

---

# Introducción al Procesamiento de Lenguaje Natural

---

Grupo de PLN – InCo

---

---

# Semántica léxica

---

---

# Análisis semántico

---

*Maharani sirve comida vegetariana.*

$\exists x \text{ sirve}(\text{maharani}, x) \wedge \text{comida}(x) \wedge \text{vegetariana}(x)$

*Maharani sirve platos vegetarianos.*

$\exists x \text{ sirve}(\text{maharani}, x) \wedge \text{plato}(x) \wedge \text{vegetariano}(x)$

*Maharani sirve menú vegetariano.*

$\exists x \text{ sirve}(\text{maharani}, x) \wedge \text{menú}(x) \wedge \text{vegetariano}(x)$

Las tres oraciones son equivalentes. ¿Cómo podemos hacer que estas tres representaciones sean iguales?

---

# Semántica léxica

---

- Enfoque basado en significados de las palabras y relaciones entre los significados (relaciones léxicas).
  - Un recurso fundamental para este enfoque es WordNet.
  - Enfoque actual: representaciones distribuidas de las palabras (word embeddings). Se verá más adelante en el curso.
-

# Significados de las palabras

---

Los diccionarios o tesauros (palabras y sus relaciones) dan definiciones textuales de los significados, llamadas glosas. Ejemplo: **banco**

- 1. m.** Asiento, con respaldo o sin él, en que pueden sentarse dos o más personas.
- 2. m.** Madero grueso escuadrado que se coloca horizontalmente sobre cuatro pies y sirve de mesa para labores de carpinteros y otros artesanos.
- 3. m.** En los mares, ríos y lagos navegables, bajo que se prolonga en una gran extensión.
- 4. m.** Conjunto de peces que van juntos en gran número.
- 5. m.** Empresa dedicada a realizar operaciones financieras con el dinero procedente de sus accionistas y de los depósitos de sus clientes.

...

---

# Diccionarios

---

- Estas definiciones no son representaciones formales del significado.
  - A veces son definiciones circulares.
  - No están explícitas en muchos casos las relaciones entre las diferentes definiciones.
  - Por más que estén en formato digital no son inmediatamente explotables en PLN.
-

# Polisemia

---

- Polisemia: múltiples significados para una sola palabra
    - banco ... de sangre / financiero
    - hoja ... de un árbol / de papel
  - Son significados relacionados (diferente a la homonimia).
  - La mayoría de las palabras muy comunes son polisémicas.
  - Para un lexema polisémico es importante establecer:
    - ¿Cuáles son los sentidos distintos que tiene?
    - ¿Cómo están vinculados estos sentidos?
    - ¿Cómo pueden ser distinguidos?
-

# Polisemia

---

## Distinción de sentidos

- Identificar tantos sentidos como sea necesario para dar cuenta de distinciones finas de significado.
- Un test para detectar sentidos distintos: combinación con una conjunción.

*Juan sintió los ruidos del jardín.*

*Juan sintió mucho pesar por su partida.*

*? Juan sintió los ruidos del jardín y mucho pesar por su partida.*

- Pero la proliferación de sentidos puede generar problemas de procesamiento.
-

# Relaciones léxicas: sinonimia

---

**Sinónimos:** Palabras que tienen el mismo significado.

lindo	bello
auto	coche
barro	lodo
fino	estrecho
solicitar	pedir

- Dos lexemas son sinónimos (perfectos) si pueden sustituirse en todos los contextos.
  - Casi no hay ejemplos de sinónimos perfectos, se habla de “sinónimos cercanos”.
-

# Relaciones léxicas: sinonimia

---

- Los sinónimos pueden tener matices de significado que hagan aceptable la sustitución solo en algún contexto.
  - Importa también la situación, tipo de lenguaje (informal o formal), etc.
    - pedir / solicitar
    - amar / querer
    - sacar / quitar (sacar una foto vs. quitar una foto)
  - La sinonimia es una relación entre significados, no entre palabras.
  - Si dos significados de dos palabras diferentes son iguales o casi iguales, esos significados son sinónimos.
-

# Relaciones léxicas: antonimia

---

**Antónimos** : Palabras opuestas en alguna componente de su significado.

oscuro      claro (extremos opuestos de una escala)

adentro    afuera (oposición binaria)

entrar      salir (movimiento en direcciones opuestas)

- En realidad, los antónimos son muy similares, comparten todos los rasgos de su significado menos uno.
  - Puede ser difícil distinguir automáticamente sinónimos de antónimos.
-

# Relaciones léxicas: hiponimia

---

- Hiponimia (Is\_a): una palabra es hipónimo de otra si denota un concepto que es subclase del denotado por la otra.
- Hiperonimia: relación inversa a la hiponimia.

*auto* es un hipónimo de *vehículo*

*vehículo* es un hiperónimo de *auto*

Hiponimia e hiperonimia son relaciones taxonómicas.

---

# Relaciones léxicas: holonimia

---

- Meronimia: relación parte-todo.
- Holonimia: relación inversa a la meronimia.

*pata* es merónimo de *mesa*

*mesa* es holónimo de *pata*

## Diferentes clases:

- Miembro de un grupo: *tripulación* → *piloto*
  - Parte o componente: *mesa* → *pata*
  - Sustancia: *destornillador* → *vodka*
-

# Relaciones léxicas: metonimia

---

**Metonimia:** Nos referimos a un concepto u objeto a través de uno de sus atributos.

*Le gusta Cortázar.* (autor por sus obras)

*La Casa Blanca no emitió comentarios.* (lugar por institución)

*Empezó el libro.* (omisión de la acción, empezó a leer, a escribir, a fotocopiar)

---

# Relaciones léxicas: propiedades

---

La sinonimia y la antonimia son simétricas.

La hiperonimia es transitiva y asimétrica:

A es hipónimo de B, B es hipónimo de C  
entonces A es hipónimo de C

La meronimia es asimétrica, **no** es transitiva

*puerta parte-de casa*

*pestitillo parte-de puerta*

*pestitillo parte-de casa ??*

---

# WordNet

---

<http://wordnet.princeton.edu/>

(demo online y descarga de los datos)

- Base de datos léxica.
  - Múltiples vinculaciones entre los datos: relaciones léxicas.
  - Incluye aspectos de un diccionario (glosa, breve explicación de un significado, ejemplos).
  - Construida en forma manual.
-

# WordNet

---

## Estadísticas de la página de Princeton

<http://wordnet.princeton.edu/wordnet/man/wnstats.7WN.html>  
(no disponible actualmente)

POS	Unique Strings	Synsets	Avg. Polysemy
Noun	117798	82115	1.24
Verb	11529	13767	2.17
Adjective	21479	18156	1.40
Adverb	4481	3621	1.25
Totals	155287	117659	

---

# WordNet

---

- Las unidades semánticas de Wordnet se llaman *synsets* (*synonym sets*), son cuasi-sinónimos, sustituibles en al menos un contexto.
  - Los significados distintos de cada palabra están diferenciados por un índice numérico.
  - Un significado (o concepto) es una lista de lemas indizados.
  - La glosa es una breve explicación de un significado, todas las acepciones de distintas palabras de un mismo synset comparten la glosa.
-

# WordNet: relaciones

---

- Las relaciones se dan entre synsets.
  - Nombres
    - hiperonimia /hiponimia
    - meronimia / holonimia
  - Adjetivos y adverbios
    - antonimia
  - Verbos
    - implicación ( )
    - troponimia ( )
-

# WordNet – Nombres

---

- Sistema léxico de herencia, dado por la relación de hiponimia.
  - Todos los nombres heredan inicialmente del concepto “entity”.
  - Las unidades semánticas asociadas a nombres están vinculadas además por relaciones de meronimia (/holonimia).
  - La meronimia puede ser de elemento (marinero/tripulación), de parte(árbol/rama), de sustancia (agua/oxígeno).
-

# WordNet: ejemplo

---

- Versión de WordNet para varios idiomas: MultiWordnet
  - Basada en WordNet 1.6 (la última versión es 3.0).
  - <https://multiwordnet.fbk.eu/english/home.php>
-

# WordNet: ejemplo

Spanish Word Word banco Search

[Word statistics](#) | [Database report](#) | [Bug report](#) | [Credits](#)

Noun

Overview

The word "**banco**" has 8 senses:

Spanish WordNet created by [Universitat de Catalunya](#) (Spain)

Noun	
▶ 1. <b>banco</b>	(Architecture, Banking) [a building in which commercial banking is transacted; "the bank is on the corner of Nassau and Witherspoon"]
▶ 2. <b>banco</b>	(Furniture) [a long seat for more than one person]
▶ 3. <b>banco</b>	(Animals, Biology) [a large group of fish; "a school of small glittering fish swam by"]
▶ 4. <b>banco</b>	(Banking) [a financial institution that accepts deposits and channels the money into lending activities; "he cashed a check at the bank"; "that bank holds the mortgage on my home"]
▶ 5. bajío, <b>banco</b>	(Factotum) [a long ridge or pile; "a huge bank of earth"]
▶ 6. <b>banco</b> , <b>banco_de_arena</b> , médano, placer, seca, secano, sirte	(Factotum) [a submerged bank of sand near a shore or in a river; can be exposed at low tide]
▶ 7. <b>banco</b> , médano	(Factotum) [a sandbank in a stretch of water that is visible at low tide]
▶ 8. <b>banco</b>	(Money) [a supply or stock held in reserve especially for future use (especially in emergencies)]

# WordNet: ejemplo

Noun

Hypernyms

Hypernyms of "banco":

Noun

- ▶ 1. **banco** -- (Architecture, Banking) [*a building in which commercial banking is transacted; "the bank is on the corner of Nassau and Withersp*]
  - > ▶ depositaria, depósito -- (Factotum) [*a place where things can be deposited for safekeeping*]
  - > ▶ instalación, servicio -- (Economy) [*something created to provide a particular service; "the assembly plant is an enormous facility"*]
  - > ▶ artefacto -- (Factotum) [*a man-made object*]
    - > ▶ cosa, objeto, objeto\_fisico, objeto\_inanimado -- (Factotum) [*a physical (tangible and visible) entity; "it was full of rackets, balls an*]
    - > ▶ entidad -- (Factotum) [*anything having existence (living or nonliving)*]
- ▶ 2. **banco** -- (Furniture) [*a long seat for more than one person*]
  - > ▶ asiento -- (Furniture) [*furniture that is designed for sitting on; "there were not enough seats for all the guests"*]
  - > ▶ mobiliario, mueble -- (Furniture) [*artifacts that make a room ready for occupancy; "they had too much furniture for the small apartme.*]
  - > ▶ moblaje, mueblaje -- (Architecture, Furniture) [*the furniture and appliances and other movable accessories (including curtains and r*]
  - > ▶ instrumental, utillaje -- (Architecture, Factotum) [*an artifact (or system of artifacts) that is instrumental in accomplishing some enc*]
  - > ▶ artefacto -- (Factotum) [*a man-made object*]
    - > ▶ cosa, objeto, objeto\_fisico, objeto\_inanimado -- (Factotum) [*a physical (tangible and visible) entity; "it was full of rackets, ball*]
    - > ▶ entidad -- (Factotum) [*anything having existence (living or nonliving)*]

▶ 3. **banca** -- (Animals, Biology) [*a large group of fish; "a school of small glittering fish swam by"*]

# Verbos

---

Dos relaciones básicas:

troponimia

*murmurar* es un modo de *hablar*

implicación léxica

*roncar* **implica** *dormir*

---

# Verbos en WordNet

---

Organizados en 15 grupos, según dominios semánticos:

cambio

comunicación

percepción

...

---

# Recursos relacionados

---

- MCR: Multilingual Central Repository.
  - SpanishWordNet 3.0.
  - MultiWordNet:  
<https://multiwordnet.fbk.eu/english/home.php>
  - BabelNet (basado en wordnet y wikipedia, entre otros recursos):  
<https://babelnet.org/>
-

# Recursos relacionados

---

- Las versiones de WordNet para idiomas distintos del inglés suelen tener menor cobertura.
  - Fueron creados con métodos semiautomáticos, curados en forma manual.
-

# Desambiguación de significados (Word Sense Disambiguation – WSD)

---

- Dados una palabra en un contexto y un conjunto de significados posibles para esa palabra, determinar el significado correspondiente a ese contexto.
  - Una extensión del problema WSD es la tarea de desambiguar todas las palabras de un texto dado (*all-words WSD*).
  - Para esta tarea se suelen utilizar corpus anotados con los sentidos de cada palabra (palabras de contenido léxico), para poder aplicar aprendizaje supervisado.
-

# Desambiguación de significados (word sense disambiguation - WSD)

---

- Ejemplo:
    - SemCor (parte del Brown Corpus)
    - 226.036 palabras
    - anotadas manualmente con significados de WordNet
-

# Desambiguación de significados (word sense disambiguation – WSD)

---

Algunas heurísticas y métodos no supervisados:

- Elegir el significado más frecuente (línea base sorprendentemente buena). En WordNet, el primer significado es el más frecuente.
  - Heurística "un significado por discurso": una palabra que aparece más de una vez en un texto siempre tiene el mismo significado.
  - Algoritmo de Lesk: enfoque más antiguo basado en conocimiento. Se elige el significado cuya glosa comparte más palabras con el contexto de la palabra a desambiguar.
-

# Desambiguación de significados (word sense disambiguation – WSD)

---

Ejemplo de aplicación de Lesk (Jurafsky & Martin, 2023)

Contexto: *The **bank** can guarantee deposits will eventually cover future tuition costs because it invests in adjustable-rate mortgage securities.*

Significados de WordNet:

## **bank1**

Gloss: a financial institution that accepts deposits and channels the money into lending activities

Examples: “he cashed a check at the bank”, “that bank holds the mortgage on my home”

## **bank2**

Gloss: sloping land (especially the slope beside a body of water)

Examples: “they pulled the canoe up on the bank”...

---

# Desambiguación de significados (word sense disambiguation – WSD)

---

Ejemplo de aplicación de Lesk (Jurafsky & Martin, 2023)

Contexto: *The **bank** can guarantee deposits will eventually cover future tuition costs because it invests in adjustable-rate mortgage securities.*

Significados de WordNet:

## **bank1**

Gloss: a financial institution that accepts deposits and channels the money into lending activities

Examples: “he cashed a check at the bank”, “that bank holds the mortgage on my home”

## **bank2**

Gloss: sloping land (especially the slope beside a body of water)

Examples: “they pulled the canoe up on the bank”...

---

# Desambiguación de significados (word sense disambiguation – WSD)

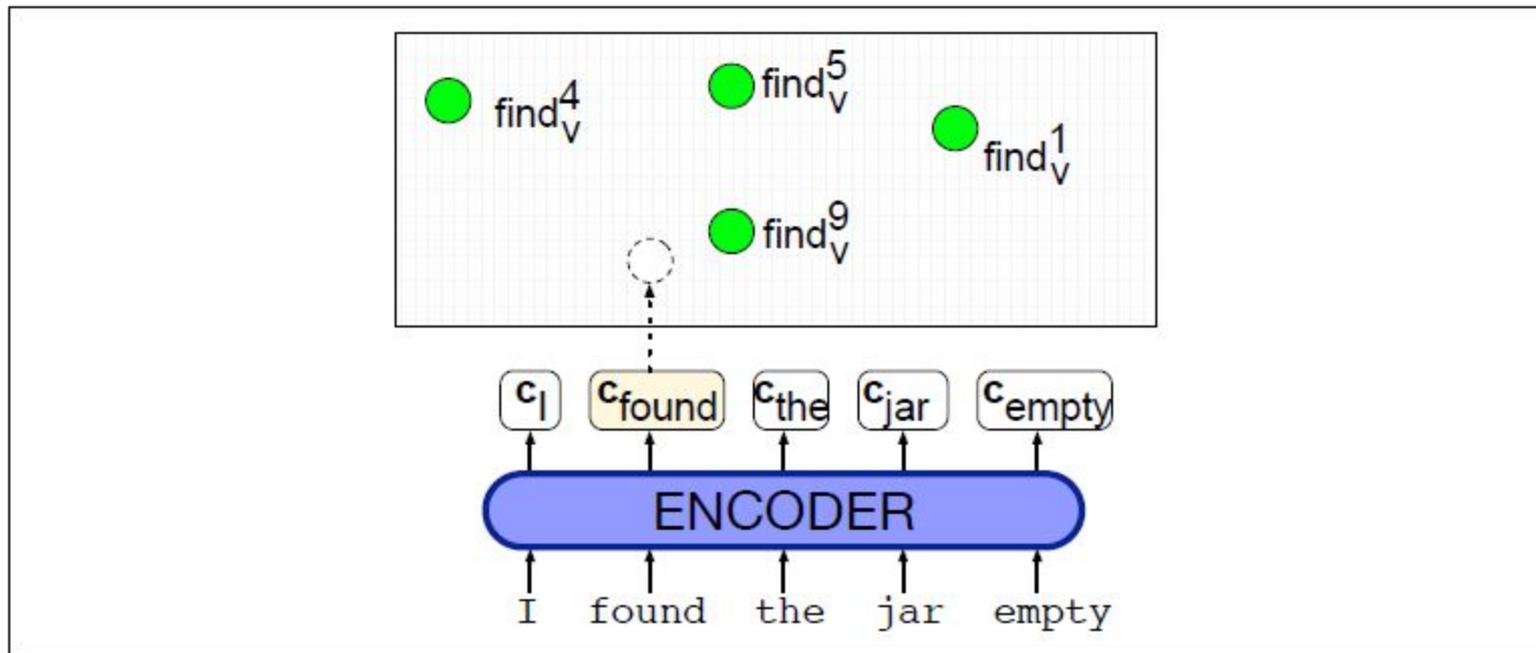
---

Enfoque actual para WSD:

- Algoritmo KNN (k-nearest-neighbor) usando word embeddings contextuales (Melamud et al., 2016; Peters et al., 2018).
    - “Entrenamiento”: Se obtiene un embedding contextual para cada palabra de SemCor (con un modelo de lenguaje neuronal) y se calcula el vector promedio de todas las palabras que tienen asociado un mismo significado.
    - Desambiguación: Dada una palabra para desambiguar en un contexto, se obtiene su embedding contextual y se elige el significado más cercano entre los embeddings de significados de SemCor.
-

# Desambiguación de significados (word sense disambiguation – WSD)

## Enfoque actual para WSD



**Figure 23.9** The nearest-neighbor algorithm for WSD. In green are the contextual embeddings precomputed for each sense of each word; here we just show a few of the senses for *find*. A contextual embedding is computed for the target word *found*, and then the nearest neighbor sense (in this case  $find_v^9$ ) is chosen. Figure inspired by Loureiro and Jorge (2019).

(Jurafsky & Martin, 2023)

# Referencias

---

*Speech and Language Processing (3rd Ed, Draft)* Capítulo 23, Daniel Jurafsky y James Martin, 2023.  
[https://web.stanford.edu/~jurafsky/slp3/ed3book\\_jan72023.pdf](https://web.stanford.edu/~jurafsky/slp3/ed3book_jan72023.pdf)

---