EXTRACCIÓN DE RELACIONES LÉXICO-SEMÁNTICAS A PARTIR DE PALABRAS DERIVADAS USANDO PATRONES DE DEFINICIÓN[1]

Eneko Agirre y Mikel Lersundi
{eneko, jialeaym}@si.ehu.es
Euskal Herriko Unibertsitatea / Universidad del País Vasco

Resumen

Este trabajo se encuadra dentro de un proyecto de extracción de relaciones léxico-semánticas a partir del análisis de un diccionario monolingüe, y se centra en las relaciones existentes entre las palabras derivadas y sus raíces. En el estudio se incluyen tres sufijos con una sola interpretación y otros 3 sufijos con más de una interpretación. La principal aportación de este artículo es un método para desambiguar las interpretaciones de los sufijos a partir de la definición que se le da en el diccionario a la palabra derivada. En otras palabras, cada interpretación de un sufijo estará unida a uno o varios patrones definitorios. El método consigue una alta precisión al escoger la interpretación correcta de la relación. Con el objetivo de integrar la información extraída en EuroWordNet, hemos explorado un nuevo método para la desambiguación automática de los sentidos tanto de la raíz como de la palabra derivada con buenos resultados. Por el contrario, la representación de las relaciones extraídas en EuroWordNet es todavía un tema abierto.

1. Introducción

Nuestro grupo de investigación IXA está dando una serie de pasos para construir una Base de Conocimiento Léxico del euskara. Este trabajo, como es sabido, es muy difícil de realizar a mano, y nuestra aproximación se basa en extraer las relaciones léxico-semánticas más importantes de un diccionario monolingüe [Boguraev eta Briscoe, 1989]; [Wilks et al., 1989]; [Artola, 1993]; [Castellón, 1993]; [Vanderwende, 1995]; [Richardson, 1997]; [Harabagiu, 1997]; [Rigau, 1998].

Las características especiales que tienen los diccionarios monolingües permiten extraer de los mismos las relaciones léxico-semánticas existentes entre las palabras. Por un lado, las palabras que son entradas léxicas se definen por medio de otras que también serán entrada léxica en el mismo diccionario. A esta propiedad se le denomina "circularidad" de las definiciones, y es muy importante a la hora de extraer información de los diccionarios monolingües. Por otra parte, para definir las entradas léxicas se utilizan patrones o esquemas similares. Por lo tanto, palabras que tengan ciertas propiedades similares estarán definidas por medio de patrones similares.

El diccionario que hemos elegido para llevar a cabo este trabajo es el *Euskal Hiztegia* de Ibon Sarasola [Sarasola, 1996]. Este es un diccionario que está basado en la literatura clásica del euskara, y que hace unos cuantos años se convirtió a formato TEI [Arriola et al., 1997]

En la siguiente sección presentamos el objetivo de este trabajo. Seguidamente se comentan los sufijos elegidos, así como sus posibles interpretaciones. En la sección 4 se presenta el método, seguido de los resultados obtenidos. La sección 6 introduce un método para incorporar las relaciones extraídas en una Base de Conocimiento Multilingüe. Finalmente se comentan trabajos relacionados y se presentan las conclusiones.

2. La derivación

Volviendo a la motivación inicial, nuestra intención era extraer el máximo de relaciones léxico-semánticas utilizando las definiciones del diccionario monolingüe. En [Agirre et al.,

[1] Este trabajo se está realizando gracias a una beca predoctoral otorgada por el Gobierno Vasco [BFI98.217, AE]. Asimismo se encuadra en los proyectos Hiztegia2002 (FEDER 2FD1997-1503) y HERMES (MCYT TIC2000-0335).

2000] describimos el proceso de extracción de relaciones de sinonimia e hiperonimia/ hiponimia, así como de relatores específicos.

Al analizar los resultados obtenidos, nos dimos cuenta que en las categorías de nombres y verbos conseguíamos relacionar muchas palabras, pero no ocurría lo mismo para los adjetivos. Como es sabido, la relación jerárquica de hiperonimia/hiponimia no se da entre los adjetivos [Fellbaum, 1999], ya que los adjetivos no se definen por medio del tipo de definición aristotélica clásica *genus et differentia specifica*.

Con el fin de completar el ámbito de los adjetivos, comenzamos a buscar entre las definiciones patrones que incluyeran unos relatores específicos (una serie de patrones sintácticos especiales). Cuando comenzamos a encontrar estos relatores específicos, nos dimos cuenta que muchos de ellos tenían una estrecha relación con el sufijo que aparecía en la entrada léxica (en los casos en que la entrada léxica fuera un término derivado). Esto es, que las entradas léxicas derivadas por un mismo sufijo tenían en la definición un mismo relator específico, un tipo de definición muy especial y siempre el mismo. Además, este relator específico tenía mucho que ver con el significado del sufijo, y, gracias a él, podríamos establecer la relación entre el derivado y su raíz.

Este artículo describe el estudio que hemos realizado sobre la posibilidad de extraer la relación léxico-semántica existente entre el derivado y su raíz, basándonos en las posibles interpretaciones del sufijo y de los patrones definitorios que sigue la definición de la palabra derivada. Aunque en un principio el objetivo era completar el escaso número de relaciones que teníamos para los adjetivos, ampliamos el estudio a sufijos derivativos nominales y verbales, dado que en euskara la derivación está muy extendida y permitirá extraer un número amplio de relaciones que no siempre están explícitas en la definición.

La relación léxico-semántica que obtenemos es la que existe entre el término derivado y la raíz sobre la que se ha construido ese término. Esta es una relación simple que viene dada por el significado del sufijo. Cuando el sufijo no es ambiguo no hace falta más procesamiento, pero en el caso de que admita varias interpretaciones nuestra hipótesis de trabajo es que podremos escoger la interpretación correcta atendiendo a la forma en que está definida la palabra derivada en el diccionario.

3. Sufijos que se han estudiado

3.1. Sufijos no ambiguos

Estos son los sufijos que no tienen más que un significado. Al aparecer unidos al radical, siempre crean un derivado que tendrá el mismo significado: el agente "-gile", el que crea nombres abstractos que muchas veces son oficios "-gintza" (oficio o local donde se ejerce), y el factitivo "-arazi" que crea verbos factitivos.

En los dos primeros, vemos que el significado del derivado tiene mucho que ver con el sufijo, y que este significado no nos da la verdadera relación léxico-semántica existente entre el derivado y el radical. La verdadera relación existente entre el derivado y el radical es la de *objeto prototípico*: en el caso de "-gile" hay una persona cuyo trabajo o acción consiste en hacer aquello que está especificado por el nombre que constituye la raíz del derivado; y en el caso de "-gintza" tenemos un oficio o acción que consiste en crear el objeto designado por el nombre que es la base del derivado:

artikulu + gile = articulista (persona que escribe artículos);
amu + gile = anzolero (persona que hace anzuelos);
aulki + gintza = sillería (oficio para construir sillas);
artikulu + gintza = producción de artículos, escritos.

El otro sufijo que tiene un solo significado es el factitivo "-arazi". Este sufijo se le une a raíces verbales, y crea un verbo factitivo:

janarazi = hacer comer.

3.2. Sufijos ambiguos

Denominamos sufijos ambiguos aquellos que tienen más de una interpretación. En el caso de nuestra investigación hemos trabajado con tres: "-garri", y "-tsu" que forman adjetivos, y "-keria" que forma nombres.

3.2.1. "-garri"

El sufijo "-garri" se le une a las bases verbales y forma adjetivos.

En cuanto al significado tiene tres valores: uno activo y dos pasivos:

- "-garri" activo: el derivado que se forma con "-garri" activo, diremos que es el *agente* de la acción del verbo; por tanto, diremos que entre

el derivado y la raíz tendremos la relación *agente*[2]:

baregarri. (calmante, tranquilizante) Baretzen duena. (Que calma).

- "-garri" pasivos hay dos, pero el caso es que los dos están unidos con ser *objeto* de la acción del verbo[3]:

  - uno *merece* ser objeto de la acción que le corresponde. Diremos que entre el derivado y la raíz existe la relación *objeto-merecer*:

    aztergarri. (investigable, analizable) Aztertua izatea merezi duena. (digno de estudio)[4]

  - el otro puede ser objeto de la acción que le corresponde. Diremos que entre el derivado y la raíz existe la relación *objeto-poder*:

    jasangarri. (soportable) Jasan daitekeena, eramangarria. (Que se puede soportar, llevadero)

3.2.2. "-keria"

Las bases de los nombres derivados construidos con el sufijo "-keria" son las siguientes:

- adjetivo que exprese una cualidad reprobable, que indique algo que no es bueno. De todos modos, hay veces que estas bases adjetivas son colores, pero entonces el significado no tiene nada que ver con los mismos (berdekeria = obscenidad, verdulería; zurikeria = falsedad, engaño) [berde = verde; zuri = blanco]
- nombres animados, pero no humanos; normalmente suelen ser nombres de animales que tienen alguna cualidad reprobable (txerrikeria = marranería, marranada, cochinada; astakeria = burrada)
- nombres con el rasgo semántico humano (umekeria = niñería, chiquillada; sorginkeria = brujería)

Los derivados formados con el sufijo "-keria" son nombres.

En cuanto a los posibles significados, tenemos dos:

- aquello que expresa una *cualidad* que se reprocha:

  nagikeria. (vagancia) Nagitasuna gaitzesgarritzat markatua. (vagancia con sentido peyorativo)

- aquello que expresa la consecuencia de una *acción*: indarkeria, bortxakeria

  maltzurkeria. (falsedad, fraude) Maltzurrari dagokion egitea (Acción propia del astuto).

Por lo que se ve, lo que conseguimos con esta distinción es separar lo que son *propiedades*, con lo que son *eventos*.

Aunque en teoría sea así, muchas veces nos será difícil saber cuál es la frontera entre estas dos interpretaciones. A veces admite las dos interpretaciones, y el lexicógrafo incluye dos definiciones diferentes separadas por un signo de ";". Cada vez que aparece este signo dentro de lo que denominamos definición, lo consideraremos como definición diferente, a pesar de que la distinción no sea lo sumamente grande como para que aparezca como acepción diferente.

3.2.3. "-tsu"

El sufijo "-tsu" se une a nombres y forma adjetivos.

El caso del sufijo "-tsu" es algo especial, ya que a lo largo de la historia se ha dicho que tiene un solo significado (*abundancia*), pero nosotros, analizando los derivados que crea, hemos pensado que pudiera tener más de una interpretación (aunque el matiz de abundancia lo tendrán todos sus derivados):

- uno que representa una *situación* o un *estado*

  ekaiztsu. (tormentoso) Ekaitzak daudena, ekaitza dagoena. (donde hay tormentas)

- y otro que expresa una *propiedad* que se tiene. Esta propiedad se puede poseer de en un número llamativo, o de un modo llamativo

  goroldiotsu. (musgoso) Goroldioz estalia. (cubierto de musgo)

4. Método

Antes de nada debemos saber cuáles son las entradas léxicas derivadas que aparecen en el diccionario. Para ello, analizamos estas entradas con el analizador MORFEUS [Urkia, 1997], que reconoce la raíz y el sufijo. Después de hacer esto, decidimos con qué sufijos vamos a

---

[2] En realidad, la relación *agente* existirá entre el nombre del cual este adjetivo es complemento y el verbo que es la base de este derivado adjetival.

[3] Esta relación también será entre nombre y verbo. Ver nota 2.

[4] Parece ser que aquí es diferente en español y en euskara, ya que en el primero aparece "poder" y en el segundo "merecer".

trabajar y estudiamos las diferentes acepciones que pueden tener estos sufijos.

Una vez llevados a cabo los pasos anteriores, para cada posible interpretación del sufijo se han estudiado los patrones definitorios que siguen las definiciones de las palabras derivadas con esa interpretación. De ese estudio se sintetizan una serie de reglas de las Gramáticas de Restricciones [Tapanainen, 1996]

A continuación expondremos la forma de las reglas y la cómo funcionan. Para ilustrarlo utilizaremos el siguiente ejemplo:

"entzungarri. Entzuna izatea merezi duena.". (escuchable. Que merece ser escuchado)

Tenemos una regla con la cual logramos poner la etiqueta *&OBJEKTU-MEREZI-20* al participio definido que aparece en la definición, en caso de que a su derecha en primer lugar aparezca la forma nominal del verbo "ser", en segundo lugar a la derecha el nombre "merezi" (merecer), y en tercer lugar un verbo auxiliar.

Al aparecer en la definición la marca *&OBJEKTU-MEREZI-20* sabremos que entre el derivado "entzungarri" (escuchable) y la base "entzun" (oír) tendremos una relación de *objeto-merecer*.

Para que estas reglas puedan funcionar, anteriormente tenemos que pasar todas las definiciones por el analizador antes citado MORFEUS, que lematiza y etiqueta con la categoría gramatical las palabras de las definiciones.

Aparte de tratar directamente las definiciones, también hemos tratado la "extensión de la sinonimia". En algunos casos el adjetivo se define mediante un sinónimo, lo que no permite desambiguar la interpretación del sufijo. Por fortuna, frecuentemente el sinónimo suele ser a su vez una palabra derivada mediante el mismo sufijo, y esto permite que escojamos la interpretación que se haya dado para el sinónimo. En otras palabras, cuando la definición se haga mediante un sinónimo con el mismo sufijo, se puede extender la interpretación del sinónimo.

Por ejemplo: la definición de "doloragarri" es el sinónimo "mingarri". De la definición de "mingarri" (que produce dolor), hemos deducido que la interpretación del sufijo es la de *agente*, y, por tanto, podemos extender esta interpretación a "doloragarri" (doloroso *agente* dolor).

5. Resultados

Una vez analizadas las definiciones de los derivados construidos hemos tomado muestras al azar para evaluar los resultados obtenidos. Las muestras rondan el 8% - 10% del total de derivados para cada sufijo, esto es, alrededor de 30/40 palabras. Primero analizamos el numero de interpretaciones desambiguadas, y al final damos los resultados de precisión.

5.1. "-garri"

A pesar de que los lemas de la muestra son 30, aparecen más términos para marcar, ya que un mismo lema puede estar definido por una definición propiamente dicha y por un sinónimo, o por más de un sinónimo, lo que puede llevar a tener más de una interpretación por palabra derivada. Conseguimos desambiguar la relación existente entre el derivado y su base o raíz para el 75% de la muestra. Unas veces conseguimos desambiguar esa relación directamente, gracias a la definición; y, otras veces, gracias a la extensión de la sinonimia.

Agente	15		
objeto-merecer	5		
objeto-poder	1		
Sinónimos	10	marca por extensión	6
		no conocidos	4
sin desambiguar	5		
Total	36	desambiguados	27

Tabla 1. Resultados del sufijo "-garri"

5.2. "-keria"

Los elementos que desambiguamos directamente, gracias a la definición son 23. En otros casos (11) el lema que es entrada léxica está definido por medio de un sinónimo, y en 9 casos conocemos cuál es la relación de este sinónimo derivado y su base. En total, desambiguamos el derivado el 87,18% de los casos.

acción	16		
cualidad	9		
sinónimos	11	marca por extensión	9
		no conocidos	2
sin desambiguar	3		
Total	39	desambiguados	34

Tabla 2. Resultados del sufijo "-keria"

5.3. "-tsu"

De estas 40 definiciones hay 4 que no tienen ninguna marca (por lo tanto no pueden ser desambiguadas), y hay otras tres que no se pueden analizar porque son definiciones erróneas. Logramos desambiguar la relación el 83,78% de los casos.

estado	2		
posesión	16		
sinónimos	14	marca por extensión	13
		no conocidos	1
no derivados	1		
sin desambiguar	4		
defs. con errores[5]	3		
total	37	desambiguados	31

Tabla 3. Resultados del sufijo "-tsu"

5.4. Precisión

Para evaluar la precisión de los resultados de este proceso de desambiguación, dos lingüistas han etiquetado a mano las muestra tomada al azar. En la tabla siguiente se dan los datos acerca de la conformidad entre lingüistas y el método. A la hora de medir los resultados hemos dejado de lado los elementos que el método no conoce, y nos hemos basado en aquellos que conoce.

	-garri	-keria	-tsu
Dos lingüistas	%90	%82,76	%83
Lingüista A y método	%100	%96,15	%82
Lingüista B y método	%100	%84,62	%80
Los tres	%100	%80,77	%72

Tabla 4. Resultados de la comparación

En el caso de "-garri" el método desambigua bien en el 100% de los casos que marca. Justo los casos en los que los lingüistas discrepan son casos que no están marcados automáticamente. Esto explica que la comparación con la máquina consiga más alta precisión.

Para el sufijo "-keria", ya hemos comentado antes que es difícil establecer la frontera entre los casos que son *cualidad* y los que expresan *acción*. De hecho un 33% de los derivados recibe las dos interpretaciones por parte de algún lingüista.

Y, finalmente, los resultados confirman que verdaderamente hay algún tipo de diferencia entre los derivados formados por el sufijo "-tsu" (83% de los casos los lingüistas están de acuerdo en las interpretaciones); pero, a su vez, ese 83% nos dice que habrá que refinar la interpretación que nosotros hemos lanzado, y que deberá ser revisada en trabajos sucesivos.

Los resultados muestran que hacer estas distinciones semánticas no es fácil para las personas. Concretamente "-keria" y "-tsu" consiguen solamente el 80% de acuerdo entre lingüistas. Con todo, nuestro método consigue resultados comparables a las personas.

6. Desambiguación de acepciones e integración de las relaciones en eurowordnet

Como ya se ha citado en la introducción, el objetivo final de este trabajo es la construcción y enriquecimiento de una Base de Conocimiento Léxica para el euskara. Esta Base de Conocimiento estará ligada a EuroWordNet [Vossen, 1998], que es una Base de Conocimiento Léxico Multilingüe que aparte del euskara incluye el inglés, castellano, italiano, holandés, alemán, francés, checo, estonio y catalán.

Para poder integrar las relaciones en dichas bases de conocimiento hacen falta dos cosas:

1. Que la relación pueda expresarse en el esquema de representación de las bases de conocimiento.
2. Que la relación se dé entre conceptos (que están basados en las acepciones de las palabras) en lugar de palabras.

En concreto hemos realizado un estudio de estos dos apartados para EuroWordNet, que exponemos a continuación.

6.1. Representación de las relaciones extraidas en EuroWordNet

Al estudiar las relaciones descritas en la sección 3, observamos que no se trata de relaciones simples. Por ejemplo, estudiemos el adjetivo "aztergarri":

[5] Estos errores provienen de la conversión del diccionario a formato TEI. No los vamos a tener en cuenta a la hora de obtener resultados, ya que es un error anterior.

aztergarri. (investigable, analizable) Aztertua izatea merezi duena (digno de estudio)

A la relación entre los dos hemos llamado *objeto-merecer*: "aztergarri" *objeto-merecer* "aztertu". En un principio parece que podría ser posible simplificar la relación y reflejarla mediante la relación *objeto* de EuroWordNet.

Por desgracia, EuroWordNet solamente dispone de relaciones binarias entre conceptos, y no es posible componer las relaciones. Además la relación de *objeto (paciente)* se da solamente entre nombres o entre nombre y verbo. En nuestro caso tenemos un adjetibo, y en todo caso la relación de *paciente* se daría entre el nombre del cual sea complemento "aztergarri" y el verbo que es la raíz de "aztergarri": "aztertu" (analizar).

Por razonamientos similares llegamos a la conclusión de que ninguna de las relaciones extraídas se puede representar mediante relaciones de EuroWordNet. En un futuro pensamos extender el esquema de representación de EuroWordNet para acomodar relaciones complejas.

6.2. Desambiguación de acepciones

Como ya hemos citado anteriormente, es necesario discriminar cuál de los sentidos de la palabra derivada y de la raíz es aplicable. Por ejemplo: la raíz de "amugile" (el que hace anzuelos) tiene al menos dos acepciones, "anzuelo" propiamente dicho y "anzuelo" en sentido figurativo, pero la relación sólo se da entre "amugile" y la primera acepción. Hay que hacer notar que si se añade a EuroWordNet una relación entre dos conceptos en euskara, esa relación también estará disponible para sus traducciones. Por ejemplo: al integrar la relación "garagardogile" → *objeto prototípico* → "garagardo" (cervecero, cerveza) esta relación también estará disponible para el castellano, e incluso para brewer en inglés, aunque en su raíz no se mencione la cerveza.

La desambiguación de acepciones es una tarea difícil y de alto índice de error. Esto se agrava en nuestro caso, dado que la parte en euskara de EWN está en construcción, y actualmente incluye numerosas acepciones erróneas synset-palabra[6] que se están limpiando. Además algunas palabras no están incluidas todavía.

Por ello, en lugar de usar otro tipo de algoritmos de desambiguación, hemos decidido explorar un nuevo tipo de método que trata de explotar el que muchas palabras derivadas tienen traducciones que a su vez son derivadas.

En otras palabras, vamos a comparar la palabra derivada y su raíz (en euskara) con sus traducciones al ingles y castellano que sean derivadas, y vamos a tomar la intersección de los conceptos para las raíces en los demás idiomas.

Por ejemplo: *zilargile – zilar*

	derivado	conceptos	raiz	conceptos
Eusk.	zilargile	<silversmith>	zilar	<silver>
Cast.	platero	<silversmith>	plata	<tableware> <silver>

Tabla 5. Comparación entre lenguas

La intersección de las acepciones para "zilargile" y "platero" es el mismo concepto, lo que confirma que el concepto ligado a "zilargile" era correcto (hay que recordar que hay errores en la versión actual de EuroWordNet para el euskara). Por otro lado, la intersección de los conceptos para la raíz permite saber que la acepción correcta de plata en platero (entendido como persona que labra la plata[7]) es la del metal, y no la de vajilla.

Para evaluar el potencial de esta técnica de desambiguación hemos analizado los resultados para un total de 100 pares derivado-raíz extraídos para el sufijo "-gile". La tabla 6 indica que el algoritmo es capaz de relacionar tanto la raíz como la palabra derivada en un 16% de los casos, y al menos uno de los dos en un 31% de los casos (la derivada en un 22% y la raíz en un 9%). La cobertura tan baja se debe sobre todo al estado preliminar de EuroWordNet para el euskara.

[6] Estas acepciones fueron añadidas automáticamente usando diccionarios bilingües.

[7] La versión de EuroWN usada en este trabajo no incorpora la acepción de "*mueble para guardar la vajilla*" ni la de "*persona que vende plata*". Si esas acepciones estuvieran incluidas no afecta el resultado, ya que la intersección para los derivados sólo daría la acepción de "*persona que labra la plata*" para platero.

	%
Ninguno	53,0
Derivada	22,0
Raíz	9,0
Ambos	16,0

Tabla 6. Cobertura de la desambiguación de acepciones de un total de 100 ejemplos

En cuanto a la corrección, un 97% de los conceptos son correctos en el caso de las palabras derivadas (de los que algo más de la mitad eran monosémicas en euskara), y un 80% de conceptos son correctos en el caso de las raíces. Adicionalmente, un 15% de casos devuelven más de un concepto, de los que al menos 1 es correcto.

	Derivado (%)		Raíz (%)	
	Total	Mono.	Total	Mono.
Correcto	97,3	52,6	80,7	7,7
Correcto entre varios			15,4	
Incorrecto	2,7		3,8	

Tabla 7. Precisión de la desambiguación de acepciones

8. Comparación con otros trabajos

Pocos trabajos actuales se han preocupado de incluir las relaciones léxico-semánticas que se dan entre derivados y sus raíces en bases de conocimiento, y no somos conscientes de ninguno que trate sobre la desambiguación automática de las interpretaciones de los sufijos derivativos.

Penthedourakis y Vanderwende [1993] realizan un exhaustivo trabajo sobre la incorporación automática de este tipo de relaciones en una base de conocimiento. Son capaces de identificar la raíz correctamente en un 94% de los casos, y de desambiguar correctamente la acepción de la raíz en un 65% de los casos polisémicos. Aunque reconocen diferentes interpretaciones para algunos sufijos (p.ej. "-er") no mencionan cómo llegan a desambiguar esas interpretaciones. El método que proponen se basa en las relaciones semánticas extraídas de las definiciones.

Un trabajo más reciente [Peñas et al., 2000] extrae relaciones para palabras derivadas que no siempre coinciden con la raíz (p.ej. admirable – admiración). Toman todas las acepciones de la raíz como buenas, lo que introduce un 10% de acepciones erróneas. No mencionan la posibilidad de interpretaciones múltiples.

Nuestro trabajo se fundamenta por ahora en el análisis de los patrones definitorios, lo que esta un nivel de sofisticación inferior al sistema propuesto por Penthedourakis y Vanderwende. Próximamente vamos a abordar la extracción automática de las relaciones presentes en las definiciones, lo que permitiría añadir más abstracción a nuestro sistema. Por otro lado, nuestro análisis indica que dado el estado actual de EuroWordNet para el euskara, el error al tomar todas las acepciones de la raíz sería muy superior al 10%.

9. Conclusiones y trabajo futuro

Este trabajo describe un método para extraer la relación léxico-semántica existente entre las palabras derivadas y sus raíces a partir del análisis de las definiciones. La principal aportación de este artículo es un método para desambiguar las interpretaciones de los sufijos a partir de la definición que se le da en el diccionario a la palabra derivada, que consigue una precisión entre el 80% y el 100%, y una cobertura de media del 82%. En total hemos extraído cerca de 1.600 relaciones a partir del análisis de tres sufijos con una sola interpretación, y otros tres sufijos con mas de una interpretación. Además hemos explorado la desambiguación automática de los sentidos tanto de la raíz (80% de precisión) como de la palabra derivada (97% de precisión), lo que permite integrar la información extraída en EuroWordNet. Ello permite enriquecer las relaciones, no solamente del euskara, sino de los otros idiomas enlazados a EuroWordNet. De la misma forma si este trabajo se llevara a cabo para otros idiomas, permitiría mayor cobertura de los fenómenos derivativos.

En cuanto a la representación de las relaciones léxico-semánticas, los sufijos estudiados reflejan relaciones complejas que no pueden representarse mediante las relaciones simples que ofrece EuroWordNet. Actualmente estamos estudiando posibles extensiones a la representación.

En un futuro también pensamos profundizar en el análisis sintáctico de las definiciones para extraer una red de relaciones semánticas complejas. Por otro lado, exploraremos el uso de técnicas de aprendizaje automático para desambiguar las interpretaciones de los sufijos. Este método puede ser útil para escoger la interpretación de palabras derivadas que no están en el diccionario.

Agradecimientos

Agradecemos a Itziar Aduriz y Jose Mari Arriola su colaboración en la desambiguación de las interpretaciones.

Bibliografia

Agirre, E.; Ansa, O.; Arregi, X.; Artola, X.; Díaz de Ilarraza, A.; Lersundi, M.; Martínez, D.; Sarasola, K.; Urizar, R.; 2000, "Extraction of semantic relations from a Basque monolingual dictionary using Constraint Grammar", *EURALEX International Congress*, Ulrich Heid, Stefan Evert, Egbert Lehmann, Christian Rohrer editores, Stuttgart, Universität Stuttgart, 9-2 (2000) 641-650.

Alegria, I., Artola, X., Sarasola, K., Urkia, M., 1996, "Automatic morphological analysis of Basque", *Literary & Linguistic Computing*, Oxford, Oxford University Press, vol. 11, n. 4 (1996) 193-203.

Arriola J.M., Artola X., Gojenola K., Soroa A., 1997, "TEI: testu-kodeketarako gidalerroak", *EKAIA* 7 (1997) 121-139.

Artiagoitia, X., 1995, "-garri atzizkiaren izaera bikoitzaz. Zergatik den maitagarria bezain mingarria", *Anuario del Seminario de Filología Vasca 'Julio de Urquijo'*, XXIX-2 (1995) 161-198.

Artola, X., 1993, *HIZTSUA: Hiztegi-sistema urgazle adimendunaren sorkuntza eta eraikuntza / Conception et construction d'un système intelligent d'aide dictionnariale (SIAD)*, Donostia, UPV-EHU, tesis doctoral.

Azkarate, M., 1990, "Euskal –garri vs. latinezko –bilis: bi balio desberdin", *Patxi Altunari omenaldia (Mundaiz-5)*, Elixabete Perez Gaztelu eta Patri Urkizu Sarasua (editores), Deustuko Unibertsitatea, Donostia, 5 (1990) 19-39.

Azkue, R. M. ,1923-1925, *Morfología Vasca*, Bilbao, La Gran Enciclopedia Vasca, 1969 (3 tomos)

Boguraev, B. and Briscoe, T. (editores), 1989, *Computational Lexicography for Natural Language Processing*, New York: Logman.

Castellón, I., 1993, *Lexicografía Computacional: Adquisición automática de conocimiento léxico*, Universidad de Barcelona. Tesis doctoral.

Felbaum, C., 1999, "Semantics via Conceptual and Lexical Relations", *Bradth an Depth of Semantic Lexicons*, editado por Eveline Viegas, Kluwer Academic Publishers, Netherlands, (1999) 247-262.

Gràcia, L.; Fullana, O.; Turon, L.; Suñer, A.; Cabré, M. T.; Lorente, M.; Bayà, M. R.; Bernal, E.; Varela, S.; Haouet, L.; Martín, J.; Azkarate, M.; Odriozola, J. C.; Zabala, I.; Perez, E.; 2000, *Configuración morfológica y estructura argumental: léxico y diccionario,* Zarautz, Euskal Herriko Unibertsitatea.

Harabagiu, S. M., 1997, *WordNet-Baed Inference of Textual Context, Cohesion and Coherence*, University of Southenrn California. Tesis doctoral.

Penthedourakis, J. and Vanderwende, L., 1993. *Automatically Identifying Morphological Relations in Machine-Readable Dictionaries.* Microsoft internal report MSR-TR-93-06.

Peñas, A., Chugur, I., Gonzalo, J. y Verdejo, F., 2000, *Incorporación de adjetivos al WordNet Español Procesamiento del Lenguaje Natural (SEPLN'2000)*, 24 (2000)

Richardson, S. D., 1997, *Determining similarity and inferring relations in a Lexical Knowledge Base*, New York. Tesis doctoral.

Rigau, G., 1998, *Automatic Acquisition of Lexical Knowledge form MRDs*, UPC. Tesis doctoral.

Sarasola, I., 1996, *Euskal Hiztegia*, Donostia, Kutxa Gizarte- eta Kultur-Fundazioa.

Tapanainen, P., 1996, *The Constraint Grammar Parser CG-2*, University of Helsinki Publications, n. 27 (1996).

Urkia M., 1997, *Euskal Morfologiaren tratamendu informatikorantz*, Gasteiz, UPV/EHU, Tesis doctoral.

Vanderwende, L.; 1995, *The analysis of noun sequences using semantic information extracted from on-line dictionaries*, Geogetown University, Washington, DC. Tesis doctoral.

Vossen, P., 1998, *Introduction to EuroWordNet*. Computers and the Humanities, special issue on EuroWordNet

Wilks, Y., D. Fass, C. Guo, J. McDonald, T. Plate, B. Slator, 1989, "A Tractable Machine Dictionary as a Resource for Computational Semantics", *Computational Lexicography for Natural Language Processing*, Boguraev, B., and T. Briscoe (editores), Longman, London, (1989) 193-228.