

Indexación automatizada de publicaciones lingüísticas: El proyecto CAPLE

Joaquín Garrido Medina

Facultad de Ciencias de la Información

Departamento de Filología Española III

Universidad Complutense de Madrid

28040 Madrid

Resumen

El proyecto de codificación automatizada de publicaciones lingüísticas en español (CAPLE) tiene como objetivo la construcción de un prototipo que indexe automáticamente las publicaciones mediante el procesamiento de sus títulos. La primera fase del proyecto consiste en elaborar, en primer lugar, un léxico terminológico y un tesoro, sobre la base de un corpus de muestra, en el dominio conceptual de la lingüística. Se desarrolla, en segundo lugar, una representación semántica en formato proposicional de dicho universo conceptual, y un diccionario provisto de descripciones semánticas compatibles con el formato del universo conceptual. Un primer examen de ejemplos de la muestra sirve para exponer los problemas a que dicho proyecto se enfrenta.

Indexación automatizada de publicaciones lingüísticas: El proyecto CAPLE

1. Introducción

1.1. Estrategias de indexación de publicaciones

Al introducir en bases de datos documentales los datos bibliográficos de un artículo o monografía de lingüística, se suelen codificar los temas que trata mediante descriptores o palabras clave que sirven posteriormente para realizar búsquedas temáticas en las bases de datos, o para clasificar las publicaciones en las secciones temáticas de las bibliografías periódicas que aparecen en varias revistas especializadas. Esta tarea de codificación temática (indexación, indiciación, o indización) requiere conocimiento experto acerca de la teoría lingüística y supone un considerable coste de tiempo. Las estrategias de indexación automatizada de publicaciones consisten bien en un proceso de emparejamiento de índices, sean términos que caracterizar de forma general el tema tratado, sean descriptores de un tesoro o de un sistema de clasificación (como el decimal; véase Moya e Hípola 1987), o de aplicación de patrones que buscan en los textos ciertas palabras (sintagmas e incluso oraciones) consideradas clave para la recuperación de información.

Frente a estas estrategias de automatización, gran parte de la investigación en lingüística computacional y en inteligencia artificial se esfuerza por comprender el significado de las oraciones mediante el análisis de sus estructuras sintácticas, semánticas, y discursivas (Walker 1987, 247). La vía intermedia, adoptada en este

proyecto, consiste en obtener las principales características semánticas de las palabras, sintagmas y oraciones que aparecen en los títulos de las publicaciones. La indexación o codificación temática se realizará a partir de estas propiedades semánticas de los títulos.

1.2. Definición del proyecto

El proyecto de codificación automatizada con representación semántica de publicaciones lingüísticas en español (CAPLE) tiene dos objetivos principales: (1) construir un prototipo que codifique automáticamente las publicaciones de un corpus inicialmente limitado mediante el procesamiento de sus títulos, y (2) elaborar para dicho prototipo (a) un léxico terminológico (reducido) en el dominio de la lingüística, y (b) una representación semántica de dicho universo conceptual (véase Garrido 1989). El prototipo se construye para estudiar la viabilidad de un codificador con representación semántica en gran escala; el proyecto tiene la duración estimada de tres años (1989-1992), y se realiza en un sistema de bases de datos relacionales en un ordenador personal.

1.3. Evaluación de la estrategia y productos

La utilidad de un sistema de codificación automatizada con representación semántica es doble: sirve para codificar las publicaciones mediante el análisis de sus títulos, y para proporcionar a cada ficha bibliográfica una descripción en unidades de representación semántica que haga posible realizar búsquedas no de términos sino de conceptos en virtud de las relaciones construidas en el universo conceptual. No se trata, por consiguiente, de un sistema de recuperación de información que trate las unidades lingüísticas como objetos desprovistos de significado, o relacionados con descriptores (véase Meya y Huber 1986, §2.3.2, especialmente pp. 62-70). En sistemas de este tipo (sin representación semántica), como observan Martí y Rodríguez Hontoria (1988, §3, p.199), "La utilización de herramientas de tipo lingüístico es rara y la intervención humana fundamental". El planteamiento aquí es el inverso: se trata de reducir la intervención humana, y de desarrollar herramientas de tipo lingüístico.

El corpus se limita a los títulos de las publicaciones por diferentes razones. En primer lugar, con esta opción se reduce el corpus, de manera que con la aplicación de los mismos recursos se

consigue mayor variedad temática y mayor número de publicaciones que si se incluyeran los textos. En segundo lugar, se dispone de inventarios bibliográficos informatizados, es decir, de los títulos, mucho más fácilmente que de las propias publicaciones en soporte legible por máquina (que requerirían el empleo de un lector óptico y de mayor cantidad de espacio de almacenamiento): gran cantidad de las consultas se realizan en inventarios bibliográficos informatizados. En tercer lugar, la indexación de los títulos plantea en sí misma retos suficientes para hacer que su análisis merezca el esfuerzo requerido. El inconveniente práctico de que en ocasiones los títulos no expresan suficiente información acerca del tema tratado puede quedar compensado por el número de títulos que el sistema logra codificar con un éxito relativamente aceptable para el usuario.

Es interesante señalar que no se trata aquí de construir un sistema que permita al usuario emplear una base de datos relacionada usando la lengua española (véase Sopena 1983a y 1983b, citados por Marcos Marín / Sánchez Lobato 1988, o los sistemas para inglés, CLOU de MicroRim y Q&A de Symantec, que menciona Halvorsen 1988, §11.1). Se trata más bien de construir un sistema que analice semánticamente los títulos de las publicaciones en español, sin que su funcionamiento se realice empleando el español. Sin embargo, y con productos adyacentes pero independientes que resultan del proyecto figuran una base de datos terminológicos en lingüística, y un tesoro de lingüística.

Los tesauros permiten búsquedas de unidades conectadas por relaciones verticales, en una jerarquía ramificada, o por relaciones horizontales, mediante conexiones previamente establecidas entre nodos del tesoro. La ventaja de un sistema con representación semántica estriba en que las relaciones que permite establecer entre las unidades son más complejas, ya que las relaciones se dan entre esquemas proposicionales, no entre términos. De este modo se hacen posibles las inferencias o deducciones, es decir, el establecimiento de nuevas relaciones. Las bases de datos terminológicas, por otra parte, tienen la finalidad de la estandarización (Cabré 1986, 79); en el proyecto se intenta determinar los usos normales en las publicaciones del corpus, mediante la elaboración de una base de datos terminológicos. Esta base puede ser utilizada independientemente, ampliándola a un corpus de publicaciones que permita estandarizar la terminología lingüística actual en España.

2. Características generales del proyecto

Aunque los objetivos finales puedan ser ambiciosos, en su primera fase el proyecto consiste en elaborar una terminología y un tesoro en el dominio de la lingüística, en formato de base de datos, y en emplear ambos, el inventario terminológico y el tesoro, como punto de partida para construir un sistema de descripción semántica de títulos de publicaciones lingüísticas (fig.1).

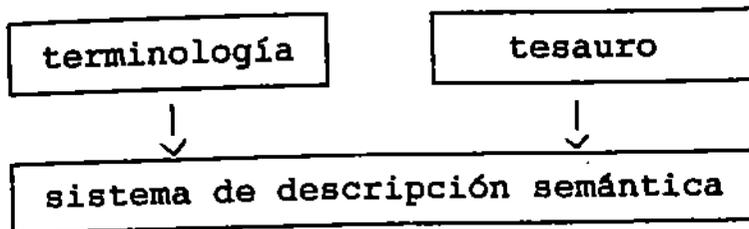


Figura 1. Primera fase del proyecto

El sistema comprende dos componentes principales, el analizador sintáctico y semántico, y el componente léxico, y un tercer componente, de organización del conocimiento no lingüístico. La enorme dificultad de la tarea del proyecto se reduce por la doble restricción de emplearse una muestra limitada (que puede ser inicialmente la de los artículos de los veinte primeros números de la "Revista Española de Lingüística", es decir, hasta el año 1990), lo que conlleva un vocabulario relativamente reducido (frente a las dimensiones, por ejemplo, de un diccionario o una base de datos léxica de entre 35.000 y 40.000 entradas; Alcalá et al. 1989 y Magro et al. 1990) y una complejidad sintáctica acotada.

Antes de abordar los problemas que plantea el componente léxico, conviene mencionar brevemente algunas características que debe reunir el sistema. En cuanto al analizador, su primera tarea es proporcionar la estructura sintáctica de la secuencia de entrada, consultando el inventario léxico, e identificando las palabras de la secuencia. El analizador tiene que basarse en la hipótesis de la correspondencia regla a regla de sintaxis y semántica, de modo que para cada estructura sintáctica el analizador proporcione una estructura semántica, expresada en un lenguaje proposicional, no necesariamente estándar, que se construya a partir de las denotaciones léxicas de manera recursiva. Por último, dado que el significado oracional proporciona sólo un esquema proposicional que debe ser completado con

datos de la situación y conocimientos generales, es necesario un componente, llamado universo conceptual, que organice dichos datos y conocimientos, y que sirva para completar las fórmulas proposicionales dando como resultado expresiones semánticamente interpretadas. Este componente no es dependiente del contexto, sino que ofrece acceso a datos generales del universo conceptual tratados en este caso de la teoría y el análisis lingüísticos. (Cf. el diccionario semántico y la base de dominio en el sistema de interpretación semántica propuesto por Castellón, Manzanera y Marín 1989, integrados con información contextual y pragmática en etapas necesariamente consecutivas, como en el generador universal automático de interfaces de Rodríguez Hontoria). En la figura aparecen representados los módulos del sistema, cuyas interconexiones no representan flujos:

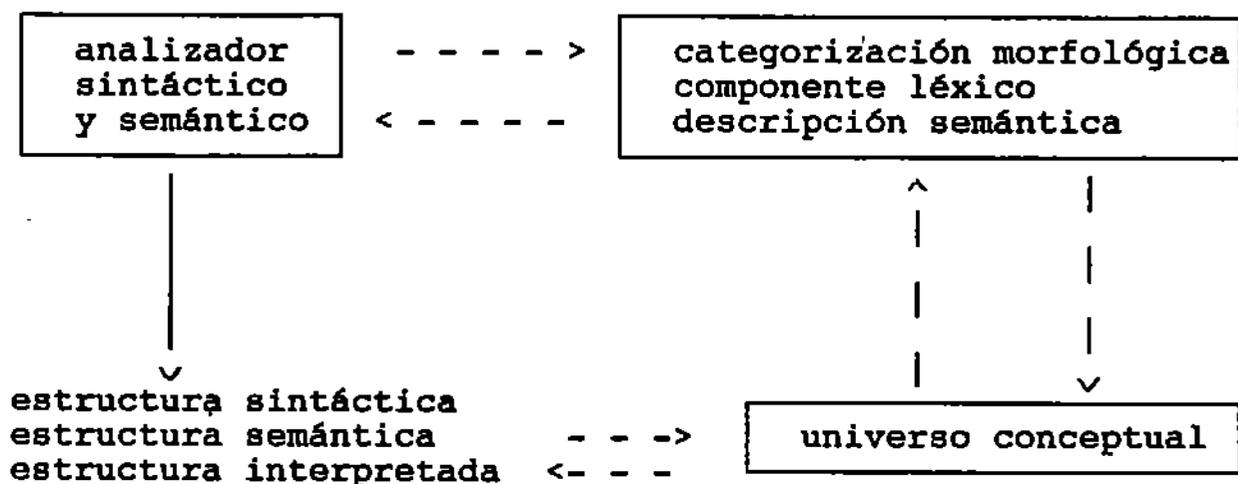


Figura 2. Esquema del sistema

Una manera de considerar el proceso de análisis consistiría en obtener la descripción sintáctica de la secuencia; en segundo lugar, la descripción semántica expresada proposicionalmente; y, en tercer lugar, la interpretación de la expresión proposicional con arreglo al universo conceptual.

3. Léxico y descripción semántica

3.1. El componente léxico

El componente léxico tiene que contener dos tipos de léxico, terminológico y común. Empleando categorías preteóricas, con valor meramente descriptivo del proceso, en el vocabulario común aparecen las unidades léxicas de las categorías de determinante, conjunción, pronombre, preposición, además de los signos de puntuación. En el vocabulario terminológico figuran los términos de lingüística. Las categorías de nombre, adjetivo, verbo y adverbio aparecen en ambos tipos de vocabulario.

La elaboración del vocabulario comprende la recogida de formas de la muestra, y la construcción de las correspondientes fichas terminológicas y un tesoro de lingüística. En la ficha terminológica (cf. Cabré 1987, 57-58) debe figurar la categoría sintáctica de cada forma, los contextos (las secuencias de títulos) en que aparece, y su definición, que incluye las definiciones en los diccionarios de lingüística existentes en español, si la forma en cuestión figura en ellos. Por último, el descriptor es una unidad del tesoro. (Según la norma internacional ISO 2788-1974, un tesoro es un vocabulario de términos, controlado y flexible, estructurado (relaciones entre los términos, semánticas y genéricas), y aplicado a un dominio de conocimientos: debe reflejar la información contenida en el conjunto de documentos a que se aplica.)

La segunda fase de la construcción del vocabulario consiste en la elaboración de un sistema de representación semántica en que aparezcan formuladas las definiciones y que recoja las relaciones establecidas mediante el tesoro. El sistema de representación debe servir para construir la interpretación de los títulos, de manera que sea la base de funcionamiento del analizador.

3.2. El problema de la descripción semántica

En la teoría semántica actual se ha atendido sobre todo a los problemas de composicionalidad y de cuantificación y referencia. Sin embargo, aquí se trata de analizar el significado léxico. Para ello contamos con la tradición estructuralista del análisis en rasgos, y con el análisis de la gramática de casos. Esta última tiene la

ventaja de ocuparse principalmente del significado léxico de preposiciones y verbos, y de establecer las relaciones de composicionalidad oracionales mediante marcos o esquemas de casos. Además, como indica Halvorsen (1988, §11.5), en el campo de la inteligencia artificial existe la tradición creada por el grupo de Schank de la teoría de la dependencia conceptual, con sus marcos análogos a los de la gramática de casos (véase, en español, Schank y Abelson 1977). Los sistemas de descripción semántica basados en la lógica, por el contrario, se ocupan de la semántica oracional, bien traduciendo las oraciones a fórmulas proposicionales, y aplicando procedimientos de inferencia lógica para sacar conclusiones, bien representando la estructura semántica oracional mediante una sintaxis proposicional, e interpretándola después procedimentalmente. Aunque se ha aplicado tanto la teoría de lógica intensional de Montague (1974; véase Meya 1987) como la de semántica de situaciones de Barwise y Perry (1983), la dificultad de estos intentos radica en los mecanismos de inferencia que se les puede aplicar, como señala Halvorsen (1988, §11.5).

En las principales teorías actuales acerca de la estructura léxica (y su integración en la estructura sintáctica oracional) sigue planteada la cuestión de la integración de los papeles o casos semánticos y las relaciones gramaticales de sujeto, objeto, etc. (cf. Andrews 1988, §3.6, y Sells 1985). Por consiguiente, en cualquier teoría gramatical de las actualmente disponibles que se emplee, parece inevitable abordar el problema describiendo, por una parte, las propiedades de subcategorización sintáctica, y, por la otra, el lugar que ocupa la unidad en un esquema proposicional o en una representación no estándar, sea de papeles semánticos, sea de marcos o redes. Aunque inicialmente el análisis esté limitado a estructuras fijas y definidas (frente a la concepción flexible del significado, por ejemplo en Garrido 1987), el sistema se podrá ampliar a otras estrategias, como el análisis mediante el concepto de relevancia (Sperber y Wilson 1986, Espinal 1988) para abordar la dependencia del contexto como proceso de interpretación (Garrido 1990).

4. Descripción de un ejemplo

4.1. Distribución de categorías en la muestra ejemplo

Para dar una idea de los problemas que conlleva la construcción del vocabulario, puede ser útil abordar como ejemplo los 17 títulos

artículos aparecidos en la "Revista Española de Lingüística" en número 18, fascículos 1 y 2, del año 1988 (véase figura 3).

Especificidad estructural, léxica y sintáctica del habla del bilingüe (dominio francoespañol). Análisis cuantitativo del discurso.

La lectura, ¿eslabón entre la lengua escrita y la hablada?.

Los paradigmas verbales en la historia del inglés: más allá del epifenómeno.

Consideraciones en torno a la categoría «verbo» en algunos comentaristas de Aristóteles.

Estudio comparado del orden de palabras en inscripciones jurídicas arcaicas, griegas y latinas.

Concepto platónico del signo.

Norma y uso y método generativo: el caso de la ley de Grassmann.

Categorías verbales, categorías enunciativas y oraciones subordinadas.

La expresión de la finalidad en las subordinadas del griego antiguo.

0 Un enfoque lexemático de la homonimia, polisemia y sinonimia.

1 Variantes conjuncionales para la expresión de la finalidad en las oraciones subordinadas latinas.

2 Normalización y uso del euskera en la administración.

3 Norma y uso según Quintiliano.

4 La metáfora como error: un aspecto cognitivo de la teoría semiótica de Walter Percy.

5 Tipología frente a reconstrucción: la hipótesis glotática.

6 El funcionamiento de 'el cual': norma y uso.

7 Fundamento y desarrollo de la gramática liminar.

Figura 3. Muestra de títulos (REL 18)

La muestra comprende 173 palabras (formas) y 36 signos de puntuación. Entre estos últimos figuran la comilla simple para indicar comienzo y final de forma de cita, y el símbolo # para indicar punto y aparte (es decir, final de un título). Las categorías empleadas y sus frecuencias correspondientes aparecen en la figura 4.

adjetivo	aj	29
adverbio	av	3
conjunción	c	11
determinante	det	33
nombre	n	60
nombre propio	npro	5
preposición	p	31
pronombre	pro	1

Figura 4. Frecuencias en las categorías léxicas de la muestra

En cuanto a la categorización léxica, en la muestra, en primer lugar, se han separado las contracciones de preposición y artículo ('del'); como ejemplos de locución preposicional, se han unido 'en torno a' y 'frente a'. Del mismo modo, podría figurar como locución preposicional 'más allá de', aunque aparece separado, 'más' y 'allá' como adverbio, y 'de' preposición. En cualquier caso, 'más' es un adverbio de diferente tipo de 'allá' (y del otro que aparece, 'como').

En la muestra aparece todas las formas en minúscula (puesto que los signos de puntuación ya indican el final de la oración o del título anteriores), salvo los nombres propios. Cabría distinguir en los nombres propios varias categorías, como antropónimos, topónimos, etc. En la muestra aparecen 5, y de ellos dos forman el conjunto de nombre y apellido ('Walter Percy'). Los nombres propios también deben recibir una descripción semántica, y, como los nombres comunes terminológicos, deben estar relacionados con diferentes conceptos. Por ejemplo, la representación semántica de 'Aristóteles' es un conjunto de propiedades, entre las cuales figura la de ser autor de conceptos de la gramática tradicional.

Las formas de naturaleza terminológica pertenecen a las categorías de adjetivo y nombre (incluyendo el nombre propio), además de las formas citadas. Por ello convendría considerar una forma citada conjuntamente con las marcas de cita, por ejemplo 'el cual'. También en este caso habría una descripción semántica que permitiera relacionar la forma citada con conceptos sintácticos (por ejemplo, pronombre relativo).

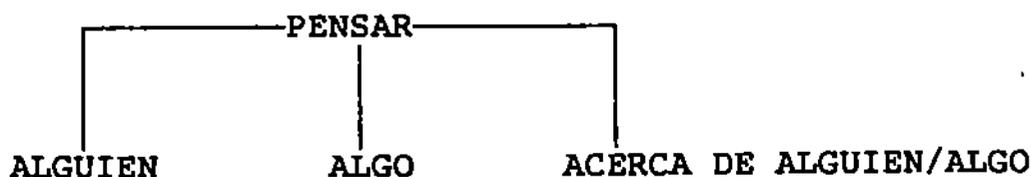
4.2. Sustantivos

El empleo de papeles semánticos o marcos o redes requiere considerar la denotación de nombres, adjetivos y verbos como predicados que relacionan argumentos. Por ejemplo, en 'Concepto platónico del signo', el significado de 'concepto' se considera construido a base de un predicado, con una estructura análoga a la de 'Lo que Platón piensa acerca del signo'. (El adjetivo 'platónico' se interpreta aquí como predicado que relaciona al individuo denotado por 'Platón' con otra entidad, denotada por el sustantivo.) El predicado que interviene en la denotación de 'concepto' se aplica a tres argumentos, el objeto acerca de lo que se piensa, el autor del pensamiento, y el resultado del pensamiento. Este último argumento es lo que es propiamente el objeto denotado por 'concepto'; los otros argumentos son denotados por complementos del sustantivo:

$\lambda x[A(a,b,x)],$

siendo 'A' el predicado, 'a' la constante denotada por 'Platón', 'b' la denotada por 'el signo', y 'x' la variable ligada por el operador lambda.

En otros términos, el significado de 'concepto' pertenece a la misma red o marco que el de 'pensar':



A su vez, el predicado etiquetado como 'PENSAR' se definirá en términos de otros predicados.

Habrán otras construcciones que tengan una interpretación análoga, por ejemplo 'El signo según Platón', 'El signo en Platón', pero que se abordarán desde el punto de vista de las correspondientes preposiciones, 'según', 'en', etc.

En 'signo' (en la lengua común), el predicado relaciona por lo menos el argumento correspondiente al significante con el del significado, y el de la persona productora del signo: 'el signo de la cruz', 'el signo de salida', 'el signo de los cristianos'. Como 'signo' pertenece al vocabulario terminológico, estará definido de acuerdo con una teoría. Supongamos que la definición sea

'representación de una información acerca de la realidad (significado) mediante un objeto físicamente perceptible y transportable (significante)'. La descripción semántica constará igualmente de un predicado, con los argumentos correspondientes a la información, al aspecto tratado de la realidad, al objeto significante; también habrá un argumento correspondiente a la persona que construye la información acerca de la realidad. Según otra definición, puede no existir este último argumento. El universo será más completo si dispone de varias definiciones de la denotación de 'signo' (por ejemplo, 'el signo según Saussure', 'el signo según Hjelmslev', etc.).

La consulta al universo conceptual amplía la información acerca de la denotación de 'signo', acerca de la de 'Platón', y conecta la expresión semántica que describe el significado del título con las denotaciones de 'signo' y de 'Platón' del universo. Ahora existirá una denotación para 'el signo según Platón'.

La ausencia de verbos en la muestra queda pues compensada por el carácter de acción o de resultado de una acción que presentan ciertos nombres, como el mencionado 'concepto', y también 'expresión', 'funcionamiento'. Ciertos nombres deverbales parecen ociosos, como 'consideraciones' y 'estudio' en la muestra.

En 'Consideraciones en torno a la categoría «verbo» ...', el término pertenece a un conjunto (como 'notas', 'reflexiones', 'aproximación') que indica únicamente la actitud del autor de considerar el trabajo como no exhaustivo (lo cual a su vez puede ser interpretado de diferentes maneras, como modestia verdadera o falsa). La misma interpretación reciben preposiciones como 'hacia', 'sobre', o locuciones prepositivas como 'en torno a', 'acerca de'.

Los casos de 'estudio' y 'análisis' presentes en la muestra son distintos. El sustantivo 'estudio' podría considerarse redundante, ya que el término de su preposición 'de' denota el tema tratado en el artículo. Por ejemplo, 'Estudio del orden de palabras' sería equivalente a 'El orden de palabras'; sin embargo, admite un adjetivo, en este caso 'comparado', que indica el método o procedimiento empleado. Del mismo modo, 'análisis' es soporte del adjetivo 'cuantitativo', que igualmente indica el procedimiento u orientación seguidos.

4.3. Adjetivos

Entre los 29 adjetivos de la muestra, hay cuatro unidades (cinco formas), 'antiguo, arcaicas, griegas, latinas, latinas', referidas a lenguas ('arcaicas' a 'inscripciones', e, indirectamente, a una época de la lengua). Por una parte, dos de clasificación temporal dentro de una lengua, 'antiguo' 'arcaico', y dos de clasificación de lenguas, 'griego' y 'latín'. El caso de 'francoespañol' se refiere al bilingüismo, de modo que indirectamente se trata también de lenguas: 'francoespañol' correspondería a 'francés y español', aplicado al bilingüe. Estos adjetivos de lenguas se tienen que poner en relación con los sustantivos: en la muestra aparecen 'euskera', 'griego', 'inglés'.

La primera tarea puede ser, por consiguiente, relacionar los adjetivos de la muestra con los sustantivos, e introducir los sustantivos y adjetivos correspondientes que no aparezcan en la muestra. Por ejemplo, 'conjuncional' y 'conjunción', 'verbal' y 'verbo' (en la muestra no aparece 'verbal' en el sentido de 'oral' o 'lingüístico'). Habrá una relación regular en muchos casos (por ejemplo 'platónico' y 'Platón'), pero en otros bien no habrá sustantivo ('liminar'), bien no habrá adjetivo ('dominio'). En cualquier caso, la relación semántica tiene la misma complejidad que la expresada mediante preposición 'de': 'platónico' y 'de Platón', 'verbal' y 'del verbo' son relaciones diferentes desde el punto de vista de los papeles o casos semánticos.

5. Conclusión

El proyecto de codificación automatizada de publicaciones lingüísticas en español (CAPLE) tiene como objetivo la construcción de un prototipo que indexe automáticamente las publicaciones mediante el procesamiento de sus títulos. La primera fase del proyecto consiste en elaborar, en primer lugar, un léxico terminológico y un tesoro, sobre la base de un corpus de muestra, en el dominio conceptual de la lingüística. Se desarrolla, en segundo lugar, una representación semántica en formato proposicional de dicho universo conceptual, y un diccionario provisto de descripciones semánticas compatibles con el formato del universo conceptual. Un primer examen de ejemplos de la muestra sirve para exponer los problemas a que dicho proyecto se enfrenta.

BIBLIOGRAFIA

- Alcalá, Alfonso /Rodríguez, Chelo /Sopeña, Luis de /Valladares, Carmen /Villar, Celia. 1989. Una base de datos léxica para el procesamiento del lenguaje natural. *Procesamiento del Lenguaje Natural* 8, diciembre 1989, 123-134.
- Andrews, Avery D. 1988. Lexical structure. En Newmeyer 1988, 1, 88.
- Barwise, Jon /Perry, John. 1983. *Situations and attitudes*. Cambridge, Mass., M. I. T. Press.
- Cabré i Castellví, M. Teresa. 1986. Els bancs de dades terminològiques. Organització per a la llengua catalana. *Procesamiento de Lenguaje Natural* 4, octubre 1986, 76-90.
- Cabré i Castellví, M. Teresa. 1987. Les bases de dades terminològiques: especificitat i organització. En C. Martín Vide (ed), *Actas del II Congreso de Lenguajes Naturales y Lenguajes Formales*, 47-64. Barcelona, Universidad de Barcelona.
- Castellón, Irene /Manzanera, Alicia /Martí, M. Antònia. 1989. Interpretación semántica en un sistema dialogado. *Procesamiento de Lenguaje Natural* 8, diciembre 1989, 45-61.
- Espinal, M. Teresa. 1988. Significat i processament de la informació lingüística. En C. Martín Vide (ed), *Actas del III Congreso de Lenguajes Naturales y Lenguajes Formales*, 291-307. Barcelona, Universidad de Barcelona.
- Garrido, J. 1987. Aproximación al significado como representación cognoscitiva y a su realización en ordenador. *Procesamiento de Lenguaje Natural* 5, octubre 1987, 87-91.
- Garrido, J. 1989. Codificación automatizada de publicaciones lingüísticas en español. En *Sociedad Española de Lingüística, Simposio (Salamanca 18-20.12.1989)*, Resúmenes de comunicaciones y ponencias, p.73.
- Garrido, J. 1990. Lexical entry format and contextual information. *Congreso Internacional de la Asociación Europea de Lexicografía (EUROLEX)*, Benalmádena, Málaga, 28.8-1.9.1990.
- Halvorsen, Per-Kristian. 1988. Computer applications of linguistic theory. En Newmeyer 1988, 2, 198-219.
- Marcos Marín, Francisco/ Sánchez Lobato. 1988. *Lingüística aplicada*. Madrid, Síntesis.
- Martí Antonín, M. Antònia /Rodríguez Hontoria, Horacio. 1989. Aplicaciones del tratamiento informático del lenguaje natural. En C. Martín Vide (ed), *Actas del III Congreso de Lenguajes Naturales y Lenguajes Formales*, 191-211. Barcelona, Universidad de Barcelona.

- Meya Llopart, Montserrat. 1987. Análisis semántico montaguiano en el proyecto SPICOS. En C. Martín Vide (ed), Actas del II Congreso de Lenguajes Naturales y Lenguajes Formales, 317-334. Barcelona, Universidad de Barcelona.
- Meya Llopart, Montserrat /Huber, Wolfgang. 1986. Lingüística computacional. Barcelona, Teide.
- Montague, Richard. 1974. Ensayos de filosofía formal. [Traducción parcial del original de 1974.] Madrid 1979, Alianza.
- Moya, Félix de /Hípola, Pedro. 1987. Problemas lingüísticos en la automatización de los sistemas de clasificación documental. Procesamiento de Lenguaje Natural 5, octubre 1987, 74-85.
- Newmeyer, Frederick J. (ed). 1988. Linguistics: The Cambridge survey, 1-4. Cambridge, Cambridge University Press.
- Nirenburg, Sergei (ed). 1987. Machine translation. Theoretical and methodological issues. Cambridge, Cambridge University Press.
- Rodríguez Magro, Consuelo /Sopeña Pastor, Luis de/ Valladares Barragán, Carmen/ Villar Rodríguez, Celia. 1990. Clasificación morfológica del léxico castellano para un analizador en ordenador. Actas del VII Congreso Nacional de Lingüística Aplicada, 491-503. Sevilla, Universidad de Sevilla.
- Schank, Roger /Abelson, Robert. 1977. Guiones, planes, metas y entendimiento. Madrid 1988, Paidós.
- Sells, Peter. 1985. Teorías sintácticas actuales. Trad. R. Cerdà. Barcelona 1989, Teide.
- Sopeña, Luis de. 1983a. Aportaciones a la comprensión del castellano en un sistema de interrogación de bases de datos relacionales. Universidad del País Vasco, Servicio Editorial.
- Sopeña, Luis de. 1983b. Interrogación de bases de datos en lengua natural. Novática IX.53, 15-21.
- Sperber, Dan /Wilson, Deirdre. 1986. Relevance. Communication and cognition. Oxford, Basil Blackwell.
- Walker, Donald E. 1987. Knowledge resource tools for accessing large text files. En Nirenburg 1987, 247-261.

