Estudio de cooperación de métodos de desambiguación léxica: Marcas de Especificidad *vs.* Máxima Entropía

Armando Suárez y Andrés Montoyo.
Grupo de Procesamiento del Lenguaje y Sistemas de Información
Departamento de Lenguajes y Sistemas Informáticos. Universidad de Alicante.
E-mail: {armando,montoyo} @dlsi.ua.es

Resumen

En este artículo se presentan dos métodos que resuelven la ambigüedad léxica, se realiza una comparación entre ellos y se demuestra que una adecuada cooperación mejoraría los resultados obtenidos.

Este artículo presenta un estudio comparativo sobre dos métodos que resuelven la ambigüedad léxica de nombres en textos escritos en inglés. Los métodos utilizados son Marcas de Especificidad y modelos de probabilidad de Máxima Entropía.

El primero se clasifica como un método basado en el conocimiento y el segundo como un método de aprendizaje supervisado basado en corpus. Esta clasificación representa las dos grandes líneas de investigación que se están desarrollando actualmente en el campo de la ambigüedad léxica en el procesamiento del lenguaje natural. Ambas tendencias tienen sus ventajas e inconvenientes.

Mediante la experimentación sobre un mismo conjunto de evaluación, el trabajo presentado muestra una mejora de un 12% cuando se combinan los dos métodos utilizados.

1. Introducción

La resolución de la ambigüedad léxica es uno de los objetivos principales de cualquier sistema de Procesamiento del Lenguaje Natural (PLN).

En el presente artículo nos centramos en la resolución de la ambigüedad léxica, la cual aparece cuando las palabras presentan diferentes significados. A esta tarea se le conoce como Desambiguación del sentido de las palabras (Word Sense Disambiguation, WSD). Esta es una "tarea intermedia" [13] que sirve de ayuda cuando necesitamos conocer el sentido de las palabras en algunas aplicaciones del PLN, como en Traducción Automática (TA), Recuperación de la Información (RI), Clasificación de Textos, Análisis del Discurso, Extracción de Información (EI), etc.

De forma más genérica, la WSD consiste en la asociación de una palabra dada en un texto con una definición o significado, el cual la distingue de otros significados atribuibles a esa palabra. La asociación de palabras a los sentidos, se cumple dependiendo de dos recursos de información (contexto[1] y recursos de conocimiento externos[2]).

Existen diferentes métodos de trabajo para WSD basándose principalmente en el conocimiento (WSD *Knowledge-Driven*) y/o en el corpus (WSD *Corpus-Based*), según puede verse en la clasificación aportada en [1].

Durante los últimos años, se han realizado muchas investigaciones sobre WSD basada en la combinación de métodos basados en el conocimiento y en el corpus. Yarowsky, en [4], deriva clases de palabras a partir de palabras en categorías comunes del *Roget´s International Thesaurus*. Karov *et al*, en [4], describen un sistema que aprende, a partir de un corpus, un conjunto de usos típicos para cada sentido de las palabras polisémicas de un MRD, proponiendo un par de fórmulas de similitud entre palabras y oraciones. Yarowsky en [19] acumula palabras preseleccionadas a partir de un diccionario electrónico (*Machine Readable Dictionary, MRD*) en el primer paso de su procedimiento, con el objetivo de reunir características locales para posteriormente realizar WSD usando corpora. Resnik, en [13], presenta un método para desambiguar el sentido de las palabras basándose en "*Information content*" reunida a partir de un corpora y usando WordNet [7]. La "*Information content*" de una clase se estima por la probabilidad de aparición de la clase en un corpus.

---

[1] El **contexto** de la palabra a ser desambiguada es considerado como un conjunto de palabras que acompañan a la palabra a desambiguar, junto con las relaciones sintácticas, categorías semánticas, etc. En este artículo el contexto se compone de oración en oración.

[2] Los **recursos de conocimiento externos** son los recursos léxicos, enciclopédicos, recursos de conocimiento léxico (WordNet) desarrollados manualmente que proporcionan datos valiosos para asociar palabras con sentidos, etc.

Rigau *el al.* en [15] combinaron varias técnicas y recursos léxicos independientes usando información de MRD´s para representar vectores de contenido y probar diferentes técnicas y medidas de similitud para signar el sentido genus hiperonimo correcto. Rigau *el al.* en [16] usan un MRD bilingüe para asignar categorías semánticas de WordNet a sentidos de palabras y realizar un proceso de entrenamiento para juntar las palabras en categorías semánticas.

El presente artículo se centra en un estudio empírico de la resolución del sentido de la ambigüedad léxica, utilizando un método basado en el conocimiento y otro en el aprendizaje supervisado a partir de corpus. El objetivo es comprobar a partir de los resultados obtenidos las posibles mejoras derivadas del desarrollo de un nuevo método híbrido que combine la desambiguación del sentido de las palabras basado en el conocimiento y en el corpus.

En la sección 2 se describe el método de Marcas de Especificidad [4] clasificado como WSD basado en el conocimiento. En la sección 3 se presenta un método basado en el principio de Máxima Entropía, el cual se clasifica como WSD basado en el corpus. En la sección 4 se presentan los distintos experimentos realizados para los dos métodos anteriores y los resultados obtenidos. En la sección 5 se comparan los dos métodos y se estudia la posible cooperación entre ambos. Finalmente, se presentan las conclusiones y los trabajos futuros.

2. WSD usando Marcas de Especificidad

Normalmente, las palabras que aparecen en un mismo contexto tienen sus sentidos muy relacionados entre sí. Por tal motivo, y para resolver el problema de la ambigüedad léxica interesa tener un recurso que tenga las palabras y los conceptos organizados alrededor de clases (jerarquías), de tal forma que describan todas sus características semánticas. Así, si dos palabras pertenecen a una misma clase quiere decir que sus sentidos están fuertemente relacionados. Las Marcas de Especificidad obtienen las ventajas de lo comentado anteriormente, y para aplicarlas se usarán las relaciones jerárquicas (hiperonimia/ hiponimia) que proporciona WordNet.

En primer lugar se explicará intuitivamente la noción de Marca de Especificidad para posteriormente aplicarla en la desambiguación. Cuanta más información común comparten dos conceptos, más relacionados estarán, y la información común que comparten esos dos conceptos es indicada por el concepto padre de ambos en la jerarquía, al cual se llamará de ahora en adelante Marca de Especificidad (ME). Estas ME se obtendrán a partir de clases semánticas de la jerarquía de WordNet.

El método recorrerá todos los subárboles de la jerarquía semántica de WordNet para el contexto de entrada y para cada ME calculará cuantas palabras del contexto de entrada se agrupan alrededor de ella. Aquella ME que agrupe el máximo número de palabras del contexto, será elegida como el sentido de la palabra. Con otras palabras, el método busca aquella ME que tenga mayor densidad de palabras debajo de su subárbol, queriendo decir que sus sentidos están fuertemente relacionados. Este método se aplica sobre las jerarquías de WordNet de la siguiente manera.

La entrada al método WSD será el grupo de palabras $W=\{W_1, W_2, ..., W_n\}$ que se obtienen de la oración y forman el contexto. Cada palabra W_i se busca en WordNet y se obtienen los sentidos asociados a ellas $S_i=\{S_{i1}, S_{i2}, ..., S_{in}\}$ y para cada sentido S_{ij} se obtendrá todos los synsets hiperónimos en su taxonomía IS-A. Inicialmente, se busca el concepto común a todos los sentidos de las palabras que forman el contexto de entrada. A este concepto se le denomina Marca de Especificidad Inicial (MEI). Si esta MEI no resuelve la ambigüedad de las palabras, se va descendiendo nivel a nivel a través de la jerarquía WordNet asignando nuevas ME. Para cada ME anterior, se calculará el número de conceptos que forman parte del contexto y que están contenidos en la subjerarquía. Aquella ME que en su subjerarquía, tenga el mayor número de palabras del contexto será la elegida, asignando el sentido que nos devuelve WordNet a cada una de estas palabras que forman parte de la ME seleccionada.

En la figura 2, se puede apreciar gráficamente como la palabra W_1 tiene cuatro sentidos diferentes y varias palabras de contexto. La MEI no resuelve la ambigüedad léxica, ya que la palabra W_1 aparece en tres subjerarquías con diferentes sentidos. Sin embargo, la ME con el símbolo (*) contiene el mayor número de palabras del contexto (tres) y, por lo tanto, será elegida para resolver el sentido S_2 de la palabra W_1. Las palabras W_2 y W_3 también son desambiguadas eligiendo el sentido S_1 para ambas. Para la palabra W_4, que no ha sido desambiguada satisfactoriamente, se le aplicarán las heurísticas especificadas a continuación y que están previstas para estos casos.

Después de evaluar el método se observó que se podía mejorar, obteniendo unos niveles de desambiguación bastante mejores. Para ello se definieron seis heurísticas llamadas: de hiperónimo, de definición, de hipónimo, de glosa hiperónimo, de glosa hipónimo y de Marca de Especificidad Común. Una explicación detallada del método WSD utilizando las ME junto con las distintas heurísticas pueden encontrarse en [10], y su aplicación a otras tareas de PLN en [13].

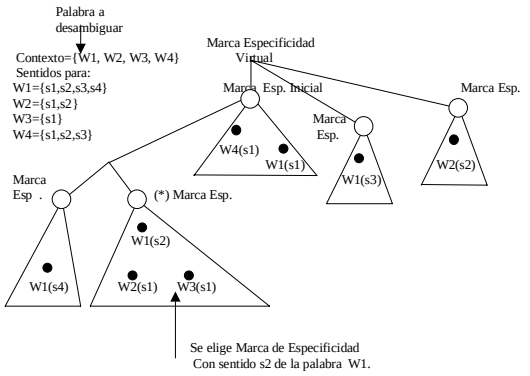

Figura 1 Marcas de Especificidad (ME)

Además, como se describe en [11], el método de ME trabaja satisfactoriamente cuando es aplicado a sistemas de clasificación, ya que se obtiene una *precision* de 96,1% en la experimentación realizada sobre *IPTC Subject Reference Systems*[3].

3. WSD usando Máxima Entropía

El método que presentamos a continuación se inscribe dentro del conjunto de métodos de aprendizaje supervisado basados en corpus. El objetivo es obtener clasificadores de sentidos de las palabras mediante un modelo de probabilidad basado en el principio de Máxima Entropía (MXE) [20], [4] y [1].

Los modelos de Máxima Entropía buscan la óptima distribución de probabilidad asumiendo la máxima ignorancia sobre los datos, esto es, no se asume ningún conocimiento que no se encuentre reflejado en el conjunto de entrenamiento. Su principal ventaja reside en su capacidad de representación de fuentes de información de contexto heterogéneas.

El propósito del método aquí presentado es aplicar estos modelos a la desambiguación léxica de nombres, verbos y adjetivos.

Se parte de un conjunto de ejemplos de aprendizaje que nos servirá para construir el clasificador. En nuestro caso, estos ejemplos son frases etiquetadas sintácticamente donde la palabra objetivo está etiquetada con su sentido correcto. Para representar la evidencia, los modelos MXE utilizan funciones binarias que indican la presencia o no de una cierta característica dentro de un contexto que acompaña al ejemplo utilizado en el aprendizaje. Definidas estas funciones y obtenido su valor al aplicarlas a todos los ejemplos de entrenamiento, se obtiene el modelo de probabilidad óptimo que maximiza la entropía. Esto se consigue mediante la estimación de unos parámetros que nos indican el peso o la importancia de cada función en el proceso de clasificación. Al final, el modelo obtenido tiene la misma distribución de probabilidad que la observada en el conjunto de entrenamiento.

Supongamos que *w* es la palabra para la que queremos obtener un clasificador, que *X* es el conjunto de contextos observados, *p'(x)* la probabilidad observada de aparición de un determinado contexto *x*, y *C* el conjunto de clases (significados de la palabra en nuestro caso). La distribución óptima de probabilidad *p** será, de todas las distribuciones posibles, aquella que maximice la entropía *H(p)* partiendo de la distribución observada *p*.

(1)

$$p^* = \arg\max_{p \in P} H(p)$$

$$H(p) = -\sum_{x \in X, c \in C} p'(x)\, p(c \mid x) \log p(c \mid x)$$

Tras la estimación de parámetros obtendremos una función *cl: X→C* que permitirá obtener la clase correcta para un contexto lingüístico mediante la elección del valor de probabilidad máximo. El cálculo de la probabilidad de que *c* sea la clase correspondiente al contexto *x* se realiza mediante la expresión (2).

(2)

$$p(x,c) = \frac{1}{Z(x)} \prod_{i=1}^{K} \alpha_i^{f_i(x,c)}$$

Cada parámetro estimado α_i está asociado a una única función f_i, y Z(x) es una constante de normalización que nos asegura que la suma de los

[3] The IPTC Subject Reference System has been developed to allow Information Providers access to a universal language independent coding system for indicating the subject content of news items. http://www.ipt.org

valores de probabilidad para el contexto x es igual a 1.

Las funciones definidas en el sistema aquí presentado son, básicamente, funciones binarias que informan de las palabras y combinaciones de palabras, y etiquetas sintácticas que se hallan en la cercanía de la palabra objetivo, así como de su posición relativa

La estimación de parámetros (o "pesos" de cada función) se realiza mediante el procedimiento denominado Generalized Iterative Scaling (GIS). Una vez obtenidos estos parámetros disponemos de un clasificador para la palabra en cuestión que nos dirá cual es su sentido en cualquier contexto lingüístico.

4. Experimentos y Resultados

En esta sección se describirá un conjunto de experimentos y los resultados obtenidos. El principal objetivo de estos experimentos es ver la efectividad de los dos métodos, descritos en las secciones anteriores, sobre un mismo conjunto de ejemplos. Posteriormente a partir de la comparación de los resultados obtenidos se estudiará la viabilidad de un nuevo método híbrido basado en la cooperación de ambos.

Para evaluar y comparar los dos métodos aquí propuestos se han utilizado todos los artículos presentes en las carpetas Brown1 y Brown2 de Semcor [5], tal y como se distribuye con WordNet1.6. Para realizar el proceso de desambiguación se han seleccionado al azar unos cuantas nombres: *account, age, art, car, child, church, cost, duty, head, interest, line, member, people, term, test,* y *work.*

En este artículo se presentan tres experimentos, que se mostrarán en las subsecciones siguientes.

4.1 Experimento sobre Marcas Especificidad

En la realización del primer experimento, se seleccionaron todas las oraciones en las que apareciera alguno de los nombres anteriormente seleccionados de entre todo el corpus Semcor. Para cada una de estas oraciones se obtenían los nombres que acompañan a la palabra a desambiguar, formando el contexto de entrada. Este contexto se le introduce al método de WSD, el cual una vez aplicado devuelve automáticamente el sentido correspondiente de WordNet para cada uno de los nombres. Una ventaja importante al aplicar este método es su proceso completamente automático, por lo que no necesita procesos de entrenamiento, codificación manual léxica de las entradas ni etiquetado manual de los nombres del texto. Sin embargo, un inconveniente encontrado en los experimentos realizados con el Semcor es que el método depende en gran medida de las relaciones semánticas (Hiperonimia/Hiponimia) y la organización jerárquica conceptual que utiliza WordNet para desambiguar el sentido de las palabras. Por eso, en la Tabla 1, la cual muestra los resultados[4] obtenidos cuando se aplica el método de ME sobre los nombres seleccionados, hay palabras que tienen un porcentaje de desambiguación tan bajo. Por ejemplo, *test* obtiene un porcentaje tan bajo de desambiguación debido a que los otros nombres que forman el contexto no están relacionados semánticamente por WordNet.

nombre	#	**P**	R	A
account	21	**0,048**	0,048	1,000
age	86	**0,523**	0,523	1,000
art	64	**0,333**	0,328	0,984
car	65	**0,734**	0,723	0,985
child	180	**0,622**	0,594	0,956
church	107	**0,539**	0,514	0,953
cost	76	**0,289**	0,289	1,000
duty	23	**0,348**	0,348	1,000
head	168	**0,204**	0,190	0,935
interest	126	**0,444**	0,444	1,000
line	118	**0,209**	0,203	0,975
member	68	**0,515**	0,515	1,000
people	244	**0,531**	0,520	0,980
term	45	**0,156**	0,156	1,000
test	34	**0,088**	0,088	1,000
work	190	**0,255**	0,253	0,989
TOTAL	1615	**0,404**	0,395	0,978

Tabla 1 Resultados de evaluación de ME en el corpus Semcor

4.2 Experimento sobre Máxima Entropía.

Para la construcción de los distintos clasificadores, y basándonos en trabajos previos como [2], [10] y [7], se han definido funciones que indican la presencia de palabras en posiciones

[4] "Attempted"(A) es obtenido del resultado entre el número total de sentidos contestados y el número total de sentidos. "Precision" (P) se define como el resultado entre los sentidos desambiguados correctamente y el número total de sentidos contestados. "Recall" (R) se define como el resultado entre los sentidos desambiguados correctamente y el número total de sentidos."#" es la cantidad de ocurrencias de la palabra en el corpus evaluado.

relativas a la palabra objetivo: *w-1, w-2, w-3, w+1, w+2, w+3, (w-1 w-2), (w-1, w+1), (w+1, w+2), (w-3, w-2, w-1), (w-2, w-1, w+1), (w-1 w+1 w+2), (w+1 w+2 w+3).*

También se han tenido en cuenta las etiquetas sintácticas en el contexto, nuevamente en posiciones relativas: *p-3, p-2, p-1, p+1, p+2, p+3.* Estas etiquetas sólo se han computado si la palabra correspondiente es una forma funcional.

Asimismo, se ha utilizado una función basada en la frecuencia de aparición de ciertas palabras con contenido léxico en contextos asociados a una clase en concreto: palabras que aparecen, en todo el corpus de entrenamiento y para cada sentido de la palabra a clasificar, 5 veces o más.

Para la evaluación del método se ha dividido el corpus en 10 partes (*10-fold cross-validation*). Se realizan 10 experimentos distintos, seleccionando, en cada uno de ellos, una décima parte del corpus como conjunto de evaluación y el resto como conjunto de entrenamiento. Los resultados obtenidos se muestran en la Tabla 2. En ella se muestran la cantidad media de ocurrencias de cada nombre en los conjuntos de evaluación y los valores promedio de *precision*, *recall* y *attempted*.

nombre	#	**P**	R	A
account	2,7	**0,285**	0,263	0,872
age	10,3	**0,313**	0,143	0,438
art	7,3	**0,596**	0,575	0,966
car	6,9	**0,959**	0,959	1,000
child	19,1	**0,957**	0,169	0,189
church	12,7	**0,558**	0,543	0,967
cost	8,4	**0,883**	0,851	0,962
duty	2,5	**0,778**	0,685	0,870
head	16,6	**0,600**	0,582	0,961
interest	13,7	**0,485**	0,454	0,932
line	12,2	**0,070**	0,067	0,946
member	7,3	**0,874**	0,874	1,000
people	27,1	**0,626**	0,359	0,530
term	5,2	**0,445**	0,430	0,951
test	3,6	**0,258**	0,252	0,938
work	20,3	**0,405**	0,392	0,962
	TOTAL	**0,586**	0,473	0,805

Tabla 2 Resultados de evaluación de MXE en el corpus Semcor

Debido al tamaño del corpus Semcor, el método no dispone de una gran cantidad de ejemplos de los que aprender, y la inmediata consecuencia de este hecho son los bajos valores de recall y attempted. Aunque el sistema podría utilizar alguna heurística (por ejemplo, frecuencia de aparición de cada clase en el corpus de entrenamiento) se asume que ante situaciones de información insuficiente (todas las clases, para un contexto concreto, tienen la misma probabilidad condicional) la clasificación no se puede llevar a cabo. Asimismo, los valores tan dispares de precisión también son imputables a la misma causa.

El nombre *interest*, que en los resultados mostrados obtiene un 48.5% de precision y un 45.4% de recall, evaluado sobre el *DSO English sense-tagged* corpus [7], mucho más extenso, se llega a obtener un 74.5% y 73.8% respectivamente. Debido a las diferencias de etiquetado entre DSO y Semcor no es posible aplicar el aprendizaje con el primero al segundo.

4.3 Estudio de cooperación

Para comparar los dos métodos se ha revisado una porción del conjunto total de datos evaluados anteriormente. Esta porción se compone de 18 artículos del Semcor que contienen 267 ocurrencias de los nombres seleccionados.

A partir de las evaluaciones de cada método, se han contabilizado aquellos casos en los que los dos métodos devuelven el mismo sentido para la misma ocurrencia. También se han contabilizado aquellos casos en que al menos uno de los métodos proporciona el sentido correcto. Un resumen de los datos obtenidos se muestra en la Tabla 3.

Casos comparados	267
Coinciden	30,71%
Al menos uno acierta	71,91%
Aciertan los dos	22,47%
Fallan los dos	28,09%

Tabla 3 Comparativa de la combinación de resultados de los dos métodos

A la vista de los datos mostrados, se puede observar que la cooperación de los dos métodos mejora la desambiguación del sentido de las palabras, ya que los porcentajes de precision, recall y attempted son mejores que los obtenidos por cada método individualmente.

En concreto, para la porción del Semcor estudiada, la cooperación entre los dos podría mejorar en aproximadamente un 12% los resultados del método con más éxito, ya que alguno de los dos métodos (o los dos) obtiene la desambiguación correcta en el 72% de los casos.

Ambos métodos coinciden y aciertan en el 22,5% de los casos. Al respecto de esto último, es remarcable el dato de que en los casos de coincidencia la mayor parte corresponde a aciertos: del 31% de los casos en que ambos proponen el mismo sentido, los aciertos suponen el 73%.

5. Conclusiones y trabajos futuros

Este artículo muestra la aplicación de dos métodos, Marcas de Especificidad (basado en el conocimiento) y Máxima Entropía (basado en corpus) que resuelven la ambigüedad léxica. Los dos métodos se probaron y se compararon sobre el mismo corpus de evaluación.

Los resultados de los experimentos realizados demuestran que la cooperación de ambos métodos mejora la desambiguación del sentido de las palabras si los comparamos con los obtenidos de sus respectivas aplicaciones individuales.

Con estos datos se puede afirmar que un nuevo método híbrido que combine las aproximaciones basadas en el conocimiento y en el corpus mejorará la desambiguación del sentido de las palabras.

Los métodos basados en corpus tienen la desventaja de su dependencia del propio corpus de entrenamiento, tanto por disponibilidad e idoneidad del mismo como por su discutible aplicación en diferentes dominios. Los sistemas basados en el conocimiento obtienen, en el momento actual, peores resultados que los primeros y exigen un gran esfuerzo de representación del conocimiento, pero son más fácilmente aplicables a dominios heterogéneos. Un método que una las dos aproximaciones aprovecharía las ventajas de cada una y minimizaría los aspectos negativos respectivos.

6. Agradecimientos

Este trabajo ha sido financiado por la Comisión Interministerial de Ciencia y Tecnología (CICYT) con referencia (TIC2000-0664-C02-02).

7. Referencias

[1] Berger A., Della Pietra S. and Della Pietra V. (1996). A Maximum Entropy Approach to Natural Language Processing. *Computational Linguistics* vol.22(1).

[2] Escudero G., Márquez L. and Rigau G. (2000). Boosting Applied to Word Sense Disambiguation. *European Conference on Machine Learning*, 129-141.

[3] Ide N. and Véronis J. (1998) Introduction to the Special Issue on Word Sense Disambiguation: The State of the Art. *Computational Linguistics*. 24 (1), 1-40.

[4] Karov Y. and Edelman S. (1996) Learning Similarity-Based Word Sense Disambiguation from Sparce Data. *Research Report CS-TR-96-05*. The Weizmann Institute of Science, Rehovot, Israel.

[5] Landes, S., Leacock C., and Tengi, R. (1998). Building a Semantic Concordance of English. In C. Fellbaum, editor, *WordNet: An electronic lexical Database and Some Applications*. MIT Press, Cambridge, MA.

[6] Manning, C. D. and Schütze, H. (1999) Foundations of Statistical Language Processing. *The MIT Press*, Cambridge, Massachusets, ISBN 0-262-13360-1.

[7] Martínez D. and Agirre E. (2000). One Sense per Collocation and Genre/Topic Variations. *Proceedings of the Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora*. Hong Kong.

[8] Miller G. A., Beckwith R., Fellbaum C., Gross D., and Miller K. J. (1990) WordNet: An on-line lexical database. *International Journal of Lexicography*, 3(4): 235-244.

[9] Montoyo, A. and Palomar M. (2000). Word Sense Disambiguation with Specification Marks in Unrestricted Texts. *In Proceedings 11th International Workshop on Database and Expert Systems Applications (DEXA 2000)*, pages 103-108. Greenwich, (London).

[10] Montoyo, A. and Palomar, M. (2001). Specification Marks for Word Sense Disambiguation: New Development. *2nd International conference on Intelligent Text Processing and Computational Linguistics* (CICLing-2001). México D.F. (México).

[11] Montoyo A., Palomar, M. and Rigau, G. (2001) WordNet Enrichment with Classification Systems. *WordNet and Other Lexical Resources: Applications, Extensions and Customisations Workshop. The Second Meeting of the North American Chapter of the Association for Computational Linguistics (NAACL-01)*. Carnegie Mellon University. Pittsburgh, PA, USA.

[12] Ng, H.T. and Lee, H.B. (1996). Integrating Multiple Knowledge Sources to Disambiguate Word Senses: An Exemplar-Based Approach. In *Proceedings of the 34th Annual Meeting of the Association for Computational Linguistics*, ACL.

[13] Palomar M., Saiz-Noeda M., Muñoz, R., Suárez, A., Martínez-Barco, P., and Montoyo, A. (2000). PHORA: NLP System for Spanish. *In Proceedings 2nd International conference on Intelligent Text Processing and Computational Linguistics* (CICLing-2001). México D.F. (México).

[14] Resnik P. (1995) Disambiguating noun groupings with respect to WordNet senses. *Proceedings of Third Workshop on Very Large Corpora.* 54-68.Cambridge, MA.

[15] Rigau G., Atserias J. and Agirre E. (1997) Combining Unsupervised Lexical Knowledge Methods for Word Sense Disambiguation. *Proceedings 35th Annual Meeting of the ACL*, 48-55, Madrid, Spain.

[16] Rigau G., Rodríguez H.. and Agirre E. (1998) Building Accurate Semantic Taxonomies from Monolingual MRD´s. *Proceedings 17th International Conference on Computational Linguistics and 36th Annual Meeting (COLING-ACL´98)*, Montreal, Canada.

[17] Wilks Y. And Stevenson M. (1996) The grammar of sense: Is word sense tagging much more than part-of-speech tagging? *Technical Report CS-96-05*, University of Sheffield, UK.

[18] Yarowsky D. (1992) Word Sense disambiguation using statistical models of Roget´s categories trainined on large corpora. *Proceedings of 14th COLING*, 454-460, Nantes, France.

[19] Yarowsky, D. (1995) Unsupervised word Sense disambiguation rivaling supervised methods. *Proceedings of 32nd Annual Meeting of the ACL*.

[20] Ratnaparkhi, A. (1998). Maximum Entropy Models for Natural Language Ambiguity Resolution. *Ph.D. Dissertation*. University of Pennsylvania.