

Introducción al Procesamiento de Lenguaje Natural

Grupo de PLN – InCo

Roles Semánticos

Estructura argumental

verbo  argumentos (complementos) (son subcategorizados)
adjuntos

adjuntos

- complementos circunstanciales

argumentos

- sujeto
- objeto directo
- objeto indirecto
- complementos de régimen (preposicionales)

Estructura argumental

cantar: [sujeto, objeto directo, objeto indirecto]

dormir: [sujeto]

llover: [-]

Estructura argumental

*Juan (le) **canta** una canción a su amada.*

*Juan **canta** una canción.*

*Juan **canta**.*

*También él empieza a comer, a dormir, como un niño que **duerme** un sueño sin miedo.*

Llueve.

***Llueven** pingüinos de punta.*

Estructura argumental

Podemos definir varias entradas diferentes para *cantar*:

cantar₁: [sujeto, objeto directo, objeto indirecto]

cantar₂: [sujeto, objeto directo]

cantar₃: [sujeto]

o definir argumentos opcionales:

cantar: [sujeto, (objeto directo), (objeto indirecto)]

Roles semánticos

Establecen una relación entre sintaxis y semántica

El niño come *la manzana*.

sujeto ↔ agente

objeto directo ↔ tema (paciente)

La manzana es comida por el niño.

sujeto ↔ tema (paciente)

complemento agente ↔ agente

Roles semánticos

El ruido desagrada *al maestro*.

sujeto ↔ tema

objeto directo ↔ experimentante

El maestro detesta *el ruido*.

sujeto ↔ experimentante

objeto directo ↔ tema

Roles semánticos

Juan corta el pan con el cuchillo.

sujeto ↔ agente

objeto directo ↔ tema

complemento preposicional ↔ instrumento

Roles semánticos

- Existe una diferencia entre la función sintáctica y el rol semántico del constituyente.
 - El análisis sintáctico nos da:
 - los constituyentes
 - a veces también la función sintáctica
 - pero casi nunca el rol semántico
-

Grilla Semántica

*[Juan]*_{AGENTE} *rompió* *[la ventana]*_{TEMA}.

*[Juan]*_{AGENTE} *rompió* *[la ventana]*_{TEMA} *[con una
piedra]*_{INSTRUMENTO}.

*[Una piedra]*_{INSTRUMENTO} *rompió* *[la ventana]*_{TEMA}.

*[La ventana]*_{TEMA} *se rompió*.

*[La ventana]*_{TEMA} *fue rota* *[por Juan]*_{AGENTE}.

La grilla semántica del predicado *romper* contiene los roles semánticos: AGENTE, TEMA, INSTRUMENTO.

Roles semánticos

¿Qué roles existen?

AGENTE – ***El mozo** derramó la sopa.*

EXPERIMENTANTE – ***Juan** tiene dolor de cabeza.*

FUERZA – ***El viento** abrió la ventana.*

TEMA – *María rompió **el hielo**.*

RESULTADO – *El obrero construyó **un muro**.*

CONTENIDO – *Carolina dijo “¿**Conocés a Juan?**”*

INSTRUMENTO – *Juan abrió la puerta **con la llave**.*

BENEFICIARIO – *Pedro compró el boleto **para su jefe**.*

ORIGEN – *Volé **desde Chile**.*

DESTINO – *Manejé **hasta Las Piedras**.*

... y más

Si tuviéramos los roles semánticos...

- Equivalencia de:
 - *Los obreros construyeron la casa*
 - *La casa fue construida por los obreros*
 - Implicación de 1 a 2 en:
 1. *La oferta de una fórmula de pago de la deuda por Argentina está siendo considerada por los acreedores.*
 2. *Argentina ofreció una fórmula de pago de la deuda.*
-

Si tuviéramos los roles semánticos...

Respuestas a preguntas:

Juan golpeó a Pedro con un palo ayer en la calle.

Quién?

A quién?

Cuándo?

Dónde?

Si tuviéramos los roles semánticos...

- Resumen automático
 - Los predicados y los núcleos de los roles argumentales resumen el contenido.
 - Extracción de información
 - Si se reconocen los roles estamos más cerca de los atributos de una BD
-

Semantic Role Labeling

- En español: Etiquetado de Roles Semánticos (ERS)
- Tarea en PLN que consiste en asignar roles semánticos a constituyentes.

*[Juan]*_{AGENTE} *rompió* *[la ventana]*_{TEMA} *[con una piedra]*_{INSTRUMENTO}

Requiere:

- Haber identificado los constituyentes sintácticos
 - Haber identificado los predicados
 - Conocer la grilla semántica.
-

Semantic Role Labeling

- Aprendizaje supervisado
 - Necesidad de corpus para entrenamiento y testeo
 - Además:
 - Qué es un predicado
 - Cuáles son los roles
 - Recursos: PropBank y FrameNet (para el inglés)
-

Sin análisis sintáctico?

Se puede reformular el problema como un etiquetado de secuencias:

Juan	agente
rompió	pred
la	tema
ventana	tema
con	instrumento
una	instrumento
piedra	instrumento
...	

Sin análisis sintáctico?

Collobert y Weston (2008, 2011) intentan resolver este problema como un caso de tagging

- Word embeddings (Wikipedia)
 - Multitask learning (POS tagging, NER, SRL y otras tareas a la vez)
 - Redes neuronales convolutivas
 - Uno de los primeros ejemplos de Deep Learning
-

Recursos léxicos para SRL

Para el inglés (parcialmente para otras lenguas) se desarrollaron dos recursos que fueron muy utilizados:

- PropBank
 - FrameNet
-

PropBank

<http://verbs.colorado.edu/~mpalmer/projects/ace.html>

Cuenta con:

- Grilla semántica para los verbos

Los argumentos se anotan con números: arg0, arg1, arg2

...

Arg0 – proto-agente

Arg1 – proto-paciente

...a partir del arg2, los roles son específicos para cada verbo.

- Corpus de oraciones etiquetadas (Penn Treebank WSJ)

Más de 100.000 proposiciones (cláusulas) anotadas.

PropBank

Ejemplos de PropBank:

increase.01: go up incrementally

Arg0: causer of increase

Arg1: thing increasing

Arg2: amount increased

Arg3: start point

Arg4: end point

FrameNet

¿Qué pasa con estos ejemplos?

[*El precio de la carne*]_{ARG1} *aumentó* [*un 5%*]_{ARG2}.

[*El precio de la carne*]_{ARG1} *subió* [*5%*]_{ARG2}.

Hubo un aumento [*de 5%*]_{ARG2} [*en el precio de la carne*]_{ARG1}.

Se usan distintos verbos, o incluso un nombre, pero hay equivalencia.

FrameNet da respuesta a esto, mientras que PropBank no lo considera.

FrameNet

<http://framenet.icsi.berkeley.edu>

- FrameNet es un proyecto de anotación manual de roles semánticos.
 - 960 frames
 - 11600 unidades léxicas
 - 150,000 oraciones anotadas
-

FrameNet

- Un frame describe una situación.
 - Se instancian los participantes (elementos del frame) en roles semánticos, de modo particular para cada frame.
 - Se codifican también relaciones entre frames (herencia, causación) y con elementos de frames.
-

FrameNet

Ingestion

Definition:

- An *Ingestor* consumes food or drink (*Ingestibles*), which entails putting the *Ingestibles* in the mouth for delivery to the digestive system. This may include the use of an *Instrument*. Sentences that describe the provision of food to others are NOT included in this frame.
-

FrameNet

Ingestion

FEs:

Core:

Ingestibles [Ingible]

The Ingestibles are the entities that are being consumed by the Ingestor.

Ingestor [Ing]

The Ingestor is the person eating or drinking.

FrameNet

Ingestion

FEs:

Non-Core:

Degree [Degr]

The extent to which the Ingestibles are consumed by the Ingestor.

*The wolves DEVoured the carcass **completely**.*

Duration [Dur]

The length of time spent on the ingestion activity.

*They 've been **EATING for hours!***

....

FrameNet

Ingestion

Inherits From: Ingest_substance, Manipulation

Is Inherited By:

Subframe of:

Has Subframes:

Precedes:

Is Preceded by:

Uses: Cause_motion

Is Used By: Food, Tasting

FrameNet

Ingestion

Lexical Units

breakfast.v, consume.v, devour.v, dine.v, down.v, drink.v, eat.v, feast.v, feed.v, gobble.v, gulp.n, gulp.v, guzzle.v, have.v, imbibe.v, ingest.v, lap.v, lunch.v, munch.v, nibble.v, nosh.v, nurse.v, put away.v, put back.v, quaff.v, sip.n, sip.v, slurp.n, slurp.v, snack.v, sup.v, swig.n, swig.v, swill.v, tuck.v

¿Y para el español?

PropBank

- No existe un equivalente en español.
- El corpus AnCora tiene anotados algunos roles semánticos y la subcategorización (AnCoraVerb).
- <http://clic.ub.edu/corpus/es>

FrameNet

- Existe un Spanish FrameNet
- 305 frames, 575 unidades léxicas, 10.000 oraciones.
- <http://gemini.uab.es/SFN>

ADESSE

- <http://adesse.uvigo.es/>
-

Referencias

- *Speech and Language Processing (3rd Ed, Draft)* Capítulo 24, Daniel Jurafsky y James Martin, 2023.
https://web.stanford.edu/~jurafsky/slp3/ed3book_jan72023.pdf
-