

Los Roles Semánticos en la Tecnología del Lenguaje Humano: Anotación y Aplicación*

Semantic Roles in Human Language Technology: Annotation and Application

P. Moreda

Grupo de Procesamiento del Lenguaje Natural - Universidad de Alicante
Carretera San Vicente del Raspeig s/n 03690 - Alicante, Spain
moreda@dlsi.ua.es

Resumen: La tesis presentada se centra en el análisis semántico de textos, y más concretamente, en el proceso de anotación de roles semánticos y su aplicación a otras tareas de Procesamiento de Lenguaje Natural. Dicho trabajo se puede resumir en tres objetivos principales: i) Investigar en los conjuntos de roles semánticos y recursos lingüísticos definidos sobre ellos que hay disponibles hasta la fecha; ii) Abordar el desarrollo de una herramienta para la anotación automática de roles semánticos, SemRol, que permita realizar un análisis de las necesidades de dichas herramientas; iii) Aportar información y conclusiones sobre la influencia de los roles semánticos en sistemas de búsqueda de respuestas.

Palabras clave: Roles semánticos, etiquetado automático de roles semánticos, aprendizaje automático supervisado, selección de características, búsqueda de respuestas

Abstract: The presented PhD focuses on semantic analysis, and more specifically in semantic role labeling processes and their application to other Natural Language Processing areas. So, the main goals of this work could be defined as follows: i) To research into semantic role sets and linguistic resources using them that are available to date; ii) To develop a semantic role labeling tool, named SemRol, to analyze the information needed for these kind of processes; iii) To contribute conclusions about the influence of semantic roles in question answering systems.

Keywords: Semantic roles, automatic semantic role labeling, supervised machine learning, feature selection, question answering

Introducción

Un rol semántico es la relación entre un constituyente sintáctico (generalmente, aunque no siempre, argumento del verbo) y un predicado (generalmente, aunque no siempre, un verbo). Ejemplos de roles semánticos son agente, paciente, beneficiario, etc. o también adjuntos como causa, manera, lugar, etc.

Por ejemplo, la oración (1), tiene cinco constituyentes cada uno de ellos con un rol semántico diferente. El constituyente sintáctico “*Mary*” tiene el rol agente, y los constituyentes, “*John*” y “*with a baseball*” tienen los roles paciente e instrumento, respectivamente. Además, los constituyentes “*in the park*” y “*yesterday*” tienen los roles lugar

y tiempo, respectivamente.

- (1) [AGENT *Mary*] hit [PATIENT *John*]
[INSTRUMENT with a baseball] [TEMP
yesterday] [LOC in the park]

Una característica importante de los roles semánticos es que aunque cambie el orden de los constituyentes o incluso la voz o el tiempo verbal de la oración, los roles semánticos de los constituyentes no varían. Por ejemplo, consideremos la oración anterior (1), si la cambiamos por la oración (2):

- (2) [TEMP *Yesterday*] [PATIENT *John*] was
hit [AGENT by *Mary*] [INSTRUMENT
with a baseball] [LOC in the park]

o incluso si la traducimos al castellano y aún habiendo diferencias sintácticas como en el ejemplo (3):

- (3) [TEMP *Ayer*] [AGENT *María*] gol-
peó [PATIENT a *Juan*] [LOC en el par-

* Esta tesis ha sido parcialmente financiada por el proyecto *TEXT-MESS*: Minería de Textos Inteligente, Interactiva y Multilingüe basada en Tecnología del Lenguaje Humano (TIN2006-15265-C06-01)

que] [*INSTRUMENT* con una pelota de béisbol]

se obtiene que en cualquiera de los casos “*Mary/María*” continúa jugando el rol agente, “*John/Juan*” el rol paciente, “*with a baseball/con una pelota de béisbol*” el rol instrumento, “*in the park/en el parque*” el rol de lugar y “*yesterday/ayer*” el rol temporal.

Por tanto, se podría concluir, que los roles semánticos permiten interpretar los textos determinando las relaciones semánticas entre las entidades y los eventos en los que participan. Es decir, los roles semánticos ayudan a determinar, capturar y representar “quién hizo qué a quién, dónde, cuando, cómo y por qué”, pudiendo con ello responder a preguntas como “quién”, “cuándo”, “dónde”, etc. Ver figura 1.

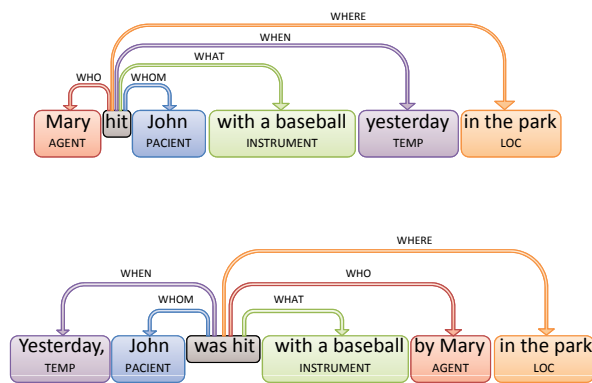


Figura 1: Utilidad de los roles semánticos en búsqueda de respuestas

Aportaciones

Atendiendo a los objetivos fijados, las principales aportaciones de este trabajo al conocimiento de la investigación en roles semánticos se pueden resumir en:

- Amplia recopilación de la gran diversidad de propuestas de conjuntos de roles semánticos realizadas hasta la fecha.
- Propuesta de un conjunto de roles semánticos propio, desarrollado atendiendo a criterios de aplicabilidad, generalidad, jerarquía y conexión con otras propuestas de anotación.
- Evaluación y estudio exhaustivos de los diferentes recursos lingüísticos basados en roles semánticos desarrollados hasta el momento.

- Análisis de los procesos de selección de características más importantes definidos hasta el momento, necesarios en enfoques basados en aprendizaje automático.
- Investigación en cuanto a sistemas de anotación automática de roles semánticos desarrollados, atendiendo al corpus utilizado, a la lengua para la que se han definido, al conjunto de roles semánticos utilizado en la anotación, a la información requerida para afrontar el proceso de anotación, y a la estrategia o enfoque seguido en dicho proceso de anotación.
- Desarrollo de una herramienta de anotación propia, denominada SemRol, que hace uso de enfoques basados en aprendizaje automático supervisado, y que permite seleccionar el clasificador y la información concreta a utilizar en función de las necesidades de la anotación.
- Experimentación y prueba de un proceso de ajuste de la información requerida por SemRol.
- Evaluación de diferentes estrategias de anotación seguidas en SemRol: i) anotación por sentidos frente a única, ii) anotación global frente a individual.
- Desarrollo de un sistema de búsqueda de respuestas modular, basado en Web, que permite extraer a partir de *snippets* posibles respuestas atendiendo a diferentes criterios: i) entidades nombradas, ii) reglas semánticas, y iii) patrones semánticos.
- Evaluación de la aportación de los roles semánticos a los sistemas de búsqueda de respuesta, desde dos perspectivas diferentes: i) comportamiento de reglas y patrones basados en roles semánticos, ii) comportamiento de entidades nombradas y roles semánticos.

Información general de la tesis

Tesis doctoral en Informática realizada en la Universidad de Alicante por Paloma Moreda Pozo bajo la dirección del Dr. Manuel Palomar Sanz. La defensa de la tesis tuvo lugar el 21 de Julio de 2008 ante el tribunal formado por los doctores Antonio Ferrández, Estela Saquete, Ruslan Mitkov, Lidia Moreno y L. Alfonso Ureña. La calificación obtenida fue *Sobresaliente Cum Laude* por unanimidad.

La tesis completa puede consultarse en la dirección <http://gplsi.dlsi.ua.es/mwgpplsi/upload/5/51/TesisMoreda.pdf>