

## Extracción de relaciones sintagmáticas de corpus anotados

Borja Navarro, Lorenza Moreno-Monteagudo y Patricio Martínez-Barco

Departamento de Lenguajes y Sistemas Informáticos

Universidad de Alicante

Carretera San Vicente del Raspeig s/n  
03690 San Vicente del Raspeig, Alicante  
{borja,loren,patricio}@dlsi.ua.es

**Resumen:** En este trabajo se presenta un nuevo recurso, diseñado con el objetivo de ser usado en desambiguación semántica de las palabras en español, basado en las relaciones sintagmáticas entre las palabras. Las relaciones sintagmáticas son relaciones entre sentidos dentro de un sintagma o dentro de una oración. En nuestro caso, estas relaciones han sido extraídas de un corpus anotado manualmente, el corpus Cast3LB. Este corpus ha sido anotado con información sintáctica (constituyentes y funciones) e información semántica (sentidos de WordNet). De este corpus se han extraído aproximadamente 3 000 patrones. Cada patrón muestra la relación sintagmática entre el sentido del verbo y el sentido de sus argumentos dentro de una oración. Sin embargo, estos patrones son excesivamente específicos para ser usados en contextos multilingües o en documentos de dominio general. Para solucionar esto es necesario obtener patrones más generales semánticamente, para lo cual se ha utilizado la ontología SUMO

**Palabras clave:** Corpus, WSD

**Abstract:** In this paper, we present a new resource, designed for being used in WSD, based on syntagmatic relations between senses for Spanish. These relations have been extracted from a corpus: the Cast3LB corpus which has been manually annotated with syntactic and semantic information (WordNet senses). From it, approximately 3 000 patterns have been extracted. These patterns show the syntagmatic relations between verb senses and its arguments within a sentence. However, these patterns can be too specific to be used in multilingual contexts or in open domain texts. Consequently, it is necessary to obtain more abstract patterns. In order to do so, we have also developed general patterns using semantic classes based on the SUMO ontology.

**Keywords:** Corpora, WSD

### 1. Introducción

La resolución automática de la ambigüedad semántica de las palabras (*Word Sense Disambiguation* (WSD)) es una tarea compleja en la que están implicados diferentes tipos de información lingüística como información morfológica y categorial (categoría gramatical y lema de las palabras), información sobre el contexto local donde aparecen las palabras (n-gramas), información sintáctica, información semántica (el sentido de palabras monosémicas, roles semánticos, preferencias semánticas, etc.)<sup>1</sup>.

Toda esta información suele ser extraída

de diferentes recursos lingüísticos como corpus o lexicones. Uno de los principales recursos léxicos utilizado es WordNet (Fellbaum, 1998). Todos los sentidos en WordNet (synsets) están relacionados entre ellos mediante relaciones paradigmáticas: sinonimia, hiponimia, hiperonimia, antonimia, etc. Todas estas relaciones lingüísticas han sido ampliamente utilizadas en WSD.

Sin embargo, para una tarea como la resolución de la ambigüedad de las palabras es necesario conocer, no sólo las relaciones paradigmáticas que una palabra mantiene con otras dentro del léxico, sino también las relaciones sintagmáticas entre sentidos dentro de una oración (Bentivogli y Pianta, 2004). De hecho, en muchas ocasiones, el sentido correcto de una palabra ambigua depende del sen-

<sup>1</sup>Este trabajo ha sido parcialmente financiado por el proyecto CES-ECE (HUM2004-21127-E) y el proyecto R2D2 (TIC2003-07158-C04-01).

tido de las palabras con las que aparece en el texto: el sentido de las palabras del contexto local, el sentido de las palabras con las que mantiene dependencias sintácticas, etc. Todas estas relaciones se pueden enmarcar dentro de las relaciones sintagmáticas.

Por ejemplo, el sentido de un verbo influye en el sentido de su objeto y su sujeto. Un verbo como “andar” normalmente necesita un sujeto de tipo “animado”, o un verbo como “comer” precisa de un objeto “comestible” (Resnik, 1993). Este es un caso específico de relación sintagmática, relacionado con las “preferencias semánticas” de (Wilks, 1972).

Además, con el uso de estas relaciones sintagmáticas es posible obtener coocurrencias entre verbos y argumentos. Por ejemplo, es muy común que el verbo “crear#00926361” aparezca con el objeto “empleo#08547608”. Este es un caso específico de relación entre sentidos con cierto grado de repetición en la lengua.

En este trabajo se presenta un nuevo recurso, con el objetivo de usarlo en WSD, basado en relaciones sintagmáticas entre sentidos en español. Para este recurso nos hemos centrado en un tipo específico de relación sintagmática: la relación entre un sentido verbal y el sentido de sus argumentos.

Estas relaciones han sido extraídas automáticamente del corpus Cast3LB, que ha sido anotado manualmente con información sintáctica y semántica. La información sintáctica anotada son constituyentes sintácticos y funciones. La información semántica es el sentido de cada nombre, verbo y adjetivo. A partir de esta información, se han extraído patrones verbales en los cuales el sentido de cada verbo está relacionado con el sentido del núcleo de cada argumento. Con ello, estos patrones muestran las relaciones sintagmáticas en una oración entre el sentido verbal y cada argumento.

Para obtener un recurso más general, de cada patrón se ha derivado un tipo de patrón más general, partiendo de la clase semántica de cada sentido del patrón.

El resto del trabajo está organizado de la siguiente forma: en la siguiente sección se presentará el corpus Cast3LB. En la sección 3 se presentará el modelo de patrón, y en la siguiente sección el proceso de extracción y generalización. La descripción del recurso obtenido se desarrolla en la sección 5. En la sección 6 se presentan trabajos relacionados y en

la última sección las conclusiones.

## 2. *El corpus Cast3LB*

En el corpus Cast3LB, todos los nombres, verbos y adjetivos han sido anotados manualmente con su sentido concreto, siguiendo una aproximación “all words”. El recurso léxico utilizado como referencia ha sido WordNet español (Navarro et al., 2003).

Los textos que forman el corpus Cast3LB han sido extraídos del corpus Lexesp (Sebastián et al., 2000) y del corpus Hermes<sup>2</sup>. Así, el corpus está formado por muestras de diferentes tipos de textos: noticias, ensayo, noticias deportivas, textos científicos, editoriales, textos literarios, etc.

Cast3LB tiene aproximadamente 13 412 palabras anotadas: 4 705 nombres, 1 498 verbos y 2 395 adjetivos.

Comparando Cast3LB con otros corpus anotados con sentidos, tiene un tamaño medio. Es más pequeño que el corpus SemCor (250 000 palabras) (Miller et al., 1993) y el corpus MultiSemCor (92 820 palabras anotadas para el italiano) (Bentivogli y Pianta, 2005), pero tiene más palabras anotadas que los corpus “all words” utilizados en SENSEVAL-3 para el italiano (5 189 palabras: 2 583 nombres, 1 858 verbos, 748 adjetivos) (Uliveri et al., 2004) o para el inglés (5 000 palabras aproximadamente) (Snyder y Palmer, 2004).

Cast3LB ha sido anotado por tres anotadores. Sin embargo, no se ha seguido anotación doble en paralelo, sino que se ha seguido una anotación transversal: cada anotador ha anotado la misma palabra a lo largo de todo el corpus al mismo tiempo. De esta manera, la consistencia de la anotación se mantiene para cada palabra. Además, el proceso de anotación es más eficaz pues el anotador debe analizar la estructura de sentidos de cada palabra ambigua sólo una vez.

Junto a la anotación semántica, el corpus Cast3LB ha sido anotado manualmente con información sintáctica. En concreto, ha sido anotado con información sobre constituyentes sintagmáticos y funciones (sujeto, objeto directo, etc.) (Civit et al., 2003)

<sup>2</sup><http://nlp.uned.es/hermes/>

### 3. *Patrones sintáctico-semánticos de relaciones sintagmáticas*

El principal objetivo de los patrones sintáctico-semánticos es representar la estructura argumento-predicado de las oraciones: la relación que se establece entre sentidos dentro de una oración. Estas relaciones, como se ha comentado anteriormente, son las relaciones sintagmáticas entre el sentido verbal y el sentido de los argumentos.

Desde un punto de vista teórico, un patrón sintáctico-semántico es la formalización de la estructura argumento-predicado relacionada con un verbo. Está formado por tres constituyentes principales:

1. El sentido específico de un verbo.
2. Los principales argumentos relacionados con el sentido del verbo. En concreto, el núcleo de cada argumento y la relación sintáctica entre los argumentos y el verbo.
3. Los rasgos semánticos de los argumentos, esto es, el sentido del núcleo de cada argumento.

El límite del patrón es la oración, y su elemento central es el verbo. La estructura argumento-predicado está relacionada directamente con el sentido verbal.

Desde un punto de vista semántico, cada argumento se identifica mediante el sentido de su núcleo. De esta manera, al final se obtienen relaciones sintagmáticas entre sentidos en cada patrón sintáctico-semántico.

Sin embargo, estos sentidos de WordNet son demasiado específicos para la caracterización semántica de los argumentos y del verbo. Nuestro objetivo es desarrollar un recurso útil para textos de dominio general y, más allá, para colecciones textuales multilingües. Por ello, se debe hacer una representación semántica del verbo y los argumentos mucho más general y abstracta.

Por esta razón, se ha generalizado el sentido de los patrones mediante clases semánticas. Hemos utilizado el conjunto de clases semánticas de la ontología SUMO (Niles y Pease, 2001), que ya ha sido utilizada previamente en WSD.

### 4. *Extracción y generalización*

El proceso de anotación está organizado en estas cuatro fases:

1. El primer paso es la localización y extracción del verbo
2. Una vez que el verbo ha sido extraído, el sistema de extracción busca sus argumentos. En este paso, el sistema se guía por la información sobre función sintáctica (sujeto, objeto, etc.) de cada argumento. Para cada argumento localizado, se extrae su núcleo, bien sea un núcleo nominal, pronominal o verbal. Se extrae, además, la etiqueta de función y la categoría sintáctica.

Hay algunas oraciones en las que es posible que aparezcan argumentos con núcleo verbal. Esto ocurre con las cláusulas subordinadas. Para estos casos, el núcleo del argumento es el propio verbo, pero, además, estos argumentos son nuevos patrones. Por tanto, este verbo es a su vez considerado núcleo de un nuevo patrón.

3. El último paso es la extracción de los rasgos semánticos para cada núcleo argumental. Una vez que un núcleo verbal o nominal es extraído, el sistema extrae su sentido marcado en el corpus. Dado que el corpus está anotado manualmente, no hay problemas con la ambigüedad de las palabras.

Este proceso de extracción está basado en la extracción de patrones de subcategorización de texto plano, como (Manning, 1993) o (Korhonen, 2002). A diferencia de estos, el corpus que hemos utilizado está etiquetado a mano, lo cual presenta clara ventajas como:

- Primero, toda la información lingüística extraída (sentidos, relaciones sintagmáticas, etc.) está anotada explícitamente en el corpus. Por ello no es necesario ningún tipo de proceso de inferencia para extraer la información lingüística.
- En segundo lugar, dado que la información lingüística ha sido revisada manualmente, en principio es correcta y no es necesario ningún tipo de filtro estadístico para rechazar patrones erróneos.
- Por último, la información es extraída de textos reales que representan el estado actual de la lengua.

Sin embargo, el uso de corpus anotados presenta dos desventajas importantes:

- El tamaño del corpus es pequeño: no hay suficientes datos para desarrollar un recurso robusto. Más aún, no es posible construir un corpus anotado lo bastante grande como para dar cuenta de todos los aspectos de las lenguas.
- Las relaciones sintagmáticas y los sentidos de las palabras son específicos del dominio y de la lengua. Por ello, no es posible utilizar este tipo de patrón en colecciones multilingües de textos.

Para resolver estos problemas se ha llevado a cabo un proceso de abstracción y generalización de los patrones. Cada patrón ha sido completado con la clase semántica correspondiente al sentido de cada palabra. Las clases semánticas utilizadas han sido las clases de la ontología SUMO.

**SUMO** (*Suggested Upper Merge Ontology*) está formada por conceptos lo suficientemente generales para cubrir, a alto nivel, un conjunto amplio de áreas. Incluye ontologías de distintos dominios como comunicaciones, países y regiones o economía y finanzas entre otras. Ha sido mapeada a las distintas versiones de WordNet. Hemos utilizado la versión mapeada a WordNet 1.6 que contiene en total unos 20 000 conceptos y 60 000 axiomas, y 687 clases diferentes (Niles y Pease, 2001).

Durante el proceso de generalización, el sentido de los verbos y de los argumentos es completado con su correspondiente clase semántica. Así, los patrones sintáctico-semánticos están formados por el sentido de verbos y argumentos, la clase semántica correspondiente a cada uno y las relaciones sintagmáticas entre ellos.

Por ejemplo, un patrón recurrente en el corpus (cinco apariciones) es

“*crear#00926361-empleo#08547608 (CD)*”

Dado que los sentidos de WordNet son muy específicos, el patrón resultante también lo es. Mediante el proceso de abstracción y generalización, se obtiene un patrón mucho más general como el siguiente:

“*Intentional process - employ (CD)*”

Estos patrones generales son independientes del dominio dado que están formados por clases semánticas genéricas y no por sentidos específicos. Además, al ser más genera-

les, pueden ser utilizados en ámbitos multilingües: todas las clases semánticas están relacionadas con los sentidos de WordNet. Así, a través del Inter-lingua Index de EuroWordNet, es posible utilizar el patrón en otra lengua.

### 5. Descripción del recurso obtenido

Del corpus Cast3LB se han extraído 3 002 patrones. Todos ellos tienen especificado el sentido del verbo y el sentido de cada uno de los núcleos argumentales.

Aquellos patrones que no tenían este sentido específico han sido rechazados. Por ejemplo,

- patrones en los que no aparece ninguna palabra como núcleo en el corpus,
- patrones en los que el sentido del núcleo verbal o argumental no aparece en WordNet español,
- patrones cuyo núcleo verbal es un verbo copulativo,
- patrones que contienen algún núcleo argumental pronombre,

Estos casos, al no tener sentido, dejan huecos en la estructura del patrón y, por tanto, han sido rechazados.

Con esto se han obtenido 1 512 patrones diferentes de 1 argumento, 853 patrones diferentes de 2 argumentos, 555 de 3, 72 de 4 y 10 patrones distintos de 5 argumentos. Como se ha comentado anteriormente un total de 3 002 patrones.

En cambio, considerando las clases semánticas, se han obtenido un total de 2 430 patrones de las cuales 953 tienen un único argumento, 840 tienen 2 argumentos, 555 tienen 3, 72 tienen 4 y 10 patrones distintos de 5 argumentos. La razón por la que se obtiene un menor número de patrones es porque existen patrones con las mismas relaciones paradigmáticas, pero con sentidos diferentes en el verbo y en los argumentos, que tienen las mismas clases semánticas. Son patrones diferentes si se consideran los sentidos, pero iguales si se consideran las clases semánticas. A medida que aumenta el número de argumentos los patrones son más específicos y el número de patrones diferentes no disminuye tanto.

Los patrones están codificados en XML. En el apéndice 1 se puede ver como ejemplo un patrón completo.

## 6. Trabajos relacionados

Otros recursos léxicos con información sobre argumentos verbales son Framenet (Ruppenhofer et al., 2005) o PropBank (Palmer, Gildea, y Kingsbury., 2004). Éstos especifican el rol semántico entre los argumentos y el verbo (Agente, Paciente, Tema, Instrumentos, etc.). El recurso aquí presentado no tiene marcado los roles semánticos, sólo las funciones sintácticas: sujeto, objeto directo, etc.

Otro trabajo centrado en la extracción de relaciones sintagmáticas para WSD es (Bentivogli y Pianta, 2004). Por un lado, este trabajo se ha centrado en las relaciones sintagmáticas dentro de sintagmas nominales, no en las relaciones sintagmáticas dentro de oraciones. Por otro lado, sólo consideran relaciones sintagmáticas con un alto nivel de lexicalización, es decir, con un uso muy estable y común en la lengua.

(Agirre y Martínez, 2001; Agirre y Martínez, 2002) proponen un sistema muy similar de extracción de relaciones sintagmáticas de corpus anotados con sentidos. El proceso, al igual que el expuesto aquí, está dividido en dos partes: una primera de extracción de patrones y una segunda de generalización.

Este trabajo presenta dos diferencias importantes con relación al nuestro. En primer lugar, se centran únicamente en relaciones sujeto - verbo y verbo - objeto. Para establecer estas relaciones, se basan en un analizador sintáctico automático (Minipar), que puede acumular errores. Los patrones que nosotros extraemos se basan en un análisis manual, por ello pueden ser cualquier tipo de relación: se dispone de información sobre las funciones sintácticas de los argumentos validada manualmente. Por esta misma razón, el error en el análisis sintáctico es menor.

Sin embargo, (Agirre y Martínez, 2001) utilizan el corpus SemCor (Miller et al., 1993), que consta de 250 000 palabras. Este corpus es mucho mayor que el Cast3LB, con 100 000 palabras. Por ello, la información de frecuencia que extraen tiene más valor que la extraída del corpus Cast3LB.

La segunda diferencia es la idea de clase. (Agirre y Martínez, 2001) parten de la palabra y generalizan a una clase, entendiendo por clase o concepto el synset de WordNet.

Nosotros partimos ya del synset (concepto o clase para (Agirre y Martínez, 2001)) y generalizamos a clases ontológicas (SUMO).

Otro trabajo similar, desarrollado dentro del proyecto MEANING es (Atserias, Villarejo, y Rigau, 2003; Atserias et al., 2003). Al igual que el presentado aquí, trabajan sobre español. Además, desarrollan un proceso de generalización de patrones con relaciones sintagmáticas haciendo uso de clases semánticas generales.

La principal diferencia es el modelo de patrón que extraen. El corpus utilizado para español es EFE, analizado automáticamente con un analizador sintáctico superficial. Por tanto, los constituyentes extraídos (sintagmas nominales) son superficiales. No especifican si la relación entre el verbo y los sintagmas argumentos es sujeto u objeto. Simplemente extraen la colocación tipo "SN verbo SN". El patrón aquí presentado es más desarrollado, ya que es extraído de un corpus anotado a mano. Esto permite ser más preciso en los constituyentes del patrón y en las relaciones entre el verbo y los argumentos.

Por último, la anotación del corpus Sensem (Catellón et al., 2006), que actualmente está en desarrollo, es muy similar al tipo de patrón aquí expuesto y extraídos del corpus Cast3LB. El objetivo del corpus Sensem es anotar el comportamiento sintáctico y semántica de 250 000 verbos en español. Dado que están centrados en los verbos, incluyen más información lingüística que la extraída en nuestro trabajo como, por ejemplo, roles semánticos o Aktionsart (aspecto). La información del corpus Cast3LB es más general, no centrada sólo en las estructuras verbales<sup>3</sup>.

## 7. Conclusiones y trabajo futuro

En este trabajo se ha presentado un recurso basado en relaciones sintagmáticas entre sentidos, orientado a la tarea de resolución de la ambigüedad de las palabras. Este recurso ha sido desarrollado automáticamente a partir de un corpus anotado manualmente: el corpus en español Cast3LB. Este corpus ha sido anotado con información sintáctica (constituyentes y funciones) y con información semántica (sentidos de palabras). Se han extraído patrones sintáctico-semánticos for-

<sup>3</sup>En esta línea, dentro del proyecto CESS-ECE se está desarrollando la ampliación y anotación del corpus Cast3LB con roles semánticos, siguiendo el modelo de Propbank

mados por el sentido del verbo, el sentido de los argumentos y las relaciones sintagmáticas entre ellos. Dado que estos patrones son muy específicos, se ha desarrollado un proceso de generalización en el que cada sentido ha sido completado con su correspondiente clase semántica según la ontología SUMO. Con ello, los patrones resultantes son más independientes del dominio y de la lengua de origen.

Nuestra intención es aplicar el recurso desarrollado dentro de un sistema de WSD basado en clases semánticas. Además de otros rasgos, se tendrán en cuenta las relaciones sintagmáticas de cada palabra ambigua dentro del patrón donde aparece: el sentido o clase semántica del verbo con el que aparece, o el sentido o clase semántica de los argumentos con los que aparece en la oración.

### Bibliografía

- Agirre, E. y D. Martínez. 2001. Learning class-to-class selectional preferences. En *Workshop on 'Computational Natural Language Learning' (CoNLL-2001)*, Toulouse, Francia.
- Agirre, E. y D. Martínez. 2002. Integrating selectional preferences in wordnet. En *1st International WordNet Conference*, Mysore, India.
- Atserias, J., M. Castillo, F. Real, H. Rodríguez, y G. Rigau. 2003. Exploring large-scale acquisition of multilingual semantic models for predicates. *Revista de Procesamiento del Lenguaje Natural*, 31.
- Atserias, J., L. Villarejo, y G. Rigau. 2003. Integrating and porting knowledge across languages. En *RANLP 2003*, Borovets, Bulgaria.
- Bentivogli, L. y E. Pianta. 2004. Extending wordnet with syntagmatic information. En *2n GWC*, Brno.
- Bentivogli, L. y E. Pianta. 2005. Exploiting parallel texts in the creation of multilingual semantically annotated resources: The multiseacor corpus. *Natural Language Engineering. Special Issue on Parallel Text*, 11(3):247–261.
- Catellón, I., A. Fernández-Montraveta, G. Vázquez, L. Alonso, y J. Capilla. 2006. The sensemap corpus: a corpus annotated at the syntactic and semantic level. En *Language Resources and Evaluation - LREC 2006*, Génova, Italia.
- Civit, Montserrat, M<sup>a</sup>Antonia Martí, Borja Navarro, Núria Bufí, Belén Fernández, y Raquel Marcos. 2003. Issues in the Syntactic Annotation of Cast3LB. En *Proceedings of 4th International on Workshop on Linguistically Interpreted Corpora (LINCOL-03)*. *EA CL03*, páginas 9 – 16, Budapest.
- Fellbaum, Christiane, editor. 1998. *WordNet. An Electronic Lexical Database*. The MIT Press, Cambridge.
- Korhonen, Anna. 2002. *Subcategorization acquisition*. Technical Report. University of Cambridge, Cambridge.
- Manning, Christopher. 1993. Automatic acquisition of a large subcategorization dictionary from corpora. En *Proceedings of the ACL*.
- Miller, G. A., C. Leacock, T. Radee, y R. Bunker. 1993. A semantic concordance. En *Proceedings of the 3rd ARPA Workshop on Human Language Technology*, San Francisco.
- Navarro, Borja, Montserrat Civit, M. Antonia Martí, Raquel Marcos, y Belén Fernández. 2003. Syntactic, semantic and pragmatic annotation in Cast3LB. En *SProLac. Proceeding of Workshop Shallow Processing of Large Corpus. Corpus Linguistics 2003*, Lancaster.
- Niles, Ian y Adam Pease. 2001. Towards a standard upper ontology. En *Proceedings of 2nd International Conference on Formal Ontology in Information Systems (FOIS'01)*, Ogunquit, USA.
- Palmer, M., D. Gildea, y P. Kingsbury. 2004. The proposition bank: An annotated corpus of semantic roles. *Computational Linguistics*.
- Resnik, Philip. 1993. *Selection and Information: A Class-Based Approach to Lexical Relationships*. Ph.D. tesis, University of Pennsylvania.
- Ruppenhofer, J., M. Ellsworth, M. Petruck, y C. Johnson. 2005. *Framenet: Theory and Practice*. <http://framenet.icsi.berkeley.edu>, June, 13.

Sebastián, N., M.A. Martí, M. F. Carreiras, y F. Cuetos. 2000. *LEXESP: Léxico Informatizado del Español*. Edicions de la Universitat de Barcelona, Barcelona.

Snyder, Benjamin y Martha Palmer. 2004. The english all-word task. En *Proceedings of SENSEVAL-3: Third International Workshop on the Evaluation of Systems for the Semantic Analysis of Text*, Barcelona. ACL.

Uliveri, Marisa, Elisabetta Guazzini, Francesca Bertagna, y Nicoletta Calzolari. 2004. Senseval-3: The italian all-words task. En *Proceeding of Senseval-3: Third International Workshop on the Evaluation of Systems for the Semantic Anlysis of Texts*, Barcelona.

Wilks, Yorick. 1972. *Grammar, meaning and the machine analysis of language*. Routledge, London.

#### A. Anexo 1: Ejemplo de patrón sintáctico-semántico

```
<patron ora=e2-0.xml1> <oracion>No espere el lector una definición
previa de la palabra cultura</oracion>
<verbo>
  <texto>espere</texto>
  <pos>vmsp3s0</pos>
  <lema>esperar</lema>
  <sentido>00405636</sentido>
  <voz>activa</voz>
  <sumo>IntentionalPsychologicalProcess</sumo>
</verbo>
<argumento id=1>
  <texto>el lector</texto>
  <sintagma>sn</sintagma>
  <funcion>SUJ</funcion>
  <lema>lector</lema>
  <pos>ncms000</pos>
  <sentido>06298882</sentido>
  <sumo>hasSkill</sumo>
</argumento>
<argumento id=2>
  <texto>una definición previa de la palabra cultura</texto>
  <sintagma>sn</sintagma>
  <funcion>CD</funcion>
  <lema>definición</lema>
  <pos>ncfs000</pos>
  <sentido>04398116</sentido>
  <sumo>Stating</sumo>
</argumento>
</patron>
```