Hiddink <u>ganó</u> cuatro Ligas</i> [1VOZ:42]
A1:Suj	A2:OInd	A3:ODir	3	<i>Nos <u>ganó</u> la guerra América</i> [MAD:278]
A1:Suj	A2:OInd	A3:Obl(a)	2	<i>A limpia no me <u>ganaba</u> nadie</i> [LAB: 269]

La observación del corpus nos muestra que es extremadamente raro que un verbo se asocie con una sola estructura sintáctica, por lo que no puede concluirse que el léxico determine unívocamente la sintaxis. Sin embargo, sobre la asociación de verbos y construcciones existen dos puntos de vista complementarios, aunque no necesariamente incompatibles. Por un lado la perspectiva lexicalista (por ejemplo, Levin 1993), que toma el punto de vista del verbo y observa *alternancias* de esquemas sintáctico-semánticos admitidos los elementos léxicos, más o menos en la línea de lo mostrado en Tabla 9 y Tabla 10. Por otro lado, tenemos la perspectiva construccionista (Goldberg 1995, 2002), que adopta el punto de vista de la construcción y busca “generalizaciones superficiales” relativas a los usos de un esquema construccionista con diferentes elementos léxicos, es decir, lo que tienen en común los ejemplos de cada una de las líneas siguientes:

(6)

Me enseñaba inglés—Me daba inglés—Me decía cosas en inglés

Me enseñó a cantar—Me ayudó a cantar—Me obligó a cantar

Cargó los tomates en el carro—Puso los tomates en el carro

Cargó el carro de tomates—Llenó el carro de tomates

Les ganó el partido—les robó el partido—le quitó el dinero—le compró el coche

Ganó al ajedrez—jugó al ajedrez

La teoría de la valencia es una teoría lexicalista, pero necesitamos construcciones, definidas con relativa independencia del léxico, para dar cuenta al menos de construcciones no previstas por el significado verbal y de actantes adicionales, no esenciales en la escena evocada por el verbo. En (7a), Goldberg utiliza un verbo básicamente intransitivo en la que llama construcción de movimiento causado [“caused motion construction”]: Subj V Obj Obl, con el significado “X causes Y move Z (by doing V)” que es aportado por la construcción y no por el verbo. En (7b) tenemos un verbo básicamente biactancial [*Romper*: <A0:Agte, A1:Pac>], que presenta un actante adicional gracias a la presencia de un dativo posesivo, lo que permite desdoblar el participante afectado con referencia al todo personal y la parte del cuerpo inmediatamente afectada:

(7)

a. *She sneezed the napkin off the table* (Goldberg 1995)

b. *Le rompió la cara*

En general, necesitaremos hacer referencia a la interacción de verbos y esquemas construccionales siempre que debamos referirnos a mecanismos de coerción en la adaptación del verbo al significado registrado para la construcción y viceversa (Welke 2009). En cualquiera de las direcciones utilizadas, verbo o construcción, encontraremos una red de sentidos relacionados, que se acomodan al significado de los elementos combinados. Los esquemas construccionales acomodan su significado primariamente al de los verbos que entran en la construcción, como podemos comprobar en la polisemia del esquema <Suj ODir OInd> en español:

(8)

- a. Vsop + N: *Eso le dio rabia*
- b. Transferencia de posesión: *Se lo entregó*
- c. Comunicación: *Se lo dijo*
- d. Causativo de proceso mental: *Se lo enseñó*
- e. Desplazamiento: *Se lo llevó*
- f. Adquisición + Origen: *Se lo robó*
- g. Beneficiario: *Se lo hizo*
- h. Dativo posesivo: *Le besó la mano*
- i. ...

Sin entrar en detalles, la base de datos sintáctico-semánticos nos podrá proporcionar también datos cuantitativos sobre qué verbos entran en cada esquema. Una técnica estadística refinada para medir esas asociaciones ha sido elaborada por Gries y Stefanowitsch, con el nombre de análisis colostruccional:

“Collostructional analysis always starts with a particular construction and investigates which lexemes are strongly attracted or repelled by a particular slot in the construction (i.e. occur more frequently or less frequently than expected); crucially, such “slots” can exist at different levels of linguistic structure” (Stefanowitsch and Gries 2003, 214)

Si observamos las estructuras argumentales desde el punto de vista del verbo, los diferentes esquemas en que encontramos un verbo suponen diferentes maneras de organizar sintáctica y semánticamente el contenido evocado. Sobre la escena evocada que sirve de base, tanto el verbo como el esquema construccional pueden focalizar o perfilar ciertas relaciones, lo que determina la probabilidad, que podemos comprobar en el corpus, de que el verbo aparezca en cierto esquema. Muchos trabajos (sobre todo desde Levin 1993) caracterizan la sintaxis de los verbos atendiendo a si admiten cierto tipo de alternancia construccional. Por ejemplo, la llamada

alternancia causativa que relaciona ciertas estructuras transitivas e intransitivas es admitida por el verbo *cambiar* pero no por el verbo *crecer*

(9)

a. Los niños cambiaron—Algo cambió a los niños

b. Los niños crecieron—*Algo creció a los niños

Pero tampoco aquí hay una dicotomía sí/no. Volvemos a tener gradaciones y cierta probabilidad de que un verbo admita alguna de varias construcciones alternativas. Lo más relevante del perfil combinatorio de un verbo consiste en los esquemas construccionales en los que entra, las realizaciones sintácticas y léxicas de sus argumentos, y las frecuencias de esquemas y realizaciones argumentales. Dos verbos de la misma clase pueden admitir combinaciones sintagmáticas similares y diferenciarse en la frecuencia relativa de cada combinación, y en consecuencia en la frecuencia relativa de sus argumentos centrales. Los cambios de estado, por ejemplo, pueden producirse de modo espontáneo por propiedades inherentes de la entidad que cambia de estado o pueden admitir en distintos grados un participante adicional que inicia el cambio de estado. En la tabla siguiente, tenemos algunos verbos de cambio de estado y su frecuencia en ADESSE con la construcción transitiva causativa (Suj=Iniciador, ObjDir=Paciente) o con la construcción intransitiva (Suj=Paciente)

Tabla 11. Transitividad de algunos verbos de cambio de estado

Verbo	Transitivo (T)	Intransitivo (I)	Ratio (T/I)	Total (T+I)
DESPERTAR	66	43	1,5	109
MEJORAR	21	14	1,5	35
AUMENTAR	31	40	0,8	71
CAMBIAR	82	140	0,6	222
EMPEORAR	2	5	0,4	7
CRECER	0	93	0,0	93

Por otro lado, debe tenerse en cuenta también que la contraposición entre verbos y construcciones es más aparente que real (Croft 2003). Ni los verbos ocurren aislados, pues de usos contextualizados abstraemos esquemas valenciales particulares y un potencial valencial, ni los esquemas sintáctico-semánticos ocurren en abstracto: de usos contextualizados abstraemos esquemas construccionales en diferentes niveles de generalidad

(esquemas valenciales de un verbo - esquemas propios de una clase - esquemas sintácticos abstractos). En cualquier nivel y vía de abstracción las realizaciones posibles muestran una distribución estadística no uniforme, de la que obtenemos indicios sobre significados centrales del esquema y posibles extensiones, es decir, sobre el potencial semántico del verbo o del esquema.

3.4. Realizaciones de argumentos

La relación entre verbos y esquemas construccionales o patrones valenciales no puede reducirse a una lista de funciones o categorías sintácticas. Determinar la valencia de un verbo consiste en especificar para cada actante cuáles son sus realizaciones sintácticas, léxicas y semánticas. Por realizaciones sintácticas entendemos además de la función sintáctica (Sujeto, Objeto, Oblicuo, ...) y la categoría sintáctica (FN, cláusula en indicativo / subjuntivo / infinitivo, ...), también cualesquiera otras propiedades formales (caso, preposición,...) que pueda presentar una distribución o restricción idiosincrática. En cuanto a las realizaciones léxico-semánticas de la valencia, habría que especificar para cada actante qué tipo de entidades, qué clases de elementos léxicos y qué elementos léxicos particulares pueden entrar en esa posición estructural.

El análisis de las combinaciones léxicas particulares (“colocaciones”) ha sido uno de los temas preferidos desde que existe la lingüística de corpus y uno en los que mejor ha demostrado su potencial (Sinclair 1991). La definición de *colocación* como “relationship of habitual co-occurrence between words” (Stubbs 1995: 23) da lugar a una metodología consistente en buscar en corpus (lematizados o no) palabras frecuentes (“*collocates*”) dentro de un cierto contexto medido en número de palabras [“span”] alrededor de un *nodo*, y en aplicar sencillos test estadísticos (MI, t-score, ...) para decidir si la asociación entre nodo y colocado es superior a lo esperado por la casualidad. El método permite obtener patrones sintagmáticos y otros aspectos bastante sutiles del perfil combinatorio de cada palabra que tomemos como nodo.

Sin embargo, hay varias facetas en las que sería mejorable el estudio de las colocaciones. En primer lugar, conviene distinguir (Bosque 2001; 2004) entre combinaciones frecuentes y combinaciones restringidas: independientemente de que la combinación *dar + un paseo* sea más frecuente de lo que se deriva de la frecuencia de cada palabra por separado, es ésa una asociación convencional en español y no **hacer un paseo*, a diferencia de lo que ocurre en otras lenguas (francés “faire une promenade”, italiano “fare una passeggiatina”). Es decir, entre las

ocupantes frecuentes de una posición argumental podemos tener conceptos típicamente asociados a una situación, derivados de nuestro conocimiento del mundo (por ejemplo, “leer”+“libro”) y palabras que se asocian convencionalmente con combinaciones restringidas. En segundo lugar, conviene ir más allá de las listas de combinaciones frecuentes de palabras y definir tipos semánticos o conjuntos léxicos [“lexical sets”] (Hanks 1996) para caracterizar las posibilidades combinatorias de verbos y otras palabras. En tercer lugar, y vinculado con lo anterior, conviene ir más allá de allá de la búsqueda de colocaciones en un entorno medido en número de palabras, para buscar más bien “structure-sensitive collocates” (Stefanowitsch and Gries 2009, § 2), es decir, colocaciones que sean sensibles a la estructura sintáctica (Suj, Obj, ...) y/o semántica (Arg1, Arg2, ...).

Este último aspecto es especialmente relevante para la descripción de la valencia verbal. Si quisiéramos buscar qué posibles nombres admiten los verbos *ganar* y *vencer* como objeto directo, podríamos utilizar un corpus con anotación morfosintáctica como el *Corpus del español* (<http://corpusdelespanol.org/>) y buscar nombres que aparecieran entre las 4 o 5 palabras a la derecha del verbo. Encontraríamos entre los resultados más frecuentes que *vencer* se combina con *dificultad*, *obstáculo*, *resistencia*, *fuerza*, *repugnancia*, *batalla*, *enemigo*, etc... Todo esto es información contextual altamente relevante, que diferencia *vencer* de verbos casi-sinónimos como *ganar* (que se combina más bien con *dinero*, *elección*, *tiempo*, *premio*, etc...)

El problema con este tipo de resultados, aparte de los habituales e inevitables errores de etiquetación y lematización automática, es que no son sensibles a la estructura sintáctica, pues los nombres a la derecha [span=4] del verbo son en general ODIR, pero también hay SUJETO y elementos no argumentales.

(10)

- a. que ha durado la guerra de tierra tres días, que en seguida han **vencido** las **fuerzas** aliadas, y claro, ...
- b. Venus Williams, que la ha **vencido** dos **veces** esta temporada ...
- c. Mañana **vence** el **plazo** que tenemos.

Además, las búsquedas en corpus que tienen anotación morfosintáctica no pueden ser sensibles a diferencias de significado (ni en la colocación ni en el colocativo). Una imagen más elaborada de la combinatoria de las palabras lo obtenemos de los “Word Sketches” (Kilgarrif *et al.* 2004), que proporciona listas de colocados en diferentes huecos [“slots”] sintácticos

(“subject_np”, “object_np”) pero finalmente encontraremos los mismos problemas si se identifican automáticamente como objeto los nombres que aparecen a la derecha del verbo y como sujeto los que aparecen a la izquierda.

Por otro lado, no basta tampoco con encontrar colocativos en cada posición argumental, tenemos que dar cuenta de otras probabilidades de coocurrencia, comprobables en corpus, de entidades de diferente nivel de generalidad, como las categorías sintácticas o qué tipos semánticos caben en una posición argumental. En la(s) tabla(s) siguiente(s) se muestra el tipo de información combinatoria que puede obtenerse directamente de ADESSE (otras propiedades combinatorias que pueden derivarse de la base de datos no aparecen reflejadas aquí).

Tabla 12. Resumen comparativo del perfil combinatorio de Ganar-II y Vencer-I en ADESSE

GANAR II

“Superar o aventajar
[alguien] ([a alguien]) ([una competición])”
80 ejemplos anotados

A1 Ganador 80/80 (100%)	A2 Perdedor 23/80 (28%)	A3 Objetivo 28/80 (35%)
SUJ 77 AGT 3	ODIR 16 OIND 5 SUJ 2	ODIR 25 OBL 2 SUJ 1
	Ac/Dat 7 Dat (<i>le</i>) 5	Acus(<i>lo</i>) 1
<i>por</i> 3	<i>a</i> 13	<i>a</i> 2
FN 36 Pro 6	FN 13	FN 25 ADJ 1 Inf 1
Anim. 75 Abstr 5	Anim 21 Abstr 2	Abstr 25 Prop. 2
<i>curiosidad</i> 1 <i>apatía</i> 1 <i>chiismo</i> 1 <i>expansión</i> 1	<i>insomnio</i> 1 <i>semana_s-</i> 1	<i>guerra</i> 7 <i>partido</i> 4 <i>batalla</i> 2 <i>carrera</i> 2 ...

VENCER I

“Superar [alguien (o algo)] [a un adversario o una dificultad]”
89 ejemplos anotados

A1 Vencedor 86/89 (96%)	A2 Vencido 75/89 (84%)	A3 Objetivo 2/89 (2%)
SUJ 79 AGT 7	ODIR 65 SUJ 10	LOC 2
	Dat (<i>le</i>) 6 Ac/Dat 6 Acus (<i>lo</i>) 6	
<i>por</i> 7	<i>a</i> 35	<i>en</i> 2
FN 39 Pro 3	FN 49	FN 2
Anim. 68 Abstr. 17 Concr. 1	Animado 52 Abstracto 17 Concreto 6	Abstracto 2
<i>sueño</i> 4 <i>fatiga</i> 2 <i>cansancio</i> 2 <i>tensión</i> 1 ...	<i>obstáculo</i> 3 <i>tanque</i> 3 <i>timidez</i> 2 <i>nieve</i> 2 <i>dificultad</i> 1 ...	<i>partido</i> 1 <i>batalla</i> 1

La Tabla 12 resume las principales diferencias combinatorias entre los verbos cuasi-sinónimos *ganar* y *vencer* (quedan aparte otros significados de *ganar* y *vencer* que corresponden a otras entradas léxicas) tal como están registrados en ADESSE. Estas diferencias comienzan por la diferencias de frecuencia de los tres argumentos básicos admitidos por cada uno, y continúan en la mayor propensión del segundo argumento de *ganar* hacia el CIND/dativo (*ganarle a alguien* vs. *vencerlo / vencerle*), en la selección de preposición para el argumento A3 (*ganar a algo* vs. *vencer en algo*) y en la rareza de un argumento A2 inanimado con *ganar*, mientras que se *vencen* no sólo personas sino entidades concretas utilizadas por las personas como los tanques, fenómenos meteorológicos como la nieve y sobre todo situaciones abstractas negativas como obstáculos y dificultades.

Con todo ello, nos aproximamos a una caracterización más precisa del perfil combinatorio de los verbos del español, entendiendo por tal el conjunto de probabilidades de coocurrencia con elementos léxicos y gramaticales de cualquier nivel de generalidad. También damos pasos en la dirección deseable de sustituir los esquemas valenciales por “marcos predicativos enriquecidos” (Butler 2001), que den cuenta de los elementos léxicos y gramaticales que se registran el corpus en combinación con un predicado junto con las probabilidades de aparición de cada elemento en un hueco estructural.

Notes

¹ El proyecto se ha beneficiado de las ayudas ministeriales de referencia BFF2002-01197 (2002-2005), HUM2005-01573 (2006-2008) y FFI2008-01953 (2009-2011)

References

- Ágel, Vilmos. “Valenzrealisierung, Grammatik und Valenz.” *Zeitschrift für germanistische Linguistik* 23 (1995): 2-32.
- Albertuz, Francisco J. “Syntaxis, semántica y clases de verbos. Clasificación verbal en el proyecto ADESSE.” In *Actas del VI Congreso de Lingüística General, Santiago de Compostela, 3-7 Mayo 2004*, edited by Pablo Cano López, vol. 2-2, 2015-2030. Madrid: Arco Libros, 2007.
- Blumenthal, Meter, and Giovanni Rovere. *Wörterbuch der italienischen Verben / Dizionario dei verbi italiani*. Stuttgart: Klett, 1998.
- Borba, Francisco da Silva. *Dicionário gramatical de verbos do Português Contemporâneo do Brasil*. São Paulo: UNESP, 1990.

- Bosque, Ignacio. *REDES. Diccionario combinatorio del español contemporáneo*. Madrid: Editorial SM, 2004.
- . "Sobre el concepto de 'colocación' y sus límites." *Lingüística Española Actual* 23/1 (2001): 9–40.
- Burchardt, Aljoscha, Katrin Erk, Anette Frank, Andrea Kowalski, Sebastián Padó, and Manfred Pinkal. "The SALSA Corpus: a German Corpus Resource for Lexical Semantics." *Proceedings of LREC 2006* (2006): 969–974.
- Busse, Winfried, and Jean-Pierre Dubost. *Französisches Verblexikon. Die Konstruktion der Verben im Französischen*. Stuttgart: Klett, 1977.
- Busse, Winfried. *Dicionário sintático de verbos portugueses*. Coimbra: Almedina, 1994.
- Butler, Christopher S. "A matter of GIVE and TAKE: corpus linguistics and the predicate frame." *Revista Canaria de Estudios Ingleses* 42 (2001): 55–78.
- Chomsky, Noam. *Aspects of the Theory of Syntax*. Cambridge: The MIT Press, 1965.
- Croft, William. "Lexical rules vs. constructions: a false dichotomy." In *Motivation in Language. Studies in honor of Günter Radden*, edited by Hubert Cuyckens, Thomas Berg, René Dirven, and Klaus-Uwe Panther, 49–68. Amsterdam: John Benjamins, 2003.
- Cuervo, Rufino J. *Diccionario de construcción y régimen de la lengua castellana* (8 vol.). Santa Fe de Bogotá: Instituto Caro y Cuervo, 1987–1994.
- Dik, Simon C. *The Theory of Functional Grammar (Part I: The Structure of the clause)*. Dordrecht: Foris, 1989.
- Domínguez, M. J., Barbara Lübke, Meike Meliss, Gemma Paredes, Pía Poulsen, and Victoria Vázquez. "Presentación del Diccionario contrastivo de valencias verbales español-alemán." In *La lexicografía bilingüe y didáctica: ayer y hoy*, edited by Mar Campos, Eva M. González, J. Ignacio Pérez, 51–61. A Coruña: Universidade da Coruña, 2008.
- Engel, Ulrich, and Helmut Schumacher. *Kleines Valenzlexikon deutscher Verben*. Tübingen: Narr, 1976.
- Fillmore, Charles J. "Corpus linguistics" or "Computer-aided armchair linguistics". In *Directions in Corpus Linguistics*, edited by Jan Svartvik, 35–60. Berlin: Mouton de Gruyter, 1992.
- . "Valency issues in FrameNet." In *Valency. Theoretical, Descriptive and Cognitive Issues*, edited by Thomas Herbst, and Katrin Götz-Votteler, 129–160. Berlin: Mouton de Gruyter, 2007.

- Fillmore Charles J., Christopher Johnson, and Miriam Petruck. "Background to FrameNet." *International Journal of Lexicography* 16/3 (2003): 235-250.
- García-Miguel, José M., and Gael Vaamonde. "Semantic role annotation: From verb-specific roles to generalized semantic roles". *Interdisciplinary Workshop on verbs. The Identification and Representation of Verbs Features*, November 4-5, 2010.
http://linguistica.sns.it/Workshop_verb/papers/Garcia-Miguel_verb2010_submission_26.pdf
- Garrudo, Francisco. *Diccionario sintáctico del verbo inglés*. Barcelona: Ariel, 1991.
- Goldberg, Adele E. *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago/London: University of Chicago Press, 1995.
- . "Surface generalizations: An alternative to alternations." *Cognitive Linguistics* 13/4 (2002): 327-356.
- . "Constructions: a new theoretical approach to language." *Trends in Cognitive Sciences* 7/5 (2003): 219-224.
- Gries, Stefan T. "Introduction". In *Corpora in Cognitive Linguistics. Corpus-Based Approaches to Syntax and Lexis*, edited by Stefan T. Gries, and Anatol Stefanowitsch, 1-18. Berlin: Mouton de Gruyter, 2006.
- Hajič, Jan. "Complex Corpus Annotation: The Prague Dependency Treebank". *Jazykovedný ústav L. Štúra, SAV*, 2004.
<http://ufal.mff.cuni.cz/pdt2.0/publications/Hajic2004.pdf>
- Halliday, Michael A. K. *An Introduction to Functional Grammar*. London: Edward Arnold, 1985.
- Hanks, Patrick. "Contextual dependency and lexical sets." *International Journal of Corpus Linguistics* 1/1 (1996): 75-98.
- Hanks, Patrick, and James Pustejovsky. "A Pattern Dictionary for Natural Language Processing." *Revue Française de Linguistique Appliquée* 10/2 (2005): 63-82.
- Helbig, Gerard. *Probleme der Valenz- und Kasus-theorie*. Tübingen: Niemeyer, 1992.
- Helbig, Gerard, and Wolfgang Schenkel. *Wörterbuch zur Valenz und Distribution deutscher Verben*. Leipzig: Bibliographisches Institut, 1969.
- Herbst, Thomas. "Valency complements of valency patterns?" In *Valency: Theoretical, Descriptive and Cognitive Issues*, edited by Thomas Herbst, and Katrin Götz-Votteler, 15-35. Berlin: Mouton de Gruyter, 2007.

- Herbst, Thomas, David Heath, Ian Roe, and Dieter Götz. *A Valency Dictionary of English. A Corpus-Based Dictionary of the Complementation Patterns of English Verbs, Nouns and Adjectives*. Berlin/New York: Mouton de Gruyter, 2004.
- Hunston, Susan, and Gill Francis. *Pattern grammar: a corpus-driven approach to the lexical grammar of English*. Amsterdam: John Benjamins, 2000.
- Kilgarriff, Adam, Pavel Rychly, Pavel Smrz, and David Tugwell. "The Sketch Engine." In *Proceedings of the EURALEX Conference*, 105-116. Lorient, France: Université de Bretagne- Sud, 2004.
- Korhonen, Anna, Yuval Krymolowski, and Ted Briscoe. "A Large Subcategorization Lexicon for Natural Language Processing Applications." *Proceedings of the 5th Language Resources and Evaluation Conference (LREC)*, May 22-28, 2006, <http://www.lrec-conf.org/proceedings/lrec2006/>.
- Leech, Geoffrey. "Corpora and theories of linguistic performance." In *Directions in corpus linguistics*, edited by Jan Svartvik, 105-122. Berlin and New York: Mouton de Gruyter, 1992.
- Levin, Beth. *English Verb Classes and Alternations: a Preliminary Investigation*. Chicago/London: University of Chicago Press, 1993.
- Luft, Celso P. *Dicionário prático de regência verbal*. São Paulo: Editora Ática, 1995.
- Manning, Christopher. "Probabilistic Syntax." In *Probabilistic Linguistics*, edited by Rens Bod, Jennifer Hay, and Stefanie Jannedy, 289-341. Cambridge, MA: MIT Press, 2003.
- Mel'čuk, Igor. *Dictionnaire explicatif et combinatoire du français contemporain. Recherches lexico-sémantiques I-IV*. Montreal: Les Presses de l'Université de Montréal, 1984-1999.
- Nižníková, Jolana, and Miloslava Sokolová. *Valenčný slovník slovenských slovies* (The Valency Dictionary of the Slovak Verbs). Prešov: Filozofická fakulta Prešovskej univerzity, 1998.
- Palmer, Martha, Paul Kingsbury, and Dan Gildea. "The Proposition Bank: An Annotated Corpus of Semantic Roles." *Computational Linguistics* 31/1 (2005): 71-105.
- Polanski, Kazimierz *et al.* *Słownik syntaktyczno-generatywny czasowników polskich*. Wrocław: Wydawnictwo Polskiej Akademii Nauk, 1980-1992.
- Rojo, Guillermo. "La explotación de la Base de Datos Sintácticos del español actual." In *Lingüística con corpus*, edited by Josse De Kock, 255-286. Salamanca: Universidad de Salamanca, 2001.

- Schulte im Walde, Sabine. "The induction of verb frames and verb classes from corpora." In *Corpus Linguistics. An International Handbook*, edited by Anke Lüdeling, and Merja Kytö, vol. 2, 952–972. Berlin: Walter de Gruyter, 2009.
- Schumacher, Helmut, Jacqueline Kubczak, Renate Schmidt, and Vera De Ruiter. *VALBU—Valenzwörterbuch deutscher Verben*. Tübingen: Narr, 2004.
- Sinclair, John. *Corpus, concordance, collocation*. Oxford: Oxford University Press, 1991.
- Stefanowitsch, Anatol, and Stefan T. Gries. "Collostructions: Investigating the interaction of words and constructions." *International Journal of Corpus Linguistics* 8/2 (2003): 209–243.
- . "Corpora and Grammar." In *Corpus Linguistics. An International Handbook*, edited by Anke Lüdeling, and Merja Kytö, (2) 933–952. Berlin: Walter de Gruyter, 2009.
- Stubbs, Michael. "Collocations and semantic profiles. On the cause of trouble with quantitative studies." *Functions of Language* 2/1 (1995): 23–55.
- Subirats, Carlos. "Spanish Framenet: A frame-semantic analysis of the Spanish lexicon." In *Multilingual FrameNets in Computational Lexicography. Methods and Applications*, edited by Hans Boas, 135–162. Berlin/New York: Mouton de Gruyter, 2009.
- Taulé, Mariona, M. Antonia Martí, and Marta Recasens. "Ancora: Multilevel Annotated Corpora for Catalan and Spanish." *Proceedings of 6th International Conference on Language Resources and Evaluation* (2008): 96–101.
- Tesnière, Lucien. *Éléments de syntaxe structurale*. Paris: Klincksieck, 1959.
- Teubert, Wolfgang. "Sinclair, pattern grammar and the question of hatred." *International Journal of Corpus Linguistics* 12/2 (2007): 223–248.
- Thompson, Sandra A., and Paul J. Hopper. "Transitivity, clause structure, and argument structure: evidence from conversation." In *Frequency and the emergence of linguistic structure*, edited by Joan Bybee, and Paul J. Hopper, 27–60. Amsterdam: John Benjamins, 2001.
- Tognini-Bonelli, Elena. *Corpus Linguistics at Work*. Amsterdam: John Benjamins, 2001.
- Vázquez, Gloria, Laura Alonso, J.A. Capilla, Irene Castellón, and Ana Fernández. "SenSem: sentidos verbales, semántica oracional y anotación de corpus." *Procesamiento del Lenguaje Natural* 37 (2006): 113–120.

- Welke, Klaus. "Valenztheorie und Konstruktionsgrammatik." *Zeitschrift für germanistische Linguistik* 37/1 (2009): 81–124.
- Zabokrtsky, Zdenek, and Markéta Lopatkova. "Valency Information in VALLEX 2.0: Logical Structure of the Lexicon." *The Prague Bulletin of Mathematical Linguistics* 87/June (2007): 41–60.

Encoding the Past, Decoding the Future:
Corpora in the 21st Century

Edited by

Isabel Moskowich and Begoña Crespo

**CAMBRIDGE
SCHOLARS**

P U B L I S H I N G

Encoding the Past, Decoding the Future:
Corpora in the 21st Century,
Edited by Isabel Moskowich and Begoña Crespo

This book first published 2012

Cambridge Scholars Publishing

12 Back Chapman Street, Newcastle upon Tyne, NE6 2XX, UK

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

Copyright © 2012 by Isabel Moskowich and Begoña Crespo and contributors

All rights for this book reserved. No part of this book may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the copyright owner.

ISBN (10): 1-4438-3581-1, ISBN (13): 978-1-4438-3581-7