Extracción de relaciones semánticas mediante una gramática de restricciones

E. Agirre, O. Ansa, X. Arregi, J.M. Arriola,
A. Díaz de Ilarraza, M. Lersundi, A. Soroa, R. Urizar
Facultad de Informática
Apdo. 649
20.080 – Donostia
eneko@si.ehu.es

Resumen

Este artículo da cuenta de los primeros resultados de un proyecto destinado a extraer relaciones semánticas de las definiciones del diccionario monolingüe en euskara *Euskal Hiztegia*. La técnica para extraer dichas relaciones utiliza el mecanismo de *mappings* de las gramáticas de restricciones, que se apoyan en la información morfosintáctica ya provista por el analizador morfológico MORFEUS. Se presentan los resultados referentes a la extracción de las relaciones de sinonimia e hiperonimia, en la que se consigue etiquetar el 85% de las definiciones de los sustantivos. Cubrimos el 91,3% de las definiciones con genus/sinónimo con un error en la extracción del 5%, aunque el 98,9% de las definiciones etiquetadas tienen al menos un genus/sinónimo correcto. Los resultados obtenidos, así como el pequeño esfuerzo necesario, nos indican que es factible extraer relaciones semánticas a partir de la información morfosintáctica proporcionada en un tiempo razonable, ya sea para sustantivos, verbos o adjetivos.

1. Introducción[1]

Este artículo se enmarca dentro de la utilización de diccionarios de uso humano en soporte informático para la construcción de lexicones y bases de conocimiento diccionariales (Amsler, 1981; Calzolari & Picchi, 1986; Boguraev & Briscoe, 1989; Castellón, 1993), y da cuenta de los primeros resultados de un proyecto destinado a extraer relaciones semánticas de las definiciones del diccionario monolingüe en euskara *Euskal Hiztegia* (Sarasola, 1996)[2]. Este diccionario ha sido previamente convertido a un formato tratable por máquina y está disponible en el formato que sigue las directrices de la Iniciativa para la Codificación de Textos (Text Encoding Initiative –TEI–) (Burnard & Sperberg-McQueen, 1990). Las definiciones han sido analizadas morfológicamente mediante el uso del analizador morfológico MORFEUS (Urkia, 1997) y la gramática de restricciones (Karlsson et al., 1995; Tapanaimen, 1996). La técnica para extraer dichas relaciones se basa en el uso de los *mappings* de las propias gramáticas de restricciones, que aprovechan la información morfosintáctica ya inferida en pasos anteriores.

El trabajo aquí presentado se centra en la extracción de las relaciones de sinonimia e hiperonimia para las definiciones de sustantivos. El proyecto, de todas formas, va a extenderse a otras relaciones semánticas, tanto para sustantivos como para verbos y adjetivos. Es de reseñar que el euskara es un idioma que marca el caso en la morfología, lo que permite al sistema disponer de información muy valiosa a la hora de extraer las relaciones semánticas.

La estructura del artículo es la siguiente. Primeramente comentaremos las características generales del diccionario *Euskal Hiztegia* y los resultados de la conversión al formato TEI. Seguidamente pasaremos a presentar el método seguido para la extracción. En la sección tercera y cuarta analizamos la estructura del

[1] Este trabajo ha recibido ayudas del CYCIT (TIC96-1243-C03-02 proyecto ITEM) y de la UPV (141.226-TA073/96).

[2] Este mismo diccionario se llamaba *Hautalanerako Euskal Hiztegia* en un edición anterior.

campo de definición (lo que permite detectar sinónimos), y del texto definitorio en si (para la extracción de los genus). La sección quinta presenta alguna de las reglas de restricción usadas para la extracción. Finalmente se presentan los resultados, y se comentan las conclusiones y el trabajo futuro.

2. Características generales del diccionario Euskal Hiztegia

Se trata de un diccionario monolingüe en euskara, normativo y repositorio del euskara estándar, cuya confección se ha sustentado principalmente en la tradición literaria. Contiene 32.139 entradas. He aquí un artículo, tal y como aparece en la versión del MRD:

biltoki. iz. (*1745, 1824). **1.** Biltzeko tokia; berezk. talde edo elkarte bateko lagunak biltzen diren etxea edo lekua. *Denen biltoki zen hiri nagusira. Santa Kruzek esan zien biltokirik hoberena Aia zutela.* Ik. **batzoki**. **2.** (*1745, 1855). *Zah*. Antzokia. **3.** *G. er.* Biltegia.

El artículo se compone de los siguientes campos: entrada, categoría, fecha, definición, relaciones, ejemplo, abreviaturas y subentradas. En las entradas con una sola acepción el campo definición se compondrá del texto de la definición. Cuando hay más de una acepción los lexicógrafos se sirven de diversos símbolos para diferenciar entre las distintas acepciones: mediante números, letras mayúsculas y letra negrita para agrupar éstas bajo sentidos generales, etc.

El diccionario fuente, al estar dirigido al uso humano, presenta la problemática de las versiones impresas, además de la carencia de una estructuración sistemática de la información. Para poder estructurar la información se ha utilizado una gramática de cláusulas definidas (DCG) que recoge la estructura de las entradas. Para representar esa estructura analizada se han seguido las directrices de la Iniciativa para la Codificación de Textos (Text Encoding Initiative –TEI–) que codifica los textos en SGML. Para más información en lo tacante a la conversión refiérase a (Arriola et al., 1995; 1996; 1997; pte.; Arriola & Soroa, 1996).

Los resultados de esta conversión han sufrido una notable mejora desde el 85% de cobertura y 10% de tasa de error presentados en los trabajos antes mencionados, debido a un cambio en la estrategia de análisis por otra más flexible y modular basada en la estructura de acepciones de cada entrada consiguiéndose los resultados expuestos en las tablas 1 y 2.

	Cobertura
Total	98.49%
Parcial	1.50%
Nada	0.01%

Tabla 1: recuperación de las definiciones.

	Precisión
Bien	77%
Aceptable	20%
Error	3%

Tabla 2: recuperación de las definiciones.

En lo que se refiere a las definiciones, esto significa que el 77% de los análisis de los artículos son totalmente correctos, y que el 97% son correctos casi en su totalidad (en cuanto a entrada, categoría, acepción, definición y ejemplo, sólo habría errores en el reconocimiento de fecha o códigos gramaticales, y en menor medida, caracteres erróneos en la definición, o parte de la definición –no relevante– elidida).

3. El método seguido

La extracción de las relaciones semánticas a partir de diccionarios monolingües es posible gracias a varias características de éstos. Por un lado, la *circularidad* en las definiciones de estos diccionarios hace que para definir una palabra se utilicen otras que a su vez tienen entrada en el diccionario. Por otro lado, dichas definiciones –ya sea consciente o intuitivamente– se expresan mediante *fórmulas* o *esquemas* relativamente fijos en los que las entradas se definen utilizando otras de significado más simple.

Esta expresión mediante fórmulas de las definiciones hace posible la extracción de diferentes tipos de relaciones interconceptuales. Entre las *relaciones paradigmáticas* extraíbles de un diccionario monolingüe cabe destacar la **sinonimia** y **antonimia** (sobre todo en adjetivos), las relaciones taxonómicas de **hiperonimia** e **hiponimia**, así como las de **taxonomía pura** (*tipo/clase de*), **meronimia** (*parte de, trozo de, grupo/conjunto de*), de **carencia** (*falta de*), **referenciales**, de **graduación** (*muy*), etc. También pueden extraerse diversas *relaciones sintagmáticas* como de **derivación** (sustantivos deverbales, sustantivos derivados de adjetivos,…), o

diferentes *relaciones de caso* del tipo **finalidad prototípica**, **agente prototípico**, etc. (Artola, 1993).

El primer paso en el proceso de extracción de las relaciones semánticas en el *Euskal Hiztegia* consiste en la **tokenización, análisis morfosintáctico** y **etiquetado** de las unidades que constituyen la definición. Al ser el euskara una lengua aglutinante de gran complejidad morfológica, resulta imprescindible un análisis morfosintáctico previo. Para ello utilizamos el analizador morfológico para el euskara MORFEUS (Urkia, 1997), desarrollado por el grupo de investigación IXA.

El siguiente paso consiste en analizar/estudiar las diferentes estructuras de las definiciones, con objeto de localizar las fórmulas o esquemas de definición. Las definiciones en el Euskal Hiztegia son a veces bastante largas, por lo que no tratamos de determinar el esquema de las definiciones en su totalidad, sino solamente la sintaxis en torno al patrón definitorio. Para ello, se extrae una muestra aleatoria de 200 definiciones que son estudiadas con objeto de establecer los esquemas o fórmulas utilizados en las definiciones. Una vez establecidas dichas estructuras, se elaboran reglas utilizando la Gramática de Restricciones (Karlsson et al., 1995; Tapanaimen, 1996) para detectar los genus y sinónimos. Dichas reglas se aplican sucesivamente en la muestra de 200 definiciones, y, al mismo tiempo, se van refinando en base a los resultados obtenidos. A continuación, se aplican las reglas sobre una nueva muestra aleatoria para evaluar la bondad de las reglas.

4. El análisis de las estructuras

Existen básicamente tres modos de definir una entrada léxica:

i. por medio de un sinónimo:

   andurreria: harrokeria.[3]

ii. mediante una descripción con la fórmula clásica *genus + differentia*:

   algara: barre oihua.

iii. a través de *relatores específicos*, es decir, fórmulas sintácticas que relacionan una palabra de la definición con la entrada base:

   herabetasun: herabe denaren nolakotasuna.[4]

El campo definición, sin embargo, no siempre incluye una única definición. El signo de puntuación ";" también separa estructuras de definición dentro de un mismo campo.

foruzaletasun: foruen zaletasuna; foruzaleen dotrina.

Incluso dentro de estas estructuras de definición existen diferentes combinaciones. En una misma estructura de definición pueden a su vez combinarse **sinónimos** (**SIN**) y **definiciones** (**DEF**) separados por comas. Las combinaciones más comunes son las siguientes:

SIN	**DEF**
SIN, SIN	**DEF, SIN**
SIN, SIN, SIN	**SIN, DEF**
	DEF, DEF

- Un sustantivo puede definirse mediante un sinónimo (**SIN**), en cuyo caso la estructura de definición consta de un único nombre común en absolutivo singular:

  andurreria: harrokeria.

- La estructura de definición puede constar de dos nombres comunes en absolutivo singular separados por comas (**SIN, SIN**):

  estekadura: esteka, lotura.

- Los sinónimos separados por comas pueden en ocasiones ser incluso tres (**SIN, SIN, SIN**):

  abeniko: hitzarmena, egiunea, konponketa.

- Una definición y un sinónimo, en este orden y separados por una coma, también pueden componer una única estructura de definición (**DEF, SIN**):

  amatzako: amatzat harten den emakumezkoa, ama-ordea.[5]

  En el caso de que la definición preceda a un nombre común en absolutivo singular éste será casi con total seguridad un sinónimo.

[3] El genus o sinónimo aparece subrayado.

[4] Traducción al castellano: "timidez: **cualidad del que es** tímido". Solamente incluiremos las traducciones más relevantes. Las traducciones son en parte literales, para mantener el paralelismo con la definición en euskara.

[5] madre adoptiva: mujer que se toma por madre, madre adoptiva.

- El sinónimo puede preceder a una definición en una misma estructura de definición (SIN, DEF):

  isipu: ihiztailua, elizetan ur bedeinkatua jaurtitzeko erabiltzen den tresna.

Un nombre común en absolutivo singular al comienzo de la estructura de definición y ante coma es a menudo un sinónimo. Pero no siempre es así, ya que entre los 200 ejemplos estudiados encontramos dos excepciones:

lotailu: besoa, zangoa, burua inguratuz, zauri bat estaltzeko ezartzen den telazko zerrenda.[6]

En el ejemplo anterior *besoa* no es sinónimo de *lotailu*.

También puede darse el caso de que dos definiciones en una misma estructura de definición separadas por una coma (**DEF, DEF**):

etanol: 20H formula duen alkohola, ardoaren eta edari bizien alkohola.

Es muy difícil establecer criterios para aislar estas definiciones. Las comas no son referentes válidos, pues también se usan en las enumeraciones, oraciones explicativas de relativo, etc.:

artatxiki: artoaren antzeko landarea, hegaztiei jaten emateko erabiltzen diren ale biribil horixkak ematen dituena.

Como hemos mencionado anteriormente, no tratamos de determinar la sintaxis de las definiciones en su totalidad, sino solamente la estructura en torno al genus y sinónimo. Esto hace que podamos localizar sin problemas varios genus y sinónimos en una misma estructura de definición.

5. Las definiciones

La definiciones, en contraposición a los sinónimos, consisten siempre en un sintagma nominal, y tienen muy variadas sintaxis. Pueden constar, por ejemplo, de un nombre compuesto "sustantivo + sustantivo":

azosto: aza hostoa.

o de un sustantivo seguido de uno o más adjetivos:

isipu: zuhaixka bizikorra, lore-urdina

En el caso de que el sustantivo vaya seguido de más de un adjetivo, la conjunción copulativa *eta* precede al último:

pipermin: piper gorri mehe **eta** luzea, oso mina.

Una estructura muy frecuente en las definiciones es aquella en la que al sustantivo genus le precede un complemento adnominal, que puede ser cualquiera de los dos genitivos (posesivo y locativo), o una oración de relativo:

biltoki: biltze**ko** tokia; berezk. talde edo elkarte bateko lagunak biltzen dire**n** etxea edo lekua.

En el caso de las oraciones de relativo, puede ocurrir que el *genus* esté elidido. En estos casos es imposible saber si el genus es cosa o persona, pero pueden extraerse otras relaciones mediante el estudio de los relatores específicos, por ejemplo, finalidad prototípica:

argazkilari: argazkiak egiten dituena, eta berezk., bizibidetzat argazkigintza duena.[7]

La forma nominal del verbo también puede ser el núcleo del sintagma nominal:

setakeria: uste, iritzi, asmo, erabaki... bati gogor eta amor eman gabe eustea.

6. La Gramática de Restricciones

[6] venda: banda de tela que se aplica para tapar una herida rodeando la cabeza, pie o brazo.

[7] Fotógrafo: el que hace fotografías, y especialmente el que tiene la fotografía por oficio.

```
MAP (&GENUS10) TARGET IZE-ABS-MUGATU-NOTGELGEN
        IF (-1 IZENLAGUN) (1 PUNT/PKOMA/KOMA/DEF_BUKA) ;
```

Ilustración 1: ejemplo de regla para etiquetar genus.

```
/<@@lema grefier.>/<ID>/
/<@@Kategoria  iz.>/<ID>/
"<Justizia>"
   "justizia"  ADI SIN NOTPART NOTDEK AORG HAS_MAI DEF_HASI
   "justizia"  IZE ARR DEK ABS MG AORG HAS_MAI DEF_HASI NOTGELGEN
   "justizia"  IZE ARR DEK ABS NUMS MUGM AORG HAS_MAI DEF_HASI
NOTGELGEN
   "justizia"  IZE ARR  ZERO AORG HAS_MAI DEF_HASI NOTGELGEN
"<administrazioko>"
   "administrazio"  IZE ARR DEK NUMS MUGM DEK GEL DEK ABS MG
   "administrazio"  IZE ARR DEK NUMS MUGM DEK GEL
"<idazkaria>"
   "idazkari"  IZE ARR DEK ABS NUMS MUGM NOTGELGEN S:124 &GENUS10
"<$.>"
   PUNT_PUNT
```

Ilustración 2: ejemplo de definición analizada con el genus marcado.

Una vez analizadas las estructuras del campo de definición y establecida la sintaxis en torno al genus, se elaboran diversas reglas utilizando la Gramática de Restricciones para detectar los *genus* y *sinónimos*. Con estas reglas se marcan, mediante *mappings*, como sinónimos (&SIN) o genus (&GENUS) los lemas cuyo análisis y contexto sintáctico coinciden con los establecidos en la regla correspondiente.

Un ejemplo de reglas para etiquetar genus puede verse en la ilustración 1. Esa regla viene a decir que si tenemos un sustantivo (IZE) en absolutivo (ABS) definido (MUGATU) –singular o plural–, sin marca de genitivo (NOTGELGEN), al que le precede algún complemento nominal o una oración de relativo (-1 IZENLAGUN), y va seguido de algún signo de puntuación (1 PUNT/PKOMA/KOMA) o marca de final de definición (DEF_BUKA), entonces lo marcamos como genus (&GENUS10 para registrar que lo ha marcado la regla 10). Esta regla es la mas productiva para los genus, ya que marca el 63% del total de genus marcados.

Cuando esta regla se aplica sobre el resultado del analisis morfológico, añade la marca de genus como se ve en la ilustración 2[8].

Actualmente disponemos de 12 reglas para la detección de sinónimos, y 28 reglas para genus. Es de resaltar que, en total, el desarrollo de las reglas no ha llevado mas de 60 horas, incluyendo la familiarización con las gramáticas de restricciones.

7. Resultados

En total se analizaron 20.453 definiciones de nombres. Para el 85% de las definiciones el sistema encontró al menos un sinónimo o un genus. Para analizar la bondad de los resultados se consideraron dos muestras, una para las definiciones etiquetadas (ese 85%) y otra para las no etiquetadas (el 15% restante).

El tamaño de la muestra para las definiciones etiquetadas era de 216, de los que solamente 197 contienen genus o sinónimos con un total de 117 genus y 107 sinónimos[9]. Los datos se encuentran en la siguiente tabla:

	num.	enc.	no enc.	mal	bien
genus	117	101	25	9	92
sin	107	106	3	2	104

Tabla 3: análisis de definiciones etiquetadas; resultados sobre genus y sinónimos[10].

Esto nos da una cobertura de 78,6%[11] del total de los genus, y 97,2% del total de sinonimos (87,5% considerando ambos). La

[8] Grefier: secretario de la administración de justicia.

[9] Hay que hacer notar que puede haber más de un sinónimo y/o genus por definición.

[10] La columna *enc* contiene aquellos genus que sí han sido extraídos, y la columna *no enc* aquellos que no.

[11] 78.6=92/117x100

tasa de error es de 8,9%[12] para genus, y del 1,9% para sinonimos (%5,3 en total).

Los resultados de la tabla 3 se refieren al numero de genus y sinónimos recuperados del total que había en las definiciones. Si nos centramos en el número de definiciones para los que conseguimos encontrar al menos un genus o sinónimo (definiciones *conectadas*) tenemos los resultados de la tabla 4. La cobertura llega al 91,4%[13], y la tasa de error está entre el 5% y el 1,1%, según demos por válidas o no las definiciones donde se han extraído un genus/sinónimo erróneo y otro correcto (columna *mal y bien* de la tabla). Los resultados de ambas tablas coincidirían si no fuera porque hay definiciones con más de un sinónimo y/o definición.

num. def.	enc.	no enc.	mal	mal y bien	bien
197	180	36	2	7	171

Tabla 4: análisis de definiciones etiquetadas; resultados sobre definiciones.

Si esta muestra nos da la bondad con la que extraemos los genus y sinónimos el análisis de la muestra de 100 definiciones sin etiquetar (tabla 5) nos permitirá analizar cómo se puede mejorar la extracción de genus y sinónimos, y evaluar, una vez extraídos también los relatores específicos, cuántas definiciones nos quedarían sin ningún tipo de etiqueta.

Def. errónea	Meta-lenguaje	rel.	Genus A	Genus B	no recup.
13	1	55	6	16	9

Tabla 5: análisis de definiciones no etiquetadas.

De la tabla 5 se desprende que un 13% de las definiciones no pudo analizarse debido a que el campo de definición se había reconocido erróneamente, y que un 1% se ha definido mediante metalenguaje, lo que quedaba fuera de los esquemas de definición comentados en la sección 3. El 55% de la muestra podría tratarse mediante reglas para relatores específicos. En un 9% de las definiciones no se ve que se puedan recuperar los genus fácilmente. Para el 22% (A+B) de los casos, se podría recuperar el genus añadiendo nuevas reglas que ahora mismo comentaremos. No se encontraron sinónimos en esta muestra, lo que indica que casi todos los sinónimos ya se consiguieron recuperar satisfactoriamente.

Al hablar de genus recuperables se han distinguido dos conjuntos en razón de la facilidad para tratarlos. En el conjunto A se encuentran, por ejemplo, algunas definiciones donde el genus está modificado por un participio, y bastaría añadir en los *mappings* que el participio puede funcionar como adjetivo. En el conjunto B se encuentran definiciones con comas y conjunciones (*eta* o *edo*) que habíamos rechazado tratar por ahora. A primera vista parece posible tratarlos añadiendo condiciones para que las categorías de las palabras en estas estructuras sean idénticas.

Si proyectamos estos resultados al 15% de las definiciones para las que no se encontró ningún sinónimo o genus, tenemos que para un 2,1%[14] no tenemos tratamiento posible por problemas con la definición (esta cifra coincide con el 3% estimado en la tabla 2). En el mejor de los casos (conjunto B recuperable), habría un 1,35%[15] de definiciones adicional sin tratar, es decir, sin que se le haya detectado sinónimo, genus o relator específico alguno. Si no se consiguen tratar los genus del conjunto B esta cifra subiría hasta el 3,75%[16].

8. Conclusiones y trabajo futuro

Este trabajo presenta un sencillo método para la extracción de relaciones semánticas a partir del diccionario en euskara *Euskal Hiztegia*. Previamente, se ha mejorado la conversión a MRD del diccionario impreso, y se ha llegado a un 99,99% de cobertura, con una corrección del 97% en los campos de entrada, categoría, definición y ejemplos.

El método ha sido utilizado para la extracción de relaciones de hiperonimia (genus) y sinonimia para sustantivos, y es fácilmente ampliable a otras categorías y relaciones lexico-semánticas. El método se basa en el análisis superficial de las definiciones mediante reglas de *mapping* proporcionadas por las gramáticas de restricciones, y utilizando la información aportada por el analizador morfológico MORFEUS. La extracción de las relaciones de sinonimia e hiperonimia consigue etiquetar el

[12] 8,9=9/101x100

[13] 91,4=180/197x100.

[14] 14% del 15%.

[15] 9% del 15%.

[16] 25% del 15%.

85% de las definiciones de los sustantivos. El análisis de una muestra de las definiciones etiquetadas indica que cubrimos el 91,3% de las definiciones con genus/sinónimo, con un error en la extracción del 5%, aunque el 98,9% de las definiciones etiquetadas tienen al menos un genus/sinónimos correcto. Del estudio de otra muestra con definiciones no etiquetadas se deduce que con una pequeña mejora en las reglas para los genus y el análisis de los relatores específicos se podría llegar a etiquetar el 96,5% de las definiciones de los nombres.

Es precisamente en los relatores específicos donde ya hemos empezado a incidir. Para ello ya se han realizado varios análisis de frecuencia. Por un lado, de las definiciones del diccionario se extraen las palabras, pares de palabras y tríos más frecuentes con objeto de localizar: (i) genus vacíos de contenido semántico y (ii) relatores específicos. Por otro lado, se realiza otro análisis de frecuencia de los genus previamente hallados en las definiciones, ya que algunos de los genus de frecuencia alta suelen ser a su vez relatores específicos. El estudio de la muestra con definiciones no etiquetadas también ha sido útil para la detección de relatores específicos. También se está trabajando en las definiciones de verbos y adjetivos.

Los resultados obtenidos, así como el esfuerzo necesario (no más de 60 horas), nos indican que es factible extraer relaciones semánticas a partir de la información morfosintáctica proporcionada en un tiempo razonable, ya sea para sustantivos, verbos o adjetivos.

Referencias

Amsler, R.A. *Taxonomy for English Noun and Verbs* Proceedings. of the 19th Annual Meeting of the ACL. 1981.

Arriola, J.M., Artola, X., Soroa, A. *Análisis automático del diccionario Hauta-Lanerako Euskal Hiztegia.* SEPLN. 1995.

Arriola, J.M., Soroa, A. *Lexical Information Extraction for Basque*. CLIM. 1996

Arriola, J.M., Artola, X., Gojenola, K., Soroa A. *TEI: testu-kodeketarako gidalerroak.* EKAIA. 1997

Arriola, J.M., Artola, X., Soroa, A. *Hauta-Lanerako Euskal Hiztegiaren analisi erdiautomatikoa.* ASJU. Pte. de publicación.

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

Boguraev, B., Briscoe, T. (eds.) *Computational Lexicography for Natural Language Processing.* Longman: New York. 1991

Burnard, L., Sperberg-McQueen, C.M. *Guidelines for the encoding and interchange of machine-readable*. Oxford University Press. 1990

Calzolari,N., Picchi, E. *Acquisition of semantic information from an on-line dictionary.* Proceedings of COLING. 1986

Castellón, I. *Lexicografia Computacional: Adquisición Automática de Conocimiento Léxico.* Tesis de doctorado. Universitat de Barcelona: Barcelona. 1993

Karlsson F., Voutilaiknen A., Heikkila J., Anttila A (Eds.). *Constraint Grammar: Language-independent System for Parsing Unrestricted Text* Mounton de Gruyter. 1995.

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

Sarasola, I. *Euskal Hiztegia*. Gipuzkoako Kutxa. 1996.

Urkia M. *Euskal Morfologiaren tratamendu informatikorantz.* Tesis Doctoral. UPV/EHU: Gasteiz. 1997.