Utilizando WordNet para Complementar la Información de Entrenamiento en la Identificación del Significado de las Palabras

L. Alfonso Ureña López[1], Manuel De Buenaga Rodríguez[2]
[1]Departamento de Informática. Universidad de Jaén
Avda. Madrid 35, 23071 Jaén. Spain
e-mail: laurena@ujaen.es
[2]Departamento de Inteligencia Artificial. Universidad Europea de Madrid
C/ Tajo s/n, Villaviciosa de Odón. 28670 Madrid. Spain
e-mail: buenaga@dinar.esi.uem.es

Resumen

La desambiguación del significado de las palabras se ha desarrollado como una subárea del Procesamiento del Lenguaje Natural (PLN), donde el objetivo es determinar el sentido correcto de aquellas palabras que tienen más de un significado, no es una tarea final en sí misma, sino una tarea intermedia necesaria en variadas aplicaciones del procesamiento del lenguaje natural. La resolución de la ambigüedad de las palabras (WSD) es identificar el sentido correcto de los relacionados en un diccionario, una base de datos léxica o similar. Es una tarea compleja, pero muy útil en variadas aplicaciones del procesamiento en lenguaje natural, como Categorización de Texto (TC); traducción automática; restauración de acentos; encaminamiento y filtrado de textos; agrupamiento y segmentación de textos, corrección ortográfica y gramatical, reconocimiento de voz y, en general, en la recuperación de información.

Nuestro enfoque integra información de una base de datos léxica (WordNet) con dos enfoques de entrenamiento a través del Modelo del Espacio Vectorial, incrementando la efectividad de la desambiguación. Probamos los enfoques de entrenamiento con los algoritmos de Rocchio y Widrow-Hoff sobre un gran conjunto de documentos con una fina granularidad de sentidos, como son los de WordNet. Consiguiendo una alta precisión en la resolución de la ambigüedad léxica, así como una gran efectividad en su ejecución.

Palabras clave: Word Sense Disambiguation (WSD), Text Categorization (TC), WordNet, SemCor, Information Retrieval (IR), Machine Translation (MT), Contextual Windows (CW).

1 Introducción

Un problema importante del Procesamiento del Lenguaje Natural (PLN) es determinar el significado de una palabra en un contexto particular. Las diferentes acepciones de una palabra son recogidas como varios sentidos en un diccionario. La tarea de desambiguación del sentido de las palabras (WSD) es identificar el sentido correcto de una palabra en un contexto. Esta tarea es compleja, pero muy útil en variadas aplicaciones del procesamiento en lenguaje natural [Kilgarriff 97a], como Categorización de Texto (TC) [Buenaga 97]; traducción automática [Brown 91]; restauración de acentos [Yarowsky 94]; encaminamiento y filtrado de textos; agrupamiento y segmentación de textos y, en general, en la recuperación de información [Kilgarriff 97a, 97b, Sanderson 96].

Presentamos un nuevo enfoque automático para WSD basado en el uso de varios recursos léxicos. Actualmente, muchos recursos, como colecciones de entrenamiento [Yarowsky 92, Yoshiki 94, Ureña 97] y bases de datos léxicas [Resnik 95, Agirre 96, Xiaobin 95] o thesaurus

Architecture Tagger WSD

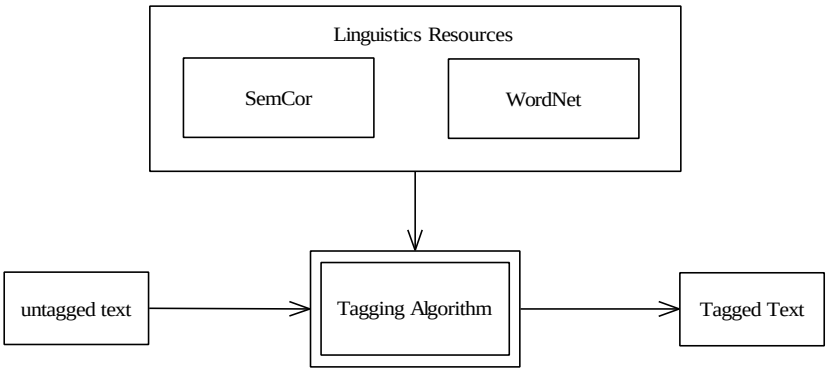

Ilustración 1. Modelo WSD utilizado

[Yarowsky 92], han sido satisfactoriamente empleados para la resolución de la ambigüedad léxica sobre todo de manera aislada. Creemos que la idea clave para la mejora de la desambiguación es incrementar la cantidad de información de la que hace uso el sistema. Planteamos un modelo integrador para WSD, empleando de manera combinada corpus de entrenamiento y bases de datos léxicas. Este modelo combinado consigue mejorar la precisión de los basados en recursos léxicos aislados.

Entre los muchos enfoques de entrenamiento que han sido bien empleados, hemos seleccionado los algoritmos de Rocchio y Widrow-Hoff. Hemos combinado la utilización de cada algoritmo con WordNet, utilizando el Modelo del Espacio Vectorial (MEV) para esta tarea y utilizamos el concepto de *Ventana Contextual* de tamaño variable [Ureña 98]. Los vectores de pesos son calculados para cada ventana contextual, empleando la base de datos léxica WordNet y el subconjunto de entrenamiento SemCor. Hemos calculado el vector de pesos para: un enfoque basado solamente en WordNet, otro basado sólo en la Colección de entrenamiento y uno basado en la integración de los recursos ambos recursos.

Comparamos la similitud término-sentido, eligiendo el significado de mayor similitud, estudiando el ángulo que forman los vectores. Hemos realizado una serie de experimentos sobre un conjunto de prueba de la colección de evaluación SemCor, mostrando que un enfoque combinado puede mejorar la efectividad de la resolución de la ambigüedad léxica.

El objetivo de un sistema desambiguador, dado un conjunto de documentos, es decidir el significado correcto de los nombres y verbos que componen los documentos. El sistema hace uso de la información contenida en los textos para computar el grado de pertenencia del término a cada significado.

El recurso más ampliamente utilizado para WSD es la colección de entrenamiento. Una colección de entrenamiento es un conjunto de documentos con los significados etiquetados manualmente, que permite al sistema asignar los significados a nuevos documentos, de acuerdo con su similitud a otros documentos de la colección de entrenamiento. Actualmente, son varios los corpus de los que pueden ser obtenidos un conjunto de entrenamiento y otro de prueba. Hemos seleccionado SemCor por su amplia utilización y disponibilidad, lo que facilita la comparación de resultados.

Una base de datos léxica[1] es un sistema con información léxica de uno o varios lenguajes. Hemos elegido WordNet dada su libre distribución, amplia cobertura y frecuencia de uso. Proponemos la integración de bases de datos léxicas y colecciones de entrenamiento para mejorar la efectividad del proceso WSD (Ver Ilustración 1).

Este trabajo está organizado como sigue. En primer lugar, introducimos la tarea de desambiguación y los recursos utilizados. Seguidamente, describimos el modelo en el que estos elementos son integrados, y examinamos ambos enfoques y la integración de los dos recursos. Después de esto, presentamos nuestra evaluación y los resultados obtenidos, y finalmente, describimos nuestras conclusiones y trabajos futuros.

[1] Desde este punto de vista, los diccionarios electrónicos pueden ser considerados como bases de datos léxicas. Actualmente, las bases de datos léxicas incluyen WordNet, EDR y Roget's Thesaurus.

2 El Modelo del Espacio Vectorial para la Desambiguación de Términos

El Modelo del Espacio Vectorial (MEV) [Salton 83] fue originalmente desarrollado para la Recuperación de Información, pero provee un soporte muy adecuado para realizar otras tareas como WSD o TC. También, el modelo está avalado por muchas experiencias en recuperación de texto [Lewis 92,Salton 89]. De hecho, el MEV es un entorno muy adecuado para expresar nuestro enfoque de WSD, pues permite la integración de múltiples fuentes de conocimiento para la desambiguación, y hace más fácil identificar el papel de cada fuente de conocimiento involucrada en la operación de desambiguación.

La clave del MEV para la Recuperación de Información (IR) es representar las expresiones del lenguaje natural mediante vectores de pesos. Cada peso representa la importancia de un término, en relación con un determinado sentido en la expresión del lenguaje natural. Una hipótesis fundamental en WSD es que cada palabra se utiliza con un único significado en un contexto concreto [Yarowsky 93]. 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*, (s_{ji}) queda representado con el peso del término, así como con los pesos de los términos circundantes. El vector de peso es $\langle ws_{j1}, ws_{k1}, ..., ws_{kn} \rangle$ donde ws_{kc} representa el 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 los textos de la colección de entrenamiento. La representación de una consulta de un término ck, se realiza mediante un vector de pesos asociados a los términos. El vector es $\langle wc_1, wck_1, ..., wc_{kn} \rangle$ donde wc_{kc} de manera análoga representa el peso de la palabra circundante c al término c_k.

La similitud semántica entre el término i con sentido j y el término viene dada por el coseno del ángulo que forman sus vectores, con arreglo a la formula:

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

Calculamos los pesos para los distintos términos de manera análoga a [Salton83] $ws_{ji}=t_{ji}*\log_2(n/f_i)$. 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*.

3 Utilización de una Colección de Entrenamiento para representar significados

La idea básica en un enfoque basado en entrenamiento para la tarea WSD, es que un conjunto de términos manualmente etiquetados (conjunto de documentos) con el sentido correcto pueden ser usados para predecir el significado de nuevos términos. Hemos utilizado SemCor[2] por estar etiquetado semánticamente con los significados de WordNet y ser dominio público.

La hipótesis clave cuando utilizamos una colección de entrenamiento para la resolución de la ambigüedad léxica es que un término aparece con un particular sentido en un determinado contexto. Los términos que constituyen ese contexto pueden ser buenos para predecir el sentido con que aparece el término. El conjunto de predictores del significado de un término, y su importancia, son computados estadísticamente por las ventanas contextuales [Ureña 98], como un paso inicial del proceso de entrenamiento. Para ello, se representa cada término del corpus de entrenamiento con un vector, cuyas componentes son: el peso del término en el párrafo y los pesos de los términos que constituyen la ventana contextual. Así, para cada uno de los nombres de la colección de entrenamiento, calcularemos su ventana contextual, construyendo tantas como palabras con diferentes sentidos existan en la colección.

[2] SemCor es un subconjunto del Brown Corpus. La colección es heterogénea cubriendo temas políticos, deportivos, musicales, etc., Sin embargo, SemCor no es banco de pruebas óptimo para la resolución de la ambigüedad, debido fundamentalmente a la fina granularidad de significados que utiliza.

Los algoritmos de entrenamiento proveen una manera de calcular los vectores de pesos para los distintos significados de una palabra. Básicamente, el proceso de entrenamiento asigna un peso a un término en un vector clase, en proporción al número de ocurrencias del termino con un determinado significado y proporción a la importancia del término en la colección.

Hemos elegido los algoritmos de Rocchio [Rocchio 71] y Widrow-Hoff [Widrow 85] para computar los pesos de los términos para un determinado significado en nuestro enfoque, como se muestra a continuación. Ambos dan la oportunidad de integración a través de una representación inicial computada por la utilización de un recurso externo como WordNet [Buenaga 97].

3.1 Algoritmo de Rocchio

El algoritmo de Rocchio produce un nuevo vector de pesos wc_k de uno existente wc^0_k y una colección de documentos de entrenamientos. El componente i del vector wc_k es calculado por la fórmula:

$$wc_{ik} = a wc^0_{ik} + b \frac{\sum_{l \in C_k} wd_{il}}{n_k} + g \frac{\sum_{l \notin C_k} wd_{il}}{P - n_k}$$

Donde wc_{0k} es el peso inicial del término i para el significado k, wt_{il} es el peso del término i para el item l de entrenamiento, C_k conjunto de índices de items asignados al significado k, y n_k el número de estos items. Los parámetros α, β y γ controlan el relativo impacto de los pesos inicial, positivo y negativo respectivamente en el nuevo vector.

Como Lewis [Lewis 96], hemos usado los valores $\beta = 16$ y $\gamma = 4$. El valor de α se establece a 20, para equilibrar la importancia de los pesos iniciales y de entrenamiento. Restringimos el clasificador para no hacer uso de pesos negativos, así al final el peso wc_{ik} será positivo, o retornará a 0 si es negativo.

El vector inicial wc^0_k es tomado frecuentemente como vector nulo, pero esto puede ser instanciado con un conjunto de pesos iniciales calculados por la utilización de un recurso externo. En la siguiente sección, veremos como se hace esto empleando WordNet.

3.2 Algoritmo de Widrow-Hoff

El algoritmo de Widrow-Hoff comienza con un vector de pesos existente wc^0_k y secuencialmente se va actualizando una vez para cada item de entrenamiento. El componente i del vector wc^{l+1}_k es obtenido del lth item y del the lth vector por la fórmula:

$$wc^{l+1}_{ik} = wc^l_{ik} + 2h(wd_l \cdot wc^l_k - y_l)wd_{il}$$

Donde wc^l_{ik} es el peso del término i en el lth vector para la clase k, wt_{il} es el vector de pesos del término i para el item l, wc_{ik} es el lth vector para la clase k, y_l es 1 si el lth item es asignado a la clase k y 0 en otro caso, y wt_{il} es el peso del término i en el lth item. La constante η es ratio de aprendizaje, el cual controla cómo de rápido le está permitido cambiar el vector de pesos y cuanto influye cada nuevo item sobre éste. Un valor típicamente usado para η is $1/4X^2$, siendo X el valor máximo de los vectores que representan los items de entrenamiento.

Como en el algoritmo de Rocchio, un vector inicial de pesos se puede producir utilizando un recurso independiente. Sin embargo, la importancia de este peso se reduce proporcionalmente al número de items de entrenamiento disponibles para una clase. Cuando hay muchos ejemplos de entrenamiento, el peso inicial es dominado por el peso obtenido de estos ejemplos. Sin embargo cuando hay peso inicial tiende a mantener sus valores.

4 Usando una base de datos léxica para complementar la información de entrenamiento

Las bases de datos léxicas contienen muchos tipos de información (conceptos, sinónimos y otras relaciones léxicas, hiponimia y otras correspondencias conceptuales, etc.). WordNet [Miller 95] es un lexicón que representa conceptos como conjuntos de sinónimos, o *synsets* (elementos básicos de WordNet*).*

El sentido o significado no es un concepto bien definido, ofreciendo frecuentemente finas distinciones dependiendo de la colocación, contexto, etc. Para nuestro propósito consideramos los sentidos de las palabras presentes en WordNet. Y, seleccionamos la información de sinonimia como "categorías de sentidos", de esta manera, un término es consultado en WordNet, donde se obtiene la información de sus *synsets* o *conceptos* asociados.

Cada *synset* se trata como un sentido distinto para cada término, del que se obtiene un conjunto de palabras sinónimas para cada término. Se construyen ventanas contextuales con cada *synset* para cada uno de los términos. De esta manera, un término, en un determinado contexto, puede desambiguarse calculando la similitud entre dicho término y las ventanas contextuales asociadas a cada uno de sus *synsets* en WordNet.

4.1 Integración de SemCor y WordNet

La efectividad de un sistema debe mejorar si está mejor informado. Partiendo de esta hipótesis, incorporamos información proveniente de WordNet a la colección de entrenamiento. Con lo que la efectividad de la desambiguación mejora.

La integración se ha realizado, como sigue. En la fase de entrenamiento se construyen primeramente los vectores conforme a lo relatado en el enfoque basado en el entrenamiento, obteniendo un conjunto de vectores, uno por cada uno de los términos que constituyen la colección de entrenamiento. A continuación, cada uno de los términos, que representan a los vectores, se consulta en WordNet, si dicho término tiene un *synset* asociado para el sentido consultado se "une" dicho *synset* al vector, recalculándolo con mayor peso, en caso contrario se elimina dicho *synset*. La fase de prueba se realiza confrontando la consulta, por medio del cálculo de la similitud, con los vectores creados en el entrenamiento.

Esta técnica de integración identifica claramente el papel de cada recurso en este enfoque de desambiguación. Por un lado WordNet proporciona información relativa a la relación de sinonimia, ampliando el número de términos en relación con un determinado sentido, cuando los datos de entrenamiento no son grandes o no son seguros. Esto directamente contribuye con los términos usados en la representación del vector. Por otro lado, la colección de entrenamiento proporciona mayor información contextual para aquellos términos mejor entrenados.

Para el algoritmo de Rocchio, nosotros hemos considerado el valor de proximidad semántica previamente producido cómo un número de ocurrencias de un término con un significado, así este valor es multiplicado por el peso de el término en la colección. Por otro lado, la inserción del peso de un término para un significado en el algoritmo de Widrow-Hoff es normalizado por la constante η.

5 Evaluación

La evaluación de la tarea WSD es muy heterogénea. Se han utilizado varias métricas y colecciones de prueba en distintos trabajos con variados enfoques. Esto ha producido un problema en lo que se refiere a la comparación de resultados entre distintos enfoques. Para minimizar este problema, hemos seleccionado para nuestro trabajo, un conjunto de métricas muy extendidas y frecuentemente utilizadas en el campo de la evaluación de los sistemas de recuperación de información [Salton 83, Frakes 92].

```
<contextfile concordance=brown>
<context filename=br-a01 paras=yes>
<p pnum=1>
<s snum=1>
<wf cmd=ignore pos=DT>The</wf>
<wf cmd=done rdf=group pos=NNP lemma=group wnsn=1
lexsn=1:03:00:: pn=group>Fulton_County_Grand_Jury</wf>
<wf cmd=done pos=VB lemma=say wnsn=1
lexsn=2:32:00::>said</wf>
<wf cmd=done pos=NN lemma=friday wnsn=1
lexsn=1:28:00::>Friday</wf>
<wf cmd=ignore pos=DT>an</wf>
<wf cmd=done pos=NN lemma=investigation wnsn=1
lexsn=1:09:00::>investigation</wf>
....
....
<wf cmd=ignore pos=DT>any</wf>
<wf cmd=done pos=NN lemma=irregularity wnsn=1
lexsn=1:04:00::>irregularities</wf>
<wf cmd=done pos=VB lemma=take_place wnsn=1
lexsn=2:30:00::>took_place</wf>
<punc>.</punc>
</s>
</p>
```

Ilustración 2. Fragmento de un documento de SemCor

5.1 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* [Lewis 92].

La Precisión puede ser definida como el cociente entre el número de términos desambiguados satisfactoriamente y el número de términos desambiguados.

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

Ilustración 3. Fragmento de un documento de SemCor sin marcas SGML

La precisión *macroaveraging* consiste en calcular la precisión para cada uno de los términos, y luego calcular la media para cada uno de ellos, como sigue:

$$P_{macroavg} = \frac{\sum_{i=1}^{n} P_i}{n}; \qquad P_i = \frac{dc_i}{dc_i + di_i}$$

P_i = Precisión del término 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, la precisión *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.

5.2 Colección de prueba

Como colección de prueba hemos tomado un subconjunto de ficheros de SemCor seleccionados aleatoriamente. SemCor [Miller 93] consta de un total de 103 ficheros de texto, como se resume en la estadística de la Tabla 1. Además de un corpus de texto, SemCor 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 que las frases de ejemplo pueden ser encontradas por varias definiciones. SemCor abarca el Brown Corpus donde sólo los nombres, verbos, adjetivos y adverbios son etiquetados semánticamente con los sentidos de WordNet. Las palabras (tales como preposiciones, determinantes, pronombres, verbos auxiliares, etc.) y caracteres no alfanuméricos, interjecciones y términos coloquiales no son etiquetados.

Para la evaluación hemos tomado como colección de prueba, un subconjunto de ficheros de SemCor, seleccionados aleatoriamente. Un ejemplo de un fragmento de un fichero de SemCor se muestra en la Illustracion 2. 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 Ilustración 3.

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.

		Subcolección		
		Entrenamiento	Prueba	Total
Ficheros	Número	68	35	103
Párrafos	Número	1934	1122	3056
Palabras etiquetadas semánticamente	Occurrencias	67363	33682	101045
	Párraf. Media.	34.83	30.02	33.06
Sentidos (Nombre)	Occurrencias	41483	1786	43269
	Párraf. Media	14.30	11.45	14.16

Tabla 2. Resultados de nuestros experimentos

Precisión	**Recursos Léxicos**					
	WordNet		***SemCor***		***WordNet+SemCor***	
	Rocch.	**WHoff.**	**Rocch.**	**WHoff.**	**Rocch.**	**WHoff.**
Microaveraging	55.4%	54.7%	78.6%	77.1%	81.3%	80.5%
Macroaveraging	57.7%	57.2%	83.0%	83.3%	85.1%	84.9%

Tabla 2. Resultados de nuestros experimentos

5.3 Resultados e interpretación

Para nuestros experimentos hemos seleccionado aleatoriamente cuatro documentos de SemCor considerados individualmente. Éstos documentos (sin etiquetas) han representado el papel de ficheros de entrada.

Los resultados para nuestra primera serie de experimentos son resumidos en la Tabla 2. Esta tabla muestra las medias micro y macroaveraging para la precisión utilizando ambos algoritmos de entrenamiento. Los valores obtenidos por el enfoque integrado muestran una apreciable ventaja sobre el enfoque basado en WordNet y el basado en SemCor.

6 Conclusiones y Futuros Trabajos

En este artículo hemos presentado un nuevo enfoque para la resolución de la ambigüedad léxica basado en la integración de varios recursos léxicos de libre distribución para mejorar la efectividad (bases de datos léxicas y córporas de entrenamiento), empleados hasta ahora de manera aislada. Este enfoque integra la información proporcionada por la base de datos léxica WordNet, dentro de los algoritmos de entrenamiento de Rocchio y Widro-Hoff a través del MEV para WSD.

Hemos probado nuestro enfoque combinado, obteniendo medidas que avalan, que la precisión es mayor que la obtenida, tanto por el enfoque basado en el entrenamiento de SemCor, como por el basado en WordNet. A pesar de la complejidad de la tarea, se ha obtenido una buena precisión teniendo en cuenta la fina granularidad de sentidos definida en WordNet.

Actualmente, estamos aplicando nuestro enfoque WSD (utilizando múltiples recursos léxicos) a tareas propias del PLN, donde la desambiguación puede ser muy útil.

Referencias

[Agirre 96] Agirre E., Rigau G. *Word sense disambiguation using conceptual density.* In Proceedings of COLING 1996.

[Buenaga 97] Buenaga Rodríguez M., Gómez Hidalgo J .M., Díaz Agudo B. *Using WordNet to Complement Training Information in Text Categorization.* Second International Conference on Recent Advances in Natural Language Processing, 1997

[Broown 91] Brown P. B., Pietra S. A., Pietra V. J.*Word Sense Disambiguation Using Statiscal Methods.* In Proc. Of ACL, pp. 264-270, 1991.

[Frakes 92] Frakes, W., Baeza, R., *Information retrieval: data structures and algorithms*, Prentice Hall, London. 1992.

[Kilgarriff 97a] Kilgarriff A *What is word sense disambiguation good for?* Proc. Natural Language Processing Pacific Rim Symposium. Phuket, Thailand. December 1997. pp 209-214.

[Kilgarriff 97b] Kilgarriff A. *Foreground and Background Lexicons and Word Sense Disambiguation for Information Extraction* Proc. International Workshop on Lexically Driven Information Extraction. Frascati, Italy. July 1997. pp 51-62.

[Lewis 92] Lewis, D., *Representation and learning in information retrieval.* Ph.D. Thesis, Department of Computer and Information Science, University of Massachusetts. 1992.

[Lewis 96] Lewis, D.D., Schapire, R.E., Callan, J.P., Papka, R. *Training algorithms for linear text classifiers*. In Proceedings of the ACM SIGIR, 1996.

[Miller 93] Miller G. Leacock C., Randee T. and Bunker R. *A Semantic concordance. In Proceedings of the 3rd DARPA Workshop on Human Language Technology*, New Jersey 1993.

[Miller 95] Miller G. *WordNet: lexical database.* Communications of the ACM Vol 38, No. 11.

[Resnik 95] Resnik P. *Disambiguating Noun Groupings with Respect to WordNet Senses* Proceedings of the 3rd Workshop on Very Large Corpora, MIT, 30 June 1995.

[Rocchio 71] Rocchio, J.J. Jr. *Relevance feedback in information retrieval*. In G. Salton, editor, *The SMART Retrieval System: Experiments in Automatic Document Processing*. Prentice-Hall, 1971.

[Salton 83] Salton G., McGill, M.J.*Introduction to modern information retrieval.* McGraw-Hill.1983.

[Salton 89] Salton, G., *Automatic Text Processing: the transformation, analysis and retrieval of information by computer*. Addison Wesley. 1989

[Sanderson 96] Sanderson, M., *Word sense disambiguation and information retrieval*. Ph.D. Thesis, Department of Computing Science, University of University of Glasgow. 1996.

[Ureña 97] Ureña López, L. A., García Vega, M., Buenaga Rodríguez, M., Gómez Hidalgo, J. M. *Resolución de la ambigüedad léxica mediante información contextual y el modelo del espacio vectorial.* Séptima Conferencia de la Asociación Española para la Inteligencia Artificial. CAEPIA. 1997.

[Ureña 98] Ureña López, L. A., García Vega, M., Buenaga Rodríguez, M., Gómez Hidalgo, J. M. *Resolución automática de la ambigüedad léxica fundamentada en el modelo del espacio vectorial usando ventana contextual variable.* Asociación Española de Lingüística Aplaicada. AESLA. 1998.

[Widrow 85] Widrow, B., Sterns., S. *Adaptative Signal Processing*. Prentice-Hall, 1985.

[Xiaobin 95] Xiaobin Li; Szpakowicz S.; Matwin S. *A WordNet-based algorithm for word sense disambiguation.* Proceedings of the Fourteenth International Joint conference on Artificial Intelligence. pp. 1368-74, vol 2. 1995

[Yarowsky 92] Yarowsky D. W*ord-sense disambiguation using statistical models of Roget's categories trained on large corpora*. In Proceedings of the 15th International Conference onr Computational Linguistics. 1992.

[Yarowsky 93] Yarowsky, D. *One Sense Per Collocation*. In *Proceedings, ARPA Human Language Technology Workshop.* Princeton, pp. 266-271, 1993.

[Yarowsky 94] Yarowsky D. *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).1994.

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

[Yoshiki 94] Yoshiky N., Yoshihiko Nitta *Co-ocurrence vectors from corpora vs. distance vectors from.* In Proceedings of COLING94.