

## Fundamentos teóricos y metodológicos para el etiquetado semántico de CESS-CAT y CESS-ESP

**Mariona Taulé**

Universitat de Barcelona  
Gran Via, 585, 08007-Barcelona,  
mtaule@ub.edu

**M. Antonia Martí**

Universitat de Barcelona  
Gran Via, 585, 08007-Barcelona,  
amarti@ub.edu

**Joan Castellví**

Universitat de Barcelona  
Gran Via, 585, 08007-Barcelona,  
joan.castellvi@ub.edu

**Juan Aparicio**

Universitat de Barcelona  
Gran Via, 585, 08007-Barcelona,  
juan.aparicio@thera-clic

**Resumen:** El objetivo de este artículo es establecer los fundamentos teóricos para la anotación semántica de los corpora CESS-ESP y CESS-CAT, mediante la asignación a cada predicado de una estructura léxico-semántica que determina la estructura argumental, las relaciones temáticas de los argumentos respecto al verbo y las alternancias de diátesis que admite. El proceso de asignación de información semántica se realiza de manera semiautomática. Como resultado se obtiene una clasificación de los predicados que permite establecer generalizaciones sobre su comportamiento.

**Palabras clave:** estructura léxico-semántica, estructura eventiva, estructura argumental, diátesis, papel temático, anotación semántica, lingüística de corpus.

**Abstract:** The aim of this article is to establish the theoretical basis of the semantic annotation of the CESS-ESP and CESS-CAT corpora, by means of assigning to each predicate a Lexicosemantic Structure. The lexicosemantic patterns determine the argument structure, the correspondence between the arguments and the verb, and the diathesis alternations. The semantic annotation is carried out semi-automatically. As a result, a classification of predicates will be obtained and generalizations on the behavior of predicates will be drawn.

**Keywords:** lexicosemantic structure, event structure, argumental structure, diatheses, thematic roles, semantic annotation, corpus linguistics.

### 1 Introducción

El objetivo de este artículo es presentar los fundamentos teóricos para la anotación semántica de los corpus CESS-ESP y CESS-CAT<sup>1</sup>, la metodología que se adopta y los primeros resultados obtenidos.

---

<sup>1</sup> El etiquetado sintáctico (constituyentes y funciones) de estos corpora se inició en el proyecto 3LB (FIT-150-500-2002-244), en el se etiquetaron 100.000 palabras para cada lengua, y actualmente tiene su continuación en el proyecto CESS-ECE (Corpus Etiquetados sintáctica y semánticamente

La caracterización semántica de los predicados se fundamenta desde un punto de vista teórico en la propuesta de descomposición léxica de Levin & Rappaport-Hovav (1995) y Rappaport-Hovav & Levin (1998), de donde se toma el concepto de estructura léxico-semántica que, en definitiva, determina el número de argumentos exigidos por el predicado verbal y el tipo de papel temático. En este sentido, se

---

para el Español, Catalán y Euskera) (HUM 2004-21127-E). Este último proyecto tiene como objetivo final la ampliación de los corpora CESS-ESP y CESS-CAT hasta 500.000 palabras.

sigue la línea iniciada por Kipper *et al.* (2002) y Kingsbury *et al.* (2002) en la construcción de *VerbNet*. También se han tenido en consideración los trabajos de Demonte (2003) para su aplicación al castellano y Mateu (2004), para el catalán. Por otra parte, se adopta la propuesta de *PropBank* (Palmer *et al.* 2005) en la caracterización de los argumentos. Finalmente, nos basamos en Vázquez *et al.* (2000) para el tratamiento de las alternancias de diátesis de ambas lenguas.

La anotación semántica de los corpora CESS-ECE se lleva a cabo en dos etapas: en una primera fase, el etiquetado se realiza de manera semiautomática para un subconjunto de 100.000 palabras de cada corpus (Civit *et al.* 2005a) y, en una segunda fase, se etiquetan las 400.000 palabras restantes de manera automática, aplicando técnicas de aprendizaje automático (Surdeanu 2006) y mediante el etiquetado automático incremental de papeles temáticos (Busser & Morante 2005).

El artículo se organiza en los siguientes apartados: en primer lugar (apartado 2) se exponen los fundamentos teóricos a partir de los cuales se caracterizan los predicados verbales. A continuación (apartado 3) se presenta la metodología para la obtención de los datos lingüísticos a partir del corpus etiquetado sintácticamente. En el siguiente apartado (4) se presentan los primeros resultados obtenidos a partir del análisis de un subconjunto de predicados, así como la tipología de papeles temáticos y la correspondencia con las funciones sintácticas. Finalmente (apartado 5) se presentan las conclusiones y se presentan las líneas futuras de investigación.

## **2 Fundamentos teóricos en el análisis de los predicados**

La correspondencia entre las funciones sintácticas y los papeles temáticos se realiza a partir de una caracterización de los predicados según la propuesta de Levin & Rappaport-Hovav (1995) y Rappaport-Hovav & Levin (1998). Consideramos que se trata de una propuesta adecuada a nuestro propósito por diferentes razones. En primer lugar, porque cohesionan en un mismo modelo la información léxico-semántica y la información eventiva con la información argumental y las alternancias de diátesis. En segundo lugar, porque en el área de la lingüística de corpus y la lingüística

computacional se han realizado ya para el inglés experiencias similares siguiendo esta aproximación. En esta línea, destacan los trabajos de Kipper *et al.* (2002) en la construcción de *VerbNet*, un léxico donde se representa la información léxico-semántica, argumental y diatética para los predicados del inglés, vinculado a su vez con la clasificación semántica propuesta en Levin (1993) y que se utiliza para el etiquetado semántico de *PropBank* (Palmer *et al.* 1995). Finalmente, existen para el catalán (Vázquez *et al.* 2000 y Mateu 2004) y para el castellano (Vázquez *et al.* 2000 y Demonte 2003) propuestas de análisis de los predicados siguiendo este modelo teórico.

En este sentido, el trabajo que realizamos parte de Vázquez *et al.* (2000), en dos aspectos: se adopta en gran medida su clasificación diatética y se toma como hipótesis de trabajo la clase de verbos de cambio. En nuestra propuesta se incorpora la caracterización léxico-semántica de los predicados que viene a ser una formalización de los componentes de significado tan solo esbozados en el modelo de partida. Por otra parte, el objetivo último de nuestra investigación no es tanto la definición de clases sino disponer de una metodología y unos criterios de análisis de los predicados que sea informativamente relevante y que permita su posterior agrupación según los diferentes criterios que se han tenido en consideración.

Los papeles temáticos vienen determinados por la estructura léxico-semántica (ELS) y la estructura eventiva. Para la caracterización de los predicados partimos de un conjunto relativamente restringido de estructuras léxico-semánticas y eventivas, que corresponden a las tres clases básicas de eventos: estados, actividades y realizaciones (Vendler 1967, Dowty 1991). Estas clases generales (apartado 2.1) son susceptibles de subclasificaciones, como se verá en el apartado 4. Los papeles temáticos vienen determinados por la pertenencia de los predicados a una de estas clases y por el tipo de alternancia diatética en que participan. De este modo, no sólo especificamos los papeles temáticos sino que caracterizamos semánticamente el predicado tanto desde el punto de vista aspectual como argumental. Esta información queda asociada a las oraciones enriqueciendo de este modo la información lingüística que podrá ser utilizada por los sistemas automáticos de etiquetación temática y por cualquier aplicación de PLN.

Esta información queda recogida en los léxicos CESS-ESP y CESS-CAT que actualmente se están desarrollando.

### 2.1 Estructuras léxicosemánticas

A diferencia de Kipper *et al.* (2000), que desarrollan estructuras léxico-semánticas (ELS) de gran complejidad, hemos preferido adoptar unas clases más generales que se subespecifican en función de los papeles temáticos y las diátesis. En este sentido, partimos de tres grandes tipos de estructuras que se corresponden con los tres tipos ontológicos de eventos (Vendler 1967, Dowty 1979), estados, actividades (o procesos) y realizaciones:

- (1) [x <ESTADO>]
- (2) x ACTUAR <MANERA/  
INSTRUMENTO> y]
- (3) [x CAUSAR [DEVENIR [y  
<ESTADO/ COSA/LUGAR>]]]

Los componentes semánticos ACTUAR, CAUSAR y DEVENIR corresponden a predicados primitivos que, entre otros factores, determinan el tipo de evento. Las constantes (*MANERA*, *INSTRUMENTO*, *LUGAR*, etc.), se expresan en cursiva y constituyen el enlace con las piezas léxicas verbales. Por último, las variables corresponden a los argumentos. Cuando las variables van ligadas a un predicado primitivo su expresión sintáctica es obligatoria, mientras que cuando van ligadas a una constante su realización es opcional. Por ejemplo, la *x* asociada a la primitiva ACTUAR debe realizarse sintácticamente como un argumento con el papel temático de Agente, mientras que la realización sintáctica del argumento semántico y ligado a la constante es opcional. El papel temático de los argumentos se infiere, por lo tanto, a partir de su posición en la ELS y del tipo de diátesis en las que participa el predicado.

La ELS de (1) se corresponde con el tipo ontológico de los estados, tiene sólo una entidad implicada en el evento, que focaliza en el estado resultante. La ELS de (2) corresponde a las actividades o procesos, suele tener sujetos de carácter agentivo y admite como diátesis la pasiva y la pasiva refleja. Finalmente, la ELS de (3) corresponde a las realizaciones, que expresan estados resultantes de un proceso

causado externamente, el sujeto de las cuales suele ser de tipo causativo.

### 2.2 Alternancias de diátesis y papeles temáticos

En nuestra propuesta se asume la clasificación de diátesis de Vázquez *et al.* (2000), aunque extendemos el modelo relacionando cada una de las diátesis con las ELS. De hecho, se parte de la hipótesis de que las diátesis admitidas por un determinado predicado derivan directamente de su ELS. Por lo tanto, consideramos que todas las realizaciones diatéticas de un mismo predicado corresponden a una misma ELS, aunque cada una de ellas focaliza en determinados componentes de la misma. En definitiva, las diátesis son estructuras superficiales que resultan de focalizar de un modo u otro los predicados de la ELS.

Por ejemplo, los predicados caracterizados como realizaciones (ELS del tipo (3)) admiten las diátesis causativa, anticausativa y resultativa, que focalizan en la primitiva CAUSAR, DEVENIR y la constante *ESTADO* respectivamente:

- (4) abrir: [x CAUSAR [DEVENIR [y  
<ESTADO>]]]
- (4a) Juan abre la puerta: [Juan CAUSAR  
[DEVENIR [puerta <ABIERTA>]]]  
(Causativa)
- (4b) La puerta se abre: [x CAUSAR  
[DEVENIR [puerta <ABIERTA>]]]  
(Anticausativa)
- (4c) La puerta está abierta: [x CAUSAR  
[DEVENIR [puerta <ABIERTA>]]]  
(Resultativa)

En el ejemplo se puede observar que el predicado ‘abrir’ tiene una única ELS (4) -una *x* que causa que *y* cambie de estado- que se realiza sintácticamente en estructuras argumentales diferentes –causativa (4a), anticausativa (4b) y resultativa (4c)- como resultado de focalizar aspectos distintos de la ELS (el componente ‘causa’, ‘cambio’ (devenir) o ‘estado’ respectivamente).

Los predicados que admiten estas ELS expresan el papel temático CAUSA en la variable *x* y TEMA en la variable *y*. Por lo tanto, la asociación de papeles temáticos a

funciones sintácticas viene determinada no sólo por las posiciones argumentales de la ELS sino también por el tipo de alternancias de diátesis admitidas, independientemente del tipo semántico de los argumentos, es decir de si se trata de un argumento [+/-humano], [+/-volitivo], [+/-afectado], etc.

Puede ocurrir que algunos predicados pertenecientes a una misma ELS no acepten las mismas alternancias de diátesis y que opcionalmente puedan presentar alguna de específica. Se trata, como es habitual en el análisis lingüístico, de casos prototípicos y casos periféricos dentro de una misma categoría.

### 3 Metodología

#### 3.1 Extracción del léxico verbal

En este trabajo se parte de la hipótesis de que el análisis de la ELS de los predicados verbales es el punto de partida para la descripción semántica de la oración. La estructura argumental expresa la aridad del verbo y de la ELS se deriva la relación entre el predicado y sus argumentos, que expresamos con papeles temáticos. Las relaciones temáticas permiten la interpretación semántica de la frase, mediante la combinación de los significados de los verbos con los argumentos.

Para la anotación semántica de CESS-ESP y CESS-CAT se ha definido una metodología que permite realizar este proceso de manera semiautomática (autores 2005). A partir de la información sintáctica expresada en los corpus de 100.000 palabras de 3LB (Vease nota 1) se deriva de forma automática un léxico verbal que contiene para cada sentido del verbo todos los esquemas sintácticos en los que aparece. A partir de esta información se crea manualmente CESS-LEX, un léxico verbal con información sintáctica (constituyentes y funciones) e información semántica (*synset* de *EuroWordNet* y ELS con sus papeles temáticos correspondientes).

#### 3.2 Anotación semántica del corpus

La asignación manual de ELS, estructura argumental y de alternancias de diátesis se realiza en los léxicos CESS-LEX-ESP y CESS-LEX-CAT. Para cada predicado verbal se

especifican, además de la ELS correspondiente, los diferentes sentidos del verbo con sus respectivas estructuras argumentales y los diferentes esquemas funcionales (diátesis) etiquetados semánticamente con papeles temáticos (autores 2006). Dicho léxico se extrae de manera automática a partir del corpus 3LB y en él se incluyen todos los verbos distintos que aparecen en el corpus, 834 para el catalán y 1.070 para el castellano<sup>2</sup>. Es a partir de estos léxicos derivados, CESS-LEX-ESP y CESS-LEX-CAT, que se etiqueta semánticamente y de manera automática un subconjunto de 100.000 palabras de cada lengua (autores 2005a). Una vez revisado manualmente este subconjunto, de manera que se garantice la calidad del mismo y el éxito de las técnicas automáticas, se procederá a la etiquetación automática de las 400.000 palabras restantes. En esta segunda fase se aplicarán técnicas de aprendizaje automático (Surdeanu 2006) y el sistema de etiquetado automático incremental de papeles temáticos (Busser & Morante 2005). En la figura 1, se muestra de manera esquemática el procedimiento seguido:

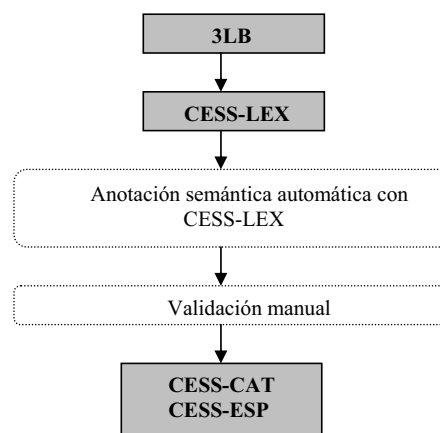


Figura 1: Proceso de derivación de CESS-ESP y CESS-CAT

<sup>2</sup> El número total de formas verbales del corpus castellano es de 7.127 y, en el caso del catalán, es de 7.033.

#### 4 *Análisis de los resultados*

##### 4.1 ELS y papeles temáticos

Como se ha indicado anteriormente (apartado 2) la parrilla temática viene determinada por la ELS del predicado y de las alternancias de diátesis en las que aparece. Para la anotación de los argumentos hemos seguido la propuesta de *PropBank* (Palmer *et al.* 2005) donde se distingue entre los argumentos obligatorios (Arg0, Arg1, Arg2, Arg3, Arg4 que refleja la proximidad del argumento respecto al verbo) y adjuntos expresados como argumentos opcionales (ArgM). Para cada argumento y dependiendo del tipo de la ELS, y por lo tanto del tipo de evento, del verbo resultan una serie de papeles temáticos que presentamos seguidamente.

- Arg0:** Arg0-AGT (Agente), Arg0-CAU (Causa), Arg0-EXP (Experiment.)
- Arg1:** Arg1-PAT (Paciente), Arg1-TEM (Tema), Arg1-ATR (Atributo), Arg1-EXT (Extensión)
- Arg2:** Arg2-ATR (Atributo), Arg2-BEN (Beneficiario), Arg2-INS (Instrumento), Arg2-EXT (Extensión), Arg2-EFI (Estado Final)
- Arg3:** Arg3-ATR (Atributo), Arg3-BEN (Beneficiario), Arg3-INS (Instrumento), Arg3-ORI (Origen)
- Arg4:** Arg4-DES (Destino)
- ArgM:** ArgM-LOC (Locativo), ArgM-TMP (Temporal), ArgM-EXT (Extensión), ArgM-FIN (Finalidad), ArgM-CAU (Causa), ArgM-MNR (Manera), ArgM-DIR (Dirección), ArgM-ADV (Adverbial)

A continuación, presentamos con mayor detalle el tratamiento dado a los argumentos Arg0 y Arg1 en los papeles temáticos de Agente, Causa, Experimentador, Tema y Paciente, por ser los que plantean una problemática más compleja.

**Arg0** Papeles temáticos: Agente, Causa y Experimentador  
Funciones sintácticas: Sujeto, C. Agente

**Arg0-AGT** (Agente)

- ELS: [x ACTUAR < MANERA / INSTRUMENTO>]

Verbos inergativos de tipo agentivo: 'Juan<sub>Arg0-AGT</sub> corre'

*caminar, nadar, gritar, ...*

Diátesis: Transitiva (Objeto cognado)<sup>3</sup>

- ELS: [x ACTUAR < MANERA / INSTRUMENTO> y]

Verbos transitivos agentivos: 'Juan<sub>Arg0-AGT</sub> lee una novela histórica'

'Clara es amada por todos<sub>Arg0-AGT</sub>'

*amar, odiar, desear, abordar, comer, beber, fregar, barrer, cepillar, peinar...*

Diátesis: Pasiva

**Arg0-CAU** (Causa)

- ELS: [x CAUSA [DEVENIR [y <ESTADO>]]]

Verbos causativos: 'Juan/la lluvia/el martillo<sub>Arg0-CAU</sub> rompió la ventana'

*aclarar, abrir, enamorar, purificar, sacralizar, construir, hundir...*

Diátesis: Causativa, Anticausativa, Resultativa

- ELS: [x CAUSA [DEVENIR [y <LUGAR>]]]

Verbos causativos: 'El médico<sub>Arg0-CAU</sub> hospitalizó al paciente'

*almacenar, encarcelar, empaquetar...*

Diátesis: Causativa, Resultativa

- ELS: [x CAUSA [DEVENIR [<COSA> EN y]]]

Verbos causativos: 'Juan<sub>Arg0-CAU</sub> ensilla el caballo'

*enharinar, empapelar, ensillar...*

Diátesis: Causativa, , Resultativa

**Arg0-EXP** (Experimentador)

- ELS: [x <ESTADO>]

Verbos inergativos estativos: 'Juan<sub>Arg0-EXP</sub> sueña' *dormir, llorar, roncar, brillar, sudar, oler, chorrear, palidecer, vivir...*

Diátesis: Transitiva (Objeto cognado)

- ELS: [x <ESTADO> y]

Verbos transitivos estativos: 'Juan<sub>Arg0-EXP</sub> tiene dos hipotecas'

*poseer, medir, pesar, tener...*

Diátesis: -

<sup>3</sup> No se han tratado todas las diátesis posibles.

**Arg1** Papeles temáticos: Tema, Paciente, Experimentador  
 Funciones sintácticas: OD, C.Régimen, Predicativo, Sujeto(+PASS), Sujeto (+INAC)

**Arg1-TEM** (Tema)

- ELS: [y <ESTADO>]  
 Verbos inacusativos: 'Los niños<sub>Arg1-TEM</sub> llegaron tarde'  
*caer, crecer, desaparecer, emerger, entrar, llegar, morir, salir, venir...*  
 Diátesis: -

- ELS: [y <ESTADO>]  
 Verbos inacusativos estativos: 'Hay cuatro personas<sub>Arg1##TEM</sub> esperando'  
*acabar, empezar, existir, haber, vivir...*  
 Diátesis: -

**Arg1-PAT** (Paciente)

- ELS: [x ACTUAR < MANERA / INSTRUMENTO> y]  
 Verbos transitivos agentivos: 'Juan lee una novela histórica<sub>Arg1-PAT</sub>'  
 'Clara<sub>Arg1-PAT</sub> es amada por todos'  
*abandonar, amar, odiar, desear, abordar, comer, beber, leer, fregar, barrer, cepillar, peinar...*

#### 4.2 Clasificación de predicados

Como ya se ha indicado anteriormente la clasificación de los predicados será consecuencia del análisis que se está llevando a cabo. No obstante, por el hecho de existir una iniciativa de clasificación verbal basada a partir de criterios diatéticos, que aquí se comparten, hemos iniciado el análisis de predicados partiendo de aquellos verbos que presentan una mayor frecuencia en el corpus y que coinciden con la clase de cambio analizada en Vázquez *et al.* (2000). La existencia de dicha clase es una hipótesis de partida de este estudio.

En nuestro marco de trabajo estos verbos se tratan desde un punto de vista eventivo como realizaciones, que se corresponden con la ELS:

- (5) [x CAUSA [DEVENIR [y <ESTADO>]]]

y admiten prototípicamente tres diátesis: la causativa (5a), la anticausativa (5b) y la resultativa (5c).

- (5a) El fuerte viento apagó las velas.  
 (5b) Las velas se apagaron.  
 (5c) Las velas están apagadas.

El componente CAUSA en estos predicados, además de las entidades prototípicamente causativas, que corresponden normalmente a un evento (6a), puede realizarse argumentalmente por entidades de carácter agentivo (6c) e instrumental (6b). En los tres casos el argumento se etiqueta como CAUSA aunque puede conllevar consecuencias diatéticas distintas, por ejemplo, en el caso de las causas de tipo agentivo pueden admitir también la diátesis de pasiva (6c).

- (6a) El viento<sub>Arg0-CAU</sub> rompió el cristal<sub>Arg1-TEM</sub>.  
 (6b) La piedra<sub>Arg0-CAU</sub> rompió el cristal<sub>Arg1-TEM</sub>.  
 (6c) El niño<sub>Arg0-CAU</sub> rompió el cristal<sub>Arg1-TEM</sub> /fue roto por el niño<sub>Arg0-CAU</sub>.

Algunos de estos verbos admiten además otras diátesis que no consideramos definitorias, como son la holística (7a-7b), la posibilidad de expresar la gradación, y el inicio o el final del cambio con sus argumentos correspondientes (8a).

- (7a) Los gritos de los niños<sub>Arg0-CAU</sub> aturdían a los invitados<sub>Arg1-TEM</sub>.  
 (7b) Los niños<sub>Arg0-CAU</sub> aturdían a los invitados<sub>Arg1-TEM</sub> con sus gritos<sub>ArgM-MNR</sub>.  
 (8a) El gobierno<sub>Arg0-CAU</sub> aumentó el precio del petróleo<sub>Arg1-TEM</sub> en un 5%<sub>Arg2-EXT</sub> hasta los 80\$ por barril<sub>Arg4-EFI</sub>.

De este modo, en los léxicos CESS-LEX-ESP y CESS-LEX-CAT donde se representan

las correspondencias entre funciones sintácticas y papeles temáticos de los argumentos, la información queda como sigue:

**Aturdir**

Arg0-CAU<sup>4</sup>

Arg1-TEM

+CAUSATIVA EJ: “Los gritos de los niños aturdieron al profesor”

SUJ Arg0-CAU

CD Arg1-TEM

+ANTICAUSATIVA EJ: “El profesor se aturdió”

SUJ Arg1-TEM

+RESULTATIVA EJ: “El profesor está aturdido”

SUJ Arg1-TEM

ATR Arg2-ATR

+HOLÍSTICA EJ: “Los niños aturdieron al profesor con sus gritos”

SUJ Arg0-CAU

CD Arg1-TEM

CC ArgM-MNR

+HOLÍSTICA EJ: “Los gritos de los niños aturdieron al profesor”

SUJ Arg0-CAU

CD Arg1-PAT

Estos datos constituyen la información lingüística para el etiquetado automático del corpus.

**4.3 Interfaz de consulta**

Actualmente se dispone ya de una interfaz de consulta en fase de pruebas donde se pueden realizar búsquedas basadas en la correspondencia entre papeles temáticos y funciones sintácticas de los corpus tratados hasta el momento.

Las búsquedas relacionadas con los papeles temáticos que actualmente están implementadas son del tipo: '¿Qué papeles temáticos realiza la función X?'; '¿Qué funciones sintácticas están

<sup>4</sup> Funciones sintácticas: SUJ=Sujeto, CD=Complemento directo, CC=Complemento Circunstancial, ATR=Atributo.

Papeles temáticos: CAU = Causa, TEMA = Tema, ATR = Atributo.

asociadas al papel temático Y?' y '¿Qué oraciones contienen el papel temático Y y Z?'

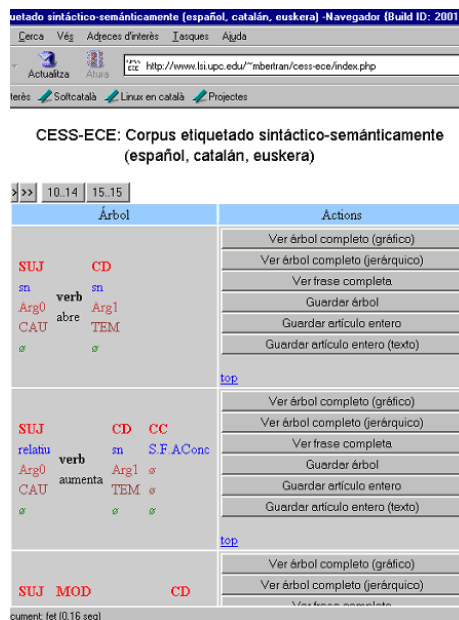


Figura 2: Resultado de una consulta al corpus etiquetado semánticamente

**5 Conclusiones**

En este artículo se han presentado las bases teóricas y metodológicas para el desarrollo de los corpus CESS-ESP y CESS-CAT, etiquetados semánticamente con papeles temáticos. Se ha integrado investigaciones teóricas sobre el análisis de predicados en catalán y en castellano con una metodología de implementación y etiquetado ya desarrollada para corpora del inglés. El objetivo último es disponer de un marco teórico y metodológico que garantice la consistencia y sistematicidad en la anotación semántica de los corpora. En nuestro enfoque se prioriza el análisis de predicados sobre su posible clasificación: en todo caso ésta será una consecuencia del análisis. Actualmente se dispone ya de una guía de etiquetado (autores 2006) para la anotación semántica de los predicados y se van a iniciar una serie de experimentos sobre acuerdo entre anotadores que permitirán mejorarla y adecuarla. Una vez etiquetado el corpus se utilizará como fuente de información para sistemas de etiquetado automático con papeles temáticos.

### **Bibliografía**

- B. Busser, R. Morante (2005). "Designing an active learning based system for corpus annotation". *SEPLN*, Granada.
- V. Demonte (2003). "Preliminares de una clasificación léxico-sintáctica de los predicados verbales del español". En Sybille Grosse y Axel Schönberger (eds.) *Ex oriente lux: Festschrift für Eberhard Gärtner zu seinem 60. Geburtstag*. Frankfurt am Main: Valentia.
- D. Dowty (1991). 'Thematic proto-roles and argument selection'. *Language*, 67.
- P. Kingsbury, M. Palmer, M. Marcus (2002). "Adding semantic annotation to Penn TreeBank". 2002. En *Proceedings of the 2002 Conference on Human Language Technology*, San Diego, CA.
- K. Kipper, H T. Dang, M. Palmer (2000). "Class-Based Construction of a Verb Lexicon". *AAAI-2000 Seventeenth National Conference on Artificial Intelligence*, Austin, TX.
- K. Kipper, M. Palmer, O. Rambow (2002). "Extending PropBank with VerbNet Semantic Predicates". *Workshop on Applied Interlinguas, held in conjunction with AMTA-2002*. Tiburon, CA.
- B. Levin (1993). *English Verb Classes and Alternations: A preliminary investigation*. University of Chicago Press, Chicago.
- B. Levin, M. Rappaport-Hovav (1995). *Unaccusativity. At the Syntax-Lexical Semantics Interface*. Cambridge, MA: MIT Press.
- J. Mateu (2004). "Relational Construal at the Interface", *Anuari de Filologia*, xxVI, Barcelona: Universitat de Barcelona.
- M. Palmer, P. Kingsbury, D. Gildea (2005). "The Proposition Bank: An Annotated Corpus of Semantic Roles", *Computational Linguistics*, 21 (1). USA: MIT Press.
- M. Rappaport Hovav, B. Levin (1998). "Building Verb Meanings", in M. Butt and W. Geuder, eds., *The Projection of Arguments: Lexical and Compositional Factors*, CSLI Publications, Stanford, CA, 97-134.
- M. Surdeanu (2006): <http://www.lsi.upc.edu/~surdeanu/swirl.html>  
Etiquetador automático de roles semánticos de software libre.
- G. Vázquez, A. Fernández, Martí, M.A. (2000). Clasificación verbal. Alternancias de diátesis. Edicions de la Universitat de Lleida: Lleida.
- Z. Vendler (1967). *Linguistics in Philosophy*. Cornell University Press: New York