Resolución de la Ambigüedad Léxica Mediante Información Contextual y el Modelo del Espacio Vectorial

L. Alfonso Ureña López Manuel García Vega
Departamento de Informática.
Universidad de Jaén
Avda. Madrid 35, 23071 Jaén. Spain
e-mail: {laurena,mgarcia}@ujaen.es

Manuel de Buenaga Rodríguez José María Gómez Hidalgo
Departamento de Inteligencia Artificial
Universidad Europea de Madrid
28670 - Villaviciosa de Odón. Madrid. Spain
e-mail: {buenaga,jmgomez}@dinar.esi.uem.es

Palabras clave: desambiguación del sentido de las palabras (WSD), ventana-contextual, WordNet, SemCor, córpora de texto, bases de datos léxicas.
Keywords: word sense disambiguation, context-windows, WordNet, SemCor, text corpus, lexical database.

RESUMEN. *La resolución automática de la ambigüedad léxica de términos polisémicos es una tarea útil y compleja para muchas aplicaciones del procesamiento del lenguaje natural. Presentamos un nuevo enfoque basado en la utilización del modelo del espacio vectorial y una colección de entrenamiento, de amplia cobertura, como recurso lingüístico. Este enfoque utiliza un conjunto variable de términos como contexto local. Hemos probado nuestro programa desambiguador del sentido de las palabras, sobre un gran conjunto de documentos, consiguiendo una alta precisión en la resolución de la ambigüedad léxica.*

ABSTRACT. *The resolution of lexical ambiguity of polysemics words is a complex and useful task for many natural language processing applications. We present a new approach for word sense disambiguation based in the vector space model and a widely available training collection as linguistic resource. This approach uses a variable set of terms like local context. We have tested our disambiguator algorithm on a large documents collection, achieving high precision in the resolution of lexical ambiguity.*

1.- INTRODUCCIÓN

Un problema importante en el procesamiento del lenguaje natural es determinar el sentido o significado de palabras ambiguas en un determinado contexto. Esta tarea de desambiguación es compleja para la mayoría de las aplicaciones del procesamiento del lenguaje natural. El perfeccionamiento en la identificación del sentido correcto de una palabra puede mejorar el comportamiento de sistemas de traducción automática [Wilks90], sistemas de recuperación de información [Sanderson96]; o utilizarse en tareas especificas como la restauración de acentos de palabras en el procesamiento de textos [Yarowsky94].

El problema de la desambiguación no es nuevo: Gale, Church y Yarowsky [Gale92] citan trabajos que se remontan a los años 50, si bien su posible utilización en aplicaciones sobre textos de

dimensiones reales, apenas sí se plantea. Sin embargo, en 1986, Lesk [Lesk86] construye un desambiguador con técnicas similares a las utilizadas en los sistemas de recuperación de información, con mayores posibilidades de escalabilidad. Desde el trabajo de Lesk hasta la actualidad, se han ideado enfoques muy diversos para la desambiguación automática (p. e. [Agirre96], [Rigau97], [Hwee96], [Yarowsky92-94], [Miller94] y [Bruce94]).

Podemos realizar una clasificación genérica de las aproximaciones seguidas en la desambiguación, dependiendo del *recurso léxico* utilizado: diccionarios, o córpora de entrenamiento. Los sistemas basados en diccionarios utilizan la información que aparece en las definiciones de las distintas acepciones del término a desambiguar [Lesk86], y otros recursos léxicos como thesaurus (p. e. Roget's Thesaurus) [Yarowsky92] o bases de datos léxicas (p. e. WordNet) [Agirre95]. En los enfoques basados en córpora de entrenamiento, se utilizan córpora etiquetados (p. e. SemCor) [Bruce94] [Yarowsky96]. En estos enfoques las técnicas de análisis varían según los casos, empleando técnicas bayesianas, probabilísticas e incluso redes neuronales.

En nuestro enfoque nos hemos basado en la utilización de córpora de entrenamiento, ya que, con este tipo de información existe una cierta evidencia experimental que proporciona mejores resultados [Yoshiki94]. En concreto, utilizamos SemCor (Semantic Concordance), debido a su disponibilidad y amplia cobertura. SemCor está compuesto por el Brown Corpus, etiquetado manualmente con los sentidos de las palabras definidas en WordNet [Miller90,95].

Algunos de los problemas que se presentan en los enfoques basados en córpora de entrenamiento son: primero, la mayoría de ellos incluyen un gran número de decisiones "ad-hoc", segundo, la escalabilidad de los métodos no queda clara en algunos casos, y finalmente, la evaluación de su comportamiento se realiza para un número reducido de términos.

En este trabajo presentamos un nuevo enfoque para desambiguar el sentido de las palabras (WSD)[1] utilizando un corpus de entrenamiento, basado en el Modelo del Espacio Vectorial (MEV). Este modelo, ampliamente fundamentado, tanto teórica como experimentalmente en el campo de la Recuperación de Información [Salton83,89] [Frakes92], permite el desarrollo de sistemas prácticos y eficientes. Presentamos además, una evaluación de nuestro método de desambiguación para un conjunto extenso de documentos y de términos diferentes del Brown Corpus, con los sentidos afinados de WordNet.

Además, nuestro enfoque puede facilitar en el futuro, la integración de ambos recursos léxicos (diccionarios y córpora de entrenamiento), como ya hemos realizado en categorización de textos [Gómez97].

Este trabajo está organizado como sigue. Primero, introducimos la tarea para la resolución de la ambigüedad léxica y los recursos que utilizamos. Seguidamente, describimos el modelo en el que integramos estos elementos. Después de esto, estudiamos el proceso de entrenamiento, seguido de la descripción del proceso WSD. A continuación, presentamos nuestra evaluación estudiando e interpretando los resultados, y finalmente, describimos nuestras conclusiones y líneas futuras.

2.- DESCRIPCIÓN DE LA TAREA

La entrada de nuestro programa WSD consta de un conjunto de documentos en lenguaje natural. En la salida, los términos que componen dichos documentos de entrada, son etiquetados con el significado correcto, de acuerdo con los sentidos proporcionados por WordNet. El sistema hace uso de la información contenida en los documentos para determinar el significado. Como es lógico, no todos los términos tendrán el mismo número de acepciones, sino que éste será variable,

[1] Denominado también resolución de la ambigüedad léxica, discriminación del sentido de las palabras, selección del sentido de las palabras o identificación del sentido de las palabras.

dependiendo de la palabra en cuestión. Los sentidos están representados con etiquetas numéricas que codifican el sentido en WordNet.
El recurso más ampliamente usado para WSD, es la colección de entrenamiento. Una *colección de entrenamiento* es un conjunto de documentos etiquetados manualmente, con cada palabra acompañada del significado con que es utilizada. La colección de entrenamiento permite al sistema deducir en nuevos documentos los significados de las palabras. En nuestro trabajo hemos utilizado SemCor, dada su libre disposición y su utilización en trabajos relacionados, lo que nos ha facilitado la interpretación del comportamiento de nuestro sistema.

3.- MODELO DEL ESPACIO VECTORIAL PARA LA RESOLUCIÓN DE LA AMBIGÜEDAD LÉXICA

Nuestro modelo de desambiguación se basa en el modelo del espacio vectorial [Salton83]. El modelo del espacio vectorial ha sido utilizado ampliamente en muchos trabajos de recuperación de información [Lewis92] [Salton83,89], así como en otros de categorización de textos [Gómez97]. Basados en estas experiencias, presentamos una adaptación del modelo del espacio vectorial a la desambiguación de textos y los caminos seguidos para calcular algunos elementos del modelo.
Nosotros utilizamos el MEV para representar el lenguaje natural por medio de vectores de pesos. Cada peso representa la importancia de un término, en relación con un determinado sentido en la expresión del lenguaje natural. Cada término s_{ji} queda representado o indexado por un vector de dimensión *m*, con los pesos asignados a cada uno de los términos de indexación. El término *i* con sentido *j*, queda representado con el peso del término, así como con los pesos de los términos circundantes.

$s_{ji}=\langle ws_{j1}, ws_{k1}, ..., ws_{kn}\rangle$ $\quad ws_{kc}$ peso de la palabra circundante c al término s_{ji}

Para el procesamiento de los textos a desambiguar, se obtienen los términos de indexación aparecidos en ellos, de una forma análoga al de lo textos de la colección de entrenamiento. La representación de una consulta de un término c_k, se realiza mediante un vector de pesos asociados a los términos.

$c_k=\langle wc_1, wck_1, ..., wc_{kn}\rangle$ $\quad wc_{kc}$ peso de la palabra circundante c al término c_k

El MEV conlleva la suposición básica, de que la similitud semántica entre los objetos representados viene dada por el coseno del ángulo que forman sus vectores.
Para desambiguar un término c_k, calculamos la similitud entre el contexto en el que aparece y las definiciones de los términos. Se selecciona el término que presenta mayor similitud con el contexto, con arreglo a la fórmula:

$$sim(s_{ji}, c_i) = \frac{\sum_{i=1}^{m} ws_{ji} \cdot wc_i}{\sqrt{\sum_{i=1}^{m} ws_{ji}^{\,2} \cdot \sum_{i=1}^{m} wc_i^{\,2}}}$$

4.- ENTRENAMIENTO CON VENTANA CONTEXTUAL

Representamos cada término del córpora de entrenamiento, por medio de un vector, cuyas componentes son: el peso del término en el párrafo; y los pesos de los términos que constituyen lo

que hemos denominado *ventana contextual*. Con este concepto, hacemos referencia a las palabras que circundan al término a desambiguar, es decir, a las palabras que están en dicho contexto, ya que pueden suministrar información acerca del sentido utilizado.
Así, para cada uno de los nombres de la colección de entrenamiento, calcularemos su *ventana contextual*. El programa desplaza la ventana desde el principio de todos los documentos que contiene el córpora de entrenamiento, hasta el final de los mismos, considerando en cada desplazamiento un nombre, y como palabras de contexto, cada una de las palabras circundantes. De esta manera, se construyen tantas ventanas contextuales como términos con diferentes sentidos existan en la colección. El tamaño de la ventana contextual, será variable y estará en función del número de términos que contenga el párrafo en cuestión.
Una vez realizado esto, se construyen los vectores para la colección de entrenamiento y se calculan los pesos para los distintos términos de manera análoga a Salton [Salton83]:

$$ws_{ji} = t_{ji} * w_i \qquad w_i = \log_2(n/f_i)$$

Donde t_{ji} es la frecuencia del término j con sentido i en la ventana contextual, *n* es el número de sentidos de término i y f_i es el número de ventanas contextuales donde aparece el término *i*. La dimensión del espacio vectorial es variable y está en función del número de sentidos que tenga la palabra a desambiguar.
Se suman los vectores que representan el mismo término y el mismo sentido, dada la susceptibilidad de repetición de algunos de ellos, para que cada término i con sentido j quede representado por un solo vector. Esto se realiza como consecuencia de la construcción de los vectores s_{ji}, para cada uno de los nombres i con sentido j que tiene el corpus.

5.- DESAMBIGUACIÓN DEL SENTIDO DE LAS PALABRAS

Hemos utilizado un algoritmo basado en el aprendizaje, fundamentado en el modelo del espacio vectorial. La entrada a nuestro programa de desambiguación consta de una colección de textos de prueba. En la salida, a cada palabra se le asigna el significado correcto (inferido por el contexto) etiquetado según WordNet.
La técnica utilizada para la resolución automática de la ambigüedad léxica, se enmarca dentro de la adquisición, con representaciones contextuales, del significado de las palabras. Una representación contextual es una caracterización del contexto lingüístico en el que una palabra expresa un determinado sentido [Miller91]. Por tanto, nuestro algoritmo trata de encontrar y representar características contextuales, a través, como ya se ha comentado, de ventanas contextuales de tamaño variable. En nuestros experimentos hemos considerado el párrafo, tal y como está definido en SemCor, como ventana contextual.
Con la consulta (que constituye el córpora de prueba a desambiguar) actuamos análogamente. Primero, calculamos las ventanas contextuales de la colección de prueba y, construyendo después los vectores, calculamos la similitud. Ésta nos proporcionará el sentido correcto en relación con el término a desambiguar de la colección de prueba.

6.- EVALUACIÓN

La evaluación de sistemas WSD ha supuesto un problema en la investigación de la desambiguación, de hecho, algunos desambiguadores han sido evaluados sólo a través de pruebas manuales sobre un grupo reducido de palabras.

La evaluación es muy heterogénea. En numerosos trabajos se han utilizado diferentes métricas y colecciones de prueba. Hemos adoptado las basadas en el campo de la evaluación de los sistemas de recuperación de información [Salton83] [Frakes92], así como una colección de prueba de amplia cobertura y libre disposición para nuestro trabajo.
Para evaluar el rendimiento de nuestro algoritmo WSD, hemos realizado un test sobre un conjunto de documentos, que será comparado con las pruebas realizadas sobre un algoritmo *línea base*. Éste asigna al término a desambiguar el sentido más frecuente en el córpus [Boguraev96].

6.1. Métricas de evaluación

Hemos utilizado la precisión como métrica básica para computar la efectividad de nuestros experimentos. El cálculo puede ser realizado utilizando *macroaveraging* y *microaveraging* [Lewis92].
La Precisión puede ser definida como el cociente entre el número de términos desambiguados satisfactoriamente y el número de términos desambiguados.
El macroaveraging consiste en calcular la precisión para cada uno de los términos, y luego calcular la media para cada uno de ellos, como sigue:

$$P_{macroavg} = \frac{\sum_{i=1}^{n} P_i}{n} \qquad P_i = \text{Precisión del término } i \qquad P_i = \frac{dc_i}{dc_i + di_i}$$

Donde dc_i es el número de desambiguaciones correctas del término i, di_i el número de desambiguaciones incorrectas del término i y n el número de términos desambiguados.
Por otro lado, el microaveraging consiste en calcular un sólo valor de precisión medio para todos los términos, según:

$$P_{microavg} = \frac{tdc}{tdc + tdi}$$

Siendo *tdc* el número de términos desambiguados correctamente y *tdi* el número de términos desambiguados incorrectamente.

6.2. Colección de prueba

Para nuestros experimentos, como ya se ha comentado, hemos utilizado SemCor [Miller93], que además de un corpus de texto, como ya se ha indicado, es un lexicón, donde cada palabra en el texto hace referencia a su correcto significado en él. Puede definirse bien como un corpus en el que las palabras han sido etiquetadas sintáctica y semánticamente, o como un lexicón en el cual las frases de ejemplo pueden ser encontradas por varias definiciones. SemCor es el Brown Corpus donde sólo los nombres, verbos, adjetivos y adverbios son etiquetados semánticamente con los sentidos de WordNet. Las palabras etiquetadas no tienen punteros a WordNet. Las palabras (tales como preposiciones, determinantes, pronombres, verbos auxiliares, etc.) y caracteres no alfanuméricos, interjecciones y términos coloquiales no son etiquetados.
Hay un total de 103 ficheros de texto en SemCor, constituyendo un total de 11.628 frases y un total de palabras de 229.370 distribuidas como se resume en la Tabla 1.
Al ser la evaluación más completa que conocemos la de Agirre [Agirre96], hemos seguido un procedimiento análogo para nuestros experimentos, con vistas a facilitar la comparación. De esta manera, hemos tomado como colección de prueba, un subconjunto de ficheros de SemCor,

compuesto por cuatros documentos seleccionados aleatoriamente. Un ejemplo de un fragmento de un fichero de SemCor se muestra en la Figura 1. Después de borrar de los ficheros fuentes de SemCor la información no relevante (etiquetas y marcas SGML, y palabras ignoradas), obtenemos las palabras como se muestra en la
Figura 2.

```
<contextfile concordance=brown>
<context filename=br-k18 paras=yes>
<p pnum=1>
<s snum=1>
<wf cmd=ignore pos=PRP>She</wf>
<wf cmd=done pos=VBD ot=notag>was</wf>
<wf cmd=done pos=VB lemma=get wnsn=2 lexsn=2:30:00::>getting</wf>
<wf cmd=done pos=RB lemma=real wnsn=1 lexsn=4:02:00::>real</wf>
<wf cmd=done pos=JJ lemma=dramatic wnsn=1 lexsn=3:00:00::>dramatic</wf>
<punc>.</punc>
</s>
<s snum=2>
<wf cmd=ignore pos=PRP>I</wf>
<wf cmd=ignore pos=MD>'d</wf>
<wf cmd=done pos=VB ot=notag>have</wf>
<wf cmd=done pos=VBN ot=notag>been</wf>
<wf cmd=done pos=RB lemma=more wnsn=1 lexsn=4:02:00::>more</wf>
<wf cmd=done pos=VB lemma=impress wnsn=1 lexsn=2:37:00::>impressed</wf>
<wf cmd=ignore pos=IN>if</wf>
<wf cmd=ignore pos=PRP>I</wf>
<wf cmd=done pos=VBD ot=notag>had</wf>
<wf cmd=done pos=RB lemma=n't wnsn=1 lexsn=4:02:00::>n't</wf>
<wf cmd=done pos=VB lemma=remember wnsn=1
lexsn=2:31:00::>remembered</wf>
<wf cmd=ignore pos=IN>that</wf>
<wf cmd=ignore pos=PRP>she</wf>
<wf cmd=ignore pos=MD>'d</wf>
<wf cmd=done pos=VB lemma=play wnsn=2 lexsn=2:36:02::>played</wf>
<wf cmd=done rdf=person pos=NNP lemma=person wnsn=1 lexsn=1:03:00::
pn=person>Hedda_Gabler</wf>
<wf cmd=ignore pos=IN>in</wf>
<wf cmd=ignore pos=PRP$>her</wf>
<wf cmd=done pos=NN lemma=highschool wnsn=1
lexsn=1:14:00::>highschool</wf>
<wf cmd=done pos=NN lemma=dramatics wnsn=1 lexsn=1:04:00::>dramatics</wf>
<wf cmd=done pos=NN lemma=course wnsn=1 lexsn=1:04:01::>course</wf>
<punc>.</punc>
</s>
```

Figura 1. Fragmento del fichero br-k18 de SemCor

was getting real dramatic. have been more impressed hadn't remembered played Hedda_Gabler highschool dramatics course.

Figura 2. Fragmento del fichero br-k18 de SemCor utilizado para la prueba

Total de Palabras	
Total palabras etiquetadas	105092
Total palabras no-etiquetadas	124278
	229370

Total frases	11628
Total símbolos no-alfanuméricos	19424
Total nombres propios	4786

Diferentes etiquetas semánticas	
Nombres	11174
Verbos	5662
Adjetivos	5028
Adverbios	1436
	23300

Tabla 1**.** Estadística de SemCor**.**

El algoritmo produce un fichero de resultados, para cada uno de los documentos seleccionados aleatoriamente, con los sentidos inferidos para que puedan ser comparados automáticamente con los ficheros originales.

6.3. Resultados e interpretación

Para nuestros experimentos hemos seleccionado aleatoriamente cuatro documentos o textos de SemCor considerados individualmente: br-a14, br-j09, br-k11 y br-k14. Estos textos han representado el papel de ficheros de entrada (sin etiquetas).

Los resultados presentados aquí pueden parecer pobres comparados con otros trabajos relevantes, ya que muchos de estos últimos, se centran en seleccionar un conjunto reducido de palabras, generalmente con un par de sentidos de muy diferentes significados, para los cuales su algoritmo muestra gran precisión (lo que hace la comparación dificultosa). Por el contrario nosotros probamos nuestro algoritmo con **todos** los nombres en un subconjunto del córpora no restringido de dominio público, haciendo distinción entre el gran número de sentidos de WordNet.

Para comparar nuestro algoritmo, como hemos comentado anteriormente, hemos decidido implementar un algoritmo línea base [Boguraev95], y confrontarlo con nuestros experimentos. La precisión obtenida por nuestro algoritmo es alta, como se puede observar en la Tabla 2, donde se encuentran resumidos los resultados para nuestra primera serie de experimentos. Esta tabla muestra las medias macroaveraging y microaveraging para la precisión.

Algoritmos WSD	*Macroaveraging*	*Microaveraging*
	Precisión	
Basado MEV	83%	78,6%
Línea Base	81%	76%

Tabla 2. Resultados totales de nuestros experimentos.

La Figura 4 representa la relación entre la precisión y el número de sentidos para ambos algoritmos. Se puede observar, que la precisión utilizando nuestro algoritmo basado en el MEV, aumenta comparado con el algoritmo línea base.

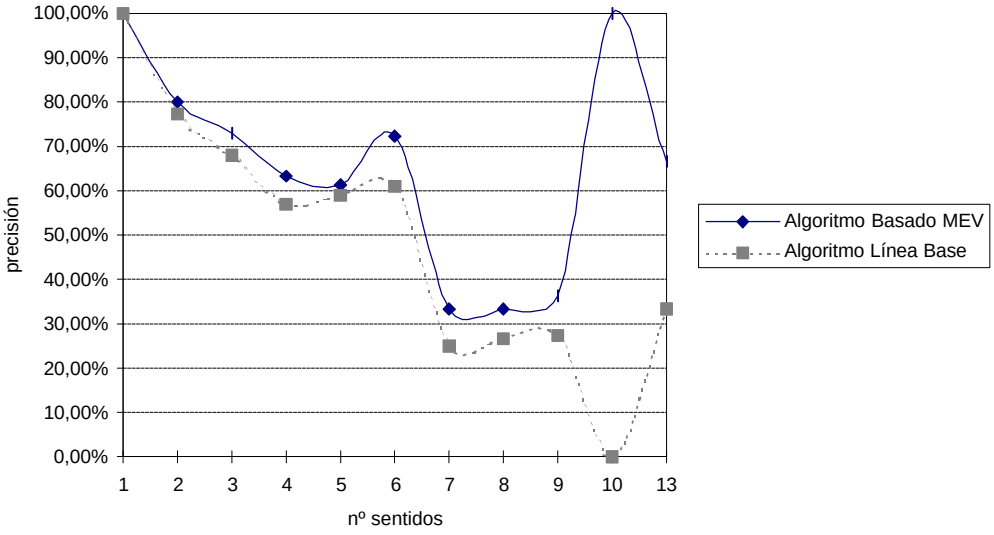

Figura 3. Representación de la precisión de nuestro algoritmo basado en el MEV y el algoritmo línea base en relación con el número de sentidos.

Hay que destacar, que el incremento en la precisión de nuestro algoritmo en relación con aquellos términos que tienen de diez y trece sentidos representados en abscisas, se produce como consecuencia de la menor frecuencia de ocurrencias de términos con muchos sentidos. Así, utilizando los ficheros seleccionados aleatoriamente para nuestros experimentos, podemos representar la distribución de la frecuencia de las ocurrencias en relación con el número de sentidos, como ilustra la Figura 4.

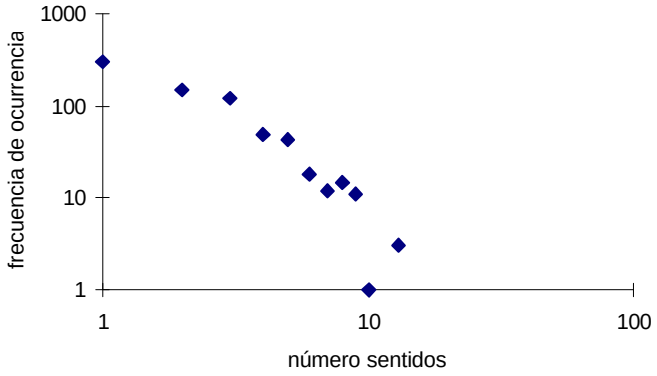

Figura 4. Distribución de la frecuencia de las ocurrencias en relación con el número de sentidos. La gráfica ha sido representada sobre una escala logarítmica.

Las palabras polisémicas, evidentemente, son más significativas que las monosémicas en relación con la precisión. La precisión para las palabras con pluralidad de significados puede ser observada en la Tabla 3. La precisión utilizando macroaveraging de nuestro algoritmo basado en el MEV para nombres polisémicos con los sentidos de WordNet es del orden del 69%, frente al algoritmo con el que se compara que tiene una precisión del 65%. Y utilizando microaveraging es del orden del 64% para nuestro algoritmo y del 59% para el línea base.

Polisémicas		
Algoritmos WSD	*Macroaveraging*	*Microaveraging*
	Precisión	
Basado en MEV	69%	64%
Línea Base	65%	59%

Tabla 3**.** Resultados de nuestros experimentos para términos polisémicos.

Se pueden extraer de nuestros experimentos, términos que tienen gran número de sentidos en SemCor, como se resume en la Tabla 4 (el número medio de sentidos de las palabras es de 8), y observar la gran precisión que obtiene nuestro algoritmo frente al línea base. Hemos seleccionado aleatoriamente palabras con distinto número de sentidos (con más de tres). Así por ejemplo, nuestro algoritmo obtiene una precisión del 100% frente al 0% del línea base, para la palabra *field* cuyo conjunto de sinónimos (*synsets)* asociados a los sentidos de WordNet es 15.

Palabra	*Precisión Ventana Contextual*	*Precision línea base*
credit	85%	0%
formation	83%	25%
title	100%	50%
time	66%	33%
study	0%	25%
system	100%	50%
field	100%	0%
thing	100%	0%

Tabla 4**.** Precisión de términos con varios sentidos que son usados en la desambiguación.

7.- CONCLUSIONES Y FUTUROS TRABAJOS

En este trabajo hemos presentado un nuevo enfoque para desambiguar el sentido de las palabras de SemCor, combinando el modelo del espacio vectorial con un algoritmo basado en el aprendizaje. Hemos utilizado el Modelo del Espacio Vectorial, ampliamente usado en recuperación de información, y el concepto de ventana contextual de tamaño variable. Hemos obtenido buenos resultados en la resolución de la ambigüedad de términos polisémicos, como son los términos de SemCor etiquetados con los sentidos de WordNet.

El algoritmo está teóricamente fundamentado y ofrece una medida general de la relatividad semántica para algún número de nombres en un texto. En el experimento, el algoritmo

desambiguó cuatro textos de SemCor, un subconjunto del Brown Corpus. Los resultados fueron obtenidos automáticamente comparando las etiquetas de SemCor con los calculados por el algoritmo, lo cual permite la comparación con otros métodos de desambiguación.
La evaluación del comportamiento del algoritmo proporciona resultados positivos, a pesar de la dificultad de la tarea y de la gran variedad de sentidos definidos en WordNet por palabra.
Actualmente, se está estudiando la influencia del tamaño de la ventana contextual en relación con la precisión. Por otro lado, nuestro enfoque puede facilitar en el futuro, la integración de varios recursos léxicos, como diccionarios y córporas de entrenamiento. Creemos que combinando las definiciones de los términos del diccionario con las ventanas contextuales construidas a partir de córporas de entrenamiento, obtendremos mejores resultados.

REFERENCIAS.

Agirre E., Rigau G. (1996): *Word sense disambiguation using conceptual density.* In Proceedings of COLING 1996.

Boguraev B., Pustejovsky J.(1996): *Corpus processing for lexical acquisition*, A Bradford Book, Language, Speech and Communication Series, The MIT Press.

Bruce R., Wiebe J. (1994): *Word sense disambiguation using decomposable models*. In Proceedings of the 32th Annual Meeting of the Association for Computational Linguistics, (ACL'94).

Frakes, W., Baeza, R. (1992): *Information retrieval: data structures and algorithms*, Prentice Hall, London.

Gale W., Church KW., Yarowsky D.(1992): *Estimating upper and lower bounds on the performance or word-sense disambiguation programs*. In Proceedings of the ACL92.

Gómez J.M., Buenaga De M. (1997): *Integrating a lexical database a training collection for text categorization.* ACL/EACL Workshop on Automatic Extraction and Building of Lexical Semantic Resources for Natural Language Applications.

Hwee Tou Ng, Hian Beng Lee (1996): *Integrating multiple knowledge sources to disambiguate word sense: an exemplar-based.* In Proceedings of ACL96.

Lesk M. (1986): *Automatic sense disambiguation: how to tell a pine cone from an ice cream cone.* In Proceedings of the SIGDOC Conference.

Lewis, D. (1992): *Representation and learning in information retrieval*. Ph.D. Thesis, Department of Computer and Information Science, University of Massachusetts.

Miller G. (1990): *WordNet: An on-line lexical database*. An International Journal of Lexicography.

Miller G., Charles G. (1991): *Contextual correlates of semantic similarity*. Language and Cognitive Processes.

Miller G. Leacock C., Randee T. and Bunker R. (1993): *A Semantic concordance. In Proceedings of the 3rd DARPA Workshop on Human Language Technology*, New Jersey 1993.

Miller G. Chodorow M., Landes S., Leacock C. and Thomas R. (1994): *Using a semantic concordance for sense identification.* In Proceedings of the ARPA Human Language Technology.

Rigau G., Atserias J., Agirre E. (1997): *Combining unsupervised lexical knowledge methods for word sense disambiguation.* In Proceedings of ACL'97.

Salton G., McGill, M.J. (1983): *Introduction to modern information retrieval*. McGraw-Hill.

Salton, G., (1989): *Automatic Text Processing: the transformation, analysis and retrieval of information by computer*. Addison Wesley.

Sanderson, M. (1996): *Word sense disambiguation and information retrieval*. Ph.D. Thesis, Department of Computing Science, University of University of Glasgow.

Wilks Y., Fass D., Guo C., McDonald J., Plate T., Slator B. (1993): *Providing machine tractable dictionary tools*. Machine Translation.

Yarowsky D.(1992): W*ord-sense disambiguation using statistical models of Roget's categories trained on large corpora*. In Proceedings of the 15th International Conference on Computational Linguistics.

Yarowsky D. (1994): *Decision lists for lexical ambiguity resolution: Aplication to accent restoration in Spanish and French*. In Proceedings of the 32th Annual Meeting of the Association for Computational Linguistics, (ACL'94).

Yarowsky D. (1995) *Unsupervised word sense disambiguation rivaling supervised methods*. In Proceedings of the 33th Annual Meeting of the Association for Computational Linguistics, (ACL'95).

Yoshiky N., Yoshihiko Nitta (1994): *Co-ocurrence vectors from corpora vs. distance vectors from.* In Proceedings of COLING94.