

# Subcategorización verbal vasca: propuesta inicial y herramienta de validación<sup>1</sup>

I. Aldezabal, 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, por un lado, el trabajo realizado en la creación de una base de datos para el análisis teórico de la subcategorización verbal del euskara y, por otro lado, la herramienta informática que ha sido desarrollada como fuente de conocimiento adicional a partir de corpus, tanto para verificar los datos que se han propuesto en el estudio manual previo como para ofrecer otras sugerencias que puedan tener interés desde el punto de vista subcategorial. El formalismo de subcategorización a definir es capaz de representar fenómenos propios de la lengua vasca como orden libre, sistema rico en casos e información sobre la concordancia de los casos ergativo, absolutivo y dativo representados en el verbo auxiliar, pero además deja abierta la posibilidad de recoger otro tipo de aspectos relevantes (componentes semánticos, estructura eventual, ...). Para lograr conocimiento adicional, la herramienta de validación presenta resultados con características cuantitativas (frecuencia de aparición por elemento y por combinación de elementos) y cualitativas (qué elementos aparecen y formando qué combinaciones).

## *1 Introducción*

El estudio de la subcategorización verbal se presenta como un paso previo fundamental para otros estudios tales como análisis sintáctico y semántico, tanto desde un punto de vista teórico

como desde el de sus aplicaciones en el procesamiento del lenguaje natural.

En este artículo presentamos, por un lado, el trabajo realizado en la creación de una base de datos para el análisis teórico de la subcategorización verbal del euskara y, por otro lado, la herramienta informática que ha sido desarrollada como fuente de conocimiento adicional a partir de corpus, tanto para verificar los datos que se han propuesto en el estudio manual previo como para ofrecer otras sugerencias que puedan tener interés desde el punto de vista subcategorial. El formalismo de subcategorización a definir ha de ser capaz de representar fenómenos propios de la lengua vasca como orden libre, sistema rico en casos e información sobre la concordancia de los casos ergativo, absolutivo y dativo representados en el verbo auxiliar, pero además deberá dejar abierta la posibilidad de recoger otro tipo de aspectos relevantes que surgieran conforme vaya estableciéndose el marco teórico (componentes semánticos, estructura eventual, ...). Para lograr conocimiento adicional, la herramienta de validación habrá de presentar resultados con características cuantitativas (frecuencia de aparición por elemento y por combinación de elementos) y cualitativas (qué elementos aparecen y en qué combinaciones).

En el punto 2 hablaremos del marco teórico de subcategorización verbal así como del punto de partida establecido para el estudio del euskara. En el punto siguiente se describe el procedimiento para la selección de los 78 verbos estudiados en una primera fase y se

---

<sup>1</sup> Este trabajo ha sido realizado en el marco de proyectos financiados por el Gobierno Vasco y la Universidad del País Vasco.

especifica el conjunto de características a definir para cada verbo. En el punto 4 nos centraremos en las herramientas informáticas de validación desarrolladas y en los resultados obtenidos. Finalmente expondremos las conclusiones y los trabajos futuros.

## 2 Marco de subcategorización verbal

El estudio de la subcategorización se basa en atribuir a cada verbo aquellas características necesarias para formar una oración correcta. En cualquier caso, el hecho de delimitarlas es lo que teóricamente acarrea mayores problemas, ya que a medida que se ha ido avanzando en la teoría lingüística, el concepto de subcategorización se ha visto replanteado y han pasado a ser objeto de estudio no sólo la realización sintagmática de las piezas léxicas sino también, y sobre todo, la relación argumental y temática (Grimshaw 90), lo cual tampoco está totalmente establecido. Por lo tanto, la tarea inicial ha sido delimitar el marco de estudio de lo que será para nosotros la subcategorización.

Si nos referimos al caso concreto del euskara, al ser una lengua muy rica en casos y de orden no fijo, la importancia recae sobre todo en marcar el caso mediante el cual se proyectan dichos argumentos, ya que en las alternancias estructurales, muchas veces, son los casos y no las categorías sintagmáticas o el orden de éstas los que cambian; Por ejemplo, el verbo *ahaztu* ('olvidar(se)'), presentando la estructura SN-SN-V, puede realizarse como "tema\_absolutivo, experimentador\_dativo", o "experimentador\_ergativo, tema\_absolutivo".

A estas dos características (orden libre y sistema de casos rico) se les une otra que a la hora de definir patrones supone una dificultad añadida. La forma verbal auxiliar en euskara ofrece información sobre los argumentos que el verbo adopta en una oración dada. Esto posibilita que frecuentemente la presencia léxica de estos argumentos no sea obligatoria – a no ser que estén funcionando como elemento focalizado de la oración– y se dan por sobreentendidos. Al contrario, hay otros elementos que aunque no presenten

concordancia con el auxiliar ni se consideran argumentos –por no desempeñar función de sujeto, objeto u objeto indirecto–, aparecen como obligatorios en ciertos verbos (p. ej., caso adlativo para el verbo *joan* 'ir'). En estos casos, por tanto, nos encontramos con la misma problemática que encuentran el resto de las lenguas, es decir, ante la duda a la hora de establecer la frontera entre los elementos subcategorizados y los no subcategorizados o adjuntos. Al parecer, otros aspectos, como es el de la estructura eventiva del verbo, guardan estrecha relación con la presencia de algunos de estos elementos, cuestión que, por ahora, no hemos tratado en nuestro trabajo.

Otro aspecto a subrayar es que nuestro interés se centra en analizar los verbos en su uso contextual, de una manera tratable computacionalmente, y no sólo en su forma canónica (lema y argumentos con el rol temático correspondiente). Así, cada verbo será estudiado según las alternancias o diferentes estructuras sintácticas que pueda presentar. Además, hay autores que afirman que la conducta sintáctica de un verbo está condicionada por su semántica y que, por lo tanto, los verbos que siguen los mismos patrones sintácticos pertenecen a un mismo grupo semántico (Levin, 1993). En este sentido, por lo tanto, el estudio de las alternancias sería una vía de acceso a la semántica.

Con todo ello, nuestro estudio lingüístico actual de la subcategorización se basa en definir aquellos elementos imprescindibles para cada verbo, señalando los casos y roles temáticos correspondientes a sus diferentes alternancias. También se han añadido otras características que son relevantes desde el punto de vista de otros trabajos como el estudio de los errores sintácticos y las colocaciones léxicas.

En lo que respecta a los resultados obtenidos en este primer experimento de validación, únicamente hemos tenido en cuenta los casos que presenta el verbo en sus diferentes patrones de subcategorización, dejando a un lado el resto de componentes.

### 3 Construcción de una base de datos

El extracto de corpus utilizado para la selección y análisis de los verbos procede del corpus EEBS (Urkia y Sagarna 1991), que refleja el euskara del presente siglo y cuenta con más de 3.000.000 de palabras lematizadas.

El primer paso en la selección de verbos a tratar fue su clasificación según la frecuencia de aparición en el corpus, para luego elegir manualmente aquellos que, por previa experiencia, parecían abarcar un mayor abanico de clases verbales. Si bien inicialmente correspondían a 78 verbos, a medida que hemos ido analizando sus diferentes acepciones y usos hemos ampliado el número de entradas a 130.

En un primer paso definimos el verbo en su forma de participio y marcamos las alternancias estructurales que le corresponden. Entre ellas incluimos aquéllas que hasta ahora hemos considerado pertinentes, pero no descartamos propuestas adicionales a raíz de trabajos posteriores. Por ahora contamos con las siguientes alternancias: incoativa, impersonal, recíproca, reflexiva, intransitiva con objeto indirecto, causativa, transitiva con objeto indirecto, transitiva con objeto proposicional, pasiva, antipasiva, de objeto oblicuo, de objeto inespecificado y de objeto dativo.

Una vez definidas las alternancias correspondientes, asociamos a cada una de ellas un registro que contiene los siguientes campos y valores:

- **Verbo:** Es el campo mediante el cual se establece la conexión.
- **Auxiliar:** Se especifica la clase de verbo auxiliar que admite el verbo en la alternancia. Como es sabido, en la lengua vasca la concordancia se da con los casos absoluto, ergativo y dativo. Las marcas de concordancia pueden presentarse en el verbo auxiliar de la siguiente manera: sólo absoluto (DA), absoluto-dativo (ZAI0), absoluto-ergativo (DU) y absoluto-dativo-ergativo (DIO).
- **Estructura Temática-Casual:** Aquí se definen el caso y el rol temático de los

elementos subcategorizados. Por ejemplo, caso absoluto con valor de tema (tema\_ABS) o caso ergativo con valor de agente (agente\_ERG). Se incluyen también las proposiciones subordinadas subcategorizadas, las cuales se agrupan en cuatro clases según las características [ $\pm$  interrogativa] y [ $\pm$  conjugada]. Así, P1 serán las enunciativas conjugadas, P2 las enunciativas no conjugadas, P3 las interrogativas conjugadas y P4 las interrogativas no conjugadas.

- **Otros casos:** Aquí se definen igualmente los sintagmas posposicionales que aun no considerándose imprescindibles, aparecen comúnmente cerca del verbo. Por ejemplo, caso inesivo con valor locativo (locativo\_INE).
- **Colocaciones:** Se incluyen las unidades complejas de las cuales es parte el verbo, es decir, las unidades léxicas que o bien aparecen como subentradas en los diccionarios o bien han adquirido, en nuestra opinión, un estatus de unidad a pesar de carecer de referencia alguna en los diccionarios.
- **Error:** Este campo se ha definido de cara a especificar las incorrecciones sintácticas que pueden derivarse del uso incorrecto de un verbo dado (p. ej., el uso del caso dativo en lugar del absoluto al hacer referencia al objeto de los verbos transitivos).
- **Ejemplos :** Se especifican los ejemplos utilizados para obtener la información de los campos mencionados.

La figura 1 muestra, a modo de ejemplo, el registro correspondiente a la alternancia causativa del verbo *ikusi* ('ver').

### 4 Herramienta para la verificación subcategorial sobre corpus

Existen diferentes trabajos relativos a la obtención de información sobre subcategorización verbal. El trabajo de (Gross, 1987) estudia la información a describir en un diccionario de subcategorización, especificando si son obtenibles manual o automáticamente,

Verbo	ikusi
Auxiliar	DU
Estructura temático-casual	experimentador_ERG, tema_ABS
Otros casos	locativo_INE, temporal_INE, sociativo_SOC, locativo_ABL, adv_temp, adv_mod, sub_mod, predicativos_INE/ABS/PAR/P2
Colocaciones	<i>ikusi arte (hasta la vista)</i>
Error	abs PRO dat
Ejemplos	@@185543,2594,0061,03: Ba. orduantxe bertan izan da den-dena ikusi dudanean: hantxe zeuden, izkina batean, erdigordeta eliza ondoan musika kioskoak zertxobait ezkutatzen zituela. @@185846,2634,0009,02: Francine-k ikusi zuen gizona sartzen.

Figura 1. Entrada para la alternancia causativa del verbo *ikusi* ('ver')

aunque en este último caso el uso de herramientas lingüísticas se limita a una mera búsqueda de palabras en el texto fuente. El trabajo de (Boguraev *et al.*, 1987) para el sistema Alvey NL Tools, y el de (Grishman *et al.*, 1994) para COMLEX se basan en el tratamiento de información obtenida bien manualmente o bien automáticamente de versiones electrónicas de diccionarios convencionales. (Briscoe y Carroll, 1997) señalan el inconveniente de que estas aproximaciones se basan en un trabajo manual, con el consiguiente esfuerzo e inexactitudes. Por este motivo presentan un sistema automático basado en un etiquetador, lematizador y analizador LR que trata de buscar la clase de subcategorización de un verbo dado y de clasificarlo según las frecuencias relativas de los patrones en los que aparece.

Siguiendo la línea planteada por Briscoe y Carroll, una vez descritas las entradas verbales en la base de datos, nuestro objetivo es validar y enriquecer esta información basándonos en un corpus muchomayor que el usado en el estudio manual. La potencia de la herramienta automática diseñada está avalada por el uso de recursos básicos de amplia cobertura:

- El lematizador/etiquetador Euslem (Aduriz *et al.*, 1996), que asigna a cada forma un conjunto de posibles interpretaciones sin tener en cuenta el contexto, dando una amplia información morfosintáctica para cada una de ellas (lema, categoría, subcategoría, aspecto, caso, número, ...), obtenida de la base de datos léxica EDBL (Agirre *et al.*, 1995).

- Una gramática de restricciones desarrollada para el euskara (Aduriz *et al.*, 1997), que intenta asignar a cada forma una única interpretación, eliminando aquellas propuestas generadas por el lematizador que son incorrectas en el contexto dado. Actualmente, se asigna la categoría gramatical correcta en un 99,12% de los casos, reduciéndose este porcentaje a un 97,86% al considerar toda la información morfosintáctica. La ambigüedad remanente es del 7,57% y del 25,85% respectivamente.

- Una gramática de unificación para el euskara (Gojenola y Sarasola, 1994). Esta gramática se ha definido utilizando PATR II, uno de los formalismos básicos de unificación. Actualmente la gramática consta de cerca de 150 reglas que, como cabe esperar, no son capaces de analizar el amplio abanico de posibilidades presentes en textos reales, pero que sí pueden reconocer los elementos relevantes de cara al estudio de la subcategorización verbal. Entre los fenómenos cubiertos por la gramática, cabe destacar la amplia descripción de los sintagmas nominales y posposicionales y de distintas clases de oraciones subordinadas.

Para una primera prueba, se han seleccionado cinco verbos de diferentes características: *agertu* ('mostrar(se)'), *atera* ('salir'), *erabili* ('utilizar'), *joan* ('ir') y *ikusi* ('ver'). Los pasos seguidos para la obtención de los resultados han sido los siguientes (ver Figura 2):

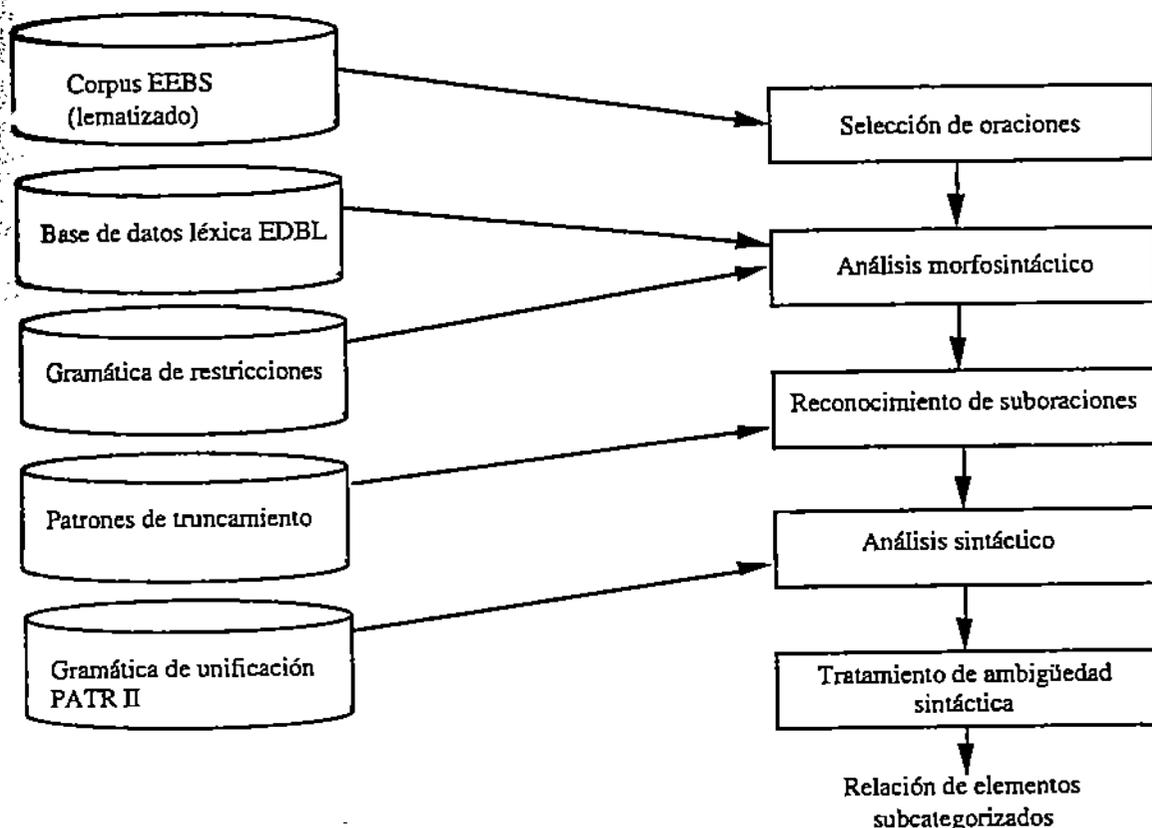


Figura 2. Fases para la obtención de resultados

1) Selección del conjunto de oraciones correspondientes a cada verbo. Se ha aprovechado la lematización parcial ya realizada en el corpus EEBS donde todas las formas verbales han sido etiquetadas con el infinitivo del verbo al que pertenecen. El hecho de que la lematización de EEBS haya sido verificada manualmente confiere una gran fiabilidad a la selección obtenida. Se extrajeron 500 oraciones para cada uno de los verbos.

2) Lematización, asignación de información morfosintáctica a las oraciones escogidas, y posterior aplicación de la gramática de restricciones. Al utilizarse toda la información morfosintáctica queda, como se ha comentado, una ambigüedad del 25,85%. Esto no ha supuesto un problema importante en el tratamiento posterior, pero en alguna ocasión, como se describirá más adelante, se ha hecho necesario el uso de heurísticos o de desambiguación manual.

3) Reconocimiento de sub-oraciones. Debido a la aparición de oraciones complejas con fenómenos de coordinación y

subordinación, cada una de las oraciones escogidas contiene apariciones de diversos verbos, siendo uno de ellos el objeto de estudio en cada caso. El objetivo de este paso es delimitar la sub-oración que comprende la aparición del verbo a tratar y todos los elementos que subcategoriza. De cara a este filtrado las fronteras entre oraciones se han detectado por medio de expresiones regulares. El número medio de palabras ha pasado de 21 en las frases originales a 6 en las sub-oraciones reconocidas. Aunque los patrones utilizados son todavía muy simples, la precisión obtenida es suficiente para poder acometer la tarea planteada. Únicamente un 5% de oraciones obtuvo un filtrado inadecuado. De cara a minimizar el efecto de este porcentaje en el resultado final, se ha preferido que se produjeran errores por omisión en lugar de por exceso, es decir, hemos considerado preferible que algunos componentes no fueran detectados, antes que incluir elementos no pertenecientes al marco de la sub-oración correspondiente al verbo actual.

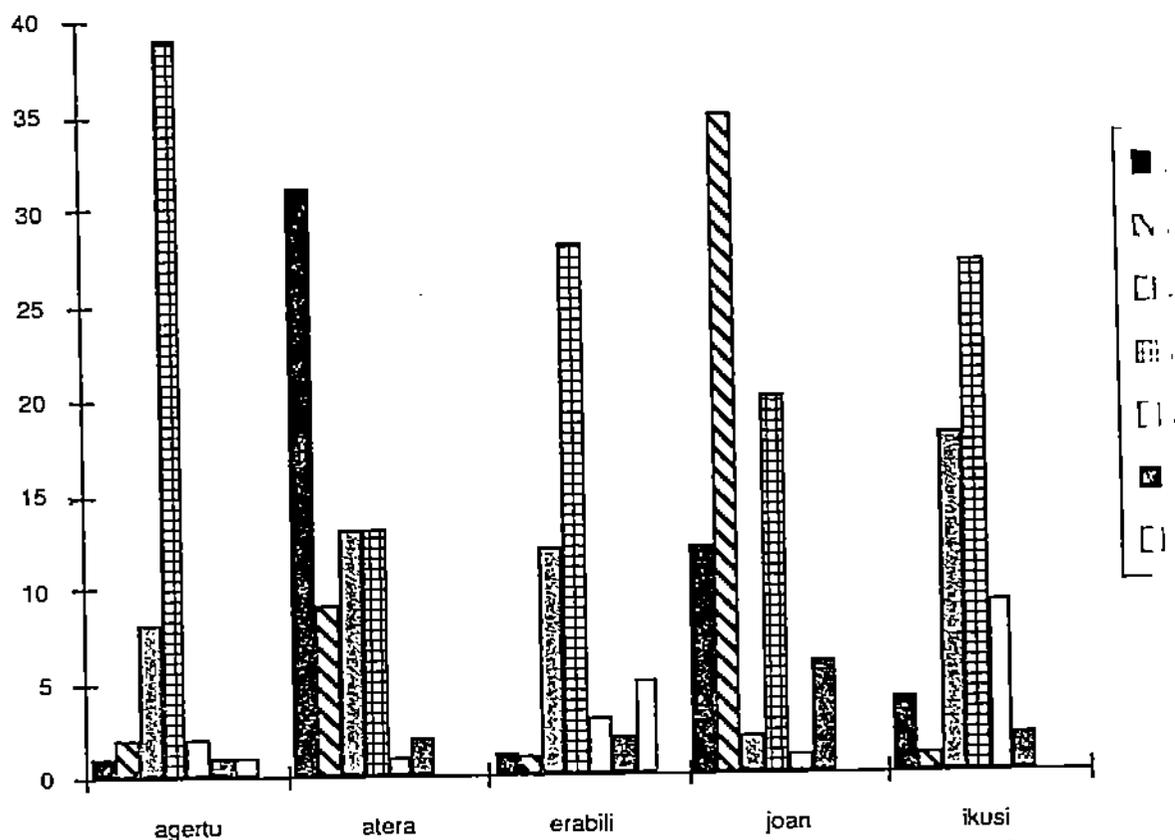


Figura 3. Frecuencias (%) de aparición de cada caso respecto al total de oraciones

4) Análisis sintáctico parcial. El analizador sintáctico se aplica a la sub-oración reconocida con objeto de identificar los elementos subcategorizados. Al ser el análisis de tipo ascendente, se obtiene una tabla con absolutamente todos los componentes encontrados. En este punto se han detectado varios problemas, debidos principalmente a ambigüedades de tipo sintáctico:

- Múltiples posibilidades para un mismo elemento. Por ejemplo, al analizar "mendiko lore polita" ("la flor hermosa del monte") tanto "lore" como "mendiko lore", "lore polita" y "mendiko lore polita" representan sintagmas nominales en caso absoluto. Se ha utilizado como heurístico la elección del componente que comprende la subcadena más larga, habiéndose obtenido buenos resultados.
- Dos subcadenas con el mismo tipo de análisis. Por ejemplo: *gorria* (el rojo) presenta una ambigüedad nombre/adjetivo, pero en ambos casos es un sintagma nominal en caso absoluto. Ya que es esta última la

única información que nos interesaba, se ha mantenido un único elemento representando ambas interpretaciones.

- Ambigüedad entre los casos ergativo/absolutivo. Ésta no ha podido ser resuelta en un alto número de ocasiones (ni por la gramática de restricciones, ni por la basada en unificación), y ha exigido la desambiguación manual, al tener gran relevancia de cara al estudio de los patrones de subcategorización.

5) Obtención de los resultados iniciales. Inicialmente estos resultados han sido agrupados por casos (ver figura 2). Posteriormente se han detectado y contado las diferentes combinaciones de los elementos subcategorizables (ver tabla 1), por ejemplo, "abs", "abs ala" o "abs ine erg" (donde *abs*, *ine*, *erg* y *ala* representan los casos absoluto, inesivo, ergativo y adlativo, respectivamente). El estudio de estos resultados ha servido para refinar la definición de los patrones de subcategorización definidos manualmente.

atera		agertu		ikusi		erabili		joan	
abl abs	56	abs	160	abs	112	abs	168	abs	78
abl	32	ine abs	72	ine abs	32	ine abs	48	ala	50
ine abs	23	ine	56	ine	28	ine	38	ala abs	35
ala abs	15	abs erg	15	erg	18	abs erg	23	ine abs	24
abs erg	14	ins	7	abs erg	15	tzeko	11	ine	23
erg	10	erg	7	la	9	abs tzeko	9	abl	22

Tabla 1. Combinaciones más frecuentes detectadas en el corpus

### 5 Valoración de los resultados

La figura 3 muestra los porcentajes de aparición de cada elemento con respecto al total de oraciones (el caso absoluto se ha obviado en esta figura debido a que, como se verá más adelante, tiene una frecuencia de aparición muy alta y uniforme para todos los verbos). Si nos centramos en los demás casos frecuentes de cada verbo, podemos apreciar que en todos aparece un caso que no pertenece a los considerados como argumento. Concretamente el caso inesivo, cuenta con presencia en todos los verbos y destaca notablemente en los verbos *agertu*, *erabili* y *ikusi*. En el caso de los verbos *atera* y *joan* aparecen frecuentemente los casos de referencia de fuente (ablativo *-tik* 'desde') y meta (adlativo *-ra* 'a'), respectivamente. Todo esto nos lleva a sugerir que algunos verbos muestran una estrecha relación con algunos casos no considerados argumentos, pero que plantean la duda de si nos encontramos ante elementos subcategorizados por exigencias semánticas del verbo. En tal caso, cuando estos elementos no aparecieran léxicamente representados se considerarían también elementos sobreentendidos de la oración.

Si pasamos a considerar los casos con una menor presencia pero aún significativa, nos encontramos con casos relacionados más exclusivamente con el uso de cada verbo. Así, el verbo *erabili* aparece con la subordinada de finalidad *-tzeko* ('para que'); el verbo *ikusi* con la subordinada sustantiva *-la* ('que') y el caso ablativo *-tik* ('desde'); *joan* con el caso sociativo *-rekin* ('con') y con el mencionado ablativo; y finalmente, *atera* con el caso adlativo *-ra* ('a').

Curiosamente, vemos que en los verbos de uso transitivo la presencia del ergativo es muy inferior a la del absoluto. En lo que se refiere al dativo, en esta relación de proporciones no aparece ni siquiera como relevante.

Para acabar con esta valoración inicial de los resultados, mencionaremos la posible causa de la presencia del caso inesivo en estos verbos (y probablemente también en otros). Siendo un caso que hace referencia bien al tiempo bien al espacio, parece reflejar las coordenadas temporales o espaciales en las que se da la actividad expresada por el verbo. Es decir, que la mención anterior sobre la importancia del aspecto verbal puede tener aquí un punto de partida para su futura investigación.

La tabla 1 muestra las combinaciones de elementos más frecuentes detectadas en el corpus. Se observa claramente que el argumento que se expresa léxicamente con mayor frecuencia es el absoluto, independientemente del tipo de verbo. Estos datos están de acuerdo con las frecuencias asociadas a las alternancias que habfan sido observadas manualmente en un corpus más pequeño. Se confirma que las alternancias causativas e incoativas son las más comunes. De todas formas, hay que recalcar que alternancias como las reflexivas, recíprocas, pasivas o antipasivas, aún no han podido ser analizadas como tal mediante esta herramienta, por lo que, por ahora, no tienen lugar como objeto de comparación.

### 6 Conclusiones y trabajo futuro

Se ha definido una base de datos con información subcategorial para un conjunto de

130 entradas verbales. Ésta incorpora aspectos propios del euskara junto con otros no tan usuales relacionados con la subcategorización, como son el tratamiento de errores y las colocaciones léxicas.

Se ha implementado una herramienta de validación. Sus resultados fundamentalmente confirman las intuiciones preliminares establecidas en la base de datos, pero también ponen de manifiesto la necesidad de definir en la estructura subcategorial del verbo otros elementos distintos del sujeto, objeto u objeto indirecto de la oración, dando pie al estudio de otros aspectos pertinentes además de los roles temáticos de los argumentos, como son el aspecto verbal o los componentes semánticos del verbo. El formalismo teórico presentado para el análisis de la subcategorización verbal del euskara permite incluir la definición de esos elementos adicionales.

Como continuación de este trabajo planteamos las siguientes líneas:

- Ampliación del estudio a todos los verbos descritos teóricamente en la base de datos, y a medio plazo definición de una herramienta general que pueda generar automáticamente una propuesta de subcategorización para cualquier otro verbo, a partir de los tipos de complementos observados en corpus.
- Enriquecimiento de la gramática. Aunque en este momento la gramática ofrece una cobertura aceptable de los principales elementos subcategorizables, es necesaria todavía la definición de reglas para tratar cierto tipo de fenómenos como algunos tipos de oraciones subordinadas, o sufijos de flexión compuestos (p. ej., el sufijo *-ri buruz*, que indica 'acerca de').
- Preveamos la inclusión de información sobre el foco (elemento previo al verbo) y los casos de concordancia presentes en la forma verbal, que pueden ser elípticos. La representación sintáctica utilizada, en forma de estructuras de rasgos, permite la fácil ampliación o modificación de la información a tratar.
- Comparación con otros trabajos. (Arriola y Soroa, 1996) desarrollan otra herramienta

para el estudio de la subcategorización verbal del euskara utilizando como corpus definiciones diccionarias 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.

### Referencias

- Aduriz I., Aldezabal I., Alegria I., Artola X., Ezeiza N., Urizar R. (1996) "EUSLEM: A Lemmatizer / Tagger for Basque" EURALEX'96, Suecia.
- 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" RANLP'97, Bulgaria.
- Agirre E., Arregi X., Arriola J.M., Artola X., Díaz De Ilarraza A., Insausti J.M., Sarasola K. (1995) "Different issues in the design of a general-purpose Lexical Database for Basque" First Workshop on Applications of Natural Language to Databases, Francia.
- Arriola J.M., Soroa A. (1996) "Lexical Information Extraction for Basque" Student Conference in Computational Linguistics, Canadá.
- Briscoe T., Carroll J. (1997) "Automatic Extraction of Subcategorization from Corpora" ANLP-97, EEUU.
- Gojenola K., Sarasola K. (1994) "Aplicación de la Relajación Gradual de Restricciones para la Detección y Corrección de Errores Sintácticos" IX Congreso de SEPLN, Córdoba.
- Grimshaw J. (1990): *Argument Structure* MIT Press: Cambridge.
- Grishman R., Macleod C., Meyers A. (1994) "Complex syntax: building a computational lexicon" COLING-94, Japón.
- Gross M. (1987) "Lexicon-grammar: the representation of compound words".
- 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.