

# Utilización y evaluación de la desambiguación en tareas de clasificación de texto

José María Gómez Hidalgo

Depto. de Inteligencia Artificial, Universidad Europea de Madrid  
jmgomez@dinar.uem.es

Alberto Díaz Esteban

Depto. de Programación e Ingeniería del Software, Universidad Europea de Madrid  
alberto@dpris.uem.es

Luis Alfonso Ureña, Manuel García

Depto. de Informática, Universidad de Jaén  
{laurena,mgarcia}@ujaen.es

## Resumen

La evaluación de la desambiguación puede realizarse tanto de manera directa como indirecta, es decir, en el marco de otra tarea de procesamiento de lenguaje natural que hace uso de ella. La evaluación directa de la desambiguación está próxima a su estandarización en el marco de competiciones como SENSEVAL. En cambio, la evaluación indirecta ha sido poco utilizada, pero es muy importante porque la desambiguación se utiliza fundamentalmente como ayuda a otras tareas. En este trabajo presentamos dos métodos de desambiguación basados en la integración de recursos, aplicados a una tarea de categorización de documentos, que se basa en la misma idea de integración. Realizamos una evaluación directa e indirecta de las técnicas de desambiguación utilizadas, logrando resultados muy positivos para ambas técnicas. Los resultados son comparables con los que obtendría un desambiguador manual, e indican que es preciso hacer uso de la desambiguación para el método de categorización propuesto.

## 1 Introducción

Un problema importante en el procesamiento del lenguaje natural es determinar el sentido o significado de una palabra en un determinado contexto (desambiguación). La desambiguación es importante, fundamentalmente, como ayuda para otras tareas de procesamiento del lenguaje natural como la traducción automática [3], el restablecimiento de acentos [30], la extracción de información [8] y la generación automática de resúmenes [13]. La desambiguación es especialmente interesante en tareas de clasificación de texto, como la recuperación de información [7, 21, 26], la recuperación de información multilingüe [5] y la categorización de texto [4].

En los últimos tiempos se han presentado numerosos enfoques para realizar la desambiguación [16]. Lamentablemente, existe falta de acuerdo en los criterios de evaluación, especialmente en la elección de las definiciones de las palabras y las colecciones de evaluación, en la consistencia en el etiquetado manual de dichas colecciones y en las métricas empleadas para medir la efectividad.

Se ha producido un importante esfuerzo en la comunidad lingüística por llegar a una estandarización

en la evaluación de la desambiguación. Se ha celebrado recientemente una "competición" al estilo de la agencia norteamericana ARPA sobre desambiguación (SENSEVAL) [10], en la que se han conseguido resolver parcialmente los problemas mencionados. Esta competición propone una colección de evaluación, un conjunto de definiciones y una serie de métricas concretas, y logra una alta consistencia en el etiquetado de la colección.

Todo este esfuerzo se centra en una evaluación directa de la desambiguación, es decir, en evaluar la efectividad en la asignación de los significados correctos a las palabras a desambiguar. Sin embargo, la desambiguación sirve fundamentalmente como ayuda a otras tareas. Por tanto, es deseable realizar una evaluación indirecta que mida la efectividad de la tarea a la que se aplica, dependiendo del método de desambiguación empleado [28].

En este trabajo, se aplica la desambiguación a una tarea de clasificación concreta, la categorización de documentos, y se realiza una evaluación de la desambiguación tanto directa como indirecta. La categorización es una tarea de clasificación muy útil y compleja, para la que proponemos un modelo basado en la integración de recursos como colecciones de entrenamiento y bases de datos léxicas, con el fin de lograr mejoras en la efectividad de la tarea [4]. Demostramos también que, para sacar el máximo partido de la información de la base de datos léxica WordNet [14], es preciso realizar un proceso de desambiguación. Para la tarea de desambiguación aplicamos la misma idea de integración de recursos que en la categorización, presentando dos métodos distintos que se evalúan tanto directamente como indirectamente. Los resultados de ambas evaluaciones no muestran diferencias significativas entre los dos métodos de desambiguación estudiados, produciendo resultados muy parecidos a los logrados por un experto manualmente.

El resto de este trabajo se organiza del modo siguiente. En primer lugar, se exponen las características de las evaluaciones directa e indirecta y sus diferencias y problemas. A continuación describimos con detalle los métodos de categorización y desambiguación desarrollados para este trabajo. Después se presentan los resultados de la evaluación, realizada

tanto de manera directa como indirecta. Finalmente, se exponen las conclusiones y se describen las líneas de trabajo futuro.

## 2 Evaluación de la desambiguación

En 1997, Resnik y Yarowsky realizaron una serie de propuestas orientadas a la estandarización en la evaluación de la desambiguación [17]. Estas propuestas se han plasmado en la celebración de una "competición" sobre desambiguación llamada (SENSEVAL) [10]. SENSEVAL se concentra en la evaluación directa de la desambiguación. Nosotros consideramos igualmente importante la evaluación indirecta de esta tarea.

### 2.1 Evaluación directa

La evaluación directa mide la efectividad en la asignación de los significados correctos a las palabras a desambiguar. La evaluación directa es fundamental para cuantificar la calidad de los distintos enfoques de desambiguación. Las evaluaciones directas realizadas suelen presentar los siguientes problemas:

- La falta de acuerdo en la elección de las definiciones de las palabras: diferentes diccionarios suelen dar distintos conjuntos de sentidos para la misma palabra.
- La escasez de colecciones de evaluación: el etiquetado semántico de un corpus es una tarea difícil y costosa.
- La inconsistencia en el etiquetado de las colecciones de evaluación: distintas personas pueden asignar diferentes significados a la misma palabra en el mismo contexto.
- Una cierta falta de acuerdo en las métricas utilizadas: diversos autores presentan diferentes formas de medir la efectividad de la desambiguación.

SENSEVAL ha resuelto parcialmente estos problemas. En primer lugar, se ha definido una colección

de evaluación única para la "competición"; y paralelamente un conjunto de significados adaptados a la colección. El corpus HECTOR [2] ha sido adaptado a la "competición" re-etiquetando sus documentos con los nuevos significados definidos a partir del propio corpus. Por otra parte, la construcción de nuevos significados ha mejorado la consistencia en el etiquetado de la colección [9]. Por último, las métricas utilizadas han sido recall y precisión

Desafortunadamente, la evaluación realizada en SENSEVAL es demasiado genérica, independiente de la tarea a la que se aplica la desambiguación. Cada tarea posee su propia idiosincrasia, lo que impide utilizar los recursos aportados por SENSEVAL si no son adecuados a la tarea. Este es el caso de nuestro trabajo de desambiguación para la categorización.

Por otra parte, algunos métodos de desambiguación son capaces de decidirse siempre por algún significado. Para estos métodos, el recall es siempre 1 y sólo tiene sentido utilizar la precisión como métrica de evaluación. En estos casos se calcula habitualmente la media de la precisión para cada palabra (macroaveraging) y la precisión media sobre todas las palabras (microaveraging) [12]. Los métodos de desambiguación presentados en este trabajo siempre toman una decisión sobre el significado a seleccionar y por tanto se evalúan utilizando esta métrica.

## 2.2 Evaluación indirecta

La evaluación indirecta de la desambiguación mide la efectividad de la tarea a la que se aplica, en función del método de desambiguación empleado. Cada tarea se evalúa de una manera distinta, con sus propias métricas y colecciones. La evaluación indirecta es fundamental para cuantificar la calidad de los distintos enfoques de desambiguación, sobre la tarea a la que se aplica.

La desambiguación se ha aplicado en otros trabajos a diversas tareas de clasificación de texto, donde los criterios de evaluación (colecciones, métricas) son los propios de cada tarea. Existe un extenso trabajo realizado en recuperación de información, donde existe la intuición de que una comparación semántica en lugar de léxica entre consultas y documentos puede producir mejores resultados [8, 26, 24]. Sin embar-

go, esta intuición no se ha visto corroborada hasta trabajos muy recientes [7, 22].

La desambiguación es una fase fundamental en la recuperación de información en un entorno multilingüe. Una posibilidad consiste realizar una indexación conceptual de los documentos de múltiples idiomas para poderlos recuperar utilizando consultas en un idioma distinto [5]. Lamentablemente, la escasez de recursos léxicos multilingües ha dificultado la realización de experimentos hasta la fecha. La opción estándar es traducir la consulta al lenguaje de los documentos de la colección [6].

La categorización de documentos también puede verse beneficiada de la desambiguación [4]. Otros autores han propuesto la utilización de recursos léxicos para la categorización, sin hacer uso de la desambiguación, y los resultados han sido bastante negativos [23].

Los trabajos presentados en clasificación de texto hasta la fecha presentan los siguientes inconvenientes:

- Sólo se comparan los resultados de la clasificación usando desambiguación manual y sin usar ningún tipo de desambiguación. No se estudian métodos prácticos de desambiguación.
- Las mejoras que potencialmente debe producir la desambiguación en la clasificación son muy dependientes del modo en que la primera se utilice al servicio de la última.
- Algunas tareas pueden ser más sensibles que otras a los errores producidos en la desambiguación.

Centrándonos en estos problemas, presentamos dos métodos concretos de desambiguación como ayuda a una tarea de clasificación, la categorización de documentos. El modo en que se aplica la desambiguación permite obtener mejoras substanciales en la categorización, aunque queda por estudiar hasta que punto es sensible el método de categorización presentado a los errores producidos en la desambiguación.

### 3 Utilización de la desambiguación en categorización de documentos

La categorización de texto consiste en la clasificación de documentos dentro de un conjunto de categorías predefinido. La categorización automática de texto es una tarea compleja de clasificación frecuentemente aplicada a la asignación de descriptores de contenido a documentos, al encaminamiento y filtrado de texto, o empleada como parte de otros sistemas de procesamiento del lenguaje natural [12].

Los enfoques más habituales de categorización de documentos se basan en la utilización de una colección de documentos previamente etiquetados (colección de entrenamiento) para predecir la asignación de categorías a nuevos documentos (ver, por ejemplo, la comparativa de [29] o las actas del AAAI/ICML Workshop on Learning for Text Categorization [19]). La idea general es obtener una representación de cada categoría a partir de los documentos de entrenamiento, con la que se compara la representación de cada nuevo documento y se decide si se incluye este último en la categoría. Sin embargo, es corriente la existencia de categorías con pocos documentos de entrenamiento, lo que dificulta la obtención de representaciones eficaces de las mismas.

La utilización de recursos léxicos adicionales puede mejorar la representación de las categorías, y en consecuencia la eficacia de los sistemas de categorización [4]. Se puede utilizar de la base de datos léxica WordNet [14] como recurso complementario a la colección de entrenamiento. Para la utilización de WordNet de manera eficaz, es preciso efectuar un proceso de desambiguación de las categorías que aparecen en la colección de documentos, como se detalla en las dos próximas secciones.

#### 3.1 Uso de WordNet en la categorización

Las bases de datos léxicas son sistemas de referencia que acumulan información sobre los elementos léxicos de uno o varios idiomas. Entre las bases de datos léxicas actuales se encuentran WordNet, EDR y

el Roget's Thesaurus. Hemos seleccionado WordNet para nuestro estudio debido a su amplia cobertura y a su frecuente utilización en otras tareas de clasificación de texto, como la recuperación de información [26, 24], o la propia desambiguación [25].

WordNet incluye información sobre diversas relaciones léxicas (sinonimia, antonimia) y conceptuales (generalización y especialización, pertenencia y posesión, etc.). En este trabajo solamente hacemos uso de la relación de sinonimia. La utilización de otras relaciones como la generalización se ha explorado con resultados negativos en otros trabajos [23].

WordNet representa cada significado por medio de un conjunto de palabras sinónimas o synset. Esta propiedad se utiliza del siguiente modo:

- Se busca el código o expresión asociada a cada categoría en WordNet, obteniéndose una lista de synsets (significados) asociados a la misma.
- Se selecciona el synset más adecuado para representar la categoría, y se incorporan todas las palabras del mismo a la representación de las categorías.

Las categorías y los documentos se representan por medio de vectores de pesos de términos, como lo realizan la mayoría de los trabajos de categorización [29]. Existen diversos algoritmos de entrenamiento que admiten la utilización de una representación inicial del objeto a entrenar (en este caso, una categoría). De entre ellos hemos seleccionado el algoritmo de Rocchio [18] para el presente trabajo. Se construye un vector inicial de pesos con los términos obtenidos de WordNet para cada categoría, y se utiliza este vector como representación inicial de la categoría. Los pesos se calculan usando una fórmula tipo  $tf \cdot idf$  [20].

#### 3.2 Uso de la desambiguación para la categorización basada en WordNet

La selección del synset más adecuado para representar una categoría es un proceso de desambiguación. Intuitivamente, cada categoría tiene un solo significado en la colección de documentos, y los experimentos

realizados en este trabajo demuestran que es preciso realizar la desambiguación para sacar el máximo partido a la información de WordNet.

La desambiguación puede realizarse de manera manual, pero es conveniente mecanizar este proceso a fin de obtener un sistema de categorización completamente automático. En los últimos años se han presentado numerosas propuestas de desambiguación automática. La mayoría de las propuestas hacen uso de uno de los recursos siguientes:

- Una colección de entrenamiento donde cada palabra está etiquetada con su significado en un diccionario. Por ejemplo, cada aparición de la palabra "line" está asignada a un significado de WordNet. Las palabras que rodean a "line" en cada aparición se pueden usar para construir una representación de cada uno de sus significados, de modo que esta representación se compare con los contextos de las nuevas apariciones de "line" para asignarles un significado [16].
- Un recurso léxico como un diccionario o una base de datos léxica. Por ejemplo, se puede comparar el contexto de la palabra a desambiguar con las definiciones del diccionario [11], o se pueden definir y utilizar distancias conceptuales haciendo uso de las relaciones de la base de datos WordNet [1].

Como en la categorización, nosotros proponemos la integración de colecciones de entrenamiento y bases de datos léxicas para la desambiguación, utilizando el mismo enfoque. Para cada significado o synset de la palabra a desambiguar, se toman todas las palabras que aparecen en el synset y se usan como representación inicial del significado. Esta representación inicial se incorpora al entrenamiento, que se realiza en este caso usando dos algoritmos distintos, el de Rocchio [18] y el de Widrow-Hoff [27]. Los significados de las palabras se representan utilizando vectores de pesos de términos, donde cada peso se calcula utilizando una fórmula tipo  $tf.idf$  [20].

## 4 Resultados e interpretación

### 4.1 Descripción del entorno experimental

Para los experimentos de categorización presentados en este trabajo, se utilizan la colección de evaluación Reuters-21578, y la base de datos léxica WordNet 1.5. La colección Reuters ha sido la más utilizada en la evaluación de sistemas de categorización, y permite la comparación indirecta con otros enfoques [29]. Lamentablemente, esta comparación está limitada por el número de categorías usadas en cada trabajo.

Reuters consta de 21578 noticias breves de carácter económico, clasificadas dentro de 135 categorías o topics como "wheat" o "acquisitions/mergers". Para la evaluación de un enfoque de categorización, se suele dividir la colección en dos partes, una de entrenamiento y otra de prueba. Nosotros hemos usado la partición de Lewis [12], una de las más populares, que reserva dos terceras partes de los documentos para entrenamiento y una tercera parte para evaluación. En esta partición existen 89 categorías que poseen documentos en las subcolecciones de entrenamiento y de evaluación.

Existen varias medidas utilizadas para evaluar la efectividad de la categorización, que se pueden clasificar en medidas para clasificadores basados en ranking, y en medidas para clasificadores binarios [29]. Nuestro categorizador produce un ranking de documentos para cada categoría, que se puede convertir en una asignación binaria (el documento se introduce en la categoría o no). Nosotros presentamos los resultados utilizando tanto medidas basadas en ranking (precisión a 11 niveles de recall) como medidas para categorizadores binarios (medida  $f_1$  calculada realizando macroaveraging y microaveraging) [29].

La desambiguación se ha realizado utilizando la colección de documentos Semcor 1.5 [15]. Semcor es un subconjunto de 103 documentos extraídos del Brown Corpus, que tratan sobre diversos temas como política, deportes, música, cine, filosofía, etc. Cada palabra que aparece en Semcor está etiquetada con su significado en WordNet 1.5. La medida que hemos utilizado para realizar la evaluación directa de la desambiguación es la precisión promediada haciendo

	<i>Rocchio</i>	<i>Whoff</i>
Prec. media	0.865	0.910

Tabla 1: Precisión media de los dos algoritmos de desambiguación.

micro y macroaveraging.

## 4.2 Resultados e interpretación de la evaluación directa

El propósito de la evaluación directa es determinar cuál de los dos algoritmos propuestos para la desambiguación produce mejores resultados. Para ello, realizamos dos evaluaciones directas, ambas centradas solamente en las categorías de Reuters.

En primer lugar, se seleccionan al azar 4 documentos de Semcor y se entrena sobre los 99 restantes. Se desambigua cada aparición de una categoría de Reuters como palabra en los 4 documentos de evaluación y se calcula la precisión media, obteniéndose los resultados que se muestran en la Tabla 1. En estos resultados no se aprecian apenas diferencias de eficacia entre los dos algoritmos propuestos, siendo ambos muy buenos para el problema de desambiguar las categorías como palabras en un documento.

En nuestro enfoque, se utilizan todos los documentos de Semcor para el entrenamiento, y se realiza la evaluación sobre Reuters. Cada categoría de Reuters se utiliza con un solo significado en la colección, que se representa a partir de los documentos clasificados en la categoría. La representación obtenida para cada significado posible de la categoría en Semcor se compara con la representación proveniente de Reuters, asignando a cada categoría un solo significado en Reuters. Este significado se compara con el seleccionado por el experto humano, obteniéndose un sólo valor de precisión para el total de categorías.

Los resultados de este experimento se muestran en la tabla 1. La diferencia de precisión es pequeña, pues corresponde a que Widrow-Hoff realiza 3 aciertos más que Rocchio sobre las 67 categorías de Reuters que poseen alguna aparición en Semcor, y por tanto se pueden desambiguar con nuestro enfoque. Esta pe-

<i>Rec</i>	<i>SWN</i>	<i>TWN</i>	<i>Rocc</i>	<i>WHoff</i>	<i>Man</i>
0.0	0.805	0.880	0.900	0.900	0.889
0.1	0.777	0.851	0.881	0.881	0.872
0.2	0.733	0.822	0.857	0.857	0.844
0.3	0.683	0.764	0.818	0.818	0.808
0.4	0.623	0.708	0.770	0.770	0.778
0.5	0.570	0.658	0.731	0.731	0.744
0.6	0.500	0.580	0.608	0.608	0.624
0.7	0.404	0.496	0.532	0.532	0.550
0.8	0.337	0.415	0.466	0.466	0.474
0.9	0.256	0.317	0.346	0.346	0.360
1.0	0.124	0.190	0.217	0.217	0.227
Med	0.528	0.608	0.648	0.648	0.652

Tabla 2: Precisión en 11 niveles de recall y precisión media.

queña diferencia puede deberse a que Widrow-Hoff realiza una integración más homogénea entre los dos recursos (Semcor y WordNet) utilizados [4].

## 4.3 Resultados e interpretación de la evaluación indirecta

El propósito de la evaluación indirecta es determinar si es preciso efectuar la desambiguación de las categorías, y cuál de los métodos de desambiguación propuestos en este trabajo produce mejores resultados. La evaluación se realiza sobre las 51 categorías que poseen apariciones en Semcor, además de poseer documentos en la subcolección de entrenamiento y en la de evaluación de la partición Lewis de Reuters.

En las tablas 2 y 3 se presentan los resultados de la evaluación de la categorización realizada sin WordNet (SWN), utilizando todos los synsets por categoría (TWN), desambiguando con el algoritmo de Rocchio (Rocc), desambiguando con el algoritmo de Widrow-Hoff (Whoff), y desambiguando manualmente (Man). En la Tabla 2 se muestran los resultados de la categorización utilizando la medida de la precisión en 11 niveles de recall (Rec) y la precisión media (Med). En la Tabla 3 se muestran los resultados de la categorización medidos utilizando la *f1* calculada tanto efectuando macroaveraging como microaveraging.

$f_1$	<i>SWN</i>	<i>TWN</i>	<i>Rocc</i>	<i>WHoff</i>	<i>Man</i>
Mac.	0.464	0.538	0.571	0.571	0.576
Mic.	0.661	0.664	0.674	0.674	0.678

Tabla 3:  $F_1$  calculada por medio de macro y microaveraging.

En cada tabla se compara la categorización sin WordNet, la categorización utilizando todos los synsets para cada categoría, la categorización con desambiguación usando Rocchio, la categorización con desambiguación usando Widrow-Hoff, y la categorización con desambiguación manual. Se pueden realizar las siguientes observaciones:

- Globalmente, los resultados apoyan la tesis de que para sacar máximo partido de WordNet en el enfoque de categorización propuesto, es conveniente utilizar algún método de desambiguación. Si no lo hay disponible, también da buenos resultados utilizar todos los synsets asociados a cada categoría.
- Los dos métodos empleados en la desambiguación producen exactamente los mismos resultados. Esto se produce porque, dado el pequeño número de categorías, ambos algoritmos presentan pocas discrepancia en la elección de los synsets de WordNet. Como además los synsets potenciales poseen muchas palabras en común, no existe ninguna discrepancia en la elección de los términos para representar a las categorías entre ambos algoritmos. Por ejemplo, la categoría "inventories" posee 4 significados ((1) inventory, stock list; (2) stock, inventory; (3) inventory; (4) armory, armoury, inventory; y (5) inventory, inventorying, stocktaking). El significado elegido por el algoritmo basado en Widrow-Hoff es el segundo, que es correcto. Sin embargo, aunque Rocchio elige el primer significado, los términos aportados por ambos enfoques son iguales salvo por "list", que aparece unas 780 veces en la colección de entrenamiento, y tiene un peso tan pequeño que no afecta a la eficacia de la categorización.

La situación planteada por la segunda observación indica probablemente que el modo en que nosotros utilizamos la desambiguación para la categorización produce que ésta sea poco sensible a los errores de la desambiguación, aunque esto es materia de estudio en trabajo futuro.

Es interesante observar que las diferencias considerando  $f_1$  calculada haciendo microaveraging son apenas apreciables. Dado que las medidas haciendo microaveraging dan más peso a las categorías con más documentos, esto demuestra que las mayores mejoras se producen precisamente para las categorías con menos documentos (de entrenamiento).

## 5 Conclusiones y trabajo futuro

En este trabajo hemos presentado dos métodos de desambiguación utilizados en el marco de una tarea de clasificación, la categorización de documentos. Tanto los métodos de desambiguación como el de categorización propuestos se basan en la integración de recursos como colecciones de entrenamiento (Semcor, Reuters) y bases de datos léxicas (WordNet).

Se ha realizado una evaluación directa e indirecta de los métodos de desambiguación, obteniéndose resultados similares para ambas técnicas, muy próximos a los obtenidos por una desambiguación manual. Los resultados demuestran que es preciso hacer uso de la desambiguación para sacar el máximo partido a la integración de recursos en la categorización, pero que la integración sin desambiguación es asimismo un enfoque muy competitivo.

En este estudio surge la cuestión de hasta que punto es sensible el método de categorización presentado a los errores producidos en la desambiguación. Las dos técnicas de desambiguación utilizadas poseen distintos márgenes de error y sin embargo producen los mismos efectos en la desambiguación. Otras tareas como la recuperación de información tienen una sensibilidad alta a los errores de la desambiguación [7], por lo que consideramos como principal línea de trabajo futuro el estudio sistemático de la sensibilidad de la categorización a los errores de la desambigua-

ción.

## Referencias

- [1] E. Agirre and G. Rigau. Word sense disambiguation using conceptual density. In *Proceedings of COLING*, 1996.
- [2] S. Atkins. Tools for computer-aided lexicography: the hector project. In *Papers in Computational Lexicography: COMPLEX*, 1993.
- [3] P.B. Brown, S.A. Pietra, and V.J. Pietra. Word sense disambiguation using statistical methods. In *Proceedings of ACL*, 1991.
- [4] M. Buenaga, J.M. Gómez, and B. Díaz. Using wordnet to complement training information in text categorization. In *Proceedings of RANLP*, 1997.
- [5] J. Gilarranz, J. Gonzalo, and F. Verdejo. Using the eurowordnet multilingual semantic database. In *Proceedings of AAAI-96 Spring Symposium Cross-Language Text and Speech Retrieval*, 1996.
- [6] J. Gonzalo, F. Verdejo, and I. Chugur. A concept-based approach to cross-language retrieval using eurowordnet. *Applied Artificial Intelligence, Special Issue on Multilinguality in the Software Industry: the AI contribution*, 1999. En prensa.
- [7] J. Gonzalo, F. Verdejo, I. Chugur, and J. Cigarrán. Indexing with wordnet synsets can improve text retrieval. In *Proceedings of the COLING/ACL'98 Workshop on Usage of WordNet for Natural Language Processing*, 1998.
- [8] A. Kilgarriff. What is word sense disambiguation good for? In *Proceedings of Natural Language Processing Pacific Rim Symposium*, 1997.
- [9] A. Kilgarriff. Gold standard for evaluating word sense disambiguation programs. *Computer Speech and Language, Special Issue on Evaluation*, 1998. En prensa.
- [10] A. Kilgarriff. Senseval: An exercise in evaluating word sense disambiguation programs. In *Proceedings of LREC*, 1998.
- [11] R. Krovetz and W.B. Croft. Word sense disambiguation using machine-readable dictionaries. In *Proceedings of ACM SIGIR*, 1989.
- [12] D.D. Lewis. *Representation and Learning in Information Retrieval*. PhD thesis, Dept. of Computer and Information Science, University of Massachusetts, 1992.
- [13] M. Maña, M. Buenaga, and J.M. Gómez. Diseño y evaluación de un generador de resúmenes de texto con modelado de usuario en un entorno de recuperación de información. *Procesamiento del Lenguaje Natural*, (23), 1998.
- [14] G. Miller. Wordnet: A lexical database for english. *Communications of the ACM*, 38(11), 1995.
- [15] G. Miller, C. Leacock, T. Randee, and R. Bunker. Using a semantic concordance for sense identification. In *Proceedings of the 3th DARPA Workshop on Human Language Technology*, 1994.
- [16] R.J. Mooney. Comparative experiments on disambiguating word senses: An illustration of the role of bias in machine learning. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 1996.
- [17] P. Resnik and D. Yarowsky. A perspective on word sense disambiguation methods and their evaluation. In M. Light, editor, *Tagging Text with Lexical Semantics: Why, What and How?*, *ACL SIGLEX*, 1997.
- [18] J.J. Rocchio Jr. Relevance feedback in information retrieval. In G. Salton, editor, *The SMART Retrieval System: Experiments in Automatic Document Processing*. Prentice Hall, 1971.
- [19] M. Sahami, editor. *Proceedings of the AAAI'98/ICML'98 Workshop on Learning for Text Categorization*, 1998.

- [20] G. Salton and M.J. McGill. *Introduction to Modern Information Retrieval*. McGraw-Hill, 1983.
- [21] M. Sanderson. Word sense disambiguation and information retrieval. In *Proceedings of SIGIR*, 1994.
- [22] H. Schutze and J. Pedersen. Information retrieval based on word senses. In *IV Annual Symposium on Document analysis and Information Retrieval*, 1999.
- [23] S. Scott and S. Matwin. Text classification using wordnet hypernyms. In *Proceedings of the COLING-ACL'98 Workshop in Usage of WordNet in Natural Language Processing Systems*, 1998.
- [24] A. Smeaton, F. Kellely, and R. O'Donnell. Trec-4 experiments at dublin city university: Thresholding posting lists, query expansions with wordnet and pos tagging of spanish. In *Proceedings of TREC*, 1995.
- [25] L.A. Ureña, M. García, Buenaga M., and J.M. Gómez. Resolución de la ambigüedad léxica mediante información contextual y el modelo del espacio vectorial. *Procesamiento del Lenguaje Natural*, (23), 1998.
- [26] E.M. Voorhees. Query expansion using lexical-semantic relations. In *Proceedings of the ACM SIGIR*, 1994.
- [27] B. Widrow and S. Sterns. *Adaptive Signal Processing*. Prentice-Hall, 1985.
- [28] Y. Wilks. Is word sense disambiguation just one more nlp task? In *Proceedings of the SENSEVAL Conference*, 1998.
- [29] Y. Yang. An evaluation of statistical approaches to text categorization. *Information Retrieval Journal*, 1998. En prensa.
- [30] D. Yarowsky. Decision list for lexical ambiguity resolution: Application to accent restoration in spanish and french. In *Proceedings of ACL*, 1994.

