

Índice

- 1 Conceptos de PLN: Análisis Morfológico y Etiquetación
- 2 Conceptos de PLN: Análisis Sintáctico Superficial
- 3 Conceptos de PLN: Semántica Léxica**
- 4 Extracción de Información
- 5 Búsqueda de Respuestas

Relaciones Semánticas: Homonimia

- Palabras con la misma forma pero significados diferentes (y no relacionados).
- Pueden degradar el rendimiento del sistema al confundir las palabras (buscadores, procesamiento del habla, etc.)
- Podemos distinguir:
 - **Homógrafas:** se escriben (y suenan) igual

$$vino: \begin{cases} \text{de uva} \\ \text{de venir} \end{cases}$$

- **Homófonas:** se suenan igual pero se escriben diferente

$$\begin{cases} vaca: \text{animal} \\ baca: \text{de coche} \end{cases}$$

Relaciones Semánticas: Polisemia

- Palabra con varios **sentidos/acepciones** (relacionados por tener origen común)

araña { bicho
lámpara

- Será necesario identificar el sentido/acepción concreto de la palabra en el contexto en el que aparece
 - Técnicas de Desambiguación del Sentido de las Palabras (*WSD*, *Word Sense Disambiguation*)

Juan tiene una araña { **bicho**
lámpara *en un terrario como mascota.*

Relaciones Semánticas: Sinonimia

- Palabras diferentes con el mismo significado
 - Para esa acepción y ese contexto (una palabra polisémica puede tener diferentes sinónimos para diferentes acepciones)

{ *listo = inteligente*
listo = preparado
listo = ingenioso

Relaciones Semánticas: Hiponimia/Hiperonimia

- *Taxonomía* en base a una relación ES_UN:
 - **Hipónimo:** el término MÁS ESPECÍFICO. P.ej.:
clavel es un hipónimo de *flor*
rosa es un hipónimo de *flor*
 - **Hiperónimo:** el término MÁS GENERAL. P.ej.:
flor es un hiperónimo de *clavel*
flor es un hiperónimo de *rosa*

Relaciones Semánticas: Oposición

- Palabras de significado "*contrario*"
- Estrictamente, existen 3 tipos:
 - **Complementariedad:** negar uno implica afirmar el otro
P.ej. *seco/mojado*
 - **Antonimia:** existe una gradación entre ambos términos
P.ej. *caliente/frío*
 - **Reciprocidad:** un término implica el otro
P.ej. *dar/recibir*

Relaciones Semánticas: Meronimia/Holonimia

- *Taxonomía* en base a una relación PARTE_DE:

- **Merónimo:** designa una PARTE. P.ej.:

dedo es un merónimo de *mano*

barco es un merónimo de *flota*

- **Holónimo:** designa el TODO. P.ej.:

mano es un holónimo de *dedo*

flota es un holónimo de *barco*

WordNet

- Base de datos léxica para *inglés*:
 - La más usada: **gratuita**, múltiples APIs
 - <http://wordnet.princeton.edu/> (también versión online)
 - v3.0 (Unix/Linux) / v2.1 (Windows)
- En base a **synsets**: conjunto de palabras (sinónimas) que pueden emplearse para expresar el *concepto* que el *synset* representa
 - Sus palabras son intercambiables (sinónimas) en un contexto dado
 - i.e., **1 synset = 1 sentido/acepción**

sentidos	palabras (formas)			
	F ₁	F ₂	...	F _n
S ₁	E ₁₁	E ₁₂		
S ₂		E ₂₂		
...				
S _m				

sinonimia (synset) (indicated by a red box around E₁₁ and E₁₂)
polisemia (indicated by a blue box around E₁₂ and E₂₂)

WordNet (cont.)

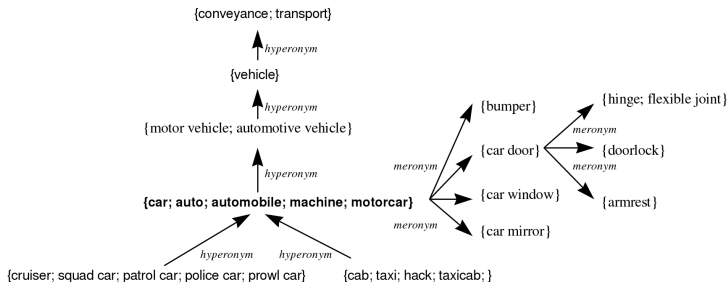
- Entrada WordNet:
 - Conjunto de sinónimos (*synset*)
 - Definición (*glosa*)
 - Ejemplo: car

The noun car has 5 senses (first 3 from tagged texts)

1. (598) car, auto, automobile, machine, motorcar -- (a motor vehicle with four wheels; usually propelled by an internal combustion engine; "he needs a car to get to work")
2. (24) car, railcar, railway car, railroad car -- (a wheeled vehicle adapted to the rails of railroad; "three cars had jumped the rails")
3. (1) cable car, car -- (a conveyance for passengers or freight on a cable railway; "they took a cable car to the top of the mountain")
4. car, gondola -- (the compartment that is suspended from an airship and that carries personnel and the cargo and the power plant)
5. car, elevator car -- (where passengers ride up and down; "the car was on the top floor")

WordNet (cont.)

- Relaciones semánticas **a nivel de synset**, no de palabra:
 - Sinonimia: inherente al propio synset
 - Antonimia
 - Hiperonimia / hiponimia: "ES_UN"
 - Troponimia: "hiponimia verbal", verbo caso particular de otro
 - P.ej., caminar \Leftrightarrow pasear
 - *Entailment*: relación de implicación–causa
 - P.ej., roncar \Leftrightarrow dormir
 - Meronimia / holonimia: "PARTE_DE"



WordNet (cont.)

- Modo de acceso:
 - Intérprete de comandos
 - Interfaz (navegador)
 - APIs: C, C++, Java, Perl, Prolog, ...
- Ampliamente usado en NLP:
 - Desambiguación de sentidos
 - Resolución de anáfora
 - Recuperación y extracción de información
 - ...
- Estadísticas (v3.0):

Categoría	#Palabras	#Synsets	#Palabra-sentido
Sustantivo	117,798	82,115	146,312
Verbo	11,529	13,767	25,047
Adjetivo	21,479	18,156	30,002
Adverbio	4,481	3,621	5,580
Totales	155,287	117,659	206,941

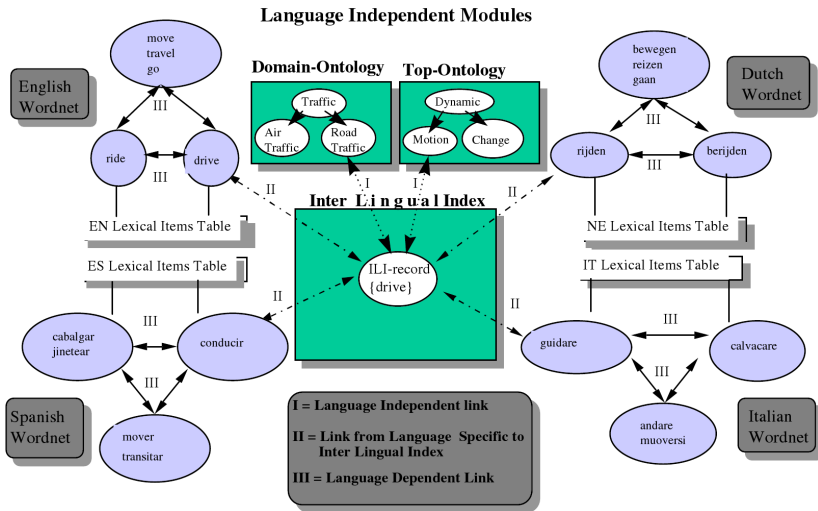
EuroWordNet

- WordNet para lenguas europeas
- Bases de datos separadas para cada idioma
 - Proyecto principal (1999): español, francés, italiano, holandés, alemán, checo y estonio
 - En desarrollo: gallego, catalán, vasco, sueco, noruego, danés, griego, portugués, rumano, lituano, ruso, búlgaro, esloveno, ...
- Generalmente de pago
- Demos online de la UPC:
 - <http://garraf.epsevg.upc.es/cgi-bin/wei4/public/wei.consult.perl>
 - <http://ixa2.si.ehu.es/cgi-bin/mcr/public/wei.consult.perl>

EuroWordNet (cont.)

- Estructuras comunes independientes del idioma
 - *Inter-Lingual-Index (ILI)*
 - Índice central independiente del idioma
 - 1 registro/sentido (sentidos WordNet original)
 - Cada synset de un idioma se corresponde con un registro del ILI
 - Permite establecer correspondencias entre synsets de diferentes idiomas:
P.ej., synset ES → registro ILI correspondiente → synset FR
 - *Top [Concept] Ontology*
 - Jerarquía de conceptos [muy] básicos (63 distinciones semánticas) establecida sobre los 1024 conceptos más importantes del ILI
 - Transferible a cualquier idioma vía el ILI
 - P.ej., *dynamic vs. static, location*, etc.
 - *Domain [Labels] Ontology*
 - Jerarquía de etiquetas de dominios/temas
 - Transferible a cualquier idioma vía el ILI
 - P.ej., *traffic, sports*, ...

EuroWordNet (cont.)



Referencias

- [EUROWORDNET] Official site:
<http://www.illc.uva.nl/EuroWordNet/>
- [WORDNET] Official site: <http://wordnet.princeton.edu/>
- [Fellbaum, 1998] Fellbaum, C. (1998). *WordNet: An electronic lexical database*. The MIT Press.
- [Vossen, 1998] Vossen, P. (1998). *EuroWordNet. A Multilingual Database with Lexical Semantic Networks*. Kluwer Academic Publishers.