

Incorporación de Adjetivos al WordNet Español

Anselmo Peñas, Irina Chugur, Julio Gonzalo y Felisa Verdejo

Dpto. Lenguajes y Sistemas Informáticos
ETS Ingenieros Industriales, UNED

{anselmo,irina,julio,felisa}@ieec.uned.es

Resumen

EuroWordNet es una base de datos léxica multilingüe con relaciones semánticas entre nombres y verbos de ocho idiomas europeos (inglés, holandés, español, italiano, alemán, francés, checo y estonio). En este artículo se describe la incorporación de synsets adjetivales al wordnet español mediante técnicas de extracción automática a partir de los datos originales de EuroWordNet y de diccionarios electrónicos.

1. Introducción

Nuestro trabajo se enmarca en el proyecto EuroWordNet (EWN) (Vossen, 1998) cuyo objetivo ha sido desarrollar de forma semiautomática una base de datos léxica multilingüe con relaciones semánticas entre palabras de varios idiomas europeos (inglés, holandés, español, italiano, alemán, francés, checo y estonio). Cada una de esas redes está formada por unidades denominadas *synsets*. Un *synset* es un conjunto de sinónimos que corresponde a un concepto lexicalizado. Entre los *synsets* se establecen las relaciones semánticas fundamentales: hiponimia, antonimia, meronimia, implicatura, etc..

Además de las relaciones entre los *synsets*, uno de los elementos importantes de la estructura de estas redes de palabras (wordnets) es el enlace al Inter-Lingual-Index (ILI). Este índice representa una lista no estructurada de conceptos o ILI-records (Vossen, Peters and Gonzalo, 1999) construida con el objetivo de unir wordnets de diferentes idiomas. El ILI es un superconjunto de todos los conceptos que aparecen en los distintos wordnets. Por ejemplo, dos *synsets* pertenecientes a dos lenguas distintas, ligados mediante la relación de EQ_SYNONYM a un mismo ILI-record se consideran equivalentes en su significación semántica. De esta forma, mediante el ILI es posible pasar de un wordnet a otro, aspecto que es de indudable interés para aplicaciones como la traducción automática, la recuperación de información multilingüe o la extracción de información.

EuroWordNet hereda la estructura de

WordNet (WN) (Miller et al., 1990) y la enriquece con nuevas relaciones semánticas y relaciones semánticas entre distintas categorías sintácticas. Las conexiones entre distintas categorías son uno de los rasgos característicos de EWN que resulta útil para muchas aplicaciones, como, por ejemplo, la Recuperación de Información (Gonzalo et al., 1998a), (Gonzalo et al., 1998b).

Las únicas categorías sintácticas contempladas actualmente en el wordnet español son nombres y verbos. Nuestro objetivo consiste en introducir en el wordnet español de forma automática el mayor número posible de *synsets* adjetivales. La metodología empleada es consistente con la aproximación utilizada en la construcción de la red semántica para el español, que se basa principalmente en enlazar palabras del español con los *synsets* del WordNet 1.5.

En un primer estudio que se describe en el apartado 2, la introducción de adjetivos se planteó en términos de creación de relaciones entre adjetivos extraídos del diccionario electrónico monolingüe y nombres/verbos correspondientes que ya formaban parte del wordnet español. Sin embargo, resultó bastante problemático enlazar automáticamente al ILI los adjetivos obtenidos, requisito indispensable para que los nuevos candidatos entren a formar parte de la estructura multilingüe de EWN. Para resolver esta dificultad se planteó un nuevo enfoque en el que la introducción de adjetivos se realiza en términos de creación de enlaces entre adjetivos e ILI-records. Los criterios y heurísticas utilizados con este fin se detallan en el apartado 3. Finalmente se establecen unas conclusiones y

se señalan futuras extensiones del trabajo.

2. Enlaces de adjetivos a nombres y verbos del wordnet español

En un primer acercamiento, la introducción de adjetivos en el wordnet español se planteó como creación de enlaces entre adjetivos extraídos del diccionario electrónico monolingüe, y nombres/verbos que ya formaban parte de la red semántica para el español, con la posterior creación de synsets para los adjetivos enlazados.

2.1. Identificación de relaciones de EWN mediante definiciones lexicográficas

En primer lugar, se realizó un estudio de las definiciones del diccionario electrónico. Algunos de los patrones recurrentes en las definiciones de adjetivos ofrecen la posibilidad de extraer de forma automática una serie de relaciones contempladas en EWN. Por ejemplo, los adjetivos definidos como “que causa + N/V” o “capaz de causar + N/V” son buenos candidatos para captar la relación de EWN “CAUSES / IS_CAUSED_BY”, mientras que patrones como “que implica + N/V” apuntan a la relación “ROLE / INVOLVED”. A su vez, los adjetivos definidos en términos de otros adjetivos proporcionan agrupaciones sinonímicas o antonímicas.

Sin embargo, tras el estudio manual de una muestra, se comprobó que este método de identificación de las relaciones contempladas en EWN, en primer lugar, es poco productivo, y, en segundo lugar, requiere un filtro manual adicional considerable, ya que para la mayoría de los patrones de definición no existe una correspondencia unívoca entre el patrón y la relación semántica. Para otros trabajos con este método se puede consultar un estudio realizado en paralelo (Soler, próxima publicación).

2.2 Patrones de definición y criterios morfológicos

A partir de este estudio preliminar, se hizo patente la necesidad de considerar otras fuentes de información para seleccionar candidatos y ampliar el número de patrones lexicográficos, teniendo en cuenta relaciones y patrones que diesen un número mayor de enlaces para adjetivos.

Un estudio del diccionario permite distinguir los siguientes tipos de definiciones:

1. sinonímicas (por ejemplo, “*censurable* –

reprobable, condenable”);

2. antonímicas (por ejemplo, “*irreal* – *no real, falto de realidad*”);

3. analíticas, que definen adjetivos en términos de nombres o verbos (por ejemplo, “*dubitativo* – *que implica o denota duda*”).

El primer tipo de definición proporciona candidatos para formar synsets adjetivales. El segundo permite establecer relaciones de antonimia (adjetivo-adjetivo). Las definiciones analíticas proporcionan criterios para crear enlaces entre adjetivos y nombres y/o verbos.

Asimismo, se decidió introducir criterios morfológicos explotando la relación que existe entre los adjetivos denominales y/o deverbales y sus bases de derivación.

A partir de una veintena de sufijos productivos utilizados en los procesos de adjetivación (Lang, 1992; Bosque y Demonte, 1999) hemos seleccionado para una exploración inicial un subconjunto formado por los siguientes morfemas: -able, / -ible, -dor, -ense, -ista, -torio. La selección de los sufijos ha sido motivada por consideraciones de productividad y mayor cobertura en cuanto a la obtención de parejas de adjetivos y nombres/ verbos del wordnet español.

Con estas consideraciones, el trabajo se ha centrado en:

- generación de candidatos mediante la aplicación de patrones de definición sobre un subconjunto restringido de adjetivos denominales y deverbales.
- la extracción automática de tres tipos de:
 - 1) relación de sinonimia;
 - 2) relación de antonimia;
 - 3) relación genérica adjetivo-nombre; adjetivo-verbo sin especificar el tipo concreto de relación según EWN.

Sobre la lista de adjetivos con los sufijos seleccionados se han aplicado los patrones que permiten identificar la palabra que constituye el núcleo conceptual de la definición. El analizador morfosintáctico (Rodríguez et al., 1998; Márquez y Padró, 1997) permite obtener la forma canónica y la categoría de la palabra extraída de la definición y, de esta manera, reconocer tanto las formas flexivas de los nombres, como las diferentes formas verbales.

En el ejemplo 1 se pueden ver algunos de los patrones de definición con ejemplos de entradas del diccionario que se ajustan a los mismos.

A continuación, los nombres y verbos

participantes de las relaciones identificadas se buscan en el wordnet español para filtrar aquellos sustantivos y verbos que no forman parte de éste.

Finalmente, se obtienen parejas adjetivo/nombre, adjetivo/verbo, así como agrupaciones sinonímicas y antonímicas (adjetivo/adjetivo).

<p>“que + Verbo 3ªp. S” abastecedor_i_1 <adj.-s> abastecedor, que abastece abrasador_i_1 <adj.> abrasador, que abrasa abrumador_i_1 <adj.> abrumador, que abruma</p> <p>“que implica + (Det) + Nombre” rogatorio_i_1 <adj.> rogatorio, que implica ruego usurpatorio_i_1 <adj.> usurpatorio, que implica usurpación</p>

Ejemplo 1. Patrones de definición

2.3 Creación de synsets

Algunos nombres y verbos, identificados como término de enlace, son polisémicos en el wordnet. La polisemia plantea un problema sobre la elaboración de criterios para decidir a cuál de los sentidos del nombre o del verbo hay que enlazar el adjetivo dado.

La hipótesis inicial consistía en suponer que, contando con una fuerte relación morfológica y semántica que une un adjetivo denominativo o deverbial con su base y con una definición lexicográfica que hace derivar el significado del adjetivo a partir del significado del nombre o verbo correspondiente (por ejemplo, “traducible que se puede traducir”), podíamos postular la posibilidad de relacionar el adjetivo derivado con todos los sentidos del nombre o verbo base de derivación.

Por lo tanto, se plantean dos posibles maneras de actuar:

1. crear un synset adjetival por cada sentido del nombre o verbo correspondiente, o bien
2. construir un único synset para el adjetivo denominativo o deverbial que reuniese todas las acepciones derivadas de los sentidos del término de enlace.

La primera opción lleva consigo distinciones innecesarias, por lo que optamos por la segunda.

Para evaluar el número de conexiones erróneas entre adjetivos y sentidos de los verbos o nombres correspondientes, revisamos de forma manual si el adjetivo de la entrada guardaba una relación semántica con todos los sentidos del

wordnet del nombre o del verbo que figura en su definición. El resultado de la comprobación de una muestra de un 30% nos llevó a estimar un margen de error del 10%. Las causas de los posibles enlaces incorrectos son variadas y complejas. Una de ellas es el hecho de que en ocasiones, mediante los patrones, en lugar de bases de derivación, se identifican otros derivados. Por ejemplo, el adjetivo “admirable” se define como “digno de admiración”, lo cual lleva a que se proponga para enlace la pareja “admirable”- “admiración”. A la hora de relacionar admirable con todos los sentidos de admiración se observa un enlace incorrecto entre “admirable” y “admiración” en el sentido de “signo de admiración”. En estos casos, no hay una verdadera relación de derivación entre los términos, sino entre cada uno de ellos y un primitivo no identificado, por lo que no se puede garantizar la corrección del enlace.

2.4 Resultados y discusión

La aplicación conjunta de los patrones de definición y del criterio morfológico dio como resultado agrupaciones sinonímicas, antonímicas y parejas de adjetivo-nombre, adjetivo-verbo con el mismo lexema y con lexemas diferentes.

Existe una posibilidad de aumentar el número de relaciones extraídas combinando las definiciones sinonímicas con las definiciones analíticas. Por ejemplo, “audible” se define sinonímicamente como “oíble”. A su vez, “oíble” se define como “que se puede oír”, relacionando el adjetivo “oíble” con el verbo “oír”. De esta manera, a través de la definición sinonímica, también es posible relacionar el adjetivo “audible” con el verbo “oír”.

La tabla 1 recoge los resultados obtenidos desglosados por cada sufijo. Se puede apreciar que el mayor porcentaje de relaciones corresponde a las relaciones entre adjetivos y verbos. Un 47.23% de los adjetivos se ha podido relacionar con algún verbo. En cuanto a las relaciones adjetivo – nombre, el porcentaje es tan sólo de 16.51%. Esta diferencia se debe fundamentalmente a la selección de sufijos. Como era de esperar, los sufijos que aportan un mayor número de relaciones adjetivo – verbo son aquellos que se utilizan para derivar adjetivos a partir de verbos (-able, -íble, -dor). Sin embargo, los sufijos como -ense, -ista proporcionan un mayor número de enlaces entre adjetivos y nombres.

	A-N	A-V	Sinónimos	Antónimos	Resto
-able 1055	49 4.65%	382 36.21%	86 8.15%	153 14.5%	194 36.49%
-dor 1616	50 3.04%	1251 77.41%	112 6.93%	0	203 12.56%
-ense 263	30 11.41 %	1 0.38%	56 21.29%	0	176 66.92%
-ista 612	378 61.76%	3 0.49%	53 8.66%	3 0.49%	175 28.60%
-torio 217	84 8.71%	49 22.58%	24 11.06%	3 1.38%	57 26.27%
-ible 8	0	5 62.50%	1 12.50%	1 12.50%	1 12.50%
Total 3580	591 16.51%	1691 47.23%	332 9.27%	160 4.47%	806 22.51%

Tabla 1. Relaciones entre palabras

Se han obtenido un total de 2282 relaciones entre adjetivos y nombres/verbos (relaciones entre parejas de palabras tal como refleja la tabla 1). Estas relaciones se traducen en 2191 synsets sin correspondencia con el ILI.

2.5 Conclusiones preliminares

El trabajo realizado bajo este primer enfoque permite:

1. Obtener un número considerable de adjetivos candidatos para ser incorporados en el wordnet español.
2. Relacionar estos adjetivos con nombres y/o verbos del wordnet español.
3. Crear agrupaciones de sinónimos y antónimos para adjetivos.

Sin embargo, como hemos apuntado en la introducción, la segunda e importante componente de la tarea de introducción de adjetivos (o de cualquier otra categoría sintáctica) en el EWN consiste en enlazar los candidatos con el ILI, asignándoles un número de ILI-record y haciéndolos, de esta manera, participar en la estructura global de los wordnets que integran EWN. En la práctica, el enlace consiste, en la mayoría de los casos, en añadir el candidato a un synset de WN inglés existente. Lo cual equivale a postular que el sentido determinado del adjetivo español dado es un equivalente de cierto adjetivo inglés considerado en su sentido correspondiente.

Sin estos enlaces, a pesar de tener un número de synset que se les podría asignar dentro del wordnet español, los adjetivos españoles no tendrían una correspondencia directa con los adjetivos de otras lenguas, sino que tendrían que ser recuperados a través de relaciones entre categorías sintácticas. Nos encontramos con las dificultades originadas fundamentalmente por la

polisemia: un coste elevado y el considerable trabajo manual que suponía encontrar la combinación adecuada de adjetivos españoles e ingleses bajo este método. Por ello, hemos mantenido –de momento– estos enlaces como un complemento externo a la información en el EWN español, que nos permite relacionar palabras (adjetivos) con los synsets nominales y verbales. Hemos reservado para la base de datos la introducción de synsets adjetivales que podemos ligar automáticamente al ILI, tal como se explica en el siguiente apartado. Queda para un futuro próximo el establecimiento de heurísticas para hacer corresponder las relaciones adjetivo-nombre/verbo con los synsets adjetivales ligados al EWN.

3. Enlace de adjetivos al Índice InterLingua

El nuevo enfoque se centra en el objetivo de enlazar el mayor número posible de adjetivos del español a las acepciones o conceptos que componen el Índice Inter-Lingua (ILI) de EuroWordNet.

Los procedimientos, criterios y heurísticas que se exponen a continuación han sido utilizados para enlazar los adjetivos del español. Sin embargo, pueden ser utilizados de igual manera sobre cualquier otra categoría gramatical y cualquier idioma para el que se disponga de diccionarios bilingües en formato electrónico.

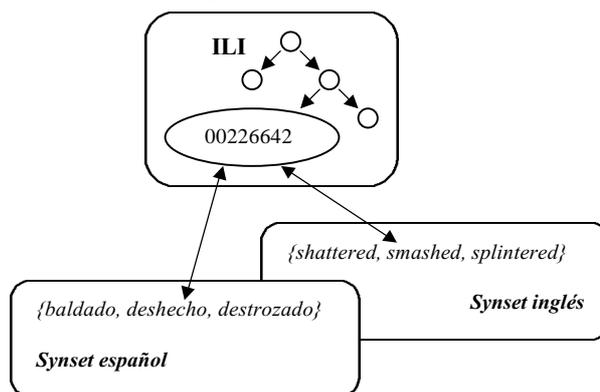


Figura 1

Para llevar a cabo la tarea, en el caso concreto de los adjetivos del español, además del propio ILI, se han utilizado los siguientes recursos:

1. adjetivos del inglés ya enlazados al ILI (synsets de WordNet 1.5),
2. diccionarios bilingües inglés-español y español-inglés en formato electrónico.

Para relacionar los adjetivos del español con el ILI se dispone de dos perspectivas.

1. los términos ingleses contenidos en el WN inglés, tomando las variantes que componen un synset y traduciéndolas al español, y
2. los adjetivos españoles contenidos en el diccionario bilingüe, buscando sus traducciones en los synsets del inglés.

A continuación se dan dos criterios para la primera perspectiva y cuatro para la segunda.

3.1 Traducción al español de adjetivos ingleses

3.1.1 Traducción de adjetivos ingleses monosémicos

La primera heurística consiste en tomar los adjetivos monosémicos del WN. Como estos adjetivos únicamente aparecen en un synset, la traducción española del adjetivo inglés se puede enlazar con el synset, puesto que al menos una de las acepciones del adjetivo español ha de corresponderse con el mismo.

cynical → 01865165
cynical → cínico
enlace: cínico / 01865165
damp → 01935444
damp → húmedo, mojado
enlaces: húmedo/01935444
mojado/01935444

Ejemplo 2

3.1.2 Traducción al español de las variantes de un synset inglés

Un synset inglés está compuesto por un conjunto de sinónimos o variantes del synset. Si cada una de estas variantes se traduce al español, la co-ocurrencia de un adjetivo español en las traducciones de variantes diferentes indica la relación entre el adjetivo español y el synset.

00235058 gruff hoarse husky
gruff: bronco malhumorado rudo
hoarse: ronco áspero
husky: ronco
enlace: ronco/00235058

Ejemplo 3

3.2 Traducción al inglés de adjetivos españoles

En esta ocasión se toman como punto de partida todos los adjetivos del español contenidos

en el diccionario bilingüe. Mediante el mismo, se obtienen las traducciones posibles al inglés para cada uno de los adjetivos.

La palabra que sirve de traducción al inglés de un adjetivo español puede tener otros usos, además del que tiene en común con el adjetivo español origen. Por tanto, el adjetivo español no se puede asociar a todas las acepciones de la palabra inglesa que sirve para su traducción.

El objetivo de la tarea es establecer enlaces entre los adjetivos del español y sus acepciones correctas en el interlingua. Para identificar y crear dichos enlaces, se utilizaron los siguientes criterios y heurísticas atendiendo a cuatro situaciones:

3.2.1 Traducción al inglés única

Si la traducción del adjetivo español es única, es decir, se corresponde con una única palabra inglesa según el diccionario, pueden darse dos casos:

- que sea monosémica según WN, el adjetivo inglés está asociado a una única acepción del WN inglés.
- que sea polisémica según WN, el adjetivo inglés está asociado a varias acepciones del WN inglés.

(a) Traducción al inglés única y monosémica

Si la traducción del adjetivo español al inglés es monosémica, es decir, se enlaza a un único concepto, entonces se puede enlazar el adjetivo español a dicho concepto automáticamente. Por ejemplo, 'abochornado' del español se traduce por 'ashamed' del inglés, y 'ashamed' está ligado únicamente al índice 00125050, por lo que 'abochornado' queda ligado al mismo de forma automática.

(b) Traducción al inglés única y polisémica

En este caso, la traducción del adjetivo español es una palabra inglesa pero que tiene varias acepciones (ver ejemplo 4).

Se trata de dilucidar si se puede enlazar el adjetivo español a todas las acepciones a las que se enlaza el adjetivo inglés, si sólo se puede enlazar a alguna de ellas o si, por el contrario, no se dispone de información suficiente para establecer relación alguna.

El estudio de las posibles relaciones se realiza para cada uno de los synsets candidatos. Para saber si se puede relacionar el concepto o synset con el adjetivo español original, se ha utilizado el

método de verificación mediante traducción inversa de los adjetivos ingleses miembros del synset en estudio. Los adjetivos ingleses se traducen ahora al español de forma que entre las traducciones de algunos de ellos volverá a aparecer el adjetivo español original (ver ejemplo 4). Cada ocurrencia del adjetivo español original entre las traducciones inversas del adjetivo inglés,

debe incrementar la posibilidad de enlace entre el adjetivo español y el concepto candidato.

De acuerdo con estas consideraciones se ha utilizado una medida muy sencilla que indica la posibilidad de enlazar un adjetivo español a un synset con adjetivos en inglés:

[Ocurrencias de adj en las traducciones inversas de las variantes de syn]²

$$\text{hook_ratio (adj, syn)} = \frac{\text{Número de traducciones inversas}}{\text{Número de traducciones inversas}}$$

adj: adjetivo original
syn: concepto candidato

Debido a la falta de correspondencia entre los diccionarios bilingües y el WordNet, existen adjetivos ingleses para los cuales no se dispone de traducción inversa. Esto ha ocurrido en 431 ocasiones sobre 10388 intentos de traducción (4.15%). En estos casos, para el cálculo del *hook ratio* se ignoran las variantes inglesas del synset que no tienen traducción inversa.

Una vez definida la medida que indica la posibilidad de enlace, es necesario establecer a partir de qué umbral resulta razonable enlazar el adjetivo español con un determinado concepto. Se tomó una muestra de 78 adjetivos con traducción única y polisémica y se comprobó que sin umbral (umbral igual a 0), únicamente el 77% de los enlaces eran correctos. Atendiendo a la necesidad de minimizar el número de enlaces que pueden ser dudosos o incorrectos, el umbral se elevó a 0.3 de forma empírica, de modo que únicamente aquellos enlaces cuyo *hook_ratio* fuera mayor que 0.3 fueron aceptados.

Con el fin de validar la adecuación de la medida y el umbral, se tomó una muestra de 100 enlaces y se estudiaron los resultados obtenidos. Se comprobó que las medidas permitieron aceptar 46 enlaces correctos y descartar 24 enlaces incorrectos. Únicamente se desechó 1 enlace correcto por no alcanzar el umbral, mientras que el resto de enlaces estudiados (29) correspondían a enlaces correctos que no pudieron ser detectados por falta de traducción inversa. Así pues, ningún enlace incorrecto fue aceptado como bueno.

3.2.2 Traducción al inglés múltiple

Si la traducción del adjetivo español al inglés es múltiple, es decir, hay varias palabras inglesas que sirven como traducción, se van a considerar dos casos más para los cuáles es posible

premeditado

Traducción: deliberate (4 acepciones):

1. 01016975: by conscious design or purpose: "intentional damage"; "a knowing attempt to defraud"; "a willful waste of time"
2. 00965893: "walking at the same measured pace"
3. 01834448: marked by careful consideration or reflection: "a deliberate decision"
4. 01407902: "a calculated insult"; "with measured irony"

Traducciones inversas y hook-ratio para cada synset:

01016975 deliberate intentional knowing willful wilful
deliberate: deliberado premeditado
deliberate: pausado lento
intentional: intencional
knowing: de_complicidad
willful:
wilful: voluntarioso terco
wilful: premeditado

$$h_r = 2^2 / 6 = 0.67 \rightarrow \text{enlace: premeditado / 01016975}$$

00965893 careful deliberate measured
careful: cuidadoso
careful: prudente
deliberate: deliberado premeditado
deliberate: pausado lento
measured:

$$h_r = 1^2 / 4 = 0.25 \rightarrow \text{no hay enlace}$$

01834448 deliberate
deliberate: deliberado premeditado
deliberate: pausado lento

$$h_r = 1^2 / 2 = 0.5 \rightarrow \text{enlace: premeditado / 01834448}$$

01407902 calculated deliberate measured
calculated:
deliberate: deliberado premeditado
deliberate: pausado lento
measured:

$$h_r = 1^2 / 2 = 0.5 \rightarrow \text{enlace: premeditado / 01407902}$$

Ejemplo 4

establecer un enlace del adjetivo a un concepto del ILI:

(a) Traducción al inglés múltiple y monosémica

Éste es el caso en el que alguna de las palabras inglesas que sirve como traducción al adjetivo español, está asociado a un único concepto. En este caso, como en el caso de traducción única monosémica, el adjetivo español se puede asociar con dicho concepto. Por ejemplo, 'chillón' del inglés se puede traducir por 'screaming', 'loud' y 'garish'. En este caso ocurre que 'screaming' tiene un único significado representado por el índice 00235682, por lo que el adjetivo 'chillón' queda ligado a dicho concepto. Lo mismo ocurre con 'garish', por lo que 'chillón' también queda ligado a 01818583.

(b) Traducción múltiple al inglés con synsets en común.

El último caso de enlace es aquel en el que varias traducciones diferentes están ligadas a algún concepto común. Entonces el adjetivo español original puede ligarse a dicho synset. En el ejemplo 5 puede observarse que las dos traducciones de 'encharcado' ('flooded' y 'swamped') son polisémicas. Sin embargo, tienen un significado común determinado por el índice 00817914 por lo que el adjetivo 'encharcado' se enlaza a dicho concepto.

<p>encharcado Traducciones:</p> <ol style="list-style-type: none"> 1. flooded 2. swamped <p>flooded 00816602 (esp of a boat; covered or overflowing with water) 00817914 (filled to overflowing as if with water)</p> <p>swamped 00064913 ("inundated farmlands"; "a swamped boat") 00817914 (filled to overflowing as if with water) 01398176 (rendered powerless esp by an excessive amount or profusion)</p> <p>Enlace: encharcado / 00817914</p>

Ejemplo 5

3.3. Resultados

El número de adjetivos enlazados al ILI de EuroWordNet está reflejado en la tabla 2.

El diccionario bilingüe español-inglés contiene 3741 adjetivos diferentes, de los cuales

se han podido enlazar 2531 (67.65%). El resto de adjetivos hasta los 3102 enlazados no están contenidos en este diccionario sino que provienen del bilingüe inglés-español.

	Enlaces adjetivo /synset	Número de Adjetivos	Número de Synsets
Traducción de adjetivos ingleses monosémicos	1768	1470	1199
Traducción de las variantes de un synset inglés	871	551	676
Traducción única y monosémica al inglés	592	592	530
Traducción única y polisémica al inglés	1271	474	1209
Traducción múltiple al inglés, algunas monosémicas	1166	931	730
Traducción múltiple al inglés con synsets en común	606	361	415
Datos Globales	5237	3102	3424

Tabla 2

Puede observarse, que los distintos procedimientos han aportado variantes diferentes a un mismo synset español. De hecho, se han creado 5237 enlaces y 3424 synsets diferentes, lo cual implica un ratio de 1.53 variantes españolas por synset. Este dato supera, incluso, el ratio de WordNet 1.5 para la categoría 3 (adjetivos, sin considerar satélites) que es de 1.26 (8102 enlaces a 6440 synsets diferentes).

En todos los casos se ha verificado manualmente una muestra significativa sin encontrarse enlaces incorrectos.

4. Conclusiones

El trabajo aquí descrito proporciona una metodología válida para enlazar no sólo adjetivos, sino cualquier otra categoría sintáctica de cualquier idioma para el que se disponga, además de WordNet, de diccionarios bilingües en formato electrónico.

Algunas de las técnicas que permiten crear enlaces mediante diccionarios ya han sido utilizadas con anterioridad (ver Atserias et al. 1997; Farreres et al. 1998). El tratamiento de unidades monosémicas no aporta novedad alguna. Sin embargo, el tratamiento de unidades polisémicas resulta más problemático y se suele abordar combinando evidencias de muy diversas fuentes. En este sentido, las heurísticas y criterios

que aquí se presentan son independientes entre sí y muestran un grado de efectividad óptimo para los adjetivos españoles, mejorando la precisión en el caso de enlace de unidades polisémicas.

El trabajo supone una confirmación de que el uso de diccionarios bilingües es más adecuado en la construcción de wordnets que la aproximación con diccionarios monolingües, si lo que se pretende es que haya una correspondencia directa con synsets de otros idiomas. Esto es así debido a la dificultad de ligar nuevos synsets a una estructura ya existente.

Sin embargo, el enfoque de diccionarios monolingües permite obtener otro tipo de relaciones que pueden resultar de utilidad para determinadas tareas como la Recuperación de Información. Por ello, resulta muy interesante tratar de combinar las dos aproximaciones. Nuestras primeras tentativas en este sentido muestran la complejidad de la tarea. Así, por ejemplo, se ha estudiado la forma de enriquecer los synsets creados a partir de las agrupaciones sinonímicas y antonímicas obtenidas mediante patrones de definición. Era de esperar que si una agrupación de sinónimos a partir del monolingüe incluye a todas las variantes de un synset español, entonces este synset podría enriquecerse con las variantes no incluidas. Sin embargo esto no siempre es así. En el ejemplo 6 se muestra el synset {bayo} obtenido mediante las técnicas de diccionario bilingüe y la agrupación sinonímica obtenida mediante patrones {bayo, asustado}. Como la agrupación sinonímica es un superconjunto del synset, cabría pensar que *asustado* podría incluirse en el mismo pero se observa que no es así porque el sentido de bayo en el synset y en la agrupación es distinto.

<p>Synset español 0028964 bayo (of a moderate reddish-brown color)</p> <p>Agrupación sinonímica obtenida mediante patrones { bayo, asustado }</p>

Ejemplo 6

5. Agradecimientos

Este trabajo ha sido financiado parcialmente por la CE, proyecto EuroWordNet LE #4003, y por la CICYT, proyecto TIC-96-1243-C03-01.

Agradecemos al grupo de investigación de la UPC dirigido por Horacio Rodríguez y al grupo de la UB dirigido por Toni Martí el apoyo

prestado para la realización de este trabajo.

6. Referencias

- Atserias J., Climent S., Farreres X. Rigau G. and Rodríguez H., 1997. *Combining Multiple Methods for the Automatic Construction of Multilingual WordNets*. In Proceedings of International Conference "Recent Advances in NLP (RANLP'97), Bulgaria.
- Alonge, A., Bertagna, F., Calzolari, N., Rovantini, A. y Zampolli, A., *Encoding Information on Adjectives in a Lexical-Semantic Net for Computational Applications*.
- Bosque, I., Demonte, V., 1999. Gramática Descriptiva de la Lengua Española, vol. 3., Madrid, Espasa Calpe.
- Gonzalo, J., Verdejo, F., Peters, C., and Calzolari, N., 1998^a. *Applying EuroWordNet to Multilingual Text Retrieval*. Special Issue on EuroWordNet, Journal of Computers and the Humanities, 32 (2-3), pp. 185-207. Kluwer Academic Publisher, Dordrecht.
- Gonzalo, J., F. Verdejo, I. Chugur, F. López y A. Peñas, 1998b. *Extracción de relaciones semánticas entre nombres y verbos en EuroWordNet*. Procesamiento del Lenguaje Natural, revista nº 23, pp. 97-103.
- Farreres, X., Rigau, G., and Rodríguez, H., *Using WordNet for Building WordNets*. Proceedings of COLING – ACL. Workshop "Usage of WordNet in Natural Language processing Systems". Montreal, Canadá, 1998.
- Lang, Mervyn F. 1992. *Formación de palabras en español: morfología derivativa productiva en el léxico moderno*. Madrid, Cátedra.
- Márquez, L. and Padró, L., 1997. *A flexible POS tagger using an automatically acquired language model*. In Proceedings of ACL/EACL'97.
- Miller, G., 1990. *WordNet: An Online Lexical Database*. Special Issue of International Journal of Lexicography, 3 (4), pp. 235 - 244.
- Rodríguez, H., Taulé, M., and Turmo, J., 1998. *An environment for morphosyntactic processing of unrestricted Spanish text*. In Proceedings of the 1st International Conference on Language Researches and Evaluation (LREC' 98).
- Soler, C., *Los Adjetivos en SpanishWordNet*, próxima publicación.
- Vossen, P., 1998. *Introduction to EuroWordNet*. Computers and the Humanities, Special Issue on EuroWordNet.
- Vossen, P., Peters, W., Gonzalo, J., 1999. *Towards a Universal Index of Meaning*. In Proceedings of ACL/SIGLEX'99: Standardizing Lexical Resources, University of Maryland.