

Asignación Automática de Etiquetas de Dominios en WordNet*

Mauro Castillo, Francis Real
Universidad Politécnica de Cataluña
Jordi Girona Salgado 1-3
08034 Barcelona
{castillo,fjreal}@lsi.upc.es

German Rigau
Universidad del País Vasco
649 Posta Kutxa
20080 Donostia
rigau@si.ehu.es

Resumen: En este artículo se describe un procedimiento para asignar de forma automática etiquetas de dominio a las glosas de WordNet. Una de las motivaciones principales del trabajo es enriquecer fuentes léxicas con información de WordNet. Para ello, se utilizan los WORDNET DOMAINS. Finalmente, se proponen y corrigen etiquetas de dominios para la parte nominal y verbal de WordNet.

Palabras clave: WordNet, WordNet Domains, Etiquetaje Automático.

Abstract: This paper describes a process to automatically assign wordnet domain labels to WordNet glosses. One of the main goals of this work is to enrich lexical sources with WordNet information. WordNet domains are used as knowledge source. Finally, Domain labels for nouns and verbs are suggested and verified.

Keywords: WordNet, WordNet Domains, Automatic Labeling.

1 Introducción

Aunque la importancia de WordNet ha trascendido ampliamente los propósitos para los que fue creado (Miller et al., 1990), y es ahora un recurso imprescindible para muchas aproximaciones al procesamiento semántico (McCarthy, 2001; Agirre y Martínez, 2002), en su estado actual, sigue sin ser un recurso suficientemente rico como para, directamente, dar soporte a muchas aplicaciones conceptuales avanzadas (Harabagiu, Pasca, y Maiorano, 2000).

Sin embargo, incluso ahora, desarrollar Bases de Conocimiento suficientemente grandes y ricas como para procesar semánticamente texto no restringido, sigue siendo una tarea titánica que sólo pueden abordar grandes grupos de trabajo a lo largo de largos periodos de tiempo (Fellbaum, 1998; Atserias et al., 1997; Bentivogli, Pianta, y Girardi, 2002).

Una de las motivaciones principales de este trabajo es enriquecer WordNet (o cualquier otro recurso léxico, como diccionarios, etc.) de forma sistemática con las etiquetas de dominio semántico de *WordNet Domains* (Magnini y Strapparava, 2002). Este recurso ha provado su eficacia en la desambigua-

ción de los dominios de las palabras (en inglés Word Domain Disambiguation) (Magnini y Cavaglià, 2000),

En este trabajo exploramos el etiquetaje semántico de recursos léxicos, asignando de forma automática y sistemática etiquetas de dominio a glosas y definiciones de diccionarios.

Como veremos, el método propuesto también puede servir para corregir y verificar el etiquetaje propuesto y dar algunas recomendaciones para asignar etiquetas de dominio en una definición de un diccionario o un texto libre.

El artículo se organiza de la siguiente manera. En la sección 2 se describe WN. A continuación se describen los *WordNet Domains* y sus posibles aplicaciones. En la sección 4 se detalla la experimentación realizada. Luego, en la sección 5 se detallan la evaluación y resultados. En la sección 6 se presenta la discusión y análisis de los resultados. Finalmente se presentan las conclusiones y trabajo futuro.

2 WordNet

WordNet es una base de datos léxica para el inglés que fue desarrollado en la Universidad de Princeton. Es una base de datos conceptual estructurada en forma de red semántica, que contiene nombres, verbos, adjetivos y adverbios. El *synset* (synonym sets o conjuntos

* Este artículo ha sido financiado parcialmente por la Comisión Europea (MEANING IST-2001-34460), Generalitat de Catalunya (2002FI 00648) y Universidad Tecnológica Metropolitana - Chile.

de sinónimos) es la unidad básica de WN que representa un concepto lexicalizado. Las relaciones semánticas entre *synsets* son: sinonimia, antonimia, hiponimia/holonimia, meronimia/holonimia, implicación y causa. La versión actual de WN es la 1.7.1. A efectos de este trabajo se considera la WN1.6 debido a la disponibilidad de los *WorNet Domains*. La versión de WN1.6 contiene 99.638 *synsets* y 121.962 palabras diferentes. En WN se define uno o más sentidos por cada palabra. Una palabra puede ser monosémica cuando sólo posee un único sentido y polisémica cuando tiene dos o más sentidos.

3 WordNet Domains

WordNet Domains es un recurso léxico desarrollado en ITC-IRST por (Magnini y Cavaglià, 2000) donde los *synsets* han sido anotados de forma semiautomática con una o varias etiquetas de dominio de un conjunto de 165 etiquetas organizadas jerárquicamente. La tabla 1 muestra la distribución del número de etiquetas que posee cada *synset*. En ella se puede apreciar como la gran mayoría de los *synset* sólo poseen una etiqueta de dominio.

#	nom	ver	adj	adv	%
1	56458	11287	16681	3460	88.202
2	8104	743	1113	109	10.105
3	1251	88	113	6	1.4632
4	210	8	8	0	0.2268
5	2	1	0	0	0.0030

Tabla 1: Distribución del número de etiquetas de dominios por *synset* en WordNet

Las etiquetas de dominios se complementan con la información contenida en WordNet. Un dominio puede incluir *synsets* de diferentes categorías sintácticas: por ejemplo MEDICINE puede contener sentidos a partir de nombres y verbos.

Un dominio puede incluir sentidos a partir de diferentes subjerarquías de WN: por ejemplo SPORT contiene diferentes sentidos derivados desde *lifeform#n#1*, *physical-object#n#1*, *act#n#2*, *location#n#1* entre otros.

Además los dominios pueden agrupar sentidos de una palabra en clases homogéneas, con la finalidad de reducir la polisemia de las palabras contenidas en WN. Para ilustrar tal idea veamos un ejemplo extraído de (B. Mag-

nini, 2001), para la palabra “*bank*” que posee 10 sentidos en WN.

Nro	SF	Dominio
#1	group	Economy
#2	object	Geography Geology
#3	possession	Economy
#4	artifact	Architecture Economy
#5	group	Factotum
#6	artifact	Economy
#7	object	Geography Geology
#8	possession	Economy Play
#9	object	Architecture
#10	act	Transport

Tabla 2: Sentidos de “*bank*” en WN con *Semantic File* (SF) y etiquetas de Dominios

La tabla 2 muestra los 10 sentidos de la palabra “*bank*” como nombre, los cuales se pueden reducir a 7 sentidos si se agrupa por dominios. Además si consideramos el *Semantic File* de cada *synset* claramente podemos disminuir la polisemia.

Por otro lado en aquellos *synset* que poseen más de una etiqueta de dominio, no parecen seguir ninguna patrón regular, por ejemplo:

1. **sultana#n#1** (pale yellow seedless grape used for raisins and wine)
Botany Gastronomy
2. **morocco#n#2** (a soft pebblegrained leather made from goatskin; used for shoes and book bindings etc.)
Anatomy Zoology
3. **canicola_fever#n#1** (an acute feverish disease in people and in dogs marked by gastroenteritis and mild jaundice)
Medicine Physiology Zoology
4. **blue#n#1 blueness#n#1** (the color of the clear sky in the daytime; "he had eyes of bright blue")
Color Quality

El ejemplo 1 y 3 dependen de subjerarquías de dominios diferentes. En cambio en el ejemplo 2 las etiquetas dependen de *Biology*. En el ejemplo 4 las etiquetas dependen de *Factotum*.

Este hecho, seguramente depende en gran medida del proceso que semiautomático que se siguió para su construcción. Muchas de

las etiquetas fueron asignadas en niveles altos de las jerarquías de WN y fueron propagadas automáticamente a través de la jerarquía de hipónimos y tropónimos. Debemos destacar, que hasta el momento no se ha realizado una verificación completa, ya sea manual o automática de todas las asignaciones de dominios a synsets.

3.1 Factotum

Se considera la etiqueta de dominio *Factotum* para aquellos *synsets* que no pertenecen a un dominio específico, pero pueden aparecer en muchos de ellos. Esta etiqueta de dominio incluye dos tipos de *synsets*:

Synsets genéricos: Se utiliza para marcar los sentidos de WN que no pertenecen a un dominio específico. Como por ejemplo: persona, día de un mes, etc.

Stop Senses: Aquellos *synsets* que aparecen con frecuencia en diversos contextos, como por ejemplo: números, días de la semana, colores, etc.

3.2 Jerarquía de dominios

Existen 165 etiquetas de dominios organizadas en cuatro niveles en forma jerárquica, donde cada nivel es de acuerdo al grado de especificación (ver figura 1).

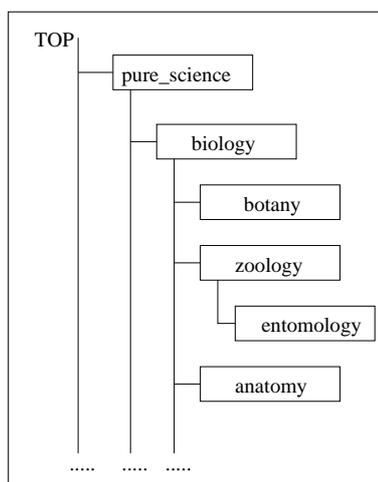


Figura 1: Parte de la Jerarquía de Dominios de WordNet

En el nivel uno de la jerarquía de dominios encontramos seis categorías:

- **Doctrines:** {art, archaeology, astrology, literature, linguistics...}
- **Free_time:** {play, sport}

- **Applied_Science:** {agriculture, alim-entation, architecture, medicine, ...}
- **Pure_Science:** {astronomy, mathematics, biology, earth, chemistry, ...}
- **Social_Science:** {anthoropology, law, military, pedagogy, sociology, ...}
- **Factotum:** {number, color, person, ...}

3.3 Aplicaciones de WordNet Domains

La importancia de las etiquetas de dominios de WN ha sido tratada en varios trabajos en los últimos años. En tareas de Word Sense Disambiguation (WSD), (Gonzalo et al., 1998) enfatiza el rol de los dominios en WSD. (Magnini y Strapparava, 2000) introducen Word Domain Disambiguation (WDD) como una variante de WSD donde para cada palabra en un texto se elige la etiqueta de dominio en vez de la etiqueta de sentido. Además en tareas como “Text Categorization” (TC) se ha utilizado WordNet Domains (Magnini et al., 2002), entre otras.

4 Experimentación

A pesar de ser un recurso muy valioso, WORDNET DOMAINS al haberse construido de forma semiautomática, requiere de un proceso de verificación, ya sea manual o automático, que nos permita validar las asignaciones realizadas de etiquetas de dominio a synsets de WordNet1.6. Más aún, nos planteamos adquirir los modelos implícitos de asignación de dominios a glosas, para poder etiquetar de forma automática otros recursos, diccionarios y versiones de WordNet que no se hayan actualmente enriquecidas con etiquetas de dominio. Por tanto, los primeros experimentos realizados se han centrado en:

- La asignación automática de dominios a glosas de WordNet1.6 (u otras versiones anteriores o posteriores de WordNet), u otras definiciones de diccionarios genéricos.
- La validación de la consistencia de las asignaciones de dominios a WordNet1.6 y en particular verificar que las etiquetas Factotum han sido asignadas de forma sistemática.

La tabla 3 muestra el porcentaje de etiquetas factotum para nombres, verbos, adjetivos

POS	con_Fac	sin_Fac	%Fac
noun	66025	58252	11.77
verb	12127	4425	63.51
adj	17915	6910	61.42
adv	3575	1039	70.93

Tabla 3: Distribución de Synsets con y sin etiquetas de dominios Factotum en WN1.6

y adverbios en WN1.6. Existe un porcentaje elevado de *synsets* etiquetados como factotum, excepto en el caso de los nombres.

Para nuestros experimentos se seleccionó aleatoriamente el 1% de los *synsets* de cada POS para realizar el test y el resto para el entrenamiento (ver tabla 4).

POS	SF	CF	%Fac
noun	572	647	11.90
verb	43	121	60.33

Tabla 4: Corpus de test para nombres y verbos sin factotum (SF) y con factotum (CF)

4.1 Método de etiquetado

El método de etiquetado automático, que se detalla a continuación, está basado en (Rigau, Atserias, y Agirre, 1997):

- Cálculo de la matriz de pesos: se toma en consideración las palabras que conforman el *synset* del corpus de entrenamiento, es decir, los *variant* o sinónimos y la glosa. Se determina la frecuencia de cada palabra con respecto a la o las etiquetas de dominio que posee el *synset*. Luego se genera un vector de pesos para cada palabra, utilizando las medidas descritas en 4.2. Como ejemplo (ver tabla 5) se presenta parte del vector de pesos para los nombres *soccer* (monosémico) y *orange* (polisémico).
- Ajuste de parámetros: Los parámetros que se tomaron en consideración fueron entre otros, el porcentaje de los *variant* (70%) y palabras de la glosa (30%). Se normalizó los vectores obtenidos para cada *synset* y se consideró como etiquetas propuestas aquellas que estuvieran en el 15% superior (rango de umbral [1..0,85]).

soccer		orange	
peso	etiqueta	peso	etiqueta
2.826	soccer	8.181	botany
2.183	play	5.129	gastronomy
1.987	football	3.019	color
1.917	sport	1.594	entomology
0.998	rugby	1.205	jewellery
...

Tabla 5: Vector de pesos para nombres con factotum (CF)

4.2 Medidas

Para calcular los pesos de las palabras asignadas a un dominio hemos probado tres funciones distintas:

M1: Fórmula de la raíz cuadrada

$$\frac{f(x, y) - \frac{1}{Nf(x)f(y)}}{\sqrt{f(x, y)}}$$

M2: Association Ratio

$$Pr(w/SC) \log_2 \left(\frac{Pr(w/SC)}{Pr(w)} \right)$$

M3: Fórmula del logaritmo

$$\log_2 \left(\frac{Nf(x, y)}{f(x)f(y)} \right)$$

5 Evaluación y resultados

Las medidas de evaluación utilizadas en los distintos experimentos se detallan a continuación:

MiA Medida i-ésima al considerar los aciertos entre la etiqueta propuesta y la etiqueta correcta

MiD Medida i-ésima al considerar los aciertos derivados de la jerarquía de dominios. Por ejemplo si la etiqueta propuesta es Zoology y la correcta es Biology, se considerará como un acierto.

AP *Accuracy* para la primera etiqueta propuesta

$$AP = \frac{\text{aciertos de la primera etiqueta}}{\text{total de synset}}$$

AT Accuracy para todas las etiquetas propuestas

$$AT = \frac{\text{aciertos de todas las etiquetas}}{\text{total de synset}}$$

Precision

$$P = \frac{(\text{etiquetas propuestas y correctas})}{(\text{total etiquetas propuestas})}$$

Recall

$$R = \frac{(\text{etiquetas propuestas y correctas})}{\text{total etiquetas correctas}}$$

F1

$$F1 = \frac{2PR}{(P + R)}$$

N	AP	AT	P	R	F1
M1A	70.94	79.75	64.74	68.25	66.45
M1D	74.50	84.85	68.88	72.62	70.70
M2A	45.75	50.39	42.73	43.12	42.92
M2D	52.09	57.50	48.75	49.21	48.98
M3A	66.77	74.50	60.86	63.76	62.27
M3D	71.56	81.45	66.54	69.71	68.09

Tabla 6: Resultados para nombres considerando factotum (CF)

N	AP	AT	P	R	F1
M1A	73.95	81.82	66.81	68.68	67.73
M1D	78.50	87.24	71.24	73.24	72.23
M2A	52.45	57.52	49.32	48.24	48.77
M2D	59.44	65.21	55.94	54.71	55.32
M3A	74.48	82.69	68.41	69.41	68.91
M3D	78.85	88.64	73.33	74.41	73.87

Tabla 7: Resultados para nombres sin considerar factotum (SF)

Para los nombres se realizaron experimentos que entrenaban con factotum y testeaban con factotum, los resultados obtenidos se muestran en la tabla 6. En la tabla 7 se muestran los resultados al realizar un entrenamiento y test de nombres sin factotum. Los mejores resultados en promedio obtenidos en ambos experimentos se obtuvieron con

V	AP	AT	P	R	F1
M1A	51.24	57.02	47.26	50.74	48.94
M1D	51.24	57.02	47.26	50.74	48.94
M2A	13.22	14.88	12.68	13.24	12.95
M2D	16.53	19.83	16.90	17.65	17.27
M3A	23.14	28.10	21.94	25.00	23.37
M3D	24.79	29.75	23.23	26.47	24.74

Tabla 8: Resultados para verbos considerando factotum (CF)

V	AP	AT	P	R	F1
M1A	69.77	76.74	64.71	55.93	60.00
M1D	74.72	83.72	69.23	61.02	64.86
M2A	20.93	25.58	19.64	18.64	19.13
M2D	41.86	51.16	38.60	37.29	37.93
M3A	41.86	55.81	39.34	40.68	40.00
M3D	53.49	67.44	46.77	49.15	47.93

Tabla 9: Resultados para verbos sin considerar factotum (SF)

la medida M1. Destacando que más del 70% de las primeras etiquetas propuestas coinciden con los *WorNet Domains*.

En la tabla 8 se muestran los resultados al realizar un entrenamiento y test de verbos con factotum. Mientras que la tabla 9 muestra los resultados de entrenamiento y test de verbos sin factotum. En ambos casos se obtienen peores resultados que para los experimentos con nombres. Una de las razones puede ser el alto número de *synsets* de verbos etiquetados con factotum (ver tabla 4). Pero en el caso de los verbos sin factotum se aprecia una mejoría de cerca del 70% de etiquetas correctas como primera propuesta.

	Train CF			
	Test CF		Test SF	
	P	R	P	R
M1A	64.74	68.25	86.15	82.35
M1D	68.88	72.62	89.23	85.29

Tabla 10: Entrenamiento con factotum para nombres usando medida M1

En la tabla 10 se hace una comparación para los nombres al realizar un entrenamiento con factotum y un test con diferentes cor-

	Train SF			
	Test CF		Test SF	
	P	R	P	R
M1A	76.37	75.66	66.81	68.68
M1D	79.04	78.31	71.24	73.24

Tabla 11: Entrenamiento sin factotum para nombres usando medida M1

pus (CF y SF). Se puede apreciar una mejoría considerable al entrenar con factotum y hacer un test sin factotum, llegando al 86.15% de precisión en la primera asignación, lo cual puede deberse a que las etiquetas distintas de factotum están correctamente asignadas.

En cambio al hacer el experimento de entrenar sin factotum y hacer test variados (CF y SF), como se muestra en la tabla 11, podemos darnos cuenta que sucede lo contrario al caso anterior. La razón puede deberse a que nuevamente la etiquetas diferentes a factotum estan bien etiquetadas.

Pero al comparar los resultados de las tablas 10 y 11 podemos decir que se alcanza en promedio una precisión cercana al 80% en las etiquetas propuestas en primer lugar.

6 Discusión

Aunque los resultados obtenidos son alentadores, si realizamos un análisis más detallado de las asignaciones erróneas realizadas por el método automático, comprobamos además que en la mayoría de los casos, las etiquetas propuestas resultan ser bastante parecidas. Esto sugiere cierta falta de sistematicidad en la asignación de etiquetas. Además, en aquellos casos donde se considera la jerarquía de dominios en la asignación (MiD), se obtiene también, en la mayoría de los casos, unos mejores resultados.

A modo de ilustración, se muestran a continuación algunos *synset* donde la etiqueta propuesta fue considerada como errónea en la evaluación, pero al analizar la glosa respectiva, podemos intuir que en muchos casos podría ser considerada como una etiqueta corregida.

1. **Palabras monosémicas.** Aquellas palabras de la glosa que sean monosémicas pueden ayudar a encontrar el dominio más adecuado.

credit_application#n#1 (an application for a line of credit)
Etiquetado con SCHOOL
Propuesta 1: Banking
Propuesta 2: Economy

OBS: *line_of_credit#n#1* es monosemico y está etiquetado como Banking.

plague_spot#n#1 (a spot on the skin characteristic of the plague)
Etiquetado con ARCHITECTURE
Propuesta 1: Physiology
Propuesta 2: Medicine

OBS: *plague#n#1* es monosemico y está etiquetado como Physiology-Medicine. Y *skin#n* posee 6 sentidos como nombre etiquetados como Anatomy, Transport y Factotum.

2. **Relación entre Etiquetas.** Existe una relación directa en la jerarquía de dominios entre las etiquetas propuestas y las correctas.

academic_program#n#1 (a program of education in liberal arts and sciences (usually in preparation for higher education))
Etiquetado con PEDAGOGY
Propuesta 1: School
Propuesta 2: University

OBS: Pedagogy es padre de School y de University.

shopping#n#1 (searching for or buying goods or services)
Etiquetado con ECONOMY
Propuesta 1: Commerce

OBS: En la jerarquía de dominios, Commerce y Economy dependen directamente de Social_science.

authorisation#n#1 (the act of authorizing)
Etiquetado con POLITICS
Propuesta 1: Law

OBS: En este caso Law y Politics dependen en forma directa de Social_science.

fire_control_radar#n#1 (radar that controls the delivery of fire on a military target)

Etiquetado con MER-
CHANT_NAVY

Propuesta 1: Military

OBS: Merchant_navy depende de Transport. Mientras que Military y Transport dependen de Social_science.

3. **Relaciones en WN.** En algunos casos los *synset* están relacionados con palabras contenidas en la glosa.

bowling#n#2 (a game in which balls are rolled at an object or group of objects with the aim of knocking them over play)

Etiquetado con BOWLING

Propuesta 1: Play

OBS: game#n#2 es hiperónimo y está etiquetado como Play. Además play#n#16 etiquetado como Play-Sport, está relacionado por holonimia con game#n#2. Además en la jerarquía de dominios, Play y Sport son hermanos; y Bowling depende de Sport.

cost_analysis#n#1 (breaking down the costs of some operation and reporting on each factor separately)

Etiquetado con FACTOTUM

Propuesta 1: Economy

OBS: La palabra “cost” perteneciente a la glosa posee 3 sentidos, etiquetados Economy, Money y Quality respectivamente.

4. **Casos inciertos y dudosos.** Existen casos en donde la etiqueta propuesta no responde a ningún patrón, pero que puede ser considerada como la etiqueta correcta.

birthmark#n#1 (a blemish on the skin formed before birth)

Etiquetado con QUALITY

Propuesta 1: Medicine

bardolatry#n#1 (idolization of William Shakespeare)

Etiquetado con RELIGION

Propuesta 1: history

Propuesta 2: literature

7 Conclusiones y trabajo futuro

Como se ha visto en los resultados, el procedimiento para asignar etiquetas de dominio a glosas de WordNet es bastante alentador, sobre todo si consideramos que es un problema difícil, al tener en consideración la polisemia existente en WN y la forma en que fueron generadas las etiquetas de dominios, generadas a través de un proceso de propagación semi-automático a través de las relaciones existentes en WN.

El procedimiento es bastante fiable en las etiquetas que propone como primera prioridad, alcanzando más del 70% de coincidencia con las etiquetas de dominio propuestas por (Magnini y Cavaglià, 2000).

También hemos realizado un primer estudio sobre la tipología de los errores del método que sugiere que en ciertos casos podríamos añadir nuevas etiquetas correctas al *synset* en cuestión o corregirlas, substituyéndolas.

También hemos visto que los *synsets* actualmente etiquetados como *factotum*, pueden ser, etiquetados con una etiqueta de dominio.

Como trabajo futuro, planeamos realizar ciertas mejoras y adaptaciones al algoritmo para etiquetar otras versiones de WordNet y para lograr etiquetar un texto libre (p.e. SemCor) con las etiquetas de dominios de WN.

Hemos planificado también realizar un extensión del método considerando las glosas de *synsets* relacionadas con el *synset* etiquetado (mediante relaciones de hiponimia o meronimia).

Por último, también queremos realizar un estudio sobre la bondad de cada asignación automática, proporcionando credibilidad sólo a aquellas que superen un cierto umbral, o cumplan determinadas propiedades.

Bibliografía

- Agirre, E. y D. Martinez. 2002. Integrating selectional preferences in wordnet. En *Proceedings of the first International WordNet Conference in Mysore, India*, 21-25 January.

- Atserias, J., S. Climent, X. Farreres, G. Rigau, y H. Rodríguez. 1997. Combining multiple methods for the automatic construction of multilingual wordnets. En *Proceeding of RANLP'97*, páginas 143–149, Bulgaria. Also to appear in a Book.
- B. Magnini, G. Cavaglià, G. Pezzulo A. Gliozzo. 2001. Using Domain Information for Word Sense Disambiguation. *Proceedings of SENSEVAL-2, France*.
- Bentivogli, L., E. Pianta, y C. Girardi. 2002. Multiwordnet: developing an aligned multilingual database. En *First International Conference on Global WordNet*, Mysore, India.
- Fellbaum, C., editor. 1998. *WordNet. An Electronic Lexical Database*. The MIT Press.
- Gonzalo, J., F. Verdejo, C. Peters, y N. Calzolari. 1998. Applying eurowordnet to cross-language text retrieval. *Computers and Humanities*.
- Harabagiu, S., M. Pasca, y S. Maiorano. 2000. Experiments with open-domain textual question answering. En *Proceedings of COLING-2000*, Saarbrücken Germany.
- Magnini, B. y G. Cavaglià. 2000. Integrating subject field codes into wordnet. En *In Proceedings of the Second International Conference on Language Resources and Evaluation LREC'2000*, Athens. Greece.
- Magnini, B. y C. Strapparava. 2000. Experiments in Word Domain Disambiguation for Parallel Texts. *Proceedings of ACL, Hong Kong*.
- Magnini, B. y C. Strapparava. 2002. User modelling for news web sites with content-based techniques. En *Proceedings WWW-2002, the Eleventh International World Wide Web Conference, Poster session*, Honolulu, Hawaii, USA,.
- Magnini, B., C. Strapparava, G. Pezzulo, y A. Gliozzo. 2002. Comparing Ontology-Based and Corpus-Based Domain Annotations in WordNet. *In Proceedings of First International WordNet Conference*.
- McCarthy, D. 2001. *Lexical Acquisition at the Syntax-Semantics Interface: Diathesis Alternations, Subcategorization Frames and Selectional Preferences*. Ph.D. tesis, University of Sussex.
- Miller, G., C. Beckwith, D. Fellbaum, D. Gross, y K. Miller. 1990. Five papers on WordNet. *Cognitive Science Laboratory, Princeton University*, Report 43.
- Rigau, G., J. Atserias, y E. Agirre. 1997. Combining unsupervised lexical knowledge methods for word sense disambiguation. En *Proceedings of joint 35th Annual Meeting of the Association for Computational Linguistics and 8th Conference of the European Chapter of the Association for Computational Linguistics ACL/EACL'97*, Madrid, Spain.