

# Similitud verbal: Análisis comparativo entre lingüística teórica y datos extraídos de corpus

*Verb similarity: Comparative analysis between theoretical linguistics and corpus data*

**Lara Gil-Vallejo**

UNIVERSITAT OBERTA DE CATALUNYA  
ESPAÑA  
lgilva@uoc.edu

**Irene Castellón**

UNIVERSITAT DE BARCELONA  
ESPAÑA  
icastellon@ub.edu

**Marta Coll-Florit**

UNIVERSITAT OBERTA DE CATALUNYA  
ESPAÑA  
mcollfl@uoc.edu

**Recibido:** 12-IX-2016 / **Aceptado:** 25-X-2017

## Resumen

La noción de similitud es crucial para un número de disciplinas del ámbito cognitivo y lingüístico. Sin embargo, se trata de un concepto amplio que se puede tratar desde diversas perspectivas de estudio. En el presente trabajo realizamos un análisis comparativo de cómo dos de estas perspectivas articulan la similitud entre los sentidos verbales en español. La primera perspectiva está fundamentada en la lingüística de corpus, específicamente en la estructura argumental asociada a sentidos verbales y capturada mediante los roles semánticos anotados en un corpus del español. La segunda perspectiva se enmarca en la lingüística teórica, y se basa en las diferentes construcciones en las que los verbos pueden participar. Mediante esta comparación se aborda la cuestión de si la estructura argumental obtenida a través de ejemplos de corpus, que refleja la expresión de los eventos, configura la similitud entre los sentidos verbales de forma congruente a como lo hacen las construcciones, a las que la teoría lingüística asigna también contenido argumental y considera piezas clave en la construcción del significado oracional. Los resultados de la comparación muestran una correlación significativa entre la caracterización de similitud verbal de ambas perspectivas, particularmente cuando se tiene en cuenta información sintáctica explícita y patrones argumentales en el corpus. Además, los resultados muestran que el eje similitud-disimilitud articulado por las construcciones está vertebrado en torno a la oposición aspectual estatividad-dinamicidad.

**Palabras Clave:** Similitud verbal, estructura argumental, construcciones, roles semánticos, eventos.

## Abstract

The concept of similarity is crucial for fields belonging to cognitive science and linguistics. However, it is a broad notion that can be approached through multiple perspectives. In this work we carry out a comparative analysis on how two of these perspectives articulate similarity between verb senses in Spanish. The first perspective is based on corpus linguistics, specifically on semantic roles annotated on a Spanish corpus, as representation of argument structure associated to verb senses. The second perspective belongs to the field of theoretical linguistics and is embodied by the constructions that a verb instantiates. With such comparison we aim to investigate whether argument structure obtained from corpus, which reflects the expression of an event, configures verb similarity in a manner congruent with the way in which constructions do it, which are argued to contain argument structure too and to constitute key elements in the generation of sentence meaning. The results of the comparison show that there is a significant correlation between the similarity formalization of both perspectives, particularly when we take into account syntactic and pattern information from corpus. Additionally, the results attribute particular importance to the aspectual opposition of stativity-dynamicity as the backbone of the similarity-dissimilarity axis for constructions.

**Key Words:** Verb similarity, argument structure, constructions, semantic roles, events.

## INTRODUCCIÓN

El concepto de similitud se emplea de forma generalizada en las teorías de la cognición, en las representaciones psicolingüísticas del lexicon y en la adquisición de modelos del lenguaje en el área del Procesamiento del Lenguaje Natural. Sin embargo, a pesar de la productividad de este concepto, existe un problema fundamental en la definición de las características que sirven de base para establecer la relación de similitud entre dos objetos. En efecto, hay un gran número de aspectos no triviales que son susceptibles de ser tenidos en cuenta para calcular la similitud entre dos palabras, conceptos, frases, etc. (De Deyne, Peirsman & Storms, 2009 para una discusión en el ámbito semántico). En consecuencia, ante la dificultad y la complejidad que supone establecer una definición de similitud en el dominio que nos ocupa, surge la necesidad de realizar estudios más exhaustivos sobre esta noción.

El objetivo de este trabajo es obtener un mayor conocimiento de los mecanismos que determinan la similitud entre sentidos verbales. Para ello, realizamos un análisis de la similitud verbal desde dos perspectivas diferentes: la perspectiva de la lingüística teórica basada en construcciones y la perspectiva de la lingüística de corpus basada en la anotación de roles semánticos. En concreto, nuestro objetivo es doble: por un lado, pretendemos evaluar hasta qué punto se da una correlación entre la similitud verbal calculada a partir del juicio de lingüistas expertos y la similitud calculada a partir de datos reales extraídos de corpus; por otro lado, pretendemos analizar qué tipo de información lingüística es relevante a la hora de determinar la similitud entre sentidos verbales. Según la revisión bibliográfica, no se ha hecho ningún estudio comparativo

de estas perspectivas hasta la fecha. Asimismo, creemos que un mejor conocimiento de los puntos de convergencia y divergencia entre estas dos perspectivas puede aportar una visión más amplia de la similitud verbal que a su vez contribuya a la creación de un modelo más robusto de similitud.

El artículo se organiza como sigue. En la sección 1 se presenta la noción de similitud como concepto fundamental para enfoques lingüísticos y cognitivos. A continuación, la sección 2 describe los materiales empleados para este estudio. Las secciones 3 y 4 se ocupan de la perspectiva de corpus y de construcciones lingüísticas respectivamente. La comparación entre ambas perspectivas y el análisis de resultados se llevan a cabo en la sección 5. Finalmente, se presentan las conclusiones que se derivan de este análisis.

## **1. Marco teórico: La similitud como noción esencial**

Desde el punto de vista de los estudios que se enmarcan en la ciencia cognitiva, la similitud representa un mecanismo explicativo básico en numerosos modelos de generación de inferencias, solución de problemas, predicción y categorización. No obstante, no hay una teoría unificada sobre cómo debe definirse y caracterizarse la similitud entre representaciones mentales. Por el contrario, hay un número de teorías divergentes acerca de cómo debe ser formalizada y medida (Goldstone & Son, 2005) que han ejercido una gran influencia sobre los modelos psicolingüísticos y computacionales de la similitud.

Por otro lado, una parte de la investigación en psicolingüística analiza la importancia de la similitud en la construcción del lexicón mental desde diferentes ámbitos entre los que podemos destacar experimentos de facilitación semántica (Neely, 1991), sintáctica (Savage, Lieven, Theakston & Tomasello, 2003) y fonética (Vitevitch & Luce, 2016).

En cuanto a los modelos computacionales del lenguaje creados mediante las herramientas del Procesamiento del Lenguaje Natural, la noción de similitud tiene una importancia crucial, ya que se emplea para cuantificar las relaciones entre unidades lingüísticas, algo esencial para una gran variedad de tareas: desambiguación de palabras (Sinha & Mihalcea, 2007), detección de paráfrasis (Fernando & Stevenson, 2008), etc. La gran variedad de relaciones entre unidades lingüísticas que pueden tenerse en cuenta para calcular la similitud pone de relieve que la falta de una definición precisa de la similitud tiene consecuencias importantes en la evaluación de los modelos computacionales.

## **2. Marco metodológico**

Para la parte de estudio en corpus realizada en este trabajo nos basamos en los datos del corpus SenSem (Fernández-Montraveta & Vázquez, 2014). Este corpus contiene 250 lemas verbales desambiguados manualmente en sus correspondientes

sentidos (un total de 980), que a su vez están anotados con el identificador de *synset* de WordNet (Miller, 1995). A cada uno de los lemas verbales están asociadas manualmente más de 100 frases anotadas a nivel sintáctico (información morfosintáctica y función sintáctica) y semántico (roles semánticos, aspecto). En total, el corpus incluye 999.712 palabras.

El material para realizar el análisis es un conjunto de sentidos verbales en español (véase en el Anexo 1 la lista con los sentidos verbales escogidos). Los sentidos verbales, a diferencia de los lemas, permiten obtener modelos más precisos de similitud, ya que diferentes sentidos de un mismo lema en muchas ocasiones se comportan de forma divergente en los planos semántico y sintáctico (Hare, McRae & Elman, 2003). Además, hay evidencia experimental que apunta al hecho de que los diferentes sentidos de una palabra tienen un impacto en la percepción de la similitud (Klein & Murphy, 2002).

En cuando a la selección de los sentidos verbales específicos usados para el presente estudio, escogemos 20 sentidos del corpus SenSem entre aquellos que cumplen los siguientes requisitos: 1) tienen más de 10 frases asociadas en el corpus; 2) pertenecen a una variedad de campos semánticos, tal y como se definen en dos recursos diferentes: *supersenses* de WordNet, en concreto el denominado *lexicographer file* (Fellbaum, 1998) y las macro-clases definidas en el corpus Adesse (García-Miguel, González Domínguez & Vaamonde, 2010), Albertuz (2007) para una explicación de las clases; 3) muestran una amplia variedad de patrones de subcategorización (SN, SP, SAdv, etc.) y roles asociados (agente, tema, experimentador, instrumento, etc.). Estos criterios están diseñados con el objetivo de optimizar la representación de diferentes características lingüísticas en la muestra seleccionada. En el Anexo 2 especificamos los diferentes campos semánticos que corresponden a cada sentido verbal según las categorías de los *supersenses* de WordNet y las macro-clases de Adesse.

Los datos que sirven de base para nuestro análisis son el conjunto de todas las posibles parejas de sentidos verbales que podemos crear combinando los 20 sentidos seleccionados previamente. Estas parejas no contienen sentidos repetidos (evitamos pares como ‘abrir-abrir’) y en ellas el orden de los sentidos no importa (‘abrir-crecer’ es el mismo par que ‘crecer-abrir’), ya que empleamos medidas simétricas para calcular la similitud. Como resultado, obtenemos 190 pares de sentidos verbales, lo que agota todas las combinaciones posibles. De esta manera es posible calcular de forma efectiva la similitud entre todos los sentidos verbales.

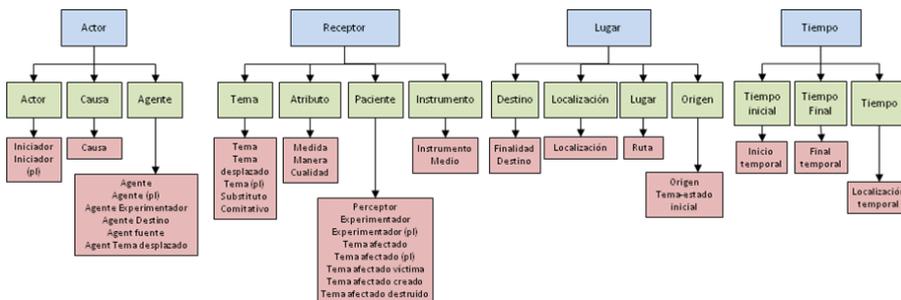
### **3. Perspectiva de corpus**

La aproximación a la similitud verbal en corpus se realiza sobre la base de la información que nos proporciona el corpus SenSem. Específicamente, emplearemos la información anotada relativa a categoría sintáctica y de roles semánticos de los

argumentos asociados a los sentidos verbales, con el objetivo de capturar la estructura argumental de las frases en las que participan los sentidos verbales seleccionados.

Consideramos que la estructura argumental es relevante en la caracterización de la similitud verbal por dos razones. Por un lado, en lingüística se considera usualmente que la estructura argumental se encuentra en la interfaz sintáctico-semántica y que es un elemento clave para los verbos, ya que hace referencia a cuántos y de qué tipo son los argumentos que se relacionan con un predicado. Por otro lado, se considera que la estructura argumental representa la codificación lingüística del evento expresado en la frase. Desde el punto de vista de los estudios sobre la organización del lexicón, la información argumental relativa a los eventos ha ido ganando importancia en su papel como cimentadora del conocimiento lingüístico (Jones, Willis & Dennis, 2015; Ferretti, McRae & Hatherell, 2001; Chwilla & Kolk, 2005).

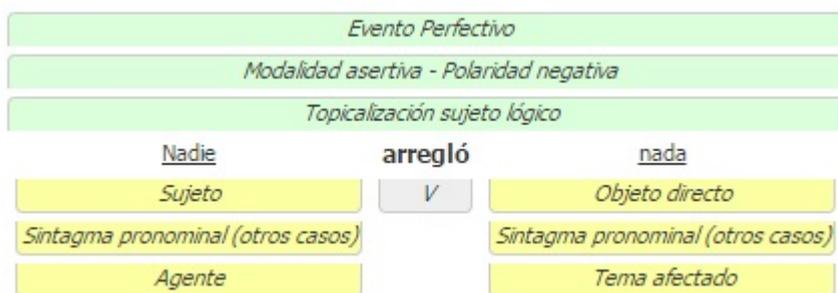
Respecto a la metodología para capturar la estructura argumental de manera empírica, en este trabajo recurrimos a las anotaciones de roles semánticos en el corpus SenSem. En cuanto a la importancia de los roles semánticos, debe mencionarse que fueron introducidos en la lingüística moderna por los trabajos de Gruber (1965), Fillmore (1967) y Jackendoff (1972), que propusieron esta categoría como mediadora entre el contenido sintáctico y el semántico en la expresión de los argumentos, capaz de definir la relación existente entre cada argumento y su respectivo predicado. Aunque se han revelado como una herramienta muy útil a la hora de caracterizar las lenguas desde una perspectiva sintáctico-semántica, existe una variedad de propuestas de conjuntos de roles que difieren en la cantidad y tipo de argumentos que se emplean a la hora de identificar participantes en eventos. Podemos nombrar como propuestas principales el trabajo de Fillmore (1967) y su influencia en la anotación de VerbNet (Kipper, 2005), la propuesta de PropBank (Palmer, Gildea & Kingsbury, 2005), la de Framenet (Fillmore, Johnson & Petruck, 2003) y finalmente la de LIRICS (Bonial, Corvey, Palmer, Petukhova & Bunt, 2011). Esta última se perfila como un estándar flexible para dar cabida a las necesidades de modelos del lenguaje enfocados a diversas tareas. Así, se propone una organización jerárquica de los roles basada en varios niveles diferentes de abstracción (los roles más abstractos como ‘actor’ subsumen roles más específicos como ‘agente’ y ‘causa’). Consideramos que este último enfoque es el adecuado para nuestras necesidades y objetivos, ya que la estructura jerárquica nos permite experimentar con diferentes grados de abstracción de la información semántica contenida en los roles.



**Figura 1.** Jerarquía de roles semánticos.

Por ello, creamos una jerarquía de tres niveles de abstracción sobre la base de los roles semánticos presentes en SenSem, que situamos en el nivel más específico. En lo referente a los niveles intermedio y abstracto, estos se generan realizando un *mapping* entre los roles de SenSem y los del nivel básico de la propuesta de LIRICS. La jerarquía obtenida se muestra en la Figura 1, donde el primer nivel corresponde a los roles más abstractos, el segundo nivel a los intermedios y el tercero a los más específicos. Como resultado obtenemos tres niveles de abstracción de roles semánticos que nos permiten formalizar la estructura argumental de manera flexible.

La información anotada del corpus que usamos para caracterizar los sentidos verbales consta, por un lado, de estos tres niveles de roles semánticos y, por otro lado, de esos mismos roles combinados con la función sintáctica asociada al argumento formando una única categoría (ej. receptor-sujeto, receptor-objeto, etc.). Estos dos tipos de caracterización capturan la estructura argumental de las frases, a la vez que permiten explorar los efectos de añadir información sintáctica explícita en el cálculo de la similitud.



**Figura 2.** Anotación de SenSem de 'Nadie arregló nada'.

**Tabla 1.** Ejemplo de atributos obtenidos del corpus.

	Constituyentes		Patrones	
	Roles	Roles + Sintaxis	Roles	Roles + Sintaxis
<b>Nivel específico</b>	Agente   Tema afectado	Agente-Sujeto   Tema afectado-Obj. Dir.	Agente+Tema afectado	Agente-Sujeto+Tema afectado-Obj. Dir.
<b>Nivel intermedio</b>	Agente   Tema	Agente-Sujeto   Tema-Obj. Dir.	Agente+Tema	Agente-Sujeto+Tema-Obj. Dir.
<b>Nivel abstracto</b>	Actor   Receptor	Actor-Sujeto   Receptor-Obj. Dir.	Actor+Receptor	Actor-Sujeto+Receptor-Obj. Dir.

Por otra parte, la información extraída de SenSem sobre los roles semánticos y sobre roles semánticos más sintaxis se formaliza de dos maneras diferentes: a partir de ‘constituyentes’, para lo cual se considera cada argumento individualmente y se obtienen categorías como ‘agente’, ‘paciente’ o ‘agente-sujeto’, ‘paciente-objeto’ si añadimos la información sintáctica; y a partir de ‘patrones’, para lo que se tiene en cuenta la anotación de la estructura argumental completa, obteniendo categorías como ‘agente+paciente’, ‘agente-sujeto+paciente-objeto’, etc. Esta formalización está ilustrada en la Figura 2, con un ejemplo de una frase con la anotación existente en SenSem, y en la Tabla 1, donde se especifican los atributos que se obtienen de ella para caracterizar los sentidos.

Por lo tanto, para cada frase tenemos 12 tipos de formalización diferentes dependiendo del grado de abstracción de los roles semánticos, la presencia o ausencia de sintaxis y la organización de la información argumental en constituyentes o patrones. De acuerdo con esto, se obtienen 12 representaciones vectoriales para cada sentido verbal, una por cada tipo de formalización. Estos vectores contienen todos los atributos existentes en la anotación de las frases en las que participan los sentidos verbales y caracterizan estos sentidos recogiendo la probabilidad de concurrencia de cada sentido verbal con cada atributo. Una vez caracterizados los verbos de esta manera, procedemos a compararlos entre sí, usando las representaciones vectoriales. Dado que previamente habíamos creado el conjunto de todos los posibles pares de sentidos verbales, obtenemos valores de similitud entre todos los sentidos. Para obtener estos valores usamos la distancia del coseno con las representaciones vectoriales. Esta distancia tiene la ventaja de ser independiente de la magnitud, es decir, la frecuencia de los sentidos no influye en el cómputo de la similitud.

En definitiva, obtenemos los valores de similitud entre todos los sentidos verbales de acuerdo con las diferentes maneras en las que la estructura argumental en corpus es capturada. Como resultado, generamos 12 rankings de similitud, uno por cada tipo de formalización de la estructura argumental. En estos rankings los pares de sentidos se ordenan de más a menos similares de acuerdo con los valores obtenidos y son

comparados con los rankings obtenidos empleando la información sobre construcciones. Los resultados de esta comparación se presentan en la sección 5.

#### **4. Perspectiva de la lingüística teórica basada en construcciones**

En esta sección presentamos el enfoque que, desde el punto de vista de la lingüística teórica, tomamos para definir y calcular la similitud verbal. La idea básica de que la sintaxis y la semántica están indisolublemente unidas es la base de enfoques tan diversos como la gramática de construcciones (Goldberg, 1995) y las teorías lexicalistas (Levin, 1993). Es precisamente este último enfoque el que tradicionalmente se ha empleado para definir grupos de verbos similares. Particularmente, Levin (1993) sostiene que la naturaleza de la relación entre la sintaxis y la semántica verbales es tal que el significado de un verbo determina en gran manera su comportamiento, entendido este como la expresión e interpretación de sus argumentos. Esta idea ha sido extensamente utilizada para definir la similitud entre verbos, agrupándolos con base a su comportamiento sintáctico para obtener clases semánticamente homogéneas. Respecto a la cuestión de qué se entiende por comportamiento verbal, Levin (1993) lo formaliza usando las llamadas alternancias de diátesis, definidas como alternancias en la expresión de los argumentos, acompañadas en ciertos casos por cambios de significado en la frase que nunca conllevan cambios en el significado verbal. Un ejemplo clásico es la alternancia causativa / incoativa ('Joana rompió el vaso / El vaso se rompió').

Sin embargo, las alternancias de diátesis presentan problemas de productividad, esto es, no siempre una estructura está en una oposición significativa con otra de tal manera que forman una alternancia de diátesis canónica. Por ejemplo, Levin (1993) presenta una serie de construcciones aisladas (bajo el epígrafe 'otras construcciones') que no se oponen a ninguna otra construcción pero que igualmente se emplean para definir las clases (construcción de objeto cognado, construcción 'x's way', etc.). Algo similar ocurre para el español (Devís Márquez, 1992). Asimismo, y de manera relacionada, no hay un acuerdo general acerca del criterio que debe usarse para definir una oposición significativa. Aunque generalmente se acepta que las diátesis representan diferentes maneras de empaquetar información similar, a modo de paráfrasis, existe una variedad de criterios para identificar el principio que sustenta la oposición entre las estructuras alternantes (cf. Para el español García-Miguel, Costas & Martínez, 2005; Vázquez, Fernández & Martí, 2000; Cifuentes Honrubia, 2006; Castellón, Martí, Morante & Vázquez, 1997). Finalmente, es importante tener en cuenta las limitaciones que las diátesis imponen a priori. Goldberg (2002) argumenta que las estructuras aisladas, sin estar combinadas en parejas diatéticas o paráfrasis, proporcionan un mayor poder de generalización en el análisis lingüístico, ya que no restringen desde el inicio los tipos de combinaciones y oposiciones que se pueden dar.

Por lo tanto, la capacidad de discriminación primaria queda depositada en las estructuras consideradas individualmente y no en las alternancias de diátesis.

Consideramos que estas objeciones se deben tener en cuenta para calcular la similitud verbal de forma sistemática. Por lo tanto, en la base de nuestro análisis no está la noción de diátesis como oposición entre construcciones, sino la de construcción. Una construcción, según la definición de Goldberg (1998), es un signo lingüístico con forma y significado, no composicional, que combina estructura sintáctica e información semántica para formar la unidad básica de la gramática. Por otro lado, las construcciones, en tanto que unidades lingüísticas, pueden poseer diferentes grados de complejidad que incluyen un amplio abanico de unidades, desde morfemas a construcciones argumentales. Goldberg (1995) asume que estas últimas, en su formulación básica, corresponden a la codificación de experiencias elementales de los hablantes (transferencia, movimiento, movimiento causado, etc.). Es este nivel de complejidad construccional precisamente el que consideramos relevante para nuestro estudio sobre la similitud verbal desde el punto de vista de la teoría lingüística.

Nuestra propuesta para calcular la similitud de los sentidos verbales se basa en el análisis de la capacidad de instanciación de una serie de construcciones por parte de los sentidos verbales detallados en la sección 2. De esta manera, aquellos sentidos que instancian las mismas construcciones son considerados similares, mientras que aquellos que instancian construcciones disjuntas son considerados disimilares. Por lo tanto, no imponemos restricciones de combinaciones a priori, sino que la combinatoria es inducida y no limitada a combinaciones de dos construcciones. Respecto a esta última cuestión, hay que mencionar que el contenido semántico típicamente asociado a las alternancias de diátesis no se pierde, sino que está integrado como parte de la construcción. Más en detalle, la gramática de construcciones asume que la instanciación de una determinada construcción por parte de un verbo viene dada por la compatibilidad de los significados de ambos (Goldberg, 1995). Específicamente, cada construcción argumental tiene un significado abstracto y un conjunto de roles argumentales asociados motivados por ese significado. Asimismo, los verbos tienen un significado léxico y un marco conceptual con una serie de participantes específicos asociados. La compatibilidad entre el verbo y la construcción se define teniendo en cuenta la compatibilidad entre los participantes del verbo y los roles argumentales de la construcción.

En cuanto a la selección de las construcciones para nuestro estudio, hemos utilizado construcciones generales descritas en múltiples gramáticas (por ejemplo Bosque & Demonte, 1999): transitiva, intransitiva, ditransitiva, predicativa y atributiva. Además, hemos empleado construcciones adaptadas de la propuesta de Levin (1993) y Goldberg (1995), teniendo en cuenta los trabajos de Cifuentes Honrubia (2006) y Vázquez, et al. (2000) para el español. Dado que la cantidad de sentidos escogidos es limitada, se han preferido aquellas construcciones que tienen un carácter más general

sobre aquellas específicas para determinados verbos. A continuación se detallan estas construcciones:

1. Causativa prototípica: es una construcción que refleja un evento complejo y en la que se explicita la causa de ese evento por medio de diferentes recursos lingüísticos, aunque en este trabajo solo tenemos en cuenta aquellos casos en los que se expresa mediante un sujeto. Este sujeto puede ser volicional (agente) o no volicional (causa). La entidad que tiene el estatus sintáctico de objeto es generalmente modificada y afectada en diferentes grados. Ej. ‘La falta de lluvias secó el río’.

2. Anticausativa prototípica (con ‘se’): es una construcción intransitiva donde la entidad afectada ocupa la posición de sujeto y el agente o causa es típicamente omitido. Ej. ‘El río se secó’.

3. Causativa de perífrasis: es una causativa en la que el predicado que expresa la noción de causatividad es el auxiliar ‘hacer’, que aparece junto al predicado que expresa el evento, en infinitivo. Adicionalmente, según Langacker (1995), este tipo de construcciones dibuja una relación directa entre los participantes. La entidad que ocupa la posición de sujeto es predominantemente una causa. No tenemos en cuenta las estructuras que se construyen con una causa indirecta que aumenta el número de participantes en el evento (como por ejemplo en ‘El juez hizo confesar al acusado’). Ej. ‘Los fuertes vientos han hecho bajar las temperaturas’.

4. Anticausativa sin ‘se’ (o anticausativa de proceso, Vázquez et al., 2000): en esta construcción, el constituyente que expresa la causa se elide. Una entidad que no está claramente afectada ocupa la posición de sujeto, aunque el grado de afectación puede variar según el verbo. Ej. ‘Las temperaturas han bajado’.

5. Media: expresa un estado o propiedad del sujeto sin necesidad de combinarse con un verbo atributivo. Generalmente va con un complemento adverbial que refuerza la lectura estativa o atemporal, a diferencia de la anticausativa prototípica, que tiene una interpretación dinámica. Esta lectura también puede ser reforzada por verbos modales. Ej. ‘La pintura se esparce con facilidad’.

6. Impersonal pronominal: el verbo aparece en tercera persona, no tienen sujeto gramatical explícito ni recuperable por el contexto. Ej. ‘Se aconseja el uso obligatorio del cinturón’.

7. Sujeto oblicuo: el iniciador del evento aparece en un sintagma encabezado por una preposición. Se suele subdividir en varios tipos de acuerdo con el significado transmitido por la preposición y el objeto (ej. origen, tiempo, instrumento, etc.), pero dada la cantidad de sentidos verbales que analizamos, no hemos tenido en cuenta estas subdivisiones. Ej. ‘La gente se beneficia de las nuevas medidas’.

8. Reflexiva: la acción realizada por el sujeto recae sobre sí mismo. Algunos autores consideran que esto implica que el referente (el sujeto sintáctico) tiene el rol de agente y paciente simultáneamente. Ej. ‘María se peina’.

9. Recíproca: esta construcción expresa eventos simultáneos. El sujeto de esta construcción es plural. Cada uno de los componentes del sujeto ejerce una acción sobre los otros, a la vez que la recibe de los demás. Ej. ‘Juan y Pedro se desafiaron’.

10. Pasiva perifrástica: el objeto ocupa una posición topicalizada y el verbo se construye con un auxiliar. Generalmente el agente se puede expresar mediante un sintagma preposicional, aunque comúnmente se omite. Ej. ‘Los bizcochos fueron comidos por los niños’.

11. Pasiva refleja: se construye con la partícula ‘se’. El iniciador de la acción no se explicita pero suele ser agentivo. Típicamente el sujeto se pospone a la partícula. Ej. ‘Se pasaron los trabajos a ordenador’.

12. Objeto cognado: en esta construcción un verbo normalmente intransitivo se combina con un objeto que mantiene una relación etimológica con el verbo o redundante con su significado, por ello las frases con esta construcción tienen un sentido tautológico. Ej. ‘Cantamos una canción’.

13. Resultativa con ‘estar’: es una construcción estativa que detalla el estado resultado del evento expresada por el verbo, es decir, expresa una propiedad de una entidad que es el resultado de un proceso sufrido por esa entidad. El iniciador de ese proceso puede ser agentivo o causativo. Ej. ‘El pan está cortado’.

Para poder llevar a cabo el análisis de similitud basado en construcciones, tres lingüistas determinaron si cada sentido verbal podía instanciar o no cada una de estas construcciones. Para ello se recogieron ejemplos de frases para cada sentido verbal y construcción con la ayuda de una gramática descriptiva (Bosque & Demonte, 1999), o bien se anotó la imposibilidad de encontrar un ejemplo. A partir de estos datos se establecieron varias sesiones de discusión en las que se valoró la adecuación de los ejemplos o la ausencia de los mismos, según el caso. En estas sesiones, tras partir de un alto grado de consenso, se llegó a un acuerdo global. Estas concurrencias de sentidos y construcciones quedan registradas en una matriz en la que cada sentido verbal es representado mediante un vector y cada construcción es una componente de ese vector, cuyo valor es 1 o 0 dependiendo de si el verbo puede instanciar la construcción o no. Un ejemplo de esta formalización puede verse en la Tabla 2.

**Tabla 2.** Muestra de la matriz de instanciación.

Sentido	construcción 1	construcción 2
Abrir	1	0
Ver	0	1

Como resultado obtenemos una representación vectorial binaria para cada sentido verbal. Para determinar la similitud existente entre todos los sentidos a partir de esta representación vectorial calculamos el coeficiente de similitud de Dice entre los vectores que caracterizan los miembros de cada par para el conjunto explicado en la sección 2 (en total, 190 pares de sentidos). La fórmula para este coeficiente puede verse en la ecuación 1, donde  $T(s_1)$  y  $T(s_2)$  son vectores que caracterizan dos sentidos ( $s_1$  y  $s_2$ ),  $|T(s_1) \cap T(s_2)|$  son las componentes comunes (construcciones para las que ambos sentidos se comportan de la misma manera) y  $|T(s_1)| + |T(s_2)|$  el número total de componentes.

**Ecuación 1.** Coeficiente de similitud de Dice

$$dist_{Dice}(s_1, s_2) = \frac{2 \times |T(s_1) \cap T(s_2)|}{|T(s_1)| + |T(s_2)|}$$

Este coeficiente es adecuado para nuestros objetivos ya que se aplica a datos binarios y oscila entre 0 (disimilitud total, los sentidos tienen un comportamiento opuesto) y 1 (similitud total, los sentidos tienen un comportamiento idéntico). Los valores de similitud de cada par obtenidos con este método se compararán con los obtenidos mediante los datos del corpus en la sección 5.

## 5. Comparación y análisis de resultados

A continuación se presentan y analizan los resultados obtenidos al comparar los datos de corpus y los del estudio basado en construcciones. Esta comparación comprende dos aspectos: un análisis cuantitativo de la correlación existente entre los valores de similitud obtenidos con datos de corpus y los obtenidos a partir de las construcciones, y un análisis de carácter cualitativo en el que examinamos las características lingüísticas que prevalecen en cada perspectiva respecto al campo semántico, el aspecto y los patrones de subcategorización que comparten los miembros de los pares de sentidos verbales.

El estudio cuantitativo consiste en un análisis de la correlación de Pearson entre los valores de similitud obtenidos a partir de cada perspectiva, teniendo en cuenta cada una de las diferentes formalizaciones de estructura argumental en corpus. El coeficiente de correlación de Pearson mide el grado de relación lineal entre dos variables aleatorias, dos perspectivas sobre la similitud verbal en nuestro caso. La

fórmula para calcular el coeficiente de correlación de Pearson (también llamado  $\rho$ ) se muestra en la ecuación 2, donde X e Y son las dos variables comparadas, el numerador corresponde a la covarianza de las dos variables y el denominador corresponde a la desviación típica de cada una de las variables.

**Ecuación 2. Coeficiente de correlación de Pearson**

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

El coeficiente de correlación de Pearson para las formalizaciones que contienen información sintáctica explícita, además de roles, se muestra en la Tabla 3; para formalizaciones que solo contienen roles se muestra en la Tabla 4 (véase Tabla 1 para una ejemplificación del tipo de información incluida). Todos los coeficientes mostrados remiten a una correlación estadísticamente significativa ( $p < 0.05$ ). En caso contrario se indica con un guión.

**Tabla 3. Correlación de similitud verbal (corpus con sintaxis, construcciones).**

Tipo de información	Tipo de representación	Correlación con construcciones
Roles específicos con sintaxis	Constituyentes	0.30
	Patrones	0.34
Roles intermedios con sintaxis	Constituyentes	0.53
	Patrones	0.54
Roles abstractos con sintaxis	Constituyentes	0.47
	Patrones	0.57

**Tabla 4. Correlación de similitud verbal (corpus sin sintaxis, construcciones).**

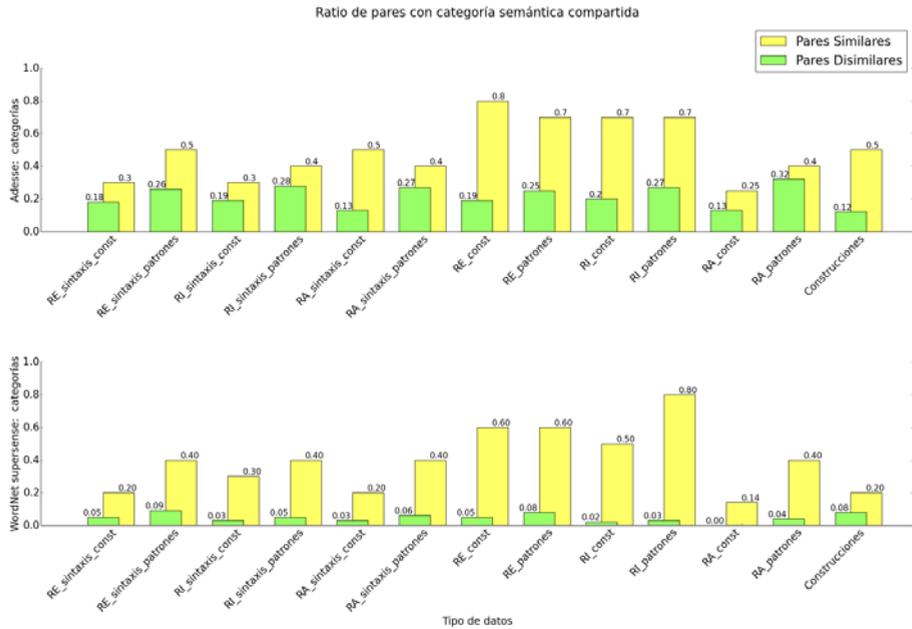
Tipo de información	Tipo de representación	Correlación con construcciones
Roles específicos	Constituyentes	0.15
	Patrones	0.34
Roles intermedios	Constituyentes	0.30
	Patrones	0.44
Roles abstractos	Constituyentes	-
	Patrones	0.24

Los resultados muestran que la presencia de sintaxis en las formalizaciones obtenidas de corpus aumenta la correlación con los valores de similitud obtenidos a partir de las construcciones teóricas (Tabla 3 vs. Tabla 4). Además, si nos fijamos en el tipo de representación de la información, observamos que las formalizaciones obtenidas de corpus basadas en patrones también dan lugar a correlaciones más altas, en comparación con las basadas en constituyentes, tanto en el caso de las formalizaciones que incluyen sintaxis (Tabla 3) como en el caso de las que no la incluyen (Tabla 4). Esta ganancia es más notoria en el caso de la formalización que no incluye sintaxis, con un aumento de la correlación de hasta 0.14 puntos. Es decir, la presencia de información de tipo estructural (sintaxis o patrones) en corpus es

importante para establecer una correlación más alta con las construcciones teóricas en cuanto a la similitud entre sentidos verbales. Respecto a la granularidad de los roles, los datos apuntan a una mayor correlación de la similitud verbal en los casos en los que el grado de abstracción de los roles es mayor.

En cuanto al análisis cualitativo, nos centramos en observar las propiedades lingüísticas más sobresalientes de cada uno de los rankings de similitud determinados por las diferentes formalizaciones de ambas perspectivas. Para analizar los aspectos que destacan en cada una de las diferentes formalizaciones procedemos a escoger los pares de sentidos verbales más similares y disimilares determinados por cada una de ellas. Para ello, para cada formalización, ordenamos los pares de sentidos de acuerdo a su valor de similitud, de menor a mayor, e inicialmente seleccionamos los 10 primeros y los 10 últimos, lo que equivale aproximadamente al 10% de los pares totales. En los casos en los que el valor de similitud del par que queda como umbral de corte es idéntico al de otros pares que quedan fuera de la selección, se incluyen estos pares en la selección. Finalmente, para cada uno de estos conjuntos de pares más similares y disimilares se hace un recuento de cuántos contienen sentidos que comparten el campo semántico según las categorías definidas en Adesse (macro-clases) y WordNet (*supersenses*). Asimismo, se computa el número de pares cuyos sentidos tienen la misma categoría aspectual. Por último, recogemos la ratio de patrones de subcategorización definidos en SenSem que son compartidos entre los sentidos de cada par. Por lo tanto, se obtiene un panorama de cuáles son las características lingüísticas que predominan en cada formalización de similitud verbal. A continuación mostramos los resultados de este análisis para cada aspecto lingüístico descrito.

En la Figura 3 presentamos la ratio de pares similares y disimilares cuyos sentidos comparten el campo semántico tal y como se define en los dos recursos ya mencionados (macro-clases de Adesse y *supersenses* de WordNet). El análisis de campos semánticos de los pares se realiza de forma separada para cada uno de estos dos recursos. El valor de la ratio de pares cuyos sentidos están asociados al mismo campo semántico se puede ver en el eje Y. En el eje X se listan de forma abreviada las formalizaciones que se han tenido en cuenta para este análisis comparativo: *RE* hace referencia a roles específicos, *RI* a los roles del nivel intermedio y *RA* a los roles del nivel abstracto, mientras que *const* remite al tipo de representación en constituyentes.



**Figura 3.** Ratio de pares con la misma categoría semántica.

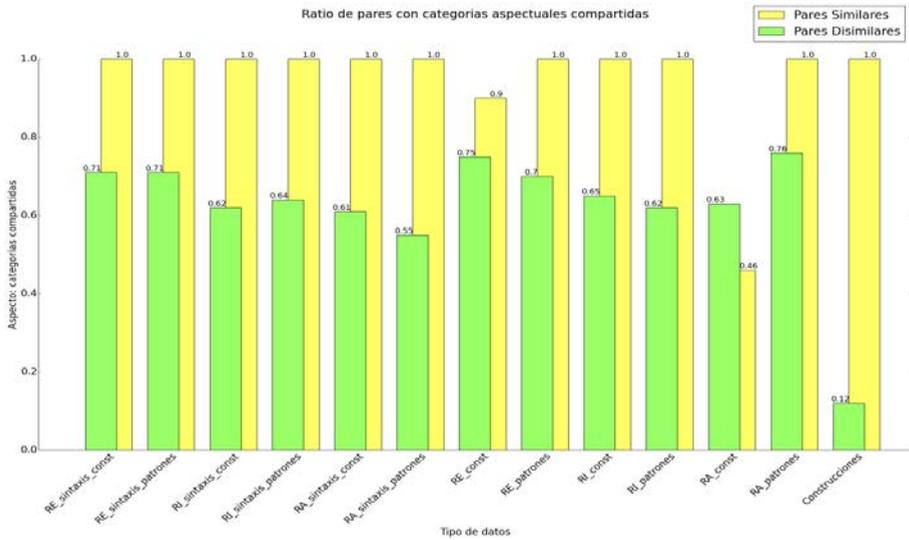
Respecto al análisis semántico, en la Figura 3 vemos que, tanto para las macro-classes de Adesse como para los *supersenses* de WordNet, la ratio de pares cuyos sentidos comparten la categoría semántica es mayor para los pares similares que para los disimilares, lo cual es esperable. No obstante, hay algunas diferencias entre las diversas formalizaciones. En cuanto a los datos de corpus, la presencia de información sintáctica explícita tiende a disminuir la ratio de pares similares con categoría semántica compartida. Por otro lado, la presencia de patrones hace aumentar la ratio de pares disimilares con categoría semántica compartida. Por ello, podemos considerar que la ausencia de información estructural (patrones o sintaxis) da lugar a valores de similitud verbal más coherentes desde el punto de vista semántico. La diferencia entre pares similares y disimilares en cuanto a la ratio de categoría semántica compartida es mayor para las formalizaciones que contienen solo roles realizadas a partir de corpus. Es decir, estas formalizaciones están más definidas por la categoría semántica que aquellas formalizaciones que contienen información sintáctica, obtenidas a partir de datos de corpus. La formalización basada en construcciones se comporta de forma similar a la que contiene roles abstractos con sintaxis organizados en constituyentes.

Además de esto, podemos ver una diferencia en los resultados según el tipo de recurso usado, macro-classes de Adesse o *supersenses* de WordNet. La diferencia entre los pares similares y disimilares es más evidente en el caso de la comparación hecha con las categorías de *supersenses*. En el caso de la comparación con las macro-classes de Adesse, los pares disimilares tienden a tener una ratio más alta de categoría semántica compartida en todas las formalizaciones, por lo que el contraste entre pares similares y

disimilares es menor. Esto puede ponerse en relación con el número y tipo de categorías que cada recurso contiene. Los *supersenses* de WordNet están constituidos por 15 campos semánticos abstractos para verbos (cambio, cognición, comunicación, etc.) (Fellbaum, 1990) mientras que Adesse contempla 8 categorías en la parte más abstracta de su clasificación verbal jerárquica (las macro-clases) basadas en dominios semánticos amplios (mental, relacional, material, etc.) que están asociados a ejemplos de corpus (García-Miguel et al., 2005).

En cuanto al aspecto, mostramos los resultados obtenidos en la Figura 4, donde podemos ver la ratio de pares similares y disimilares cuyos sentidos tienen la misma categoría aspectual. Respecto a esto, tenemos en cuenta dos categorías amplias para nuestro análisis: estados y eventos dinámicos. El uso de estas dos categorías viene motivado por el hecho de que son dos nociones que Dowty (1979) y Jackendoff (1983) proponen como esenciales y que forman una oposición mantenida en las numerosas propuestas de categorización aspectual que hay en la literatura.

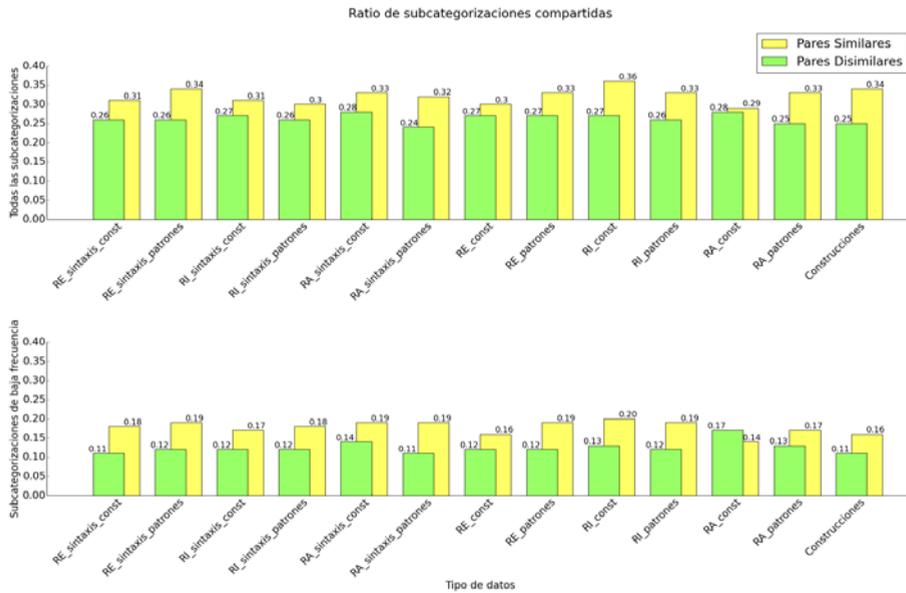
En cuanto a los resultados del análisis de categoría aspectual que podemos observar en la Figura 4, vemos como los datos de corpus se comportan en general de manera homogénea: los miembros de los pares similares comparten casi en todos los casos la categoría aspectual y los miembros de los pares disimilares muestran una ratio de categoría aspectual compartida de entre 55% y 75%. Las construcciones teóricas presentan en este caso una divergencia importante: los miembros de los pares similares comparten en todos los casos la categoría aspectual, mientras que los de los pares disimilares lo hacen solo en un 12% de los casos, lo que indica que la categoría aspectual representa un potente eje vertebrador para la similitud verbal definida según las construcciones, pero no está presente con la misma fuerza en el caso de la similitud definida usando datos de corpus. Respecto a esta diferencia de comportamiento, es relevante tener en cuenta que las construcciones poseen inherentemente una interpretación aspectual dinámica o estática. Por lo tanto, el aspecto de la construcción y el aspecto léxico del sentido verbal han de ser compatibles. Por ejemplo, Dowty (1979) y Jackendoff (1990) mencionan que las construcciones resultativas solo admiten verbos dinámicos. Así, podemos considerar que la compatibilidad aspectual juega un papel determinante a la hora de definir la similitud verbal desde esta perspectiva teórica.



**Figura 4.** Ratio de pares con la misma categoría aspectual.

Finalmente, presentamos los datos relativos a la subcategorización en la Figura 5. En ella podemos ver la ratio de patrones de subcategorización compartidos en cada una de las diferentes formalizaciones. Presentamos dos tipos de análisis de patrones de subcategorización: según los resultados obtenidos teniendo en cuenta la ratio de todos los patrones de subcategorización (parte superior de la Figura 5) y según los resultados obtenidos si solo se tienen en cuenta los patrones de subcategorización menos frecuentes, es decir, todos aquellos que no son SN-V o SN-V-SN (parte inferior de la Figura 5).

Como en los casos anteriores, los pares similares muestran una ratio mayor de patrones de subcategorización compartidos que los pares disimilares. En conjunto, no se aprecian diferencias significativas entre los datos de corpus y las construcciones. En ambos casos la ratio de patrones de subcategorización compartidos es baja, siendo menor cuando tenemos en cuenta aquellos patrones menos frecuentes.



**Figura 5.** Ratio de patrones de subcategorización compartidos.

## CONCLUSIONES

En este artículo hemos comparado dos perspectivas de análisis de la similitud verbal. Ambas tienen que ver con la estructura argumental que se asocia con los verbos, pero mientras que una recoge la visión de la teoría lingüística (las construcciones que un determinado sentido verbal puede instanciar) la otra está basada en datos empíricos, concretamente en la información sintáctico-semántica que encontramos en un corpus anotado. Nuestro análisis ha tomado como base los sentidos verbales y no los lemas porque es crucial tener en cuenta las diferencias de comportamiento entre los sentidos verbales para obtener una visión precisa de la similitud verbal.

Se han realizado dos tipos de análisis, uno de corte cuantitativo y otro de corte cualitativo, encaminados a estudiar los puntos de convergencia y divergencia relativos a la manera en la que estas dos perspectivas definen la similitud verbal.

Desde el punto de vista cuantitativo, los resultados de una correlación de Pearson entre los valores de similitud obtenidos de los datos de corpus y de las construcciones indican una correlación significativa moderada (0.57) cuando la estructura argumental de corpus se formaliza usando roles abstractos, sintaxis y patrones. Estos dos últimos factores, la presencia de sintaxis y patrones, se revelan como potenciadores de la correlación de los datos de corpus con los datos de construcciones.

En cuanto al estudio cualitativo, hemos visto que la diferencia fundamental entre la similitud que dibujan los datos de corpus y la que obtenemos mediante las construcciones es que esta última aproximación estructura la diferencia entre los polos de similitud y disimilitud en torno a la oposición aspectual dinamismo / estatividad.

Los datos obtenidos en este análisis ponen de relieve las diferencias y semejanzas existentes respecto a cómo diversas perspectivas articulan la similitud verbal. Esto no solo abre la puerta a emplear cada perspectiva por separado para medir la similitud existente entre dos sentidos verbales teniendo en cuenta qué aspectos potencia, sino que ofrece la posibilidad de combinar estas perspectivas para obtener una definición y caracterización de similitud verbal más robusta.

## REFERENCIAS BIBLIOGRÁFICAS

- Albertuz, F. J. (2007). Sintaxis, semántica y clases de verbos: Clasificación verbal en el proyecto ADESSE. *Actas del VI Congreso de Lingüística General* (pp. 2015-2030). Santiago de Compostela, España.
- Bosque, I. & Demonte, V. (Dir.) (1999). *Gramática descriptiva de la lengua española*. Madrid: Espasa.
- Bonial, C., Corvey, W., Palmer, M., Petukhova, V. V. & Bunt, H. (2011). A hierarchical unification of LIRICS and VerbNet semantic roles. Ponencia presentada en *Semantic Computing (ICSC), 2011 Fifth IEEE International Conference* (pp. 483-489). Palo Alto, California.
- Castellón, I., Martí, M. A., Morante, R. & Vázquez, G. (1997). Propuesta de alternancias de diátesis verbales para el español y el catalán. *Procesamiento del Lenguaje Natural*, 21, 31-48.
- Chwilla, D. J. & Kolk, H. H. (2005). Accessing world knowledge: Evidence from N400 and reaction time priming. *Cognitive Brain Research*, 25(3), 589-606.
- Cifuentes Honrubia, J. L. (2006). Alternancias verbales en español. *Revista Portuguesa de Humanidades*, 10(2), 107-132.
- De Deyne, S., Peirsman, Y. & Storms, G. (2009). Sources of semantic similarity. Ponencia presentada en el *31st Annual Conference of the Cognitive Science Society* (pp.1834-1839). Amsterdam: Netherlands.
- Devís Márquez, P. (1992). *Esquemas sintáctico-semánticos, el problema de las diátesis en español*. Cádiz: Servicio de publicaciones de la Universidad de Cádiz.
- Dowty, D. (1979). *Word meaning and montague grammar*. Dordrecht: Reidel.
- Fellbaum, C. (1990). English verbs as a semantic net. *International Journal of Lexicography*, 3(4), 278-301.

- Fellbaum, C. (1998). *WordNet*. Blackwell Publishing Ltd.
- Fernández-Montraveta, A. & Vázquez, G. (2014). The SenSem Corpus: An annotated corpus for Spanish and Catalan with information about aspectuality, modality, polarity and factuality. *Corpus Linguistics and Linguistic Theory*, 10(2), 273-288.
- Fernando, S. & Stevenson, M. (2008). A semantic similarity approach to paraphrase detection. Ponencia presentada en el *11th Annual Research Colloquium of the UK Special Interest Group for Computational Linguistics*, Oxford, United Kingdom.
- Ferretti, T. R., McRae, K. & Hatherell, A. (2001). Integrating verbs, situation schemas, and thematic role concepts. *Journal of Memory & Language*, 44, 516-547.
- Fillmore, C. J. (1967). The case for case. En E. Bach & R.T. Harms (Eds.), *Universals in Linguistic Theory* (pp. 1-88). Nueva York: Holt, Rinehart and Winston.
- Fillmore, C. J., Johnson, C. R. & Petruck, M. R. (2003). Background to framenet. *International Journal of Lexicography*, 16(3), 235-250.
- García-Miguel, J. M., Costas, L. & Martínez, S. (2005). Diátesis verbales y esquemas construccionales. Verbos, clases semánticas y esquemas sintáctico-semánticos en el proyecto ADESSE. En G. Wotjak & J. Cuarteto Otal (Eds.), *Entre semántica léxica, teoría del léxico y sintaxis* (pp. 373-384). Frankfurt am Main: Peter Lang.
- García-Miguel, J. M., González Domínguez F. & Vaamonde G. (2010). ADESSE. A Database with Syntactic and Semantic Annotation of a Corpus of Spanish. Ponencia presentada en *Seventh International Conference on Language Resources and Evaluation* (pp. 1903-1910). Valletta, Malta.
- Goldberg, A. E. (1995). *Constructions, a construction grammar approach to argument structure*. Chicago: University of Chicago Press.
- Goldberg, A. E. (1998). Patterns of experience in patterns of language. *The new psychology of language: Cognitive and functional approaches to language structure*, 1, 203-219.
- Goldberg, A. E. (2002). Surface generalizations: An alternative to alternations. *Cognitive Linguistics*, 13(4), 327-356.
- Goldstone, R. L. & Son, J. Y. (2005). Similarity. En K. J. Holyoak & R. G. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning* (pp. 13-36). Nueva York, US: Cambridge University Press.
- Gruber, J. S. (1965). *Studies in lexical relations*. MIT. Tesis doctoral publicada en 1976 como *Lexical Structures in Syntax and Semantics*. Amsterdam: North Holland.

- Hare, M., McRae, K. & Elman, J. L. (2003). Sense and structure: Meaning as a determinant of verb subcategorization preferences. *Journal of Memory and Language*, 48(2), 281-303.
- Jackendoff, R. S. (1972). *Semantic interpretation in generative grammar*. Cambridge, MA: The MIT Press.
- Jackendoff, R. S. (1983). *Semantics and cognition*. Cambridge, MA: The MIT Press.
- Jackendoff, R. S. (1990). *Semantic structures*. Cambridge, MA: The MIT Press.
- Jones, M. N., Willits, J. & Dennis, S. (2015). Models of semantic memory. *Oxford handbook of Mathematical and Computational Psychology* (pp. 232-254). Oxford: Oxford University Press.
- Kipper, K. (2005). *VerbNet: A broad-coverage, comprehensive verb lexicon*. Tesis doctoral, Universidad de Pensilvania, Philadelphia, Estados Unidos.
- Klein, D. E. & Murphy, G. L. (2002). Paper has been my ruin: Conceptual relations of polysemous senses. *Journal of Memory and Language*, 47, 548-570.
- Langacker, R. W. (1995). Possession and possessive constructions. En J. R. Taylor & R. E. MacLaury (Eds.), *Language and Cognitive Construal of the World* (pp. 51-79). Oxford: Clarendon Press.
- Levin, B. (1993). *English verb classes and alternations: A preliminary investigation*. Chicago: University of Chicago Press.
- Miller, G. A. (1995). WordNet: A lexical Database for English. *Communications of the ACM*, 38(11), 39-41.
- Neely, J. H. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. *Basic processes in reading: Visual word recognition*, 11, 264-336.
- Palmer, M., Gildea, D. & Kingsbury, P. (2005). The proposition bank: An annotated corpus of semantic roles. *Computational linguistics*, 31(1), 71-106.
- Savage, C., Lieven, E., Theakston, A. & Tomasello, M. (2003). Testing the abstractness of children's linguistic representations: Lexical and structural priming of syntactic constructions in young children. *Developmental Science*, 6(5), 557-567.
- Sinha, R. & Mihalcea R. (2007). Unsupervised graph-based word sense disambiguation using measures of word semantic similarity. *Proceedings of the Semantic Computing (ICSC), 2007 IEEE International Conference* (pp. 363-369). California: Irvine.

Vázquez, G., Fernández, A. & Martí, M. A. (2000). *Clasificación verbal: Alternancias de diátesis*. Lleida: Edicions de la Universitat de Lleida.

Vitevitch, M. S. & Luce, P. A. (2016). Phonological neighbourhood effects in spoken word perception and production. *Annual Review of Linguistics*, 2, 75-94.

## ANEXOS

### **Anexo 1. Sentidos verbales escogidos junto con sus definiciones y su frecuencia en el corpus entre paréntesis.**

abrir 18: recorrer el pestillo o cerrojo, desechar la llave, levantar la aldaba o desencajar cualquier otra pieza o instrumento semejante con que se cierra algo. (15)

cerrar 19: asegurar con cerradura, pasador, pestillo, tranca u otro instrumento, una puerta, ventana, tapa, etc., para impedir que se abra. (14)

crecer 1: incrementar la cantidad o la importancia de algo, desarrollarse. (116)

dormir 1: permanecer en un estado en el cual todos los movimientos voluntarios son suspendidos, generalmente para descansar. (18)

escuchar 1: poner atención a lo que se oye. (107)

estar 14: encontrarse alguien o algo en un estado determinado. (101)

explicar 1: aclarar algo, dar información sobre un asunto. (106)

gestionar 1: realizar un trámite para la consecución de una cuestión. (36)

gustar 1: encontrar atractivo o agradable alguna cosa o a alguien. (117)

montar 2: subirse alguien en un animal o un vehículo. (26)

morir 1: fallecer, dejar de existir algo o alguien. (115)

parecer 1: aparentar algo, sin serlo necesariamente. (51)

pensar 2: usar la mente alguien para examinar una idea, razonar. (25)

perseguir 1: ir detrás de alguien o algo para alcanzarle. (53)

trabajar 1: emplearse en cualquier ejercicio, obra, trabajo o ministerio. (80)

valorar 2: admitir la importancia de un hecho, cosa o acción. (70)

valer 1: tener algo un determinado valor. (45)

ver 1: recibir una imagen a través de la vista. (86)

viajar 1: ir de un lugar a otro que suele estar distante, generalmente mediante algún medio de transporte. (111)

volver 1: dirigirse hacia el lugar donde ya se ha estado. (84)

**Anexo 2.** Campos semánticos que corresponden a cada sentido según los recursos empleados.

Sentido verbal	Supersense WordNet	Macro-clase Adesse
Abrir 18	Cambio	Material
Cerrar 19	Cambio	Material
Crecer 1	Cambio	Material
Dormir 1	Actividad (corporal)	Material
Escuchar 1	Percepción	Mental
Estar 14	Estado	Relacional
Explicar 1	Comunicación	Verbal
Gestionar 1	Actividad (social)	Material
Gustar 1	Cognición	Mental
Montar 2	Movimiento	Material
Morir 1	Cambio	Existencial
Parecer 1	Estado	Relacional
Pensar 2	Cognición	Mental
Perseguir 1	Movimiento	Material
Trabajar 1	Actividad	Material
Valorar 2	Comunicación	Relacional
Valer 1	Estado	Relacional
Ver 1	Percepción	Mental
Viajar 1	Movimiento	Material
Volver 1	Movimiento	Material