

APROXIMACIÓN AL TRASVASE AUTOMÁTICO DE PREDICADOS DE FRAMENET AL ESPAÑOL MEDIANTE WORDNET

 Mario Crespo Miguel

Universidad de Cádiz, España

Resumen: Dentro del procesamiento del lenguaje, los recursos lingüísticos son descripciones estructuradas y detalladas de una determinada lengua, esenciales a la hora de estudiar el lenguaje y crear aplicaciones. Sin embargo, estos repositorios son bastantes lentos y difíciles de construir, y además la mayoría de ellos se centra en el inglés. Este trabajo trata de paliar, en cierta medida, el problema de escasez de recursos disponibles en castellano, mediante la traducción al español de las unidades léxicas de los marcos situacionales del proyecto FrameNet, un recurso *on-line* para el inglés basado en la semántica de marcos. Para ello desarrollamos un procedimiento capaz de asociar los diferentes predicados de cada marco con los *synsets* de WordNet, una base de datos léxica que organiza el vocabulario según conceptos y relaciones semánticas. Como tendremos oportunidad de comprobar, el sistema alcanza una precisión en torno al 88% y abre la puerta a su uso en estudios lingüísticos de diversa índole en español.

Palabras clave: FrameNet, análisis semántico, procesamiento semántico, WordNet, trasvase de información lingüística, desambiguación.

AN APPROACH TO THE AUTOMATIC TRANSFER OF LEXICAL UNITS FROM ENGLISH FRAMENET TO SPANISH BY USING WORDNET

Abstract: In the field of Natural Language Processing, linguistic resources are structured and detailed descriptions of a certain language. They are considered as key elements for studying languages and developing applications. However, these repositories are slow and difficult to build, and most of them focuses on English. This work tries to improve the lack of linguistic resources in Spanish by transferring part of the information encoded in the FrameNet project into Spanish. For this purpose, we developed an automatic procedure able to align the different frame predicates with the WordNet synsets that best represent them. Our system reaches an 88% precision and makes it possible to reuse this semantic resource for linguistic studies in Spanish.

Keywords: FrameNet, semantic analysis, semantic processing, WordNet, transferring of linguistic information, disambiguation.

1. INTRODUCCIÓN

El desarrollo de recursos lingüísticos resulta fundamental para la creación de sistemas de Lingüística Computacional (McCrae y Cimiano, 2015; Kurdi, 2016), si bien este tipo de aplicaciones necesita de repositorios que formalicen la información lingüística (Arano, 2005). Existe un esfuerzo continuo por aumentar los recursos disponibles para los estudios del lenguaje y el ámbito del Procesamiento del Lenguaje Natural (PNL) (Van Uytvanck et al., 2010), pudiendo ser estas fuentes textos, corpus multimodales, bases de conocimiento léxico, diccionarios, etc., pero también herramientas para crear, anotar y consultar información, así como datos procedentes de experimentos o estudios (Cristea y Pistol, 2012). En general, tienen el objetivo de formalizar y hacer explícita la información de los diferentes niveles lingüísticos de un determinado idioma. Sin embargo, la mayoría de las veces estos recursos han sido creados para hacer un análisis determinado o para describir un nivel específico del lenguaje. De todos ellos, el número de proyectos que se ocupa del análisis en el ámbito de la semántica es menor que el de otros campos como la morfología o la sintaxis. Esto es debido a que el estudio del significado requiere cierta interpretación, por lo que su estudio y descripción es mucho más complejo.

Cuando hablamos de procesamiento semántico, Martí y Taulé (2014) destacan recursos tales como redes léxicas, ontologías o diccionarios electrónicos. Entre los recursos semánticos más conocidos podemos encontrar WordNet (Miller et al., 1993), FrameNet (Baker et al., 1998; Ruppenhofer et al., 2006), PropBank (Palmer et al., 2005), o VerbNet (Kipper et al., 2000) que difieren entre sí en estructura, organización y contenido. La mayoría de estos

To cite this article: Crespo Miguel, M. (2021). "An approach to the automatic transfer of lexical units from English FrameNet to Spanish by using WordNet". *Revista de Lingüística y Lenguas Aplicadas*, 16, 49-62. <https://doi.org/10.4995/rlyla.2021.14408>

Correspondence author: mario.crespo@uca.es



proyectos se han creado para el inglés, si bien posteriormente algunos se han ampliado a otros idiomas: Spanish FrameNet (Subirats y Petruck, 2003), Japanese FrameNet (Ohara et al., 2004), German FrameNet (Burchardt et al., 2006) o 3LB-LEX (Civit et al., 2005). Sin embargo, en todos los casos, su creación necesita de mucho tiempo y esfuerzo, por lo que suelen tener alcance limitado (Burchardt, Erk y Frank, 2005).

Este trabajo tiene como objetivo el trasvase de las unidades léxicas de FrameNet de Baker et al. (1998) al español, con el fin de obtener un recurso similar al del inglés para esta lengua y poder usarlo como herramienta en estudios semánticos. Para ello, nos centraremos en la asociación de las unidades léxicas de FrameNet con los *synsets* de WordNet (como veremos en el siguiente apartado, se trata de conjuntos de elementos léxicos relacionados por sinonimia entendida en un sentido amplio). Comprobaremos cómo, mediante el uso de las diferentes relaciones semánticas que existen entre las unidades de FrameNet, se puede llegar a determinar cuál son los *synsets* de WordNet que mejor los representan. Una vez conectados ambos recursos, toda la información que proporciona WordNet puede utilizarse tanto para su traducción a otros idiomas, como para ampliar la propia cobertura original de FrameNet con nuevas unidades léxicas. Los resultados del trasvase de unidades léxicas muestran un 88% de palabras traducidas correctamente al español.

2. LOS RECURSOS SEMÁNTICOS FRAMENET Y WORDNET.

De acuerdo con CLARIN - European Research Infrastructure for Language Resources and Technology¹, la mayoría de los recursos semánticos se centra en la descripción del significado de las unidades lingüísticas (por ejemplo, un diccionario tradicional) y una minoría se ocupa de otras propiedades tales como la relación entre los valores y sentidos de las diferentes palabras, la semántica oracional y roles, o cómo se estructura lingüísticamente el conocimiento humano (por ejemplo, en un tesoro u ontología). Como ya indicábamos anteriormente, estos recursos son difíciles de construir y a veces hacen uso de un conocimiento extralingüístico que va más allá de los límites semánticos (Casas Gómez, 2014). Hilera et al. (2010) clasifican estos repositorios en función de su expresividad o riqueza semántica, y establecen cuatro niveles: Nivel 0, diccionarios, que incluyen definiciones de palabras; nivel 1, taxonomías, vocabularios y las descripciones de sus relaciones jerárquicas; nivel 2, tesauros, herramientas que contienen relaciones léxicas y de equivalencia (sinonimia, hiperonimia, etc.), y nivel 3, ontologías, que combinan las relaciones anteriores con otras conceptuales más complejas. Vilches-Blázquez, García Silva y Villazón Terrazas (2009) clasifican los recursos creados para representar la información semántica en glosarios, taxonomías, tesauros y ontologías, los cuales se diferencian atendiendo a su grado de complejidad (Gilchrist, 2003; Pieterse y Kourie, 2014).

Entre los recursos disponibles, WordNet (Miller et al., 1993) constituye una base de datos léxica que organiza el vocabulario del inglés según conceptos y relaciones semánticas. Estos grupos, llamados *synsets*, reflejan la disponibilidad del léxico inglés para un determinado concepto. Los *synsets* son conjuntos de “sinónimos cognitivos” usados en esa lengua para expresar una idea determinada. WordNet ha sido desarrollado en el *Cognitive Laboratory* de la Universidad de Princeton bajo la dirección de George A. Miller y contiene unas 150.000 palabras distribuidas entre las diferentes categorías verbales de sustantivos, verbos, adjetivos y adverbios. En la Figura 1 podemos ver los diferentes *synsets* asociados a la unidad “car.n” en inglés:

Noun: “car.n”

S: (n) car, auto, automobile, machine, motorcar (a motor vehicle with four wheels; usually propelled by an internal combustion engine) "he needs a car to get to work"

S: (n) car, railcar, railway car, railroad car (a wheeled vehicle adapted to the rails of railroad) "three cars had jumped the rails"

S: (n) car, gondola (the compartment that is suspended from an airship and that carries personnel and the cargo and the power plant)

Figura 1. *Synsets* asociados a la unidad léxica “car”²

Como se puede apreciar en la Figura 1, en cada *synset* aparecen todas las unidades léxicas que aluden al concepto representado en él, así como su definición general. WordNet está compuesto por más de 115.000 *synsets*

¹ <https://vlo.clarin.eu/?10> The Virtual Language Observatory (VLO) se desarrolló dentro de CLARIN como un medio para explorar los recursos, servicios y herramientas lingüísticas disponibles en CLARIN. Su objetivo es proporcionar una interfaz fácil de usar, que permita un proceso uniforme de búsqueda y descubrimiento de un gran número de recursos de una amplia variedad de dominios y proveedores. El objetivo de CLARIN es llevar los beneficios del acceso compartido de los recursos digitales a las Humanidades y las Ciencias Sociales (Bel et al., 2008).

² Ejemplo extraído de <https://wordnet.princeton.edu>

al modo de un tesoro u ontología, ya que estas agrupaciones no se describen de forma aislada, sino que están interconectadas de acuerdo con diferentes tipos de relaciones formales y semánticas. El resultado es una red de unidades léxicas y conceptos relacionados. Los tipos de relaciones conceptuales y léxicas que se describen en WordNet (Vossen, 1998; 2002) son:

1. Hiperonimia / Hiponimia: Y es a hiperónimo de X si X es un tipo de Y, e Y es un hipónimo de X si Y es un tipo de X.
2. Holonimia / Meronimia. Y es un holónimo de X si X es una parte de Y, e Y es un merónimo de X si Y es una parte de X.
3. Cuasi-sinonimia y Cuasi-antonimia. Y es un cuasi-sinónimo de X, si Y puede interpretarse como X en algunos contextos, e Y es un cuasi-antónimo de X si Y puede interpretarse como el opuesto a X en algunos contextos.
4. Pertenece-a y Derivado. Y pertenece a X si Y es un adjetivo perteneciente al concepto expresado por el sustantivo X, e Y se ha derivado de Z si Y es una forma derivada de la clase de palabra Z.
5. Otros tipos de relaciones:
 - Causas. Y causa X.
 - Rol_paciente. Y es el paciente de la actividad expresada por X.
 - Rol_localización. Y es la localización de la actividad expresada por X.
 - Rol_instrumento. Y es el instrumento de la actividad expresada por X.
 - Rol_agente. Y es el agente de la actividad expresada por X.

En la Figura 2 podemos apreciar los hipónimos directos y los merónimos del *synset* “*car, auto, automobile, machine, motorcar (a motor vehicle with four wheels; usually propelled by an internal combustion engine) “he needs a car to get to work”*” del ejemplo anterior. Estas relaciones están compuestas a su vez por otros *synsets*, de tal manera que WordNet constituye una red ontológica donde todos sus elementos están conectados.

S: (n) car, auto, automobile, machine, motorcar (a motor vehicle with four wheels; usually propelled by an internal combustion engine) “he needs a car to get to work”

direct hyponym / full hyponym

S: (n) ambulance (a vehicle that takes people to and from hospitals)

S: (n) funny wagon (an ambulance used to transport patients to a mental hospital)

S: (n) beach wagon, station wagon, wagon, estate car, beach waggon, station waggon, waggon (a car that has a long body and rear door with space behind rear seat)

meronym

S: (n) accelerator, accelerator pedal, gas pedal, gas, throttle, gun (a pedal that controls the throttle valve) “he stepped on the gas”

S: (n) air bag (a safety restraint in an automobile; the bag inflates on collision and prevents the driver or passenger from being thrown forward)

Figura 2. Hipónimos de uno de los *synsets* asociados a la unidad “car”.

WordNet se ha extendido a otros idiomas a partir del inglés. Este es el caso de EuroWordNet (Vossen, 1998; 2002), una base de datos como WordNet diseñada para varios idiomas europeos (holandés, italiano, español, alemán, francés, checo y estonio). Como explica Vossen (1998; 2002), cada WordNet representa el sistema único de lexicalizaciones internas de los idiomas, pero utilizando la red del inglés creada para el WordNet de Princeton como índice interlingüístico, es decir, cada nuevo idioma se combinó e interrelacionó con el WordNet original. La estructura se mantiene para todas las lenguas, pero se especifica la forma en que cada una lexicaliza un determinado concepto.

La Figura 3 nos muestra los elementos léxicos de un mismo *synset* para algunas de las lenguas en los que tiene representación. Esto nos permite utilizar esta base de datos para la recuperación de información multilingüe, ya que podemos observar cómo se lexicalizan esas agrupaciones conceptuales en otros idiomas.

Lengua	Inglés	Italiano	Español	Rumano
Synset	car, auto, automobile, machine, motorcar	auto, automobile, autovettura, vettura, macchina	auto, automóvil, coche, turismo	automobil, autovehicul, mașină

Figura 3. Equivalencias entre lenguas para un mismo synset.

FrameNet (Baker et al., 1998) es un recurso online para el inglés basado en la semántica de marcos y respaldado por el análisis de corpus (Ruppenhofer et al., 2006). Su objetivo es documentar la gama de posibilidades combinatorias tanto semánticas como sintácticas de cada palabra en cada uno de sus sentidos. Como afirma Fillmore (1977), “la gente entiende cosas realizando operaciones mentales sobre lo que ya saben. Tal conocimiento es describable en términos de paquetes de información llamados *frames*”. Un marco se basa en la idea de que ciertas palabras evocan ciertas situaciones en las que concurren determinados participantes. Las palabras que evocan un marco particular se llaman disparadores (*triggers*) o predicados. Un disparador pertenecerá a diferentes marcos si sus significados evocan diferentes situaciones (caso de las palabras polisémicas).

De igual manera, los predicados pueden considerarse como una función que describe la relación entre diferentes participantes dentro de un marco, adquiriendo cada uno de estos participantes diferentes roles semánticos en la oración. En la Figura 4 se muestran diferentes marcos, su definición, los predicados que lo activan y los roles semánticos asociados.

Exercising	<p><i>Definición:</i> An agent actively maintains or improves their level of physical fitness. <i>Predicados:</i> exercise.n, exercise.v, work out.v <i>Roles principales:</i> Agent</p>
Get_a_job	<p><i>Definición:</i> A new employee obtains a position with an employer, with which there are certain tasks associated. <i>Predicados:</i> hire on.v, sign on.v <i>Roles principales:</i> employee, employer, position, task</p>
Protest	<p><i>Definición:</i> A protester expresses a strong opinion either in support of or against an Issue, or the action(s) of a side which relate to the issue. <i>Predicados:</i> demonstrate.v, demonstration.n, protest.n, protest.v, protester.n <i>Roles principales:</i> action, issue, protester, side</p>

Figura 4. Definición, predicados y roles principales de los marcos ‘exercising’, ‘get a job’ y ‘protest’.

Al igual que WordNet, los marcos situacionales están conectados entre sí por diferentes tipos de relaciones al modo de una ontología, entre las que se encuentran:

1. *Herencia*. Relación “es un tipo de”. Esta es la relación más común para las ontologías. Determinados marcos son un tipo de otros considerados “padres”. Este sería el caso de *dead_or_alive* respecto al marco *state*. En FrameNet, esta relación se divide en *inherits_from* y *is_inherited_by*.
2. *Uso*. Un determinado marco presupone a otro como conocimiento previo. Así *medical_conditions* utiliza *observable_bodyparts*. Esta relación se divide en *Uses* y *Is_used_by*.
3. *Submarcos*. En esta relación, un marco “hijo” es un subevento de otro más complejo que funciona como marco “padre”: *criminal_process* tiene *arrest* como submarco. Esta relación puede ser dividida en *Subframe_of* y *Has_subframes*.

4. *Perspectiva sobre*. Estos marcos proporcionan visiones diferentes sobre un mismo evento. Así, *hiring* y *get_a_job* ofrecen diferentes perspectivas sobre *Employment_start*. Esta propiedad tiene dos grupos diferentes: *perspective_on* y *is_perspectivized_in*.
5. *Precede*. Estipula una relación temporal entre diferentes situaciones. *fall_asleep* precede al marco *sleep*. Se compone de dos tipos: *Precedes* y *Is_preceded_by*.
6. *Causalidad*. Un marco es la causa de una situación particular. Así *cure* causa *recovery*. Se divide en *Is_inchoative_of* y *Is_causative_of*.

FrameNet se encuentra todavía en proceso de creación. Su estado actual (versión 1.7) puede observarse en la Tabla 1 e incluye 1.222 marcos que cubren 13.640 disparadores diferentes y 10.542 roles semánticos:

Tabla 1. Estado actual de FrameNet.

	FrameNet versión: R1.7
Marcos	1.222
Predicados	13.640
Promedio de predicados por marco	12,5
Roles semánticos descritos	10.542
Promedio de roles semánticos por marco	9,7

FrameNet se ha ampliado a otros idiomas como el alemán (Burchardt et al., 2006), el francés (Candito et al., 2014), portugués (Salomão, 2009), chino (Liping y Kaiying, 2005), sueco (Friberg Heppin y Toporowska Gronostaj, 2012), japonés (Ohara et al., 2004), coreano (Kim et al., 2016), hebreo (Hayoun y Elhadad, 2016), letón (Nespore-Berzkalne et al., 2018) y español. Estos nuevos proyectos han tomado como punto de partida los marcos de Berkeley (inglés), aunque el grado de adhesión al original inglés FrameNet depende del proyecto. La Tabla 2 muestra el grado de coincidencia.

Tabla 2. Coincidencia de marcos el inglés de Berkeley.

Idioma	Número de marcos actual	Número de marcos compartidos con la versión inglesa	Porcentaje de coincidencia con la versión inglesa
Sweden FN	1.195	1.086	91%
Chinese FN	757	757	100%
French FN	121	73	60%
Spanish FN	325	280	86%

El Spanish FrameNet (Subirats y Petruck, 2003; Subirats, 2013) está desarrollado conjuntamente por la Universidad Autónoma de Barcelona y The International Computer Science Institute in Berkeley (California). Este proyecto investiga los diversos patrones de conceptualización que intervienen en el léxico español y su corpus cuenta en la actualidad con alrededor de 940 millones de palabras de diferentes géneros: principalmente periódicos, textos de noticias, reseñas de libros y ensayos de humanidades tanto de América como de España. Hasta ahora se han definido 325 marcos y 1.124 unidades léxicas.

3. EL TRASVASE DE INFORMACIÓN LINGÜÍSTICA DESDE FRAMENET.

El desarrollo de recursos léxicos requiere de un gran esfuerzo, a esto habría que sumar que estas bases de datos se centran principalmente en el inglés. Es por ello por lo que poder reutilizar o transferir la información de estos repositorios a otros idiomas se considera como una de las posibles soluciones al problema (Cristea y Pistol, 2012). Como veremos más adelante, la transferencia de unidades léxicas desde FrameNet ya ha sido abordada desde otros estudios y muchas de estas propuestas alinean los recursos de WordNet y FrameNet de tal manera que se puedan obtener correspondencias entre lenguas. En la Figura 5 se muestran dos de los diferentes predicados del marco *medical conditions*: *ailment.n* y *acromegaly.n*. Tales predicados se asocian en este marco con los *synsets* 14055408n y 14368483n de WordNet. Como ya indicamos, los *synsets* agrupan una serie de elementos léxicos relacionados por “sinonimia cognitiva”, pudiendo utilizarse estos índices para obtener su correlato en otras lenguas y, de esta manera, trasvasar los predicados del marco a todas las lenguas representadas en esta ontología léxica.

<i>Predicados del marco:</i> <i>MEDICAL CONDITIONS TRIGGERS</i>	<i>Synset asociado</i>	<i>Unidades para tal synset en inglés</i>	<i>Unidades para tal synset en español</i>	<i>Trasvase al español</i>
ailment.n	14055408n	ailment.n, complaint.n, ill.n	achaque, dolencia	achaque, dolencia,
acromegaly.n	14368483n	acromegaly.n	acromegalia	acromegalia,
Otros				

Figura 5. Proceso de trasvase usando WordNet.

La complejidad, no obstante, radica en el hecho de que muchas de estas unidades son polisémicas, por lo que se encuentran representadas en diferentes marcos y *synsets*, generando por ello un problema de desambiguación léxica que necesitamos resolver. Para entender mejor esta tarea, la Figura 6 ilustra los tres *synsets* asociados con “cold.n” y los marcos ligados a los respectivos sentidos:

<i>Synsets asociados a la unidad:</i> <i>'COLD.n'</i>	<i>Synset en inglés</i>	<i>Synset en español</i>	<i>Marcos que lo representa</i>
14145501n	cold.n common_cold.n	Catarro gripe resfriado resfriado_común	Medical conditions
05725676n	cold.n coldness.n	frialdad frío	Ambient temperature Subjective temperature
05015117n	cold.n coldness.n frigidity.n frigidness.n low_temperature.n	frío temperatura_baja	Ambient temperature Subjective temperature

Figura 6. *Synsets* asociados a la unidad ‘cold.n’ en inglés y español.

Las propuestas que buscan el trasvase de los disparadores de los marcos de FrameNet mediante WordNet se centran en encontrar el *synset* más adecuado para cada uno de los predicados del inglés. Se trata de un aspecto problemático ya que podemos llegar a traducción erróneas.

La Tabla 3 relaciona los disparadores de FrameNet con los *synsets* asociados a ellos en EuroWordNet. Como podemos ver, el 24,8% de los 13.631 predicados posee un solo sentido (monosémicos). Sin embargo, el 62.5% posee más de uno (polisémicos), lo que conlleva algún tipo de desambiguación si queremos usar WordNet para traducirlos a otras lenguas. El 12.5% no está representado en EuroWordNet, por lo que no hay posibilidad de asociarle ningún *synset*.

Tabla 3. Número de *synsets* asociados a las unidades de FrameNet.

<i>Número de marcos</i>	<i>Número total de predicados</i>	<i>Porcentaje total de synsets asociados</i>				<i>Sin representación en WordNet</i>
		1	2	3	+ 3	0
1222	13631	3393 (24.8%)	2504 (18.3%)	1582 (11.6%)	4444 (32.6%)	1708 (12.5%)

La mayoría de los trabajos que tratan de ofrecer una solución para este problema utiliza algún tipo de relación semántica y cognitiva (hiperonimia, hiponimia, sinonimia, etc.) para determinar cuál es el *synset* que mejor representa al predicado que se quiere desambiguar. La idea básica que subyace a esta forma de abordar la desambiguación es que las unidades léxicas están interconectadas, por lo que podemos llegar a determinar el valor de una unidad léxica mediante la observación de sus relaciones con el resto de elementos que aparecen con ella. Entre los principales trabajos que han desarrollado esta cuestión, encontramos a Benfeng y Fung (2004), los cuales construyen un FrameNet para el chino asociando los predicados de FrameNet a HowNet, una ontología en línea para esta lengua, que muestra las relaciones entre conceptos y sus correspondencias léxicas con el inglés. El trasvase se realiza eligiendo las clases semánticas chinas con más enlaces a un determinado marco

en Framenet. Burchardt, Erk y Frank (2005) utilizan WordNet para ampliar la cobertura de FrameNet. Para ello, analizan los sinónimos e hiperónimos de unidades que no se encuentran en esta base de datos léxica, y así determinar qué marco es el mejor candidato para cada término. En el 87% de los casos fueron capaces de asignar un marco con un 45% de precisión. Por su parte, Johansson y Nugues (2007) utilizan, entre otras medidas, conjuntos de hiperónimos para usarlos como vectores de características para cada uno de los sentidos de una determinada palabra. Posteriormente se entrenó el algoritmo SVM³ para cada marco y así decidir si un nuevo lema pertenecía a ese marco o no. Se alcanzó una precisión de 0,788 y un *recall* de 0,314 con esta técnica. Por otro lado, Pennacchiotti et al. (2008) experimentan con un modelo basado en WordNet que analiza qué *synsets* son activados por varias unidades léxicas del marco o sus hiperónimos, llegando a una precisión del 0,78 en función del número de marcos considerados. Posteriormente, Tonelli y Pianta (2009) vinculan WordNet y FrameNet utilizando la superposición de las glosas de cada *synset* y las definiciones de los marcos. De esta manera, un mayor número de vínculos entre un *synset* determinado con un marco se toma como indicador de asociación y llegan a obtener una precisión de 0,71 y un *recall* de 0,62. Por su lado también, Laparra, Rigau y Cuadros (2010) integran WordNet y FrameNet utilizando una versión diferente del algoritmo SSI⁴ que calcula el camino más corto de todos los valores posibles de una determinada unidad léxica en un marco, al conjunto de unidades ya desambiguadas en ese marco, obteniendo una precisión de 0,76 y un *recall* de 0,74. Igualmente interesante es la propuesta de Ferrández et al. (2010), los cuales explotan las relaciones semánticas en FrameNet y WordNet para calcular las correlaciones entre las unidades léxicas y *synsets*. Para ello, no solo se usan las relaciones semánticas que existen en WordNet sino también las explicitadas en FrameNet. A partir de ahí se calcula la relación entre un *synset* y un marco atendiendo a la distancia e importancia de las relaciones semánticas que existen entre los diferentes valores de la unidad y sus relaciones en ambos recursos. Obtienen una precisión de 0,772. En la misma línea de integración de recursos, López de Lacalle et al. (2014) amplían las unidades léxicas en Semlink mediante la integración de recursos de semántica léxica como FrameNet, VerbNet, PropBank y WordNet, utilizando algoritmos basados en grafos como el SSI que ya aparecía en Laparra et al. (2010) y logrando una precisión de 82,9 y un *recall* de 81,1 al mapear WordNet y FrameNet. Finalmente, en Crespo (2021) encontramos la integración de conocimiento de corpus sobre estas relaciones léxicas para el mapeo de marcos del dominio médico con WordNet, logrando un *recall* del 90% y una precisión del 83% sobre tal selección.

4. INTEGRACIÓN AUTOMÁTICA DE FRAMENET Y WORDNET: METODOLOGÍA.

La propuesta que presentamos en este trabajo incluye elementos que ya se encontraban en estudios anteriores, como la importancia de las relaciones semánticas en WordNet y FrameNet. Sin embargo, este enfoque se caracteriza por realizar un “trimming” o limpieza gradual de aquellos *synsets* menos relevantes. Se trata del proceso inverso a los considerados anteriormente, ya que el sistema considera todos los valores posibles como aceptables en un primer momento, y, paso a paso, va eliminando aquellos menos adecuados. De esta manera, en el punto de partida inicial, se asocia cada unidad con todos los posibles *synsets* en los que aparece en WordNet, y, luego, progresivamente se van recortando o eliminando los valores menos destacados respecto al marco que se está analizando. La situación se ilustra de la siguiente manera:

SITUACIÓN INICIAL	ESTADO 1	ESTADO 2	ESTADO 3
mark.v 00190023	mark.v 00190023	mark.v 00190023	mark.v 00190023
mark.v 01275762	mark.v 01275762	mark.v 00657728	mark.v 00651991
mark.v 01588493	mark.v 00662182	mark.v 00612612	mark.v 00921738
mark.v 01551195 →	mark.v 00657728 →	mark.v 00651991 →	
mark.v 01004062	mark.v 00612612	mark.v 00508032	
mark.v 01062395	mark.v 00651991	mark.v 00921738	
mark.v 00662182	mark.v 00508032		
mark.v 00657728	mark.v 00921738		
mark.v 00612612	mark.v 00800750		
mark.v 00651991	mark.v 02508245		
mark.v 00508032			
mark.v 02118476			
mark.v 00921738			
mark.v 00800750			
mark.v 02508245			

Figura 7. Proceso de eliminación de *synsets* inadecuados.

³ Support Vector Machines o máquinas de soporte vectorial son un conjunto de algoritmos de aprendizaje supervisado desarrollados por Vladimir Vapnik y su equipo en los laboratorios AT&T.

⁴ El algoritmo *Structural Semantic Interconnections* (SSI) es un aproximación para la desambiguación del sentido de unidades léxicas. Para más información ver: R. Navigli y P. Velardi (2005). “Structural semantic interconnections: a knowledgebased approach to word sense disambiguation”. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (PAMI), 27/7, 1063–1074.

El proceso consta de las siguientes partes:

Estado 1. En este primer estado se realiza una comparación de los sentidos de los predicados que queremos desambiguar y sus relaciones de hiperonimia, hiponimia y derivados, con respecto a aquellos de las unidades monosémicas (asociadas con un solo *synset*) del mismo marco. Como ya indicamos anteriormente, entre los predicados de los marcos se establecen diferentes relaciones semánticas y conceptuales. Así, por ejemplo, el marco *medical conditions* incluye disparadores como *ailment.n*, *disease.n*, *leukemia.n*, *health.n*, *healthy.a*, etc Si observamos que uno de los valores de una unidad ambigua coincide con el de una unidad monosémica (solo tiene asociado un *synset*, no necesita ser desambiguada), asumimos que ese será el sentido apropiado para ese marco. De esta manera, si el sistema encuentra coincidencias, este asume que ese es uno de los valores que busca. Por el contrario, descarta aquellos en los que no se hayan producido similitudes. En la situación en la que ninguno de los sentidos haya coincidido, se conservan todos para evaluarlos en otro estado posterior.

Marco	Predicado	Sentidos asociados	Número de coincidencias
Sign	sign.n	06791372n	12
		06876144n	11
		06811625n	11
		06793231n	11
		06646243n	17
		07286014n	2
		07276018n	1
		13856574n	2
		14301785n	8
		04217882n	0
Recovery	recuperate.v	08685677n	0
		00092690v	12
		00528339v	7
		02249438v	12

Figura 8. Número de coincidencias entre los predicados y resto de unidades

En la Figura 8 el sistema rechaza los sentidos *04217882n* y *08685677n* de la '*sign.n*'.

Estado 2. En este segundo paso, se vuelve a crear una bolsa con hiperónimos, hipónimos y derivados de cada uno de los valores que hay que desambiguar, pero amplía este mismo proceso para el resto de sentidos de las unidades que aparecen en el marco. De esta manera, para cada uno de los valores de la unidad léxica, el sistema cuenta el número de coincidencias que se producen entre ambos conjuntos. Una vez cuantificadas las coincidencias, se ordenan y el sistema se queda con los valores iguales al valor máximo.

Marco	Predicado	Sentidos asociados	Número de coincidencias
Sign	sign.n	06791372n	18
		06876144n	21
		06811625n	14
		06793231n	18
		06646243n	27
		07286014n	27
		07276018n	4
		13856574n	3
		14301785n	10
		Recovery	recuperate.v
00528339v	8		
02249438v	14		

Figura 9. Número de coincidencias entre relaciones semánticas de predicados y resto de unidades.

En la Figura 9 el sistema escogería los sentidos *06646243n* y *07286014n* de la '*sign.n*'.

Estado 3. Finalmente, se amplían las relaciones anteriores con las que aparezcan en los predicados de aquellos marcos relacionados de acuerdo con las especificadas en FrameNet. Solo se tienen en cuenta los valores restantes tras el estado 2. Una vez que se hallan las relaciones semánticas, se repite el proceso de comparación de unidades y se escogen aquellas iguales al valor máximo.

Marco	Relaciones
<i>Death</i>	Inherits from: <i>Ceasing_to_be</i> Is Inherited by: Perspective on: Is Perspectivized in: Uses: Is Used by: Subframe of: <i>Cycle_of_life_and_death</i> Has Subframe(s): Precedes: Is Preceded by: <i>Being_born, Dying, Giving_birth</i> Is Inchoative of: <i>Dead_or_alive</i> Is Causative of:

Figura 10. Relaciones entre marcos en FrameNet.

Para ejemplificar y mostrar la precisión de nuestro sistema, partimos de la selección desambiguada manualmente en Crespo (2021), que permite mostrar fácilmente los valores de precisión y *recall* de cada estado. Se trata de una selección reducida, compuesta de 11 marcos asociados al dominio de la medicina, pero que permite ilustrar bien la eficiencia de cada estado en la desambiguación de unidades. En este trabajo, entendemos *recall* como el número de valores adecuados que han sido seleccionados por el sistema (sin contar los seleccionados que no lo son), dividido por el total de los que en realidad son aceptables originalmente. La precisión será el conjunto de valores que el sistema considera como aceptables y que realmente lo son, dividido por el total de aquellos que son escogidos por el sistema. El *recall* mide el grado de pérdida de valores adecuados, y la precisión el grado de acierto en los sentidos escogidos (Powers, 2011). La Tabla 4 muestra el número de *synsets* por predicado en el marco en el estado inicial y tras aplicar cada uno de los estados:

Tabla 4. Resultados del proceso de desambiguación sobre la muestra de Crespo (2021).

Estado	Número de <i>synsets</i> asociados					Evaluación
	0	1	2	3	+ 3	
Inicial	18 (6,8%)	138 (52,4%)	39 (14,8%)	15 (5,7%)	53 (20%)	<i>Recall</i> : 100% <i>Precisión</i> : 41%
Estado 1	18 (6,8%)	173 (65,7%)	40 (15,2%)	11 (4,1%)	21 (7,9%)	<i>Recall</i> : 98% <i>Precisión</i> : 50%
Estado 2	18 (6,8%)	218 (83,6%)	21 (7,6%)	0 (0%)	6 (2,2%)	<i>Recall</i> : 88% <i>Precisión</i> : 84%
Estado 3	18 (6,8%)	223 (84,7%)	18 (6,8%)	0 (0%)	4 (1,5%)	<i>Recall</i> : 88% <i>Precisión</i> : 87%

Como se puede apreciar en los resultados, el análisis de relaciones semánticas es determinante para la determinación del valor apropiado de las unidades. En un primer momento, obtenemos un *recall* del 100% porque el sistema considera todos los sentidos como aceptables, es decir, no pierde ninguno relevante. En el estado 2 se pasa de una precisión del 50% al 84% con muy poca pérdida de *recall*. Finalmente, el estado 3 permite afinar la búsqueda, sin perder unidades relevantes. Como no podría ser de otra manera, los resultados nos indican que el sentido que una unidad léxica refleja en un marco se relaciona semánticamente con el del resto de unidades del mismo marco.

5. COMPARACIÓN CON EL SPANISH FRAMENET

A continuación, aplicamos la metodología expuesta en el apartado anterior para poder llevar a cabo la traducción de todo FrameNet de Berkeley al español. EuroWordNet será utilizado como diccionario multilingüe que permite traducir los disparadores del inglés al español o a cualquier otra lengua alineada. FrameNet contiene 13.631 disparadores en 1.222 marcos situacionales diferentes, y 8.530 de ellos (62,5%) tienen más de un sentido

conforme a WordNet. La Tabla 5 muestra la información estadística sobre los disparadores encontrados en FrameNet en el estado inicial:

Tabla 5. Synsets asociados a las unidades FrameNet en el estado inicial.

Estado	Número de synsets asociados					Total de palabras
	0	1	2	3	+ 3	
Inicial	1708 (12,5%)	3393 (24,8%)	2504 (18,3%)	1582 (11,6%)	4444 (32,6%)	13631

Tras la aplicación del proceso de desambiguación o eliminación de valores inapropiados de acuerdo con los estados anteriores los resultados son los siguientes:

Tabla 6. Synsets asociados a las unidades FrameNet en el estado final.

Estado	Número de synsets asociados					Total de palabras
	0	1	2	3	+ 3	
Final	1708 (12,5%)	10322 (75,7%)	884 (6,4%)	253 (1,8%)	464 (3,4%)	13.631

En su traducción al español se obtuvieron 7.923 unidades repartidas en los 1.222 marcos situacionales de que dispone FrameNet en esta versión⁵. Para evaluar la calidad de esta traducción se cotejó con el Spanish FrameNet original. Este recurso para el español contiene 325 marcos situacionales diferentes que se pueden consultar en su versión on-line. Sin embargo, no todos sus marcos pueden compararse, ya que solo 280 coinciden en la nomenclatura con el homólogo inglés y, de ellos, solo 144 contienen predicados que permitan la comparación.

De los 997 predicados de este repositorio para el español, 631 (63,2%) se encuentran recogidos en la traducción automática hecha para los mismos marcos. La Tabla 7 muestra la evaluación respecto a aquellos marcos más numerosos del Spanish FrameNet (aquellos con diez o más predicados). De esta selección, el 88,62% de nuestras unidades léxicas eran correctas tras una evaluación manual respecto al Diccionario de la Real Academia⁶ y al marco donde se tradujeron. Estos resultados son análogos a los presentados anteriormente para la selección de Crespo (2021). En todos los casos los predicados derivados de la traducción son más numerosos que los creados para el Spanish FrameNet.

Los errores de traducción varían en función del marco, destacando el caso del marco *likelihood* con una tasa de error del 42,17%. Además, observamos tanto marcos con una tasa de error nula pero que solo coinciden parcialmente con la versión española del Spanish FrameNet (tal es el caso de *change_of_phase* o *excreting*), como, al contrario, marcos que coinciden plenamente con nuestra traducción, pero que presentan errores (*similarity* o *taking_sides*).

Tabla 7. Evaluación de la traducción de FrameNet.

MARCOS	Tamaño original	Coincidencia	Tamaño traducción	Traducción errónea
Arriving	15	5(33%)	46	13 (28,26%)
Attaching	11	10(91%)	82	2 (2,44%)
Attempt_suasion	9	6(67%)	36	3 (8,33%)
Awareness	13	8(62%)	73	14 (19,18%)
Becoming_aware	12	7(58%)	41	12 (29,27%)
Body_movement	11	6(55%)	87	7 (8,05%)
Bragging	10	9(90%)	62	9 (14,52%)
Cause_change	10	9(90%)	33	2 (6,06%)
Cause_to_experience	9	1 (11%)	12	1 (8,33%)
Certainty	11	8(73%)	38	3 (7,89%)
Change_of_phase	14	8(57%)	33	0 (0,00%)
Change_posture	13	7(54%)	41	5 (12,20%)
Coming_to_believe	9	8(89%)	51	16 (31,37%)
Coming_up_with	9	4(44%)	25	4 (16,00%)

(Tabla 7 continúa en la página siguiente)

⁵ Esta traducción puede consultarse en la web del Instituto de Investigación en Lingüística Aplicada de la Universidad de Cádiz: <https://ila.uca.es/laboratorio-de-linguistica-computacional/>

⁶ Real Academia Española: *Diccionario de la lengua española*, 23.ª ed., [versión 23.3 en línea]. <<https://dle.rae.es>> [15 de septiembre de 2020]

(Tabla 7 continúa de la página anterior)

MARCOS	Tamaño original	Coincidencia	Tamaño traducción	Traducción errónea
Communication_manner	15	10(67%)	59	4 (6,78%)
Communication_response	10	(67%)	12	1 (8,33%)
Cotheme	11	6(55%)	32	8 (25,00%)
Departing	10	4(40%)	46	7 (15,22%)
Desiring	12	11(92%)	75	7 (9,33%)
Emanating	9	2(22%)	18	1 (5,56%)
Escaping	9	2(22%)	18	3 (16,67%)
Excreting	11	8(73%)	28	0 (0,00%)
Expectation	11	6(55%)	25	2 (8,00%)
Expensiveness	10	6(55%)	28	6 (21,43%)
Experiencer_obj	35	25(71%)	257	22 (8,56%)
Filling	15	10(67%)	84	2 (2,38%)
Giving	12	9(75%)	41	2 (4,88%)
Grasp	9	6(67%)	23	2 (8,70%)
Ingestion	13	9(69%)	49	0 (0,00%)
Judgment_communication	16	14(88%)	183	14 (7,65%)
Likelihood	9	6(67%)	83	35 (42,17%)
Locale_by_use	9	4(44%)	118	2 (1,69%)
Locative_relation	33	5(15%)	50	4 (8,00%)
Moving_in_place	12	5(42%)	41	3 (7,32%)
Path_shape	11	2(18%)	81	20 (24,69%)
Perception_active	15	12(80%)	56	5 (8,93%)
Placing	12	12(100%)	125	23 (18,40%)
Precipitation	10	4(40%)	44	0 (0,00%)
Predicting	14	10(71%)	20	0 (0,00%)
Removing	12	9(75%)	91	9 (9,89%)
Request	17	12(71%)	48	5 (10,42%)
Reveal_secret	9	7(78%)	20	0 (0,00%)
Scrutiny	16	13(81%)	64	3 (4,69%)
Sensation	14	11(79%)	49	7 (14,29%)
Similarity	10	10(100%)	71	11 (15,49%)
Taking_sides	10	10(100%)	37	1 (2,70%)

Entre los problemas observados en la traducción⁷, nos encontramos unidades correctas pero dialectales como “dragonear” en el marco *communication* o “manosear” en *experiencer_object*; términos con valores correctos, pero en desuso como “comportar” en el marco *cotheme* o “derrisión” en *judgment_communication*, así como términos correctos, pero perteneciente a contextos de uso específicos como “copar” del ámbito militar en el marco *removing*. Otros problemas observados se relacionan directamente con el propio WordNet. Así, en este repositorio se caracterizan los términos “globo ocular” o “claramente” como verbos, o se introducen elementos inexistentes en el diccionario como el verbo “conver”. Todos estos casos se contabilizaron como errores de la traducción.

Finalmente, una vez identificados los *synsets*, se puede utilizar la información de WordNet para ampliar la cobertura original de FrameNet. El proceso se ilustra en la Figura 11:

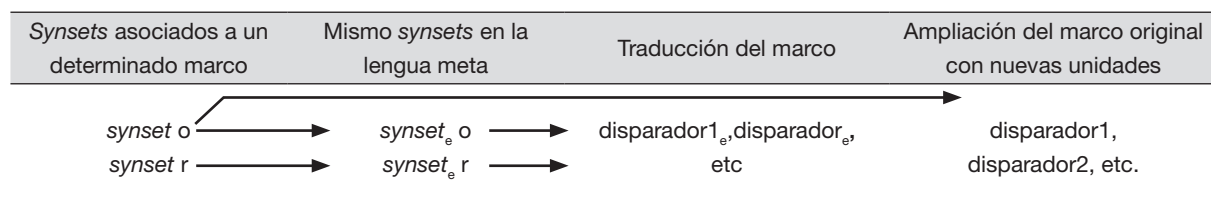


Figura 11. Traducción y ampliación del marco situacional.

Se obtiene una coincidencia de 11.923 predicados con respecto a la versión 1.7 de FrameNet, a la que se añaden 17.854 elementos nuevos que no existen en esta versión.

⁷ Al cotejar con el DRAE, *supra cit.*

6. CONCLUSIONES

El déficit de recursos del que adolecen la Lingüística y el Procesamiento del Lenguaje Natural está llevando al desarrollo de técnicas que permitan la reutilización de los ya existentes. Esta carencia de recursos se hace incluso más pronunciada para el procesamiento semántico, con lo que se hace necesario desarrollar técnicas que permitan extender su uso. Entre estos recursos, FrameNet es un proyecto cuya estructura basada en marcos situacionales permite su adaptación a cualquier lengua, ya que estos representan situaciones estereotipadas de la realidad (Minsky, 1975) (tal es el caso, por ejemplo, del marco *commercial_transaction*, que incluye elementos de marco como *buyer*, *goods*, *money*, *seller*, etc). La posibilidad de aplicar los mismos marcos a diferentes idiomas motiva el Global FrameNet (Torrent et al., 2018), que reúne los FrameNet existentes con el fin de comprender mejor las similitudes y diferencias entre los distintos idiomas. Este proyecto analiza en qué medida los marcos semánticos creados para el inglés son apropiados para analizar otros idiomas.

El trasvase de información de FrameNet a una nueva lengua supone, entre otras cosas, la traducción de sus predicados, manteniendo las relaciones entre marcos y sus roles semánticos. Las principales aproximaciones utilizan EuroWordNet como recurso léxico multilingüe. La identificación de tales *synsets* se realiza sobre las relaciones semánticas existentes entre los elementos de cada marco de FrameNet y entre marcos de este mismo recurso. Los resultados demuestran la idoneidad de esta metodología. Además, la asociación de EuroWordNet y FrameNet permite proveer a este último con una gran variedad de relaciones semánticas para hacerlo más rico y complejo.

Nuestro trabajo futuro plantea la posibilidad de creación de una red terminológica a partir del análisis automatizado de textos de especialidad y su representación en la estructura de relaciones semánticas de FrameNet. Ya en Casas Gómez (2020) se plantea que entre las unidades terminológicas se establecen diferentes relaciones de carácter lógico y conceptual. Se trata de una de las líneas de trabajo dentro de los proyectos "Comunicación especializada y terminografía: usos terminológicos relacionados con los contenidos y perspectivas actuales de la semántica léxica", "Lingüística y nuevas tecnologías de la información: la creación de un repositorio electrónico de documentación lingüística" y "Lingüística y Humanidades Digitales: base de datos relacional de documentación lingüística" sobre el ámbito de la terminología en la Semántica Léxica. El trabajo que aquí presentamos es un paso intermedio hacia este tipo de representación. Se trata, en definitiva, de dotar la investigación sobre el léxico de instrumentos eficaces que ayuden no solo a la sistematización de las unidades léxicas que conforman el inventario de una lengua concreta, sino a la determinación de sus relaciones y posibilidades de uso en situaciones comunicativas específicas.

AGRADECIMIENTOS

Esta publicación ha sido financiada por el proyecto "Comunicación especializada y terminografía: usos terminológicos relacionados con los contenidos y perspectivas actuales de la semántica léxica" (Ref. FFI2014-54609-P) del Programa Estatal de Fomento de la Investigación Científica y Técnica de Excelencia. Subprograma Estatal de Generación del Conocimiento (convocatoria 2014 del Ministerio de Economía y Competitividad) y se inscribe en el proyecto "Bases metodológicas y recursos digitales para la creación de un léxico relacional de usos terminológicos de la semántica léxica (TerLexNet)", solicitado en la Convocatoria 2020 de Proyectos de I+D+i del Ministerio de Ciencia e Innovación. Igualmente cuenta con el apoyo del proyecto "Lingüística y nuevas tecnologías de la información: la creación de un repositorio electrónico de documentación lingüística" (Ref. FEDER-UCA18-107788), perteneciente a los Proyectos de I+D+i del Programa Operativo FEDER Andalucía 2014-2020, y "Lingüística y Humanidades Digitales: base de datos relacional de documentación lingüística" (Ref. PY18-FR-2511) de la Convocatoria 2018 de Ayudas a proyectos I+D+i (Modalidad "Frontera Consolidado") en el ámbito del Plan Andaluz de Investigación, Desarrollo e Innovación (Junta de Andalucía, PAIDI 2020).

REFERENCIAS

- Arano, S. (2005). "Thesauruses and ontologies". *Hipertext.net*, 3. Disponible en <https://www.upf.edu/hipertextnet/en/numero-3/tesauros.html>
- Baker, C., Fillmore, C.J. y Lowe, J.B. (1998). "The Berkeley FrameNet project", en C. Boitet y P. Whitelock (eds.), *Proceedings of the Thirty-Sixth Annual Meeting of the Association for Computational Linguistics and Seventeenth International Conference on Computational Linguistics*. San Francisco, California: Morgan Kaufmann Publishers, 86–90. <https://doi.org/10.3115/980845.980860>
- Bel, N., Bel, S., Espeja, S., Marimon, M., Villegas, M. (2008). "El proyecto CLARIN: una infraestructura de investigación científica para las humanidades y las ciencias sociales". *Digithum*, 10.

- Benfeng, C., y Fung, P. (2004). "Automatic Construction of an English-Chinese Bilingual FrameNet". *Proceedings of HLT-NAACL 2004: Short Papers*. Boston, Massachusetts: ACL, 29-32.
- Burchardt, A., Erk, K. y Frank, A. (2005). "A WordNet detour to FrameNet". *Sprachtechnologie, mobile Kommunikation und linguistische Ressourcen*, 8, 408-421.
- Burchardt, A., Erk, K., Frank, A., Kowalski, A., Padó, S. y Pinkal, M. (2006). "The SALSA Corpus: a German Corpus Resource for Lexical Semantics". *Proceedings of Language Resources and Evaluation Conference, 2006*. Genova: LREC, 969-974. Disponible en http://www.lrec-conf.org/proceedings/lrec2006/pdf/339_pdf.pdf
- Candito, M., Amsili, P., Barque, L., Benamara, F., de Chalendar, G., Djemaa, M., Haas, P., Huyghe, R., Yannick Mathieu, Y., Muller, P., Sagot, B., y Vieu, L. (2014). "Developing a French FrameNet: Methodology and First results". *Proceedings of the 9th edition of the Language Resources and Evaluation Conference*. Reykjavik: ELRA, 1-9.
- Casas Gómez, M. (2014). "A Typology of Relationships in Semantics". *Quaderni di semantica: Rivista Internazionale di Semantica Teorica e Applicata*, 35/2, 45-74.
- Casas Gómez, M. (2020). "Conceptual relationships and their methodological representation in a dictionary of the terminological uses of lexical semantics". *Fachsprache: Internationale Zeitschrift für Fachsprachenforschung -didaktik und Terminologie*, 42/1-2, 2-26. <https://doi.org/10.24989/fs.v42i1-2.1789>
- Civit Torruella, M., Aldezabal Roteta, I., Pociello Irigoyen, E., Taulé Delor, M., Aparicio Mera, J.J., Màrquez Villodre, L., Navarro Colorado, B., Castellví Vives, J. y Martí Antonín, M.A. (2005). "3LB-LEX: léxico verbal con frames sintáctico-semánticos". *Procesamiento del Lenguaje Natural*, 35, 367-373.
- Crespo, M. (2021). *Automatic Corpus-based translation of a Spanish FrameNet medical Glossary*. Colección Lingüística. Sevilla: Universidad de Sevilla.
- Cristea, D., y Pistol, I.C. (2012). "Multilingual linguistic workflows". *Multilingual Processing in Eastern and Southern EU Languages*. Low-resourced Technologies and Translation, Cambridge Scholars Publishing, UK, 228-246.
- Ferrández, Ó., Ellsworth, M., Muñoz, R., y Baker, C.F. (2010). "Aligning FrameNet and WordNet based on Semantic Neighborhoods". *Proceedings of the International Conference on Language Resources and Evaluation, LREC 2010*. Malta: ELRA, 310-314.
- Fillmore, C.J. (1977). "Scenes and Frames Semantics", en A. Zampolli (Ed.), *Linguistic Structures Processing*. Amsterdam: North Holland, 55-82.
- Friberg Heppin, K., y Toporowska Gronostaj, M. (2012). "The Rocky Road towards a Swedish FrameNet – Creating SweFN". *Proceedings of the Eighth conference on International Language Resources and Evaluation (LREC-2012)*. Estambul: ELRA, 261.
- Gilchrist, A. (2003). "Thesauri, taxonomies and ontologies—an etymological note". *Journal of documentation*, 59/1, 7-18. <https://doi.org/10.1108/00220410310457984>
- Hayoun, A. y Elhadad, M. (2016). "The Hebrew FrameNet Project". *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*. Portorož, Slovenia: European Language Resources Association (ELRA), 4341-4347.
- Hilera, J.R., Pagés, C., Martínez, J.J., Gutiérrez, J.A., y De-Marcos, L. (2010). "An evolutive process to convert glossaries into ontologies". *Information technology and libraries*, 29/4, 195-204. <https://doi.org/10.6017/ital.v29i4.3130>
- Johansson, R., y Nugues, P. (2007). "Using WordNet to Extend FrameNet Coverage", en P. Nugues, y R. Johansson (Eds.), *LU-CS-TR: 2007-240*. Lund: Department of Computer Science, Lund University, 27-30.
- Kim, J., Hahm, Y., y Choi, K. (2016). "Korean FrameNet Expansion Based on Projection of Japanese FrameNet". *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: System Demonstrations*. Osaka: ACL, 175-179.
- Kipper, K., Trang Dang, H., Schuler, W., y Palmer, M. (2000). "Building a class-based verb lexicon using tags". *Proceedings of the Fifth International Workshop on Tree Adjoining Grammar and Related Frameworks (TAG+5)*. Paris: ACL, 147-155.
- Kurdi, M.Z. (2017). *Natural language processing and computational linguistics 2: semantics, discourse and applications* (Vol. 2). Hoboken, Nueva Jersey: John Wiley & Sons. <https://doi.org/10.1002/9781119419686>
- Laparra, E., Rigau, G., y Cuadros, M. (2010). "Exploring the integration of WordNet and FrameNet". *Proceedings of the 5th Global WordNet Conference*. Mumbai: Global WordNet Association, 1-6.
- Liping, Y., y Kaiying, L. (2005). "Building Chinese FrameNet database". *Proceedings of the 2005 International Conference on Natural Language Processing and Knowledge Engineering*. Wuhan: IEEE, 301-306. <https://doi.org/10.1109/NLPKE.2005.1598752>
- López de Lacalle, M., Laparra, E., y Rigau, G. (2014). "Predicate Matrix: extending SemLink through WordNet mappings". *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*. Reykjavik: European Language Resources Association (ELRA), 903-909.

- Martí Antonín, M.A., y Taulé Delor, M. (2014). "Computational Hispanic Linguistics". *The Routledge Handbook of Hispanic Applied Linguistics*. London: Taylor and Francis, 350-370.
- McCrae, J.P., y Cimiano, P. (2015). "Linghub: a Linked Data based portal supporting the discovery of language resources". *Proceedings of the 11th International Conference on Semantic Systems, Semantics, 1481*. New York: Association for Computing Machinery, 88-91.
- Miller, G.A., Beckwith, R., Fellbaum, C., Gross, D., y Miller, K. (ed.) (1993). *Five Papers on WordNet, cls report 43. Technical report*. New Jersey: Cognitive Science Laboratory. Princeton University.
- Miller, J.E., y Brown, K. (2013). *The Cambridge dictionary of linguistics*. Cambridge: Cambridge University Press.
- Minsky, M. (1975). "A framework for representing knowledge". *Psychology of Computer Vision*. New York: McGraw-Hill, 211-277.
- Nespor-Berzkalne G., Saulite, B., y Gruzitis, N. (2018). "Latvian FrameNet: Cross-Lingual Issue". *Human Language Technologies - The Baltic Perspective*, 307. Amsterdam: IOS Press, 96-103.
- Ohara, K., Fujii, S., Otori, T., Suzuki, R., Saito, H., y Ishizaki, S. (2004). "The Japanese FrameNet Project: An Introduction". *LREC 2004: The Fourth International Conference on Language Resources and Evaluation*. Lisbon: LREC, 249-254.
- Palmer, M., Gildea, D., y Kingsbury, P. (2005). "The Proposition Bank: An Annotated Corpus of Semantic Roles". *Journal Computational Linguistics*, 31/1. MA: MIT Press Cambridge, 71-106. <https://doi.org/10.1162/0891201053630264>
- Pennacchiotti, M., De Cao, D., Basili, R., Croce, D., y Roth, M. (2008). "Automatic induction of FrameNet lexical units". *Proceedings of the 2008 conference on empirical methods in natural language*. Honolulu: ACL, 457-465. <https://doi.org/10.3115/1613715.1613773>
- Pieterse, V., y Kourie, D. G. (2014). "Lists, taxonomies, lattices, thesauri and ontologies: paving a pathway through a terminological jungle". *KO Knowledge Organization*, 41/3, 217-229. <https://doi.org/10.5771/0943-7444-2014-3-217>
- Powers, D.M. (2011). "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation". *Journal of Machine Learning Technologies*, 2/1, 37-63.
- Ruppenhofer, J., Ellsworth, M., Petruck, M.R.L., Johnson, C. y Scheffczyk, J. (2006). *FrameNet II: Extended Theory and Practice*. Disponible en <https://framenet2.icsi.berkeley.edu/docs/r1.7/book.pdf>
- Salomão, M. (2009). "FrameNet Brasil: um trabalho em progresso". *Calidoscópio*, 7/3, 171-182. <https://doi.org/10.4013/cld.2009.73.01>
- Subirats, C., y Petruck, M. R. L. (2003). "Surprise: Spanish FrameNet!". *Proceedings of Proceedings of the Workshop on Frame Semantics at the XVII. International Congress of Linguists (CD-ROM)*. Prague: Matfyzpress.
- Subirats, C. (2013). "La integración de la semántica de marcos y la semántica de simulación: aplicaciones al procesamiento semántico automático del español", en M^a Luisa Calero and M^a Ángeles Hermosilla (eds.). *Lingüística, Poética y Cognición*. Córdoba: Servicio de Publicaciones de la Universidad de Córdoba, 307-337.
- Tonelli, S., y Pianta, E. (2009). "A novel approach to mapping FrameNet lexical units to WordNet synsets (short paper)". *Proceedings of the Eight International Conference on Computational Semantics*. Tilburg: ACL, 342-345. <https://doi.org/10.3115/1693756.1693800>
- Torrent, T.T., Ellsworth, M., Baker, C.F. y Matos, E.E. (2018). "The Multilingual FrameNet Shared Annotation Task: A Preliminary Report". *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*. Miyazaki: ELRA, 62-68.
- Van Uytvanck, D., Zinn, C., Broeder, D., Wittenburg, P., y Gardelleni, M. (2010). "Virtual language observatory: The portal to the language resources and technology universe". *Proceedings of the Seventh conference on International Language Resources and Evaluation [LREC 2010]*. Malta: European Language Resources Association (ELRA), 900-903.
- Vossen, P. (ed.) (1998). *EuroWordNet: A Multilingual Database with Lexical Semantic Networks*. Dordrecht: Kluwer Academic Publishers. <https://doi.org/10.1007/978-94-017-1491-4>
- Vossen, P. (ed.) (2002). *EuroWordNet: general document*. Disponible en <http://vossen.info/docs/2002/EWNGeneral.pdf>
- Vilches-Blázquez, L.M., García Silva, A., y Villazón Terrazas, B. (2009). "Construcción de ontologías a partir de tesauros". *Semántica Espacial y descubrimiento de conocimientos para desarrollo sostenible*. La Habana: CUJAE, 59-78.