Procesamiento del Lenguaje Natural, num. 29 (2002), pp.81-88 recibido 22-05-2002; aceptado 02-07-2002

Análisis sintáctico para el español basado en el formalismo de la teoría Significado ⇔ Texto*

Sofía N. Galicia-Haro, **Alexander Gelbukh** e **Igor A. Bolshakov**
Centro de Investigación en Computación. Instituto Politécnico Nacional
Av. Juan de Dios Batiz S/N 07738 México, D. F.
{sofia, gelbukh, igor}@cic.ipn.mx, www.gelbukh.com

Resumen Se presenta la aplicación del formalismo de la teoría Significado ⇔ Texto (*Meaning ⇔ Text Theory*) para el análisis sintáctico del español. En este método, basado en gramáticas de dependencias, el diccionario combinatorio empleado para el análisis sintáctico consta de patrones para palabras, principalmente verbos, donde se describen todas sus valencias y las formas en que ellas se realizan. No se considera un orden fijo en la oración por lo que resultan muy adecuados para el análisis del español. Los patrones del diccionario combinatorio no solamente ayudan a reducir el número de posibles variantes obtenidas por el analizador sino que incluyen información del nivel sintáctico que está conectada con la semántica de la palabra y que es requerida a niveles más profundos del análisis del lenguaje. Adicionalmente, incluimos información estadística de las realizaciones de cada valencia y de las combinaciones de valencias para cada verbo con el propósito de incrementar la eficiencia de resolución de ambigüedad en el análisis sintáctico.
Palabras clave: gramáticas de dependencias, diccionario combinatorio, análisis sintáctico.

Abstract. The application of the Meaning ⇔ Text Theory to Spanish parsing is presented. This formalism is based on dependency grammars. The combinatorial dictionary of this method is employed for the syntactic analysis; it consists of patterns for words, mainly verbs, where all its valences and the way they are realized are described. In this method, no fixed word order in the sentence is considered so it is highly adequate for Spanish parsing. The patterns of the combinatorial dictionary help not only to reduce the number of possible variants obtained from the parser but they include information of the syntactic level related to the semantic of the word which is required in deep levels of language analysis. We include statistical information of realizations for each valence and statistics of valence combinations for each verb in order to increase the efficiency of ambiguity resolution in syntactic analysis.
Keywords: dependency grammars, combinatorial dictionary, syntactic analysis.

1 *Introducción*

Para la representación de la estructura sintáctica en inglés se adoptaron las estructuras consideradas por las teorías lingüísticas derivadas del estructuralismo norteamericano, basadas en constituyentes o estructura de frase. Esas teorías lingüísticas también han sido adoptadas para la representación sintáctica de otros lenguajes (alemán, francés, etc.). En cambio las teorías lingüísticas desarrolladas a partir de los estudios de Tesnière (1969), las gramáticas de dependencias, no han sido ampliamente empleadas.

Por otra parte, la estructura de subcategorización se ha considerado como una información lingüística básica necesaria en los diccionarios para el procesamiento de lenguaje natural (EAGLES, 1996). Principalmente se ha empleado con la finalidad de restringir el número de variantes obtenidas en el análisis sintáctico y para la generación de textos. Los diccionarios de marcos de subcategorización que se han compilado con este propósito, manualmente, por ejemplo ALVEY (Boguraev, 1987) y COMLEX (Grishman et al, 1994), y automáti-

* Trabajo realizado con apoyo parcial del Gobierno de México (CONACyT y SNI), CGEPI-IPN y PIFI-IPN, México.

camente (Briscoe & Carrol, 1997), están basados en teorías de estructura de frase o de constituyentes.

En los formalismos gramaticales modernos se considera la información de subcategorización, que generalmente incluye referencias a todos los niveles de descripción gramatical: morfológico, sintáctico y semántico, aunque la cantidad de información que consideran difiere entre ellos. La forma de la descripción y el nivel donde se sitúa la descripción están definidos por el formalismo y los niveles considerados en él. Por ejemplo, los formalismos que consideran roles temáticos, describen las valencias del verbo por tipos de roles, los formalismos sin representación de nivel semántico describen la información semántica en términos sintácticos. Utilizamos el término valencia para describir semánticamente un actante del verbo, es decir, el rol semántico del actante, por lo que valencia y actante se emplean indistintamente.

Los formalismos basados en dependencias difieren de los formalismos basados en constituyentes, en cuanto a subcategorización se refiere, en que los primeros hacen una clara separación entre los complementos reales y los circunstanciales. En la *Dependency Unification Grammar* DUG (Hellwig, 1983) se consideran los complementos (dependientes de un elemento léxico que son requeridos por la semántica combinatoria inherente de la palabra), los adjuntos (aumentan la estructura de dependencias en forma arbitraria) y los conjuntados (introducidos por una conjunción). También se describen los adjuntos o predicados circunstanciales en la Gramática Funcional de Dependencias (Tapanainen *et al*, 1997). En la teoría Significado ⇔ Texto (*Meaning ⇔ Text Theory*, MTT) (Mel cuk, 1988) se describe la diátesis de cada verbo, es decir, la correspondencia entre los actantes semánticos, los de la sintaxis profunda y los de la sintaxis superficial. Por consiguiente, la información de subcategorización es específica para cada verbo; además, se separan las representaciones de una misma forma de palabra para un verbo dado con diferentes significados.

En los formalismos basados en constituyentes, esta separación no es primordial, por lo que pueden incluirse predicados cuya ocurrencia es obligatoria en el contexto local de la frase pero que no son seleccionados semánticamente por el verbo. Al no considerarse la información de subcate-gorización de una forma específica para cada verbo, generalmente se realiza una clasificación y entonces cada clase o marco de subcate-gorización es un patrón de composición de complementos que puede ser compartido por varios verbos. Estas consideraciones repercuten en la descripción de la diátesis que bajo este esquema considera que el verbo puede aparecer en una diversidad de marcos de subcategorización.

Para el español, considerando como finalidad el procesamiento de textos sin restricciones, existen dificultades al querer describir los objetos de los verbos en la forma en que se ha hecho para el inglés. Algunas características del español como su orden menos rígido, la inversión del sujeto, etc. requieren de una descripción más adecuada.

La descripción de todos los objetos de los verbos del español bajo el formalismo de la MTT que emplea los llamados Patrones de Rección (*Government Patterns*, GP) (Steele, 1990) permite definir, de una manera más adecuada esas características que difieren del inglés. En las descripciones de los GP cada palabra encabezado describe su significado, sus actantes, las palabras específicas que introducen los complementos que realizan sus valencias y las combinaciones de esos complementos, incluyendo el orden de sus ocurrencias, para las opciones permitidas y prohibidas. Este formalismo permite incluir cierto conocimiento semántico: el significado de la palabra encabezado del verbo y sus valencias, y sus características semánticas como animidad, si es necesario.

En este artículo, primero describimos brevemente la zona sintáctica en la MTT. Después discutimos las características del español que tienen una representación más adecuada bajo este formalismo mediante algunos ejemplos tomados de un corpus grande del español (LEXESP)[1]. Enseguida se describen los marcos de información de subcategorización basados en la MTT y finalmente los métodos posibles para realizar el análisis sintáctico.

2 Zona sintáctica en la MTT

La zona sintáctica del diccionario combinatorio en la teoría MTT describe, con la ayuda de una tabla de GP, la siguiente información: correspondencia entre las valencias semánticas y sintácticas de la cabecera del artículo lexicográfico, todas las formas en que se realizan las va-

[1] El corpus LEXESP fue proporcionado amablemente por H. Rodríguez, UPC-LSI, Barcelona, España.

lencias sintácticas y la indicación de obligatoriedad de la presencia de cada actante, si es necesario.

Después de la tabla de GP se presentan dos secciones: restricciones y ejemplos. Las restricciones consideradas en los GP son de varios tipos: semánticas, sintácticas o morfológicas; entre estas restricciones también se considera la compatibilidad entre valencias sintácticas. La sección de ejemplos cubre todas las posibilidades: ejemplos para cada actante, ejemplos de todas las posibles combinaciones de actantes y finalmente los ejemplos de combinaciones imposibles o indeseables.

La parte principal de la tabla de GP es la lista de valencias sintácticas de la cabecera del artículo lexicográfico. Se listan de una manera arbitraria pero se prefiere el orden de incremento en la oblicuidad: sujeto, objeto directo, objeto indirecto, etc. Cada cabecera usualmente impone cierto orden; por ejemplo, una entidad activa, sujeto, toma el primer lugar, después el objeto principal de la acción, después otro complemento (si existe), etc. También la forma de expresión del significado de la palabra encabezado influye el orden. Por el momento, en nuestro diccionario en lugar de la definición sólo usamos una descripción corta en inglés[2], por ejemplo la expresión para *acusar* es: *Person V accuses person W of action X* (persona V acusa a persona W de acción X). Este tipo de expresión precede cada GP.

Otra información obligatoria en cada valencia sintáctica es la lista de todas las posibles formas de expresión de la valencia en los textos. El orden de opciones para una valencia dada es arbitrario, pero las opciones más frecuentes aparecen normalmente primero. Las opciones se expresan con símbolos de categorías gramaticales o palabras específicas.

A continuación presentamos una descripción para el verbo *acusar* aunque una descripción más amplia de este diccionario aparece en (Galicia *et al.*, 1998). En la tabla, NP representa un grupo nominal e INF representa un verbo en infinitivo. Aquí solamente se presentan algunos ejemplos y no la totalidad. En Ejemplos de las combinaciones prohibidas solamente se considera la indicación de obligatoriedad.

[2] Empleamos el inglés para la descripción de significado puesto que no existe un lenguaje semántico sin homonimia ni sinonimia, por lo que el inglés parece más conveniente que el mismo español para lectores hispanohablantes.

1 = V	2 = W	3 = X
1. NP	2. *a* NP	1. *de* NP 2. *de* INF
Obligatoria	Obligatoria	

Ejemplos:

V + W Juan *acusa* a María.

V + W + X Juan *acusa* a María de robo.

Ejemplos de las combinaciones prohibidas:

V + X *Juan *acusa* de robo.

3 Características del español con mejor descripción bajo la MTT

3.1 Orden de palabras

El orden de palabras en el español es más libre comparado con el inglés y por lo tanto se requiere considerar los órdenes posibles de aparición de las valencias. Por ejemplo, en las frases siguientes el objeto indirecto no aparece después del verbo, de tres maneras distintas: 1) en la forma *a* NP antes del verbo, 2) como pronombre reflexivo entre sujeto y verbo, y 3) como clítico dentro del verbo. Los ejemplos son:

1. *A quienes acusan de comportamiento arrogante.*
2. *El fiscal me acusa de delito de alta traición.*
3. *Acusándole de ser el sostenedor y portavoz de Mario Segni.*

Para el inglés funciona buscar usualmente todos los objetos del verbo después de él. Sin embargo, para el español, esta información de posibles posiciones de la valencia es necesaria para el analizador sintáctico. En la sección 4 se presentan todas las combinaciones obtenidas de LEXESP para el verbo *acusar* con los correspondientes pesos estadísticos.

3.2 Sujeto y objeto directo

La inversión del sujeto es considerada como un recurso estilístico de frecuencia de aparición menor, pero a este respecto, el español presenta diferencias con otros lenguajes romances. En investigaciones sobre el orden de las palabras, (Zubizarreta, 1994) indica que el español y el italiano permiten la inversión libre del sujeto, a diferencia del francés. Por ejemplo, en las siguientes frases, el sujeto aparece después del

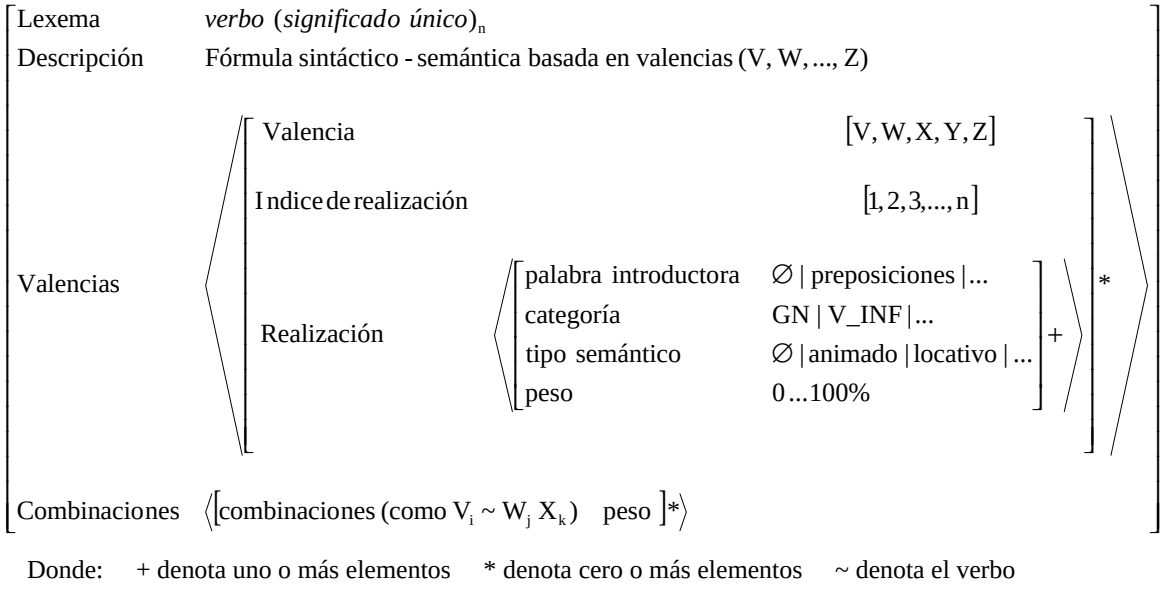

Figura 1: Estructura de la descripción de una palabra.

verbo en dos formas distintas, como nombre propio y como grupo nominal.

1. *Le acusaba Apel de desembocar en una ilusión idealista por <...>.*
2. *A quien acusaron varios testigos.*

Considerando un orden de palabras fijo, en ambos casos podría haber un reconocimiento erróneo de valencias o del significado del verbo o ambos, aunque en la primera frase el reconocimiento de nombre propio lo evitaría. En la segunda frase se requiere diferenciar entre entidad animada y grupo nominal inanimado para reconocer el sujeto de *acusar*$_1$ 'denunciar a alguien como culpable de algo . La valencia realizada como NP inanimado corresponde al verbo *acusar*$_2$ 'revelar algo, ponerlo de manifiesto (DEUM, 1996). Si *varios testigos* se reconoce como NP inanimado existe confusión entre sujeto y complemento, que resultará en una asignación de estructura incorrecta o de otro significado.

En la mayoría de los lenguajes el objeto directo está conectado con el verbo sin preposiciones; pero en español, las entidades animadas están conectadas con la preposición *a* y las no animadas directamente (*veo a mi vecina* y *veo una casa*). La animidad se considera como una personificación, por ejemplo *gobierno* en español es un sustantivo animado y al dirigirse a él se utiliza la preposición *a* (*veo al gobierno...*). Además de personas, la animidad abarca grupos de personas, animales, países, entidades abstractas (organizaciones, partidos políticos), tractas (organizaciones, partidos políticos), etc. En cambio en ruso donde también existe oposición obligatoria de palabras por animidad, los grupos de personas, los países, las ciudades no se personifican en sentido gramatical; en inglés tampoco se aplica *he* o *she* a tales entidades.

Aunque la preposición *a* también tiene otros usos, aquí nos referimos exclusivamente a su conexión con el objeto directo. Este uso sirve para diferenciar el significado de algunos verbos, por ejemplo, *querer algo* 'tener el deseo de obtener algo y *querer a alguien* 'amar o estimar a alguien . Así que la animidad es una característica evidentemente sintáctica pero con alusión semántica que se considera para la realización de las valencias y en ciertos casos permite determinar el sujeto y distinguirlo del objeto.

3.3 Valencias repetidas

Generalmente las entidades referidas por diversas valencias son diferentes. Esta es una situación normal en lenguajes naturales, cada valencia semántica puede representarse en el nivel sintáctico mediante un solo actante.

Sin embargo, existen lenguajes como el español que permite la duplicación de valencias. Los siguientes ejemplos muestran en cada oración dos grupos en negritas que representan el mismo objeto:

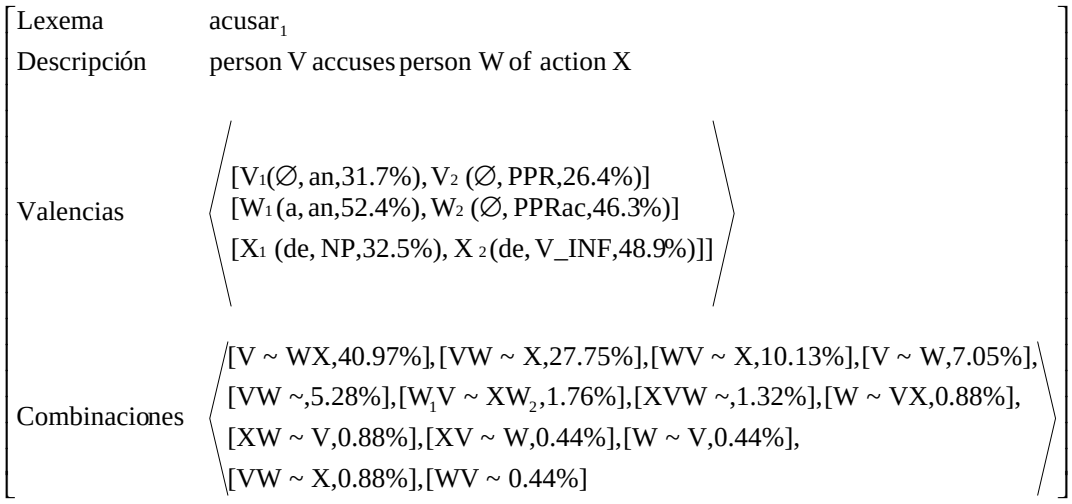

Figura 2: La representación del verbo *acusar*$_1$.

- *Arturo* ***le*** *dio la manzana* ***a Víctor****.*
- *El disfraz de Arturo, lo diseñó Víctor.*
- ***A Víctor le*** *acusa el director.*

Mientras que en la primera oración se duplica el objeto indirecto, en las siguientes oraciones se duplica el objeto directo. Mientras en la primera oración la repetición es opcional, en las siguientes es obligatoria.

Algunas veces la repetición es obligatoria. El orden de palabras y los verbos específicos imponen algunas construcciones. Por ejemplo, el cambio de los argumentos dativos y acusativos antes del verbo presenta una complicación, la repetición mostrada en los ejemplos.

Finalmente, es necesario hacer notar que todas las valencias semánticas son necesarias y suficientes, y el papel que cada una de ellas desempeña está implicado directamente por la definición lexicográfica del lexema correspondiente, pero que la situación con las valencias sintácticas es mucho más complicada y no siempre existe una correspondencia de uno a uno entre valencias sintácticas y semánticas.

3.4 Información semántica en el nivel sintáctico

Existe información semántica detectable en el nivel sintáctico requerida en niveles más profundos del procesamiento de lenguaje natural. Por ejemplo, la detección de valencias sintácticas que se enlazarán a las valencias semánticas y la distinción entre complementos reales del verbo y complementos circunstanciales realizados con la misma descripción de subcategorización.

La identificación de valencias del verbo se realiza a través de la detección de palabras específicas introductoras de los complementos. Para algunos verbos una sola palabra se identifica como la palabra que introduce los complementos que expresan una valencia, en cambio en otros verbos varias palabras se emplea*n con el mismo propósito. Por ejemplo, para acusar*, la preposición *de* en *de* NP es la preposición que introduce el complemento mediante el cual se expresa la acción por la cual se acusa a alguien. Para el verbo *expresarse* las preposiciones *en*, *de*, *con* y *mediante* en *en* NP, *de* NP, *con* NP, *mediante* NP se emplean para realizar la valencia que describe la forma en que se expresa algo. Las preposiciones también ayudan a distinguir el significado de algunos verbos. Por ejemplo, hay muchos verbos como *lanzar* que tiene diferentes significados: existe *lanzar*$_1$ (Ej. *lanzó una pelota*) y *lanzar*$_2$ (Ej. *la lanzaron de su casa*), estos dos verbos homónimos emplean diferentes preposiciones para cada significado específico.

Para algunos verbos, un marco de subcategorización describe tanto valencias del verbo como circunstancias. Por ejemplo, algunos verbos locativos (Rojas, 1988) requieren complementos con la noción de espacio, cuya marca aparece tanto en la palabra introductora del complemento como en el complemento mismo. Por ejemplo el verbo *colocar*, en la frase *coloca un libro en este momento en el espacio disponible*, el marco de subcategorización *en* NP describe tanto una valencia (*en el espacio libre*) como un complemento irrelevante para su significado (*en este momento*). En los GP, esta mar-

$$\begin{bmatrix} \text{Lexema} & acusar_2 \\ \text{Descripción} & \text{V reveals W} \\ \text{Valencias} & \left\langle \begin{matrix} [\text{V}, \langle (\varnothing, \text{NP}, 24.3\%), (\varnothing, \text{an}, 37.8\%), (\varnothing, \text{PPR}, 16.2\%) \rangle] \\ [\text{W}, \langle (\varnothing, \text{NP}, 100\%) \rangle] \end{matrix} \right\rangle \\ \text{Combinaciones} & \langle [\text{V} \sim \text{W}, 97.3\%], [\text{WV} \sim, 2.7\%] \rangle \end{bmatrix}$$

Figura 3: La representación del verbo $acusar_2$.

ca de locatividad al igual que la de animidad se introduce en la realización de las valencias.

4 Marcos Avanzados de Subcategorización

De aquí en adelante nos referimos a los *marcos avanzados de subcategorización* (MAS) en lugar de GP para evitar un nombre ligado especialmente a la MTT y para introducir información adicional que sirva al análisis sintáctico. La información contenida en estos marcos corresponde a la expuesta en la sección 2, considerada en la tabla de GP, salvo la indicación de obligatoriedad de la presencia de cada valencia. También se introduce un índice de realización para identificar las diferentes realizaciones de las valencias, y la duplicación de valencias. En un MAS, la indicación de obligatoriedad, las posibles combinaciones de actantes y las combinaciones prohibidas han sido consideradas de otra forma.

El español tiene un orden de palabras más libre que el inglés pero no totalmente libre, por lo que las posibles combinaciones de valencias son limitadas. A partir de la indicación de obligatoriedad se pueden definir algunas combinaciones no deseadas pero no la totalidad. Las combinaciones posibles y las prohibidas pueden definirse basándose en cierta experiencia pero no reflejarían los cambios en el lenguaje ni las preferencias en dominios específicos. Por lo que para especificar esta información se consideró la obtención de pesos estadísticos. Si una valencia tiene presencia en todas las oraciones extraídas del corpus para un verbo específico, se considera como una evidencia de obligatoriedad. El analizador sintáctico empleará esta evidencia para buscar las valencias aún en posiciones distantes. Por ejemplo, el verbo *acusar* requiere la presencia del objeto directo, con esta indicación el analizador sintáctico buscará este pedazo de información alrededor del verbo, considerando también las probabilidades de su aparición antes y después del verbo.

Así que se obtienen los pesos estadísticos para cada valencia referidos a las palabras introductoras de ellas y después los pesos estadísticos de las combinaciones de valencias referidas a la posición de cada valencia respecto al verbo. Esta información estadística da un rango de las descripciones de cada tipo específico de cada valencia y de sus combinaciones que permitirán incrementar la eficiencia del analizador sintáctico. En la Figura 1 se muestra la descripción general de los MAS.

En las Figuras 2 y 3 se muestran, en una presentación más práctica que la definida en la Figura 1, los MAS obtenidos a partir de un total de 227 oraciones del corpus LEXESP para el verbo *acusar*. De la información obtenida, se reconocen $acusar_1$ y $acusar_2$. La valencia realizada mediante NP que no pertenece a la expresión de $acusar_1$ marca la diferencia entre los dos verbos, siempre y cuando pueda discriminarse entre entidades animadas y no animadas. Nótese en estas figuras las diversas variantes en el orden de palabras, y la representación de valencias duplicadas en la Figura 2 con el caso $[\text{W}_1\text{V} \sim \text{XW}_2, 1.76\%]$.

Para compilar marcos de subcategorización de este tipo requerimos información sintáctica, estadística y conectada con la semántica. En la parte semántica es necesario incluir la marca de animidad y de locatividad en el corpus. Además, se requiere detectar la llamada atracción léxica (coocurrencia en estructura sintáctica) entre los verbos y las preposiciones que introducen las valencias, y diferenciar las valencias correspondientes a diversos significados del verbo.

Una aproximación para la obtención de esta información, a partir de un corpus, en cuanto a detección de las frases preposicionales y com-

plementos que realizan las valencias se describe en (Galicia et al, 2001). La detección de valencias del verbo requiere anotación manual pero la información extraída para la compilación de los MAS facilita la detección y anotación de las valencias del verbo.

5 Análisis sintáctico

Para la realización del análisis sintáctico usando los MAS, se puede emplear la estrategia natural y general de análisis sintáctico en dos pasos de las gramáticas lexicalizadas. En el primer paso se selecciona el conjunto de estructuras correspondientes a cada palabra en la oración de entrada. El segundo paso analiza sintácticamente la oración respecto a las estructuras seleccionadas.

En (Joshi, 1994) se presenta una variante, dada una oración de entrada el analizador selecciona el conjunto de llamadas *supertags* (árboles elementales) para cada elemento léxico de la oración de entrada. Enseguida se intenta combinar las *supertags* buscando todas las posibilidades para producir el análisis sintáctico de la oración. Cuando se completa el análisis existe una sola *supertag* para cada palabra, si no hay ambigüedad global. Alternativamente, se puede eliminar o reducir sustancialmente la ambigüedad de asignación de las *supertags* antes del segundo paso, usando información local como las dependencias léxicas locales o métodos estadísticos.

Otra variante es el analizador sintáctico creado dentro del proyecto ruso ETAP-2, para traducción inglés-ruso, que emplea un diccionario de patrones de manejo sintáctico. Este analizador sintáctico emplea dos tipos de reglas: de sintagmas y de preferencias. Las reglas de sintagmas son las más importantes. Establecen una relación sintáctica entre dos palabras. El sintagma contiene un conjunto de condiciones que deben cumplir estas dos formas de palabra. El algoritmo que se emplea es el siguiente. En un primer paso se crea un grafo de hipótesis posibles representadas por ligas binarias entre palabras. En un segundo paso se eliminan algunas de las hipótesis para crear un árbol. El primer paso se realiza al aplicar las condiciones lineales de todos los sintagmas y el segundo paso se logra mediante algoritmos de filtrado que consideran información lingüística.

Para abordar el problema de la ambigüedad sintáctica, proponemos un modelo para análisis sintáctico y desambiguación basado en dependencias léxicas entre palabras predicativas y en proximidad semántica. El primero sigue el enfoque de dependencias mediante patrones de manejo sintáctico descrito en este trabajo. La proximidad semántica se sustentará en el empleo de una red semántica para incorporar contexto local en un analizador que sigue el enfoque de constituyentes. Para la desambiguación sintáctica proponemos la clasificación de la totalidad de variantes de las estructuras obtenidas, por medio de un peso asignado.

En experimentos realizados, los resultados obtenidos aplicando nuestro método (Galicia et al, 2001) de compilación de información sintáctica, dependencias léxicas entre palabras predicativas, aplicados al corpus LEXESP fueron usados para analizar sintácticamente 100 oraciones y las estructuras verdaderas se encontraron clasificadas en el primer rango del 35%.

Conclusiones

Se describió la aplicación del formalismo de la teoría Significado ⇔ Texto (*Meaning ⇔ Text Theory*) para el análisis sintáctico del español, las ventajas de su uso al describir más adecuadamente características del español, la información requerida para el diccionario y la forma en que se podría realizar el análisis sintáctico.

En este método, basado en gramáticas con dependencia, el diccionario combinatorio consta de patrones para palabras, principalmente verbos aunque también pueden describirse de esta forma algunos adjetivos y sustantivos, y en ellos se describen todas sus valencias y las formas en que éstas se realizan. Se mostró la necesidad de incluir información estadística tanto para representar combinaciones de valencias y la condición de obligatoriedad de algunas valencias como para incrementar la eficiencia del analizador sintáctico

También se discutió la necesidad de detectar cierta información del nivel sintáctico (para distinguir diferentes significados de un verbo, para discriminar las valencias de los verbos, etc.) que está conectada con la semántica de la palabra y que es requerida a niveles más profundos del análisis de lenguaje natural.

Bibliografía

Boguraev, B. et al. 1987. The derivation of a grammatically indexed lexicon from the Longman Dictionary of Contemporary English. *Proc. of the 25th Annual meeting of the*

Association for Computational Linguistics, Stanford, CA.

Briscoe, E. & Carroll, J. 1997. Automatic extraction of subcategorization from corpora. *Proc. of the 5th ACL Conference on Applied Natural Language Processing*. Washington, DC.

DEUM. 1996. *Diccionario del Español usual en México*. El Colegio de México.

EAGLES. 1996. *Recommendations on Subcategorization*; www.ilc.pi.cnr.it/EAGLES 96/synlex/synlex.html.

Galicia-Haro, S., A. Gelbukh e I. Bolshakov. 1998. Diccionario de patrones de manejo sintáctico para análisis de textos en español. *Procesamiento del Lenguaje Natural*, 23:171–176.

Galicia-Haro, S., A. Gelbukh e I. Bolshakov. 2001. Obtención semiautomática de patrones de manejo para el lenguaje español. *Proc. SLPLT-2, Segundo Taller Internacional de Procesamiento Computacional del Español y Tecnologías del Lenguaje*. Editorial Club Universitario, España,, pp. 147–151.

Grishman, R., C. Macleod, and A. Meyers. 1994. COMLEX syntax: building a computational lexicon. *Proc. of the 15th Conference on Computational Linguistics*, COLING-94, pp. 268–272.

Hellwig, P. 1983. Dependency Unification Grammar, in E. Hajicova (ed.), *Functional Description of Language*, Charles University, Prague, pp. 67–84.

Joshi, A. and B. Srinivas. 1994. Disambiguation of Super Parts (or Supertags) of Speech Almost Parsing. *Proc. of the 15th International Conference on Computational Linguistics*, COLING-94, pp. 154–160.

Mel cuk, I. A. 1988. *Dependency Syntax: Theory and Practice*. State University of New York Press.

Rojas, C. 1988. *Verbos locativos en español. Aproximación sintáctico-semántica*. Universidad Autónoma de México.

Steele, J. (ed). 1990. *Meaning – Text Theory. Linguistics, Lexicography, and Implications*. University of Ottawa press.

Tapanainen, P., Järvinen, T., Heikkilä, J., Voutilainen. A. 1997. *Functional Dependency Grammar*. http://www.ling.helsinki.fi/ ~tapanain/dg/

Tesnière, L. 1969. *Éléments de syntaxe structura*le, Paris: Klincksieck.

Zubizarreta, María Luisa. 1994. El orden de palabras en español y el caso nominativo. *Gramática del Español*, edición a cargo de Violeta Demonte. El Colegio de México.