CORE
Metadata, citation and similar papers at core.ac.uk
Provided by Repositorio Institucional de la Universidad de Alicante

Extracción masiva de información sobre subcategorización verbal vasca a partir de corpus

I. Aldezabal, M. Aranzabe, A. Atutxa,
K. Gojenola, K. Sarasola
Informatika Fakultatea, 649 P. K.,
Euskal Herriko Unibertsitatea,
20080 Donostia (Euskal Herria)
e-mail: jibalroi@si.ehu.es

Patxi Goenaga
Filologia, Geografia eta Historia Fakultatea
Euskal Herriko Unibertsitatea,
01006 Vitoria-Gasteiz (Euskal Herria)

Resumen

En este artículo presentamos el trabajo realizado en la extracción automática de información sobre la aparición de complementos y adjuntos para un conjunto de 1.400 verbos a partir de un corpus periodístico de un millón y medio de palabras. Los resultados han sido evaluados, obteniéndose una precisión y cobertura satisfactorias. Estos datos se usarán para la adquisición manual y automática de información sobre subcategorización verbal.

1 Introducción

El estudio de la subcategorización verbal se presenta como un paso previo fundamental para el análisis sintáctico profundo. El gran tamaño de los corpus actualmente existentes requiere sucesivas extensiones al léxico con el objeto de incluir, por una parte, información específica para cada tipología de corpus y, por otra, nueva información léxica sobre patrones de subcategorización o restricciones de selección. Hasta ahora se han desarrollado recursos de amplia cobertura para los idiomas más extendidos, como el inglés o alemán (Grishman *et al*., 1994; Kuhn *et al*., 1998).

En el caso de idiomas de difusión más reducida como el euskara esta necesidad se agudiza debido al limitado número de investigadores y la menor disponibilidad de recursos para su tratamiento. Por ello, es interesante el estudio de métodos que permitan obtener información de manera automática o semiautomática. Asumiendo, por principio, la importancia que en el euskara tiene el estudio de los casos de declinación y los sufijos de subordinación que aparecen acompañando al verbo, nuestros experimentos se han centrado en extraer información en torno a éstos.

Previamente a éste hemos desarrollado diferentes trabajos para la obtención automática de información sobre subcategorización verbal para el euskara (Aldezabal *et al*., 1998; 2000), donde se utilizaba un analizador sintáctico parcial para la extracción de verbos junto con las unidades sintácticas adyacentes. El trabajo aquí expuesto presenta mejoras significativas en cuanto a la cobertura gramatical, el número de verbos examinados, el tamaño del corpus utilizado, y la riqueza y fiabilidad de los datos obtenidos.

En la sección 2 se revisan los trabajos que han sido realizados en el área de adquisición automática de información sobre subcategorización verbal. A continuación la sección 3 presenta la arquitectura del sistema desarrollado. La sección 4 describe los aspectos lingüísticamente relevantes para la preparación del trabajo en cuanto a número de verbos analizados, cobertura lingüística, elección del corpus de estudio y características novedosas respecto a trabajos previos. Posteriormente la sección 5 presenta los resultados que han sido obtenidos en el tratamiento de un corpus periodístico. Finalmente se exponen las conclusiones y trabajos futuros.

2 Trabajos previos

Respecto a la adquisición de información sobre subcategorización verbal existen propuestas que van desde el examen manual de corpus (Grishman *et al*., 1994) hasta la adquisición totalmente automática. Por ejemplo, Briscoe y Carroll (1997) describen un experimento basado en una gramática para la extracción automática de patrones de subcategorización y sus frecuencias de aparición asociadas, con el resultado de una precisión (número de elementos seleccionados correctamente / total de elementos devueltos

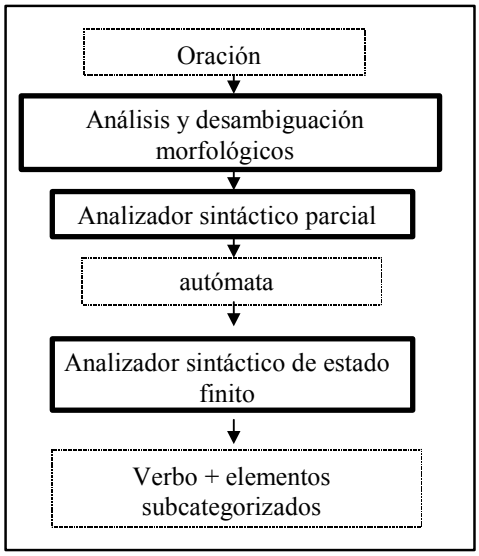

Figura 1. Arquitectura del sistema.

por el analizador) del 76,6% y una cobertura (número de elementos seleccionados correctamente / total de elementos presentes en la oración) del 43,4%.

De cualquier manera, el establecimiento de límites entre los elementos subcategorizados y los no subcategorizados o adjuntos sigue siendo un problema abierto. Es por ello por lo que se realizan trabajos en la línea del presente, haciendo uso de métodos estadísticos y así obtener resultados que ofrezcan información significativa para una aproximación a dicha distinción. Por ejemplo, Briscoe y Carroll resuelven la diferencia entre elementos subcategorizados y adjuntos mediante el establecimiento de un umbral estadístico, de forma que consideran como adjuntos aquellos elementos con frecuencia de aparición inferior al umbral.

Tal y como se ha mencionado, en trabajos previos relacionados con el euskara (Aldezabal *et al*., 1998; 2000) utilizamos un analizador sintáctico parcial para la extracción de verbos junto con las unidades sintácticas adyacentes. En ellos, el sistema se utilizó para el análisis de un total de 2.500 oraciones correspondientes a cinco verbos, y se contemplaron las siguientes mejoras para trabajos futuros:

- Ampliación del estudio a un conjunto amplio de verbos.
- Enriquecimiento de la gramática, de cara a mejorar la cobertura gramatical.

Bertara joandako guardia zibilak ere gauza bera esan zuen atzo eman zuten prentsaurrekoan adierazi zenez.
(El guardia civil que se personó allí también dijo la misma cosa tal y como se declaró en la rueda de prensa dada ayer.)

↓

(guardia zibilak) (ere) (gauza bera) (esan zuen)
(el guardia civil) (también) (la misma cosa) (dijo)
SN (*guardia*, ergativo, sing., def.) SN (*cosa*, nominativo, sing., def.) verbo(*decir*)

Figura 2. Ejemplo de oración y el resultado obtenido.

- Inclusión de información sobre los casos de concordancia presentes en la forma verbal, que pueden ser elípticos.

3 Descripción del proceso

La Figura 1 muestra la arquitectura del sistema, en el que se han combinado diferentes módulos:

- Análisis y desambiguación morfológicos. Se utiliza un analizador morfológico basado en la morfología de dos niveles (Koskenniemi, 1983; Alegria *et al*., 1996) y una herramienta de desambiguación morfológica basada en una gramática de restricciones (Karlsson *et al*., 1995; Aduriz *et al*., 1997; Ezeiza *et al*., 1998).
- Análisis sintáctico parcial. Éste reconoce todas las unidades sintácticas básicas, incluyendo sintagmas nominales, sintagmas posposicionales y varios tipos de oraciones subordinadas. Sin embargo, después de este paso se tiene una infinidad de posibles interpretaciones, como resultado de la ambigüedad morfológica (1,19 interpretaciones por palabra después de la desambiguación morfológica) y ambigüedades sintácticas recogidas por el analizador sintáctico parcial.
- Extracción de la información sobre subcategorización, escogiendo entre las múltiples alternativas que genera el analizador sintáctico parcial. Hemos definido una gramática de estado finito para, en primer lugar, aminorar la ambigüedad sintáctica y después extraer los ejemplos de subcategorización (Aldezabal *et al*., 2001). La gramática se expresa mediante expresiones y relaciones regulares, que son

implementadas mediante autómatas y transductores de estado finito. Usamos la herramienta Xerox Finite State Tool (Karttunen *et al*., 1997), que proporciona una descripción modular, declarativa y flexible, utilizando las operaciones básicas sobre expresiones regulares combinadas con los operadores de reemplazo y composición.

4 Características innovadoras del experimento

En este trabajo se mejoran diferentes aspectos respecto a trabajos previos: número de verbos tratados, tamaño y procedencia del corpus utilizado como fuente, cobertura de la gramática, y por último, procedimientos tanto lingüísticos como informáticos añadidos para la mejora de los resultados.

4.1 Número de verbos examinados

Mientras que en (Aldezabal *et al*., 1998) se analizaron únicamente 5 verbos, aquí presentamos un análisis masivo de 1.400 verbos. 400 de ellos han presentado más de 50 apariciones en el corpus, mínimo que consideramos oportuno para dar representatividad a los datos obtenidos.

4.2 Tamaño y procedencia del corpus

Otra gran diferencia es el tipo de corpus que se ha tomado como fuente. En (Aldezabal *et al*., 1998) se utilizó el corpus EEBS (Urkia y Sagarna, 1991) en el que se recogen muestras de todo tipo de textos escritos en euskara a partir de la segunda mitad del siglo XX. Un total de 2.500 oraciones fueron seleccionadas y analizadas. En el presente trabajo, por el contrario, hemos hecho uso de un corpus periodístico, que contiene todos los números de *Euskaldunon Egunkaria* desde enero de 1999 hasta mayo de 2000, y que además de ofrecer una variedad temática más rica y actual, utiliza el euskara normalizado o estándar. En este caso se han analizado 111.000 oraciones (más de 1 millón y medio de palabras en total), todas las encontradas para el conjunto de 1.400 verbos.

4.3 Cobertura gramatical

Tal y como ocurre en las lenguas romances, en euskara también encontramos las que podrían llamarse posposiciones complejas tales como *-(r)en alde* (‘a favor de’), *-(r)en kontra* (’’en contra de’), *-(r)i buruz* (‘acerca de’), ... que constan de una posposición simple –o caso de declinación– añadida al lexema y además otra palabra independiente. La función que desempeñan en la oración equivale a la de las posposiciones dependientes, por lo que se han añadido en la gramática aquéllas que más comúnmente aparecen en los textos. De esta forma, ha sido posible tratarlas como una sola unidad lingüística, evitando así análisis incorrectos. Con el tratamiento de las posposiciones complejas, la cobertura de análisis de sintagmas nominales y preposicionales es casi completa.

4.4 Evaluación

Los resultados han sido evaluados sobre un grupo de 500 oraciones (10.000 palabras), obteniéndose una precisión del 87% y una cobertura del 66%, dando estos valores una medida del grado de fiabilidad de los resultados para nuevos verbos.

4.5 Mejora del método de extracción

Tras hacer unas primeras pruebas y verificar manualmente los resultados, se consideró conveniente la aplicación de procedimientos derivados de ciertas características del euskara, con el fin de mejorar la fiabilidad de los resultados extraídos y evitar algunos errores controlables. Brevemente descritos, son los siguientes:

Agrupación de casos y sufijos de subordinación: En euskara existe un gran número de casos y sufijos de subordinación. Concretamente, en nuestra gramática hemos descrito 61 casos. Muchos de ellos, sin embargo, desempeñan una función similar respecto al verbo. Sin ahondar demasiado en lo que hemos querido definir como *función similar*, diremos que nos hemos referido a la misma función sintáctica que comúnmente cumplen (sujeto, objeto, complemento circunstancial, ...), pero tomando en cuenta también la relación semántica. Así, por citar un ejemplo, hemos agrupado los sufijos de subordinación que hacen referencia al tiempo: *-nean* (‘cuando’), *-t(z)ean* ('cuando’), *-rako* (‘para cuando’), *-terakoan* (‘cuando’), *-takoan* (‘al acabar de’), *-ino* (‘hasta que’), *-netik* (‘desde que’), *-neko* (‘para cuando’). Claro está que la agrupación podría plantearse de forma totalmente distinta según su finalidad sea más general o más específica. En nuestro caso, tras la agrupación hemos obtenido 28 grupos de casos con función similar.

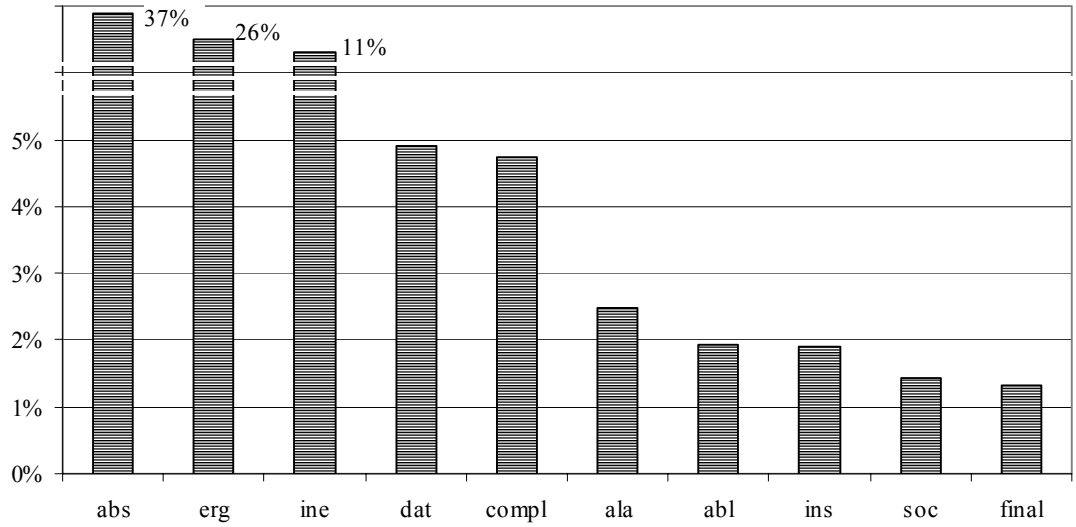

Figura 3. Frecuencia de casos y sufijos de subordinación en el corpus.

Uso del auxiliar para la recuperación de casos: El verbo auxiliar en euskara proporciona información sobre los casos llamados *gramaticales* (absolutivo, ergativo y dativo). Así, aunque en una oración no aparezca el sintagma correspondiente a uno de estos casos, el auxiliar da cuenta de ello y podemos, por lo tanto, asumir que existe ese elemento para el verbo en cuestión (característica de las llamadas lenguas *pro-drop*).

No obstante, es sabido que en los verbos inergativos el sintagma objeto (marcado por el caso absolutivo) no aparece en la oración, a pesar de que haya marcas o procesos sintácticos –en euskara la marca en el auxiliar- que lo remitan. Por lo tanto, hemos decidido no recuperar nunca el caso absolutivo y, por consiguiente, la recuperación de casos se ha hecho en las siguientes cadenas sintácticas:

- Si el auxiliar es de tipo absolutivo-ergativo (a menudo referenciado por el auxiliar en presente de indicativo de la tercera persona singular: DU), el caso ergativo se recuperará siempre. Cabe señalar que en los verbos asociados a fenómenos meteorológicos esta asunción será errónea, ya que el sujeto de la oración nunca aparecerá representado sintagmáticamente. Siendo estos verbos un grupo reducido, y perfectamente controlable a posteriori, pensamos que recuperar el caso ergativo de manera generalizada traerá más ventajas que inconvenientes.
- Si la clase de auxiliar es de tipo absolutivo-ergativo-dativo (DIO), el caso ergativo y dativo se recuperarán siempre.
- Si el auxiliar es de tipo absolutivo-dativo (ZAIO), solo se recuperará el caso dativo.

Cadenas sintácticas excluidas previamente: Ciertas combinaciones de caso y auxiliar nunca tienen lugar en la formación de oraciones; por consiguiente, las hemos eliminado previamente, ya que siempre corresponderán a un error de nuestro sistema sintáctico, habitualmente por distribuir erróneamente los casos correspondientes al verbo de la proposición principal y al de la subordinada. Las enumeramos a continuación:

- Un caso ergativo nunca es posible con un auxiliar de tipo absolutivo (DA)**,** ni de tipo absolutivo-dativo (ZAIO).
- Nunca dos casos ergativos o dos casos dativos acompañarán a un mismo verbo.
- Las estructuras sintácticas con más de cinco sintagmas no suelen ser comunes, y casi siempre son fruto de errores derivados de nuestro sistema. Por lo tanto, hemos preferido dejarlas a un lado.

5 Análisis de datos y valoración

En los dos subapartados siguientes describiremos, en primer lugar, los datos y gráficos que se han obtenido, y en segundo lugar expondremos algunas conclusiones que se pueden derivar de los mismos.

5.1 Análisis de datos

Se han hecho tres primeras aproximaciones para el análisis de los datos. En la primera aproximación, y con el fin de obtener la distribución media en el corpus, se ha medido la frecuencia relativa de cada uno de los casos.

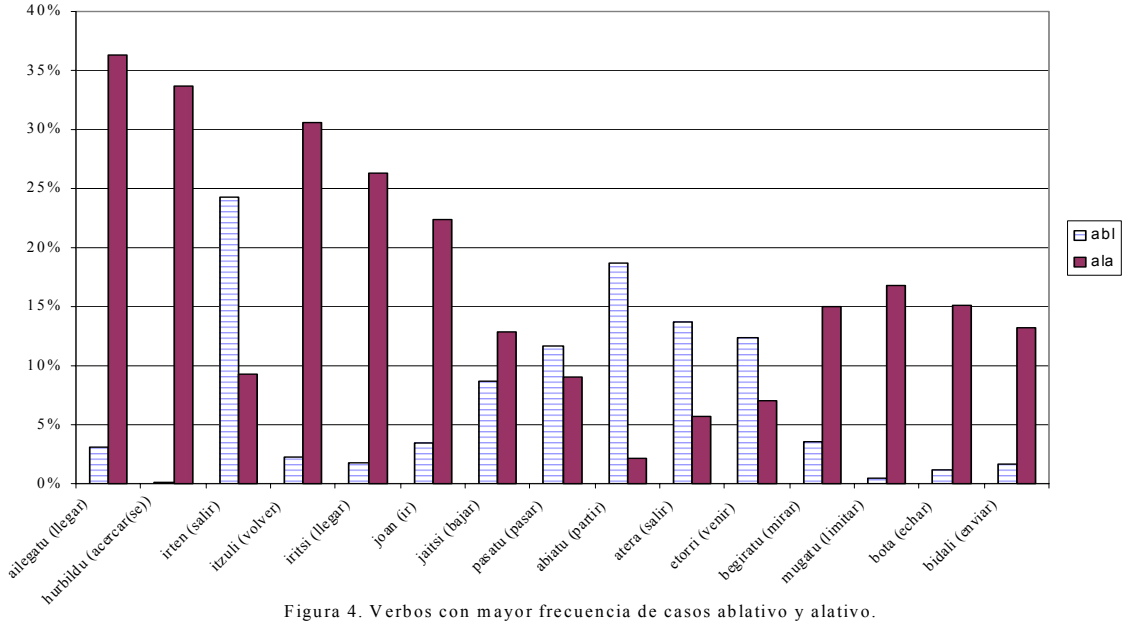

Figura 4. Verbos con mayor frecuencia de casos ablativo y alativo.

En la figura 3[1] aparecen los 10 casos que muestran una frecuencia significativa (superior al 1%). Como se puede observar, los valores son muy dispares y los casos absolutivo, ergativo e inesivo destacan con mucha diferencia.

En la segunda aproximación, hemos querido indagar sobre la validez de los datos para detectar tipos de verbos a partir de las frecuencias de los casos que le suelen acompañar. Para ello hemos preparado un experimento con objeto de detectar verbos relacionados con movimiento, obteniendo aquellos verbos que con mayor frecuencia recogen los casos ablativo (*-tik* ('de, desde, por')) y alativo (*-ra* ('a')), casos típicos que expresan movimiento. La figura 4 muestra los resultados de los 15 verbos que usan estos casos con mayor frecuencia.

Finalmente, y con el fin de comparar resultados, hemos realizado el mismo estudio de casos y verbos presentado en (Aldezabal *et al*., 1998). La figura 5 muestra los resultados que se obtuvieron para un total de 2.500 oraciones del corpus EEBS correspondientes a 5 verbos y la figura 6 los que hemos obtenido ahora (8.000 oraciones del corpus periodístico para esos mismos verbos).

[1] Las abreviaturas de la figura son las siguientes: absolutivo (abs), ergativo (erg), inesivo (ine), dativo (dat), oración subordinada completiva (compl.), alativo (ala), ablativo (abl), instrumental (ins), sociativo (soc) y oración subordinada final (final).

5.2 Valoración de los resultados

Examinando la frecuencia relativa de cada uno de los casos en todo el corpus (figura 3), se constata un claro predominio del caso absolutivo, seguido por el ergativo y el inesivo. El predominio del caso absolutivo es algo esperado, ya que es el caso que representa el sujeto de los verbos intransitivos, y el objeto de los transitivos; es decir, aparece con la inmensa mayoría de los verbos. Con el ergativo ocurre algo similar, ya que suele representar el sujeto de los verbos transitivos e inergativos, es decir, otra gran parte de verbos. La alta frecuencia del caso inesivo puede resultar sorprendente en un principio. Pero si tomamos en cuenta que el corpus utilizado como fuente es un periódico, donde mayormente se narran y describen sucesos o acciones actuales, el caso inesivo aparecerá con frecuencia porque es el que sitúa un suceso en las dos coordenadas de tiempo y espacio.

Los siguientes casos más frecuentes, el caso dativo *-(r)i* ('a') –comúnmente representativo de meta– y los sufijos de subordinación completiva, bien pueden derivarse también de la tipología del corpus, ya que se hace uso de verbos calificados como "de comunicación", donde hay un mensaje que, a veces, tiene una meta especificada, así como verbos que designan voluntad, deseo o preferencia.

En un tercer nivel aparecen los casos locativos –el ablativo *-tik* ('de, desde, por') y el alativo *-ra* ('a')– y el instrumental *-z* ('por', 'por medio de', 'mediante'...). Por último,

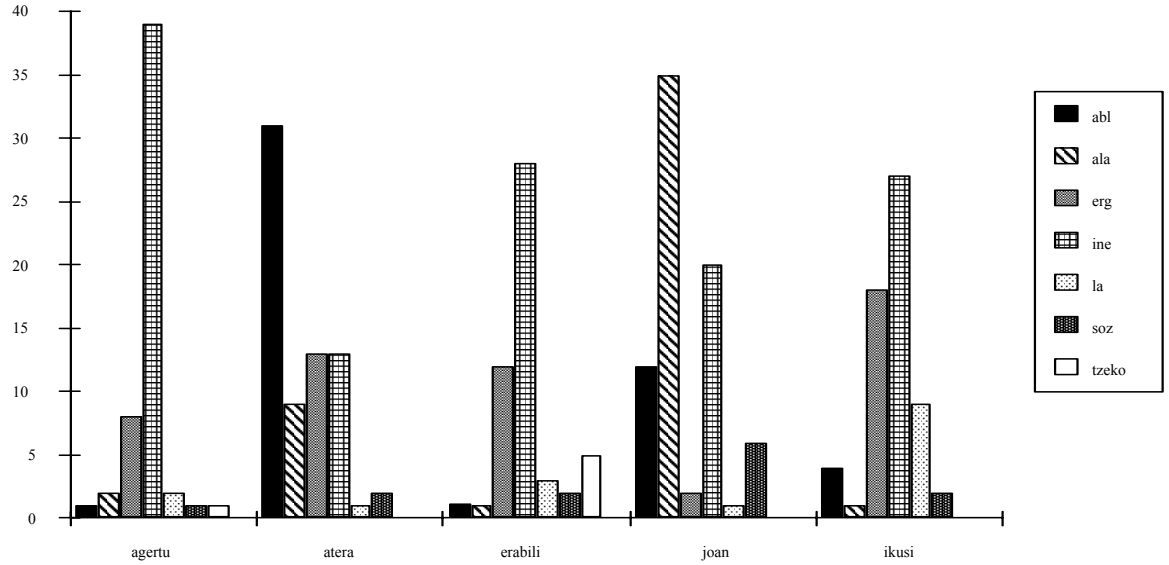

Figura 5. Frecuencias (%) de aparición de cada caso respecto al total de oraciones (Aldezabal et al. 1998).

merecen mención el caso sociativo *-kin* (‘con’) y el sufijo de subordinación *-t(z)eko*, que puede ser tanto de finalidad como completivo.

Por otro lado, los resultados de la figura 4 corroboran la validez del método para buscar verbos de una determinada tipología de casos utilizando nuestros datos. De los 15 verbos presentados, 13 sí parecen típicos verbos de movimiento pero no, en cambio, los verbos *mugatu* ('limitar') y *begiratu* ('mirar'). Esto nos lleva a pensar que o bien los casos *-tik* y *-ra* no siempre son casos relacionados con movimiento, o bien que dichos verbos con el tiempo han podido formar una unidad léxica compleja junto al sufijo.

Por otro lado se aprecia claramente que, a pesar de que los dos casos son admitidos por todos los verbos, hay unos en los que un caso destaca mucho más que el otro, y viceversa; como ejemplo extremo tenemos el verbo *hurbildu* (‘acercar(se)’) que apenas muestra presencia del caso ablativo. Esto nos da pie a proponer dos subclases:

- Verbos de procedencia que denotan el punto

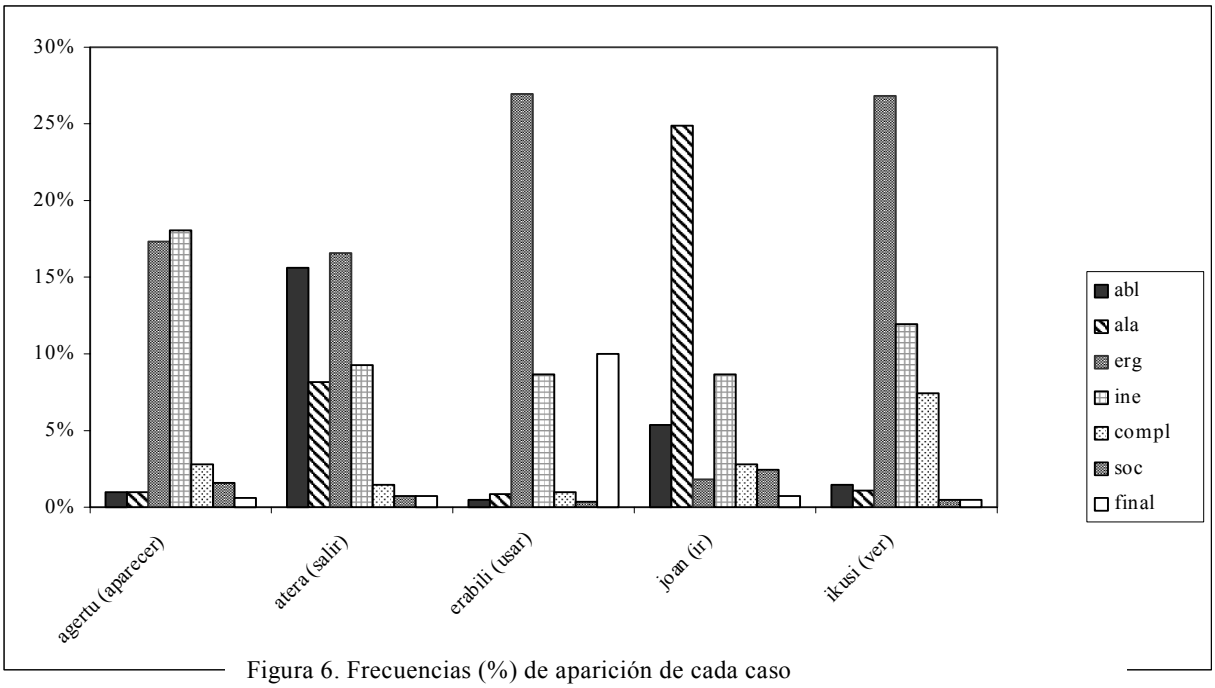

Figura 6. Frecuencias (%) de aparición de cada caso respecto al total de oraciones (corpus periodístico).de oraciones

de partida u origen, lo que viene representado por la presencia del caso ablativo; es el caso de los siguientes verbos, ordenados en base a la frecuencia de aparición de dicho caso: *irten* ('salir'), *abiatu* ('partir'), *atera* ('salir'), *etorri* ('venir') y *pasatu* ('pasar por')

- Verbos de destino que denotan con mayor claridad el punto de llegada o meta, reflejado por el caso alativo, y ordenados también por su mayor frecuencia de aparición: *ailegatu* ('llegar'), *hurbildu* ('acercar(se)'), *itzuli* ('regresar'), *iritsi* ('llegar'), *joan* ('ir'), *mugatu* ('limitar'), *bota* ('echar'), *begiratu* ('mirar'), *bidali* ('mandar'), y *jaitsi* ('descender')

Por último, observando la figura 5 y la figura 6, que comparan nuestros datos con los que obtuvimos en (Aldezabal *et al.*, 1998), podemos apreciar un aumento notable del caso ergativo, en los verbos tanto exclusivamente transitivos (*erabili* ('utilizar') y *ikusi* ('ver')), y también en los de uso dual como transitivo o intransitivo (*agertu* ('mostrar(se)') y *atera* ('salir', 'casar')). Por un lado, un motivo se puede encontrar en el tipo del corpus usado, ya que la especificación del sujeto en los sucesos es casi primordial y está claramente definida, mientras que las sensaciones, lo inesperado o lo repentino son más propios de otro tipo de textos y por consiguiente de otro tipo –o uso- de verbos menos usados en medios de prensa. Por otro lado, no hay que olvidar que el caso ergativo ha sido recuperado para los verbos de la figura 6.

Por otra parte, la presencia del caso inesivo sigue siendo relevante en todos ellos, por la misma razón mencionada en la primera aproximación.

En resumen, podemos decir que las figuras 5 y 6 muestran que no hay diferencias sustanciales respecto a las frecuencias de los casos más comunes. En cambio, sí la hay respecto a otros casos más específicos como los sintagmas posposicionales, importantes tanto para delimitar correctamente sintagmas complejos como para el estudio de subcategorización verbal. Este tipo de estructuras sintácticas no fue tratado en trabajos previos y, por lo tanto, no ha podido ser objeto de comparación.

6 Conclusiones y trabajo futuro

Este artículo presenta la aplicación de un analizador sintáctico parcial para el análisis de un corpus periodístico de gran tamaño, de cara a obtener información sobre complementos y adjuntos de un conjunto de 1.400 verbos.

Se han implementado mejoras planteadas en trabajos anteriores, tales como el mayor número de verbos analizados, la ampliación de la gramática y la inclusión de los casos gramaticales que son susceptibles de elisión. Además, se ha mejorado el módulo de extracción de información, obteniéndose una mejora cualitativa y cuantitativa.

Después de evaluar los resultados podemos decir que el grado de fiabilidad es satisfactorio. Asimismo, se ha puesto de relieve la importancia que tiene el corpus elegido como fuente.

Como futuras líneas de continuación de este trabajo podemos citar las siguientes:

- Estudio comparado con otros datos sobre subcategorización verbal que tienen en cuenta las diferentes acepciones de cada verbo. (Arriola *et al*., 1999) desarrollan otra herramienta para el estudio de la subcategorización verbal del euskara utilizando como corpus definiciones diccionariales y como analizador una gramática de restricciones que incluye la asignación de funciones sintácticas. La comparación de los resultados obtenidos en ambos estudios servirá para la validación y complemento mutuo.

- Creación de una herramienta de consulta y ayuda para la creación manual de una extensa base de datos sobre subcategorización verbal.

- Uso de técnicas estadísticas para la obtención semiautomática de patrones de subcategorización verbal.

- Definición de una herramienta general que pueda generar automáticamente relaciones semánticas y restricciones de selección.

- Integración de la información sobre subcategorización obtenida en corrección gramatical y en aplicaciones de análisis sintáctico general para la determinación de límites entre diferentes suboraciones.

Referencias

Aduriz I., Arriola J.M., Artola X., Díaz de Ilarraza A., Gojenola K., Maritxalar M. (1997) *Morphosyntactic disambiguation for Basque based on the Constraint Grammar Formalism.* Recent

Advances in Natural Language Processing, RANLP'97, Bulgaria.

Aldezabal I., Aranzabe M., Atutxa A., Gojenola K., Oronoz M., Sarasola K. (2001) *Application of finite-state transducers to the acquisition of verb subcategorization information*. Proceedings of the ESSLLI 2001 Workshop on Finite State Methods in Natural Language Processing, Helsinki, Finland.

Aldezabal I., Gojenola K., Sarasola K. (2000) *A Bootstrapping Approach to Parser Development.* International Workshop on Parsing Technologies, IWPT'2000, Trento.

Aldezabal I., Goenaga P., Gojenola K., Sarasola K. (1998) *Subcategorización verbal vasca: propuesta inicial y herramienta de validación.* SEPLN'98, Alicante.

Arriola J.M., Artola X., Maritxalar M., Soroa A. (1999) *A Methodology for the Analysis of Verb Usage Examples in a Context of Lexical Knowledge Acquisition from Dictionary Entries.* Workshop on Linguistically Interpreted Corpora, EACL'99, Bergen.

Briscoe T., Carroll J. (1997) *Automatic Extraction of Subcategorization from Corpora.* ANLP-97, Washington.

Carroll J., Minen G., Briscoe T. (1998) *Can Subcategorisation Probabilities Help a Statistical Parser?* Proceedings of the 6th ACL/SIGDAT Workshop on Very Large Corpora, Montreal.

Ezeiza N., Aduriz I., Alegria I., Arriola J.M., Urizar R. (1998) *Combining Stochastic and Rule-Based Methods for Disambiguation in Agglutinative Languages.* COLING-ACL'98, Montreal.

Grimshaw J. (1990) *Argument Structure* MIT Press: Cambridge.

Grishman R., Macleod C., Meyers A. (1994) *Comlex syntax: building a computational lexicon.* COLING-94, Japón.

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

Karttunen L., Chanod J-P., Grefenstette G., Schiller A. (1997) *Regular Expressions For Language Engineering.* Natural Language Engineering.

Koskenniemi K. (1983) *Two-level Morphology: A general Computational Model for Word-Form Recognition and Production.* PhD. thesis, University of Helsinki.

Kuhn J., Eckle-Kohler J., Rohrer. C. (1998) *Lexicon Acquisition with and for Symbolic NLP-Systems -- a Bootstrapping Approach.* Int. Conference on Language Resources and Evaluation (LREC98), Granada.

Levin B. (1993) *English verb classes and alternations*. The University of Chicago Press.

Urkia, M., Sagarna, A. (1991) *Terminología y lexicografía asistida por ordenador.* La experiencia de UZEI. VII Congreso de SEPLN, San Sebastián.