

## 3LB-LEX: léxico verbal con frames sintáctico-semánticos\*

M. Civit★, I. Aldezabal◊, E. Pociello◊, M. Taulé★, J. Aparicio★

L. Màrquez●, B. Navarro◊, J. Castellví★, MA. Martí★

★CLiC, Centre de Llenguatge i Computació  
{mcivit, mtaule, joan.castellvi, amarti}@ub.edu  
juanapariciomera@yahoo.es

◊IXA Taldea Euskal Herriko Unibertsitatea  
elisabete@si.ehu.es, izaskun.aldezabal@ehu.es

●Departament de Llenguatges i Sistemes Informàtics (UPC)  
lluism@lsi.upc.edu

◊Departamento de Lenguajes y Sistemas Informáticos (UA)  
borja@dlsi.ua.es

**Resumen:** La creación de léxicos (verbales) computacionales es larga y costosa. A partir de los corpora creados en el proyecto 3LB se deriva un léxico verbal con información sintáctica y semántica (synsets de EWN). A partir de esta información se establece la correspondencia entre funciones sintácticas y papeles temáticos para cada sentido de cada verbo. El último paso será el etiquetado de los corpora con papeles temáticos. Como resultado, los corpora de 3LB se habrán enriquecido con el etiquetado de papeles temáticos y el léxico verbal con los frames semánticos.

**Palabras clave:** Léxico verbal, anotación semántica, papeles temáticos

**Abstract:** The creation of computational (verbal) lexicons is a time-consuming task. From the corpora created at the 3LB project, a verbal lexicon with syntactic and semantic (synsets from EWN) information is being built. From this information the correspondence between syntactic functions and thematic roles for each verb sense will set. The last step will be the tagging of the corpora with thematic roles. As a result of this, the 3LB corpora will be enriched with the tagging of thematic roles and the verbal lexicon with the semantic frames.

**Keywords:** Verbal lexicon, Semantic Annotation, Thematic Roles

### 1. Introducción

En este artículo presentamos un nuevo recurso léxico, 3LB-LEX, para el euskera, catalán y castellano. Se trata de tres léxicos verbales enriquecidos con esquemas sintáctico-semánticos obtenidos a partir de los corpora del proyecto 3LB.

Los objetivos del proyecto 3LB (Palomar et al., 2004) fueron los de etiquetar sintáctica y semánticamente tres corpus correspondientes al catalán, castellano y euskera. Desde el punto de vista sintáctico se procedió al etiquetado de constituyentes y funciones para el caso del catalán y castellano y al etiquetado de dependencias para el euskera ((Civit, Bufí, y Valverde, 2004), (Civit, 2004), (Aranzabe et al., 2003)).

Para realizar este etiquetado se aplicaron pruebas de acuerdo entre anotadores para evaluar el grado de coincidencia en la anotación. En el etiquetado se mantuvo el orden superficial de constituyentes y se tuvieron en consideración las funciones sintácticas básicas: sujeto, los complementos directo, indirecto, predicativo, agente, de régimen, circunstancial y el atributo. Gracias a los controles establecidos para garantizar el acuerdo entre anotadores tanto a nivel sintáctico como semántico, los datos del corpus son altamente consistentes (un 96 % de acuerdo en la anotación sintáctica) (Civit et al., 2003), lo que redundará en la calidad de los léxicos derivados.

Para el etiquetado semántico se utilizó una versión fija de las versiones española, catalana y euskera de EuroWordNet y se asoció a cada núcleo nominal, adjetivo y verbal un único synset. El etiquetado semántico es, sin

\* El trabajo aquí presentado ha sido parcialmente financiado por los proyectos XTRACT-2 (BFF2002-04226-C03-03), CESS-CE (HUM2004-21127-E) y R2D2 (TIC-2003-07158-C04-01)

embargo, parcial en todas las lenguas: para el español están etiquetados todos los nombres y el 50 % de los verbos; para el catalán se ha etiquetado un 10 % del corpus (nombres, verbos y adjetivos), y para el euskera se han etiquetado todos los nombres.

Los corpora del catalán y el español, tienen 100.000 palabras que se corresponden respectivamente con 2.600 oraciones en catalán y 4.000 en castellano, mientras que el del euskera tiene 50.000 palabras que se corresponden con 2.750 oraciones. Hay que tener en cuenta que el euskera es una lengua aglutinante y que muchas palabras, en especial las de carácter funcional como preposiciones y artículos, se hallan sufiadas a la raíz.

La anotación de papeles temáticos en 3LB-LEX se hará siguiendo la propuesta de PropBank (Palmer, Kingsbury, y Gildea, 2005). A partir de los esquemas sintácticos de cada verbo obtenidos del corpus, se establecerá la correspondencia entre las funciones sintácticas y los roles temáticos de cada sentido verbal. El etiquetado del corpus con roles temáticos se realizará de manera automática, con un postproceso de validación manual.

La motivación para la creación de un léxico de estas características estriba, por un lado, en el interés que puede tener para los estudios lingüísticos ya que permitirá realizar estudios comparativos sobre el comportamiento sintáctico y semántico de las lenguas del proyecto, además del inglés. La compatibilidad en el etiquetado con papeles temáticos va a facilitar mucho este tipo de análisis. Por otro lado, el léxico se podrá utilizar para realizar aprendizaje automático de parsers sintáctico-semánticos, que se podrán utilizar para diversas tareas, en especial, para la extracción de información y de conocimiento.

Este artículo se estructura del siguiente modo: en la sección 2 presentamos el proceso de derivación del léxico verbal y el contenido de las entradas resultantes. En la sección 3 se presenta la metodología para el enriquecimiento de los árboles de análisis con información sobre los papeles temáticos; se justifica el tagset semántico utilizado y se analiza el caso específico de los complementos circunstanciales y de régimen. En la sección 4 se expone cómo se utilizará el léxico para el etiquetado automático de corpus así como las líneas futuras de investigación.

## 2. Derivación del léxico verbal

La construcción de un léxico es larga y costosa. En las teorías sintácticas actuales el léxico requiere una especificación completa sobre la fonología, la sintaxis y la semántica de las entradas léxicas. La construcción de un léxico de estas características puede enfocarse desde una metodología *top-down*, analizando lingüísticamente las entradas y consultando tratados gramaticales y diccionarios, o bien se puede seguir un enfoque *bottom-up*, desde los datos. La aproximación que se propone se basa en este último enfoque: a partir de los corpora etiquetados, se infiere uno o varios esquemas sintácticos de cada una de las entradas verbales. Si bien no se obtiene una cobertura total sobre los usos de cada entrada, el material obtenido constituye un punto de partida válido susceptible de ser ampliado y completado mediante el incremento de la cantidad de corpus etiquetado o bien con metodologías complementarias.

### 2.1. Proceso de derivación del léxico

Para la derivación del léxico verbal se han tenido en cuenta tanto las oraciones como las cláusulas, incluyendo los verbos en forma finita y los verbos en forma no finita. El número total de formas verbales del corpus en castellano es de 7.127 y el número de verbos distintos de 1.070. Para el del catalán el número total de formas verbales es de 7.033 mientras que el número de verbos distintos es de 834. La media de apariciones de un verbo en castellano es de 6,66 y en catalán de 8,43. El número de verbos que aparecen una sola vez en el corpus es de 392 para el castellano y 304 para el catalán, por lo que la media de aparición de los verbos, teniendo en cuenta sólo los que aparecen dos o más veces, es de 9,93 para el castellano y 12,69 para el catalán. En el corpus del euskera, de cada cuatro palabras una es un verbo, es decir, aproximadamente un 25 % del total del corpus. El número de formas verbales es de 13.261, y el número de verbos distintos es de 375. La diferencia con respecto al castellano y el catalán se debe seguramente a que el euskera lexicaliza menos en el verbo y depende más de los argumentos<sup>1</sup>.

<sup>1</sup>Por el momento, partimos de un número moderado de verbos y de ocurrencias, debido a la limitación de los recursos actualmente disponibles, pero se acaba de conceder un proyecto que permitirá aumentar

La extracción de la forma verbal se ha hecho teniendo en cuenta los nodos de los árboles que incluyen las formas verbales simples y compuestas, tanto las correspondientes a los llamados tiempos compuestos como a las perífrasis verbales. En las formas compuestas el lema corresponde a la forma más a la derecha que es la forma con mayor carga léxica. El resto de elementos de la entrada léxica está formado por todos los constituyentes que tienen una función sintáctica respecto del verbo de la entrada, por lo que no aparecen en esta relación los complementos del nombre.

## 2.2. Contenido de la entrada

Para cada aparición de un verbo en el corpus se extrae una entrada derivada que contiene, además de la forma verbal, los constituyentes que la acompañan y sus funciones sintácticas. Para el euskera las funciones son las siguientes:

Func.	Glosa
ncsubj	sujeto no oracional
ncobj	objeto no oracional
nczobj	O.I. no oracional
ncmod	modificador no oracional
cmod	modif. orac. finito
xmod	modif. orac. no finito
ccomp_subj	subordinada completiva finita en función de sujeto
ccomp_obj	subordinada completiva finita en función de objeto
ccompobj	subordinada completiva finita en función de O.I.
xcomp_subj	subordinada completiva no finita en función de sujeto
xcomp_obj	subordinada completiva no finita en función de objeto
xcomp_zobj	subordinada completiva no finita en función de O.I.
ncpred	predicativo oracional
xpred	predicativo no oracional

Las funciones sintácticas utilizadas en catalán y castellano son las siguientes:

los corpora a 500.000 palabras para catalán y castellano y a 350.000 para el euskera. Por consiguiente, el léxico verbal aumentará sensiblemente.

Func.	Glosa
SUJ	Sujeto
CD	Compl. directo
CI	Compl. indirecto
ATR	Atributo
CPRED	Compl. predicativo
CREG	Compl. regido
CC	Compl. circunstancial
CAG	Compl. agente

Por ejemplo, para la aparición del verbo *parlar* (*hablar*) en Cat3LB: *En la segona de les conferències, programada per al dia 9 de juny, Robert Brufau parlarà dels edificis en què l'estructura juga un paper decisiu en l'arquitectura a partir d'obres diverses*<sup>2</sup> se ha extraído la entrada derivada que aparece en el cuadro 1.

Func.	Texto
CC	En la segona de les conferències, programada per al dia_9_de_juny,
SUJ	Robert_Brufau
verb	parlarà
CREG	dels edificis en què l'estructura juga un paper decisiu en l'arquitectura
CC	a_partir_d'obres diverses

Cuadro 1: Entrada derivada: *parlar*

Tal como se observa en la entrada anterior se mantiene el orden superficial de las diferentes funciones sintácticas, lo que permite realizar estudios sobre el orden de aparición de sujeto y complementos<sup>3</sup>.

Los cuadros 2 y 3 muestran dos entradas derivadas de Cast3LB para dos apariciones del verbo *mover* que presenta dos esquemas distintos (las cláusulas con el verbo *mover* aparecen en negrita en las oraciones).

La oración correspondiente a *mover-1* es *\*0\* se enfocó desde muy antiguo hacia el transporte de vivos y bienes, del que \*0\* tenía ejemplo en las naves **que movía el viento por aguas y mares con estimable eficiencia***

La oración correspondiente a *mover-2* es *Este último partido, entre dos equipos que*

<sup>2</sup>En la segunda de las conferencias, programada para el día 9 de junio, Robert Brufau hablará de los edificios en que la estructura juega un papel decisivo en la arquitectura a partir de obras diversas.

<sup>3</sup>Como en los corpora de 3LB se han añadido nodos para el sujeto elíptico, éste aparece también en las entradas derivadas.

Func.	Texto
CD	que
verbo	movía
SUJ	el viento
CC	por aguas y mares
CC	con estimable eficiencia

Cuadro 2: Entrada derivada: *mover-1*

Func.	Texto
SUJ	que
verbo	mueven
CREG	a la reflexión

Cuadro 3: Entrada derivada: *mover-2*

*una vez más están entre los mejores clasificados de las votaciones semanales de entrenadores y periodistas especializados -no hay clasificación oficial al no existir una liga nacional universitaria- resultó ser de los **que mueven a la reflexión.***

De la oración de Eus3LB *Alderdiko kideei ez diegu agindu bat eman*<sup>4</sup> se ha derivado la entrada para el verbo *eman* (*dar*) que aparece en el cuadro 4.

Func.	Texto
ncsubj	pro (elipsis)
nczobj	alderdiko kideei
verbo-aux	ez-eman-diegu
ncobj	agindu bat

Cuadro 4: Entrada derivada: *eman*

Este léxico, además de la información sintáctica, contendrá, para cada entrada verbal, el synset de EuroWordNet<sup>5</sup>. De este modo, para aquellas entradas que estén ampliamente representadas, este léxico permitirá comprobar si existe o no variación en los esquemas sintáctico-semánticos para los distintos sentidos de un verbo. Se dispondrá, por lo tanto, de un recurso que extiende la información de EuroWordNet mediante la inclusión de esquemas sintáctico-semánticos.

Debido a que los corpora del 3LB están anotados con árboles sintácticos completos

<sup>4</sup>No les hemos dado ninguna orden a los miembros del partido.

<sup>5</sup>Una vez se haya completado la anotación semántica de los corpora de 3LB.

incluyendo las funciones anteriormente mencionadas, la extracción inicial de entradas léxicas verbales con sus correspondientes argumentos ha sido realizada de forma completamente automática.

A continuación presentamos cómo se enriquecerá el léxico con información sobre los papeles temáticos.

### 3. Enriquecimiento con papeles temáticos

A partir de las entradas derivadas presentadas en la sección 2.2 se asociarán a cada función, para cada sentido de cada verbo, los papeles temáticos correspondientes.

El cuadro 5 presenta los papeles temáticos que se utilizarán en el proceso de etiquetación semántica. Se corresponden con los utilizados en PropBank (Palmer, Kingsbury, y Gildea, 2005), (Pro, 2002) y su utilización en los léxicos de 3LB-LEX se fundamenta en los siguientes motivos<sup>6</sup>:

1. Aplicabilidad: se trata de un conjunto de papeles temáticos que se han definido para el etiquetado de corpus y han sido puestos a prueba en el etiquetado de PennTreeBank para la constitución de PropBank. Se trata de papeles temáticos obtenidos a partir del análisis de ocurrencias verbales en un corpus y, por lo tanto, han sido ya contrastados en una amplia casuística de ejemplos.
2. Flexibilidad: se trata de un sistema de etiquetado que contiene dos niveles: el de los argumentos y el de los papeles temáticos propiamente dichos. Así, para cada posición argumental existen diferentes posibilidades de realización de un determinado papel temático. Y, a la inversa, un determinado papel temático puede aparecer en diferentes posiciones argumentales. De este modo se preserva, por un lado, el grado de proximidad de un argumento respecto del verbo, cosa que se indica con la numeración de los argumentos, y, por otro, el tipo de relación semántica de cada argumento con el verbo.
3. Estandarización: este tipo de anotación por un lado permite una comparación directa con PropBank y, por otro,

<sup>6</sup>Una primera propuesta de anotación para el corpus 3LB aparece en (Navarro et al., 2004).

se pueden establecer equivalencias con otras propuestas. Una vez finalizado el etiquetado se podrá realizar un estudio comparativo para cuatro lenguas: inglés, catalán, euskera y español.

La propuesta de etiquetado aparece en el cuadro 5. Los argumentos numerados son los que ocupan posiciones argumentales, mientras que *ArgM* se corresponde con los adjuntos o elementos más marginales en relación al verbo. *ArgA* se utiliza en PropBank para indicar *volitional motion* como en las oraciones: *Upenn works John hard* y *Mr. Dinkins would march his staff out of board meetings*, un tipo de construcción que también aparece en las lenguas implicadas en el proyecto: en euskera hay un morfema (-*arazi*) admitido por todos los verbos y que introduce otro argumento que correspondería a *ArgA*; en catalán y castellano este argumento aparece en oraciones como *la madre pasea a la niña por el parque*, donde *la madre* es el agente inductor de la acción (*ArgA*) y *la niña* es el agente que pasea (*Arg0*). En cuanto a *Arg0* y *Arg1*, que son las posiciones argumentales más próximas al verbo, se trata de protoroles, es decir, de los argumentos que tienen mayor número de propiedades que definen un protoagente y un protopaciente o prototema (Dowty, 1991).

Role	Descripción
Arg0	agente
Arg1	tema / paciente
Arg2	beneficiario / instrumento / atributo / estado final / extensión
Arg3	origen/ beneficiario / instrumento / atributo
Arg4	destino
ArgMs	adjuntos verbales
ArgA	agente inductor

Cuadro 5: Papeles temáticos

El cuadro 6 presenta la especificación para los *ArgM* o adjuntos verbales que utilizamos para marcar los léxicos de 3LB-LEX. Dado que parte de la información que se especifica en la etiqueta funcional de *ArgM* de PropBank aparece en 3LB en la anotación sintáctica, ésta se omite. Se trata de la información correspondiente a la negación (NEG), los conectores discursivos (DIS) y los verbos modales (MOD).

LOC: locativo	CAU: causa
EXT: extensión	TMP: temporal
FIN: finalidad	MNR: manera
ADV: general	DIR: dirección

Cuadro 6: Especificación de *ArgMs*

A partir del análisis de todas las apariciones de un sentido de cada verbo se generalizará la información y se creará una entrada verbal prototípica que luego se utilizará para anotar los árboles sintácticos con los roles temáticos.

Así por ejemplo, tras el análisis de todas las apariciones del verbo *parlar* en catalán, se obtiene la siguiente entrada verbal:

Parlar	
SUJ	Arg0
CREG	Arg1-de-TEM
CC	Arg2-amb-BEN
CC	ArgM-TMP
CC	ArgM-ADV

Con esta entrada verbal prototípica se anotarán posteriormente todas las apariciones del dicho verbo en el corpus, de modo que el resultado final será:

CC-ArgM-TMP	En la segona de les conferències, programada per al dia_9_de_juny ,
SUJ-Arg0	Robert_Brufau
REL	parlarà
CREG-	dels edificis en què
-Arg1-de-TEM	l'estructura juga un paper decisiu en l'arquitectura
CC-ArgM-ADV	a_partir_d'obres diverses

Las entradas prototípicas del verbo *mover* serán:

mover-1	<i>cambiar de lugar</i>
SUJ	Arg0
CD	Arg1-PAT/TEM
CC	Arg2-EXT
CC	Arg3-ORI
CC	Arg4-DES
CC	ArgM-LOC
CC	ArgM-MNR

mover-2	<i>acción obligada</i>
SUJ	Arg0
CD	Arg1-PAT
CREG	Arg2-a

A partir de estas entradas podrán anotarse las apariciones del verbo:

CD-Arg1-TEM	que
REL	movía
SUJ-Arg0	el viento
CC-ArgM-LOC	por aguas y mares
CC-ArgM-MNR	con estimable eficiencia

SUJ-Arg0	que
REL	mueven
CREG-Arg2-a	a la reflexión

En el caso del euskera, la entrada prototípica para el verbo *eman* sería:

Eman	
ncsubj	Arg0
ncobj	Arg1
nczobj	Arg2

Y la frase anotada sería como sigue:

ncsubj-Arg0	pro (elipsis)
nczobj-Arg2	alderdiko kideei
REL	ez-eman-diegu
ncobj-Arg1	agindu bat

### 3.1. La problemática de los CC

En los corpora de 3LB, el complemento circunstancial se ha considerado como un todo inespecificado; es decir, no se han distinguido los diferentes tipos de CC. La derivación del léxico verbal a partir de corpus y la importancia de este tipo de complementos para aplicaciones como la extracción y la recuperación de información debería llevar a reconsiderar el tratamiento de los adjuntos. Debería diseñarse una nueva versión del corpus donde el complemento circunstancial estuviera subespecificado, de manera que se pudiera establecer un correlato casi directo con los diferentes roles utilizados en PropBank para los adjuntos. Con la versión actual de los corpora de 3LB la anotación temática de estos complementos deberá realizarse de modo manual si en la entrada prototípica aparece más de un tipo de ArgM.

### 3.2. La problemática del complemento de régimen

En los corpora español y catalán se ha definido la función sintáctica *complemento de régimen* para aquellos casos en que el verbo rige un complemento con preposición. En

inglés los complementos de este tipo de verbos no se tratan de un modo especial y no se contemplan en la correspondencia función sintáctica-papel temático. En nuestro léxico se ha incluido el complemento de régimen como valor posible del Arg1 y del Arg2.

### 3.3. 3LB-LEX

El léxico verbal derivado de los corpora 3LB es un conjunto de entradas verbales en las que se especifica, para cada ocurrencia verbal, el esquema funcional en que aparece. A partir de los esquemas funcionales, se deriva la entrada léxica prototípica para cada verbo. En ella se representan los diferentes sentidos del verbo con sus respectivos esquemas funcionales etiquetados semánticamente con papeles temáticos. Es decir, cada función sintáctica tiene su correspondiente papel temático asociado. A partir de este léxico *prototípico* se procederá a la anotación automática del corpus y a su posterior validación manual. En principio se realizará una validación selectiva priorizando las funciones que presenten ambigüedad en la asignación de papeles temáticos, especialmente los complementos circunstanciales. Todo este proceso aparece resumido en la figura 1.

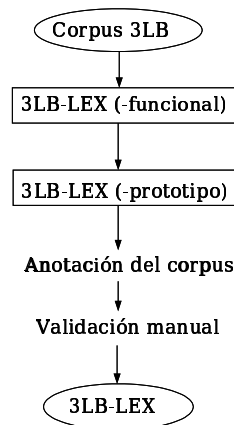


Figura 1: 3LB-LEX

## 4. Conclusiones y trabajo futuro

Hemos presentado 3LB-LEX, un recurso léxico que contiene información sobre la estructura sintáctica, la estructura argumental y los papeles temáticos de los verbos contenidos en los corpora de 3LB.

El treebank enriquecido con anotación de papeles temáticos se podrá usar para inferir analizadores sintáctico-semánticos, del estilo de los etiquetadores de roles semánticos que

existen ya para el inglés: Gildea y Jurafsky (2002); Carreras y Màrquez (2004); Litkowski (2004). Estos analizadores podrán ser de uso general en multitud de aplicaciones y, en particular, se podrán reutilizar en la fase de ampliación de los corpora 3LB, ya que se pueden usar para realizar un etiquetado automático inicial de los papeles temáticos de los nuevos ejemplos (incluyendo verbos nuevos que puedan aparecer), para su posterior revisión manual.

3LB-LEX se podrá utilizar para establecer comparaciones interlingüísticas entre las lenguas del proyecto: se comparará para un mismo verbo en las distintas lenguas los diferentes papeles temáticos que tiene y cómo éstos se realizan sintácticamente. Esta información permitirá realizar inferencias sobre la correspondencia interlingüística entre sintaxis y semántica. Como los papeles temáticos coinciden con los de PropBank, la comparación podrá hacerse extensiva al inglés. Con estos datos se pueden emprender el etiquetado sintáctico-semántico de corpora paralelos disponiendo de información de una única lengua. Por otra parte, también se pueden realizar estudios teóricos sobre las diátesis y su diferente realización interlingüística.

A medida que progrese la anotación manual de los léxicos verbales se procederá a volcar toda esta información de nuevo sobre los corpora para enriquecer las etiquetas de función sintáctica con los papeles temáticos. Este procedimiento será semiautomático, puesto que en los casos de los circunstanciales habrá que asignar el papel temático de modo manual; para el resto de las funciones la anotación será automática.

### Bibliografía

2002. PropBank Annotation Guidelines. <http://www.cis.upenn.edu/ace/>.
- Aranzabe, M., J. Arriola, A. Atutxa, I. Balza, y L. Uria. 2003. Guía para la anotación sintáctica manual de Eus3LB (corpus del euskera anotado a nivel sintáctico, semántico y pragmático). Informe técnico, UPV/EHU/LSI/TR 13-2003.
- Carreras, X. y L. Màrquez. 2004. Introduction to the CoNLL-2004 Shared Task: Semantic Role Labeling. En *Proceedings of the 8th Conference on Computational Natural Language Learning, CoNLL-2004*.
- Civit, M. 2004. Guía para la anotación de las funciones sintácticas de Cast3LB: un corpus del español con anotación sintáctica, semántica y pragmática. Informe técnico, CLiC, Centre de Llenquatge i Computació. disponible en <http://clic.fil.ub.es/personal/civit/>.
- Civit, M., A. Ageno, B. Navarro, N. Bufí, y MA. Martí. 2003. Análisis cualitativo y cuantitativo del acuerdo entre anotadores en el desarrollo de corpus interpretados lingüísticamente. *Procesamiento del lenguaje Natural*, 31.
- Civit, M., N. Bufí, y M.P. Valverde. 2004. Guia per a l'anotació de les funcions sintàctiques de CAT3LB: un corpus del català amb anotació sintàctica, semàntica i pragmàtica. Informe técnico, CLiC, Centre de Llenquatge i Computació. disponible en <http://clic.fil.ub.es/personal/civit/>.
- Dowty, D.R. 1991. Thematic proto-roles and argument selection. *Language*, 67(3).
- Gildea, D. y D. Jurafsky. 2002. Automatic Labeling of Semantic Roles. *Computational Linguistics*, 28(3):245-288.
- Litkowski, K. 2004. Senseval-3 task: Automatic labeling of semantic roles. En *Proceedings of the Senseval-3 ACL-SIGLEX Workshop*.
- Navarro, B., P. Moreda, B. Fernández, R. Marcos, y M. Palomar. 2004. Anotación de roles semánticos en el corpus 3lb. En *Taller de Herramientas y Recursos Lingüísticos para el español y el portugués, Iberamia*.
- Palmer, M., P. Kingsbury, y D. Gildea. 2005. The Proposition Bank: An Annotated Corpus of Semantic Roles. *Computational Linguistics*, 21(1).
- Palomar, M., M. Civit, A. Díaz, L. Moreno, E. Bisbal, M. Aranzabe, A. Ageno, M.A. Martí, y B. Navarro. 2004. 3LB: Construcción de una base de datos de árboles sintáctico-semánticos para el catalán, euskera y castellano. *Procesamiento del lenguaje Natural*, 33.