Creado el 11/03/2014
Última versión: 11/03/2014
M. Antònia Martí i Mariona Taulé, Universitat de Barcelona
Comunidad moocTLH
Esta obra está bajo una licencia de Creative Commons Reconocimiento-CompartirIgual 3.0 Unported.

Enfoques y recursos básicos/NIVELES DE ANÁLISIS DEL LENGUAJE

Semántica y pragmática 3

Maria Antònia Martí nos va hablar de semántica y pragmática como niveles de análisis del lenguaje. Esta lección la ha realizado conjuntamente con Mariona Taulé y ambas son doctoras de la Universitat de Barcelona.

Son muchos los sistemas de PLN que requieren tratar el significado léxico, y por lo tanto deben resolver la ambigüedad a que da lugar la polisemia y la homonimia. Es el caso de los sistemas de traducción automática o los sistemas de pregunta-respuesta. A la tarea de resolución de la polisemia se la denomina 'desambiguación semántica automática' (*Word Sense Disambiguation*, WSD).

Como podemos ver en el ejemplo, una misma forma en español debe recibir diferentes traducciones al inglés puesto que el significado es distinto.

- La **banda** tocó durante toda la velada. - *The **band** played the whole evening*.
- Le atracó una **banda** de malhechores. - *The **gang** of evildoers stolen him*.
- El barco **atracó** en el puerto. - *The ship **docked** at the port*.
- Se **atracó** de comer. - *He **swelled** to eat*.

Continuamos desarrollando el análisis semántico hablando de la desambiguación semántica automática.

La desambiguación semántica automática

La polisemia es un fenómeno muy habitual en las lenguas. Si observamos las entradas de un diccionario comprobaremos que muchas de ellas -alrededor del 50 %- tienen más de una definición o sentido.

Esta pluralidad de sentidos no interfiere, en la práctica, en la comunicación humana ya que el contexto, tanto lingüístico como extralingüístico, desambigua su significado como se puede ver en los ejemplos:

- El programa se emite en **banda** ancha
- La **banda** tocó durante toda la velada
- Le **atracó** una **banda** de malhechores
- El barco **atracó** en el puerto

- **Se atracó** de comer

Métodos de WSD

El problema fundamental de los sistemas de desambiguación semántica automática es como proporcionar al sistema el conocimiento necesario para conseguir un resultado correcto.

Existen dos aproximaciones fundamentales:

1. La desambiguación semántica basada en corpus anotados
2. La desambiguación semántica basada en fuentes de conocimiento léxico

Veremos en primer lugar la desambiguación semántica basada en corpus anotados.

Basada en corpus anotados

Un corpus anotado semánticamente a nivel de palabra tiene asociado a cada palabra su sentido o significado según una determinada fuente léxica (un diccionario, WordNet, etc.). A partir de un corpus de estas características, los sistemas de desambiguación semántica basados en aprendizaje automático 'aprenden' los contextos más habituales de los sentidos de cada palabra y utilizan posteriormente este conocimiento para etiquetar nuevos corpus.

El conocimiento que adquieren estos sistemas consiste en las palabras que en el texto rodean la palabra que se quiere desambiguar. Estas palabras del contexto son los rasgos que utiliza el clasificador para predecir el significado de nuevas ocurrencias.

Basada en fuentes de conocimiento

Los sistemas de desambiguación automática basados en fuentes de conocimiento asignan los sentidos a las palabras a partir de reglas heurísticas que relacionan la palabra que se quiere desambiguar con sus descripciones, definiciones o ejemplos en la fuente de conocimiento.

Si la fuente de conocimiento es un diccionario, las reglas utilizarán las definiciones, si es un red semántica utilizará las relaciones semánticas o los ejemplos, según el material que esté disponible.