

# Ciencia de Datos y Lenguaje Natural

## Clase 1-2

Grupo PLN - INCO

Universidad de la República

12 de agosto de 2022

## Algunos elementos en el estudio del lenguaje

- ▶ Unidades en el texto
- ▶ Niveles tradicionales de análisis
- ▶ El léxico
- ▶ La sintaxis
- ▶ La semántica
- ▶ La pragmática
- ▶ La ambigüedad

## Oración y enunciado

- ▶ Hay varias caracterizaciones de las nociones de oración y enunciado.
- ▶ El concepto de oración es gramatical, suele definirse como una unidad compuesta por un sujeto y un predicado.
- ▶ El concepto de enunciado involucra la situación concreta, quién es el hablante, a quien/es se dirige y el contexto en general.

## Oración y enunciado, ejemplos

- ▶ El gato se despereza.
- ▶ Ideas verdes incoloras duermen furiosamente.
- ▶ Me conviene que no vengas.
- ▶ Silencio !!

## Oración y enunciado, ejemplos

- ▶ (El gato)<sub>Suj</sub> (se despereza)<sub>Pred</sub>.
- ▶ (Ideas verdes incoloras)<sub>Suj</sub> (duermen furiosamente)<sub>Pred</sub>.
- ▶ (Me conviene)<sub>Pred</sub> (que no vengas)<sub>Suj</sub>.
- ▶ Silencio !!

Todas son enunciados !! Las tres primeras son además oraciones.

## Sintaxis y semántica

- ▶ La sintaxis refiere a cómo se combinan las unidades
  - ▶ Morfología, estructura al interior de la palabra por combinación de morfemas
  - ▶ Sintaxis, estructura al interior de la oración, combinatoria de las palabras
- ▶ La semántica refiere al significado
  - ▶ Semántica "literal", incluye significados de las palabras (o de unidades multi-palabra) y de su composición. Es el dominio de las relaciones lógicas, (ej. A implica B).
  - ▶ Pragmática - Significado real", considera la situación de enunciación, las implicaturas conversacionales, las presuposiciones.

## Clases de palabras

- ▶ nombre/sustantivo (casa, casas, felicidad)
- ▶ verbo (cantar, cantando, cantaremos, sabe, sé)
- ▶ adjetivo (alto, alta, lindos, inteligente, solo)
- ▶ adverbio (medio, simplemente, solo, ayer)
- ▶ determinante (incluye artículo) (el, una, unos, esos, nuestro, algún)
- ▶ preposición (a, de, por, contra)
- ▶ pronombre (yo, él, mí, se, aquello, donde, que)
- ▶ conjunción (y, o, si, que)

## Clases de palabras

- ▶ Palabras de contenido léxico (clases abiertas): nombres, verbos, adjetivos, algunos adverbios
- ▶ Palabras de contenido funcional o gramatical (clases cerradas): determinantes, pronombres, preposiciones, conjunciones, algunos adverbios.



## Ejemplo morfología

flores = flor + es — flexión

desaprobar = des + aprobar — derivación

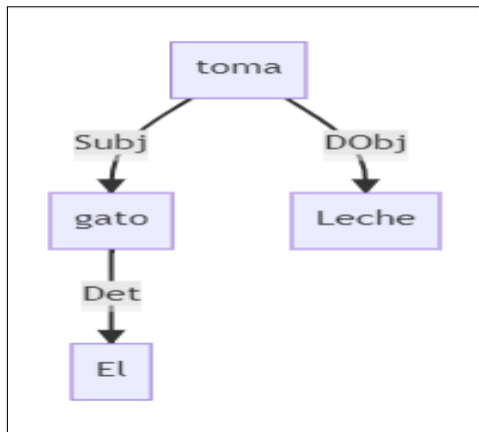
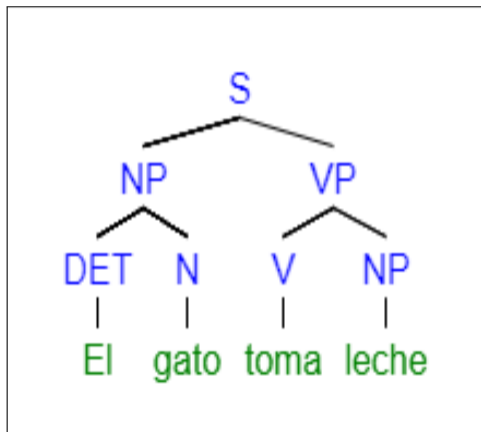
canto, cantás (cantas), canta, cantamos, cantan (cantáis), cantan,  
... + de 50 formas

# Sintaxis

- ▶ Agrupamientos de palabras al interior de la oración
- ▶ En general se pueden pronominalizar, desplazar en conjunto y la oración sigue estando bien formada.
- ▶ Constituyentes habituales:
  - ▶ grupo nominal (GN, *el vecino de mi madre*)
  - ▶ grupo adjetival (GAdj, *difícil de entender*)
  - ▶ grupo preposicional (GP, *de mi madre*)
  - ▶ grupo adverbial (GAdv, *cerca del parque*)

,

## Árbol de constituyentes y de dependencias



## Semántica

La semántica refiere al significado

- ▶ Semántica "literal", incluye significados de las palabras (o de unidades multi-palabra) y de su composición. Es el dominio de las relaciones lógicas, (ej. A implica B).
- ▶ Pragmática - Significado real", considera la situación de enunciación, las implicaturas conversacionales, las presuposiciones.

## Pragmática: ejemplo de presuposición

- ▶ El actual rey de Francia es calvo.
- ▶ ¿La proposición anterior es verdadera o falsa?
- ▶ Ni verdadera ni falsa.
- ▶ Hay una presuposición que no se cumple.
- ▶ El artículo definido *el* presupone existencia y unicidad.

## Ambigüedades en todos los niveles

La ambigüedad es la posible generación de varios significados distintos para un mismo objeto lingüístico.

- ▶ Ambigüedad fonética (segmentación en palabras)
- ▶ Ambigüedad léxica y polisemia
- ▶ Ambigüedad de categoría sintáctica
- ▶ Ambigüedad de constituyentes sintácticos

## Ambigüedad fonética

La ambigüedad deriva de segmentar en palabras de distintas formas secuencia fónica.

- ▶ *¿Me diste la caja?* frente a *¿Mediste la caja?*
- ▶ *Dámela vacía* frente a *Dame la vacía.*

## Ambigüedad léxica y polisemia

Es el caso prototípico de ambigüedad. Más de un significado para una sola palabra. Otro caso que incide también en la sintaxis es que además haya más de una categoría sintáctica.

Un ejemplo es “planta” con al menos 3 sentidos; nombre: vegetal; nombre: planta industrial y verbo.



## Ambigüedad léxica y polisemia

- ▶ homonimia : significados no vinculados , ej. vela de barco y vela de cera
- ▶ polisemia : significados vinculados , ej. cena – 1- reunión para la comida de la noche / "¿vas a la cena?"; 2- conjunto de platos que componen la comida / "la cena está servida"
- ▶

## Ambigüedad léxica y polisemia

WordNet es una Base de Datos Léxica, hecha inicialmente para el Inglés y con versiones para varios idiomas. Se recoge información que permite notar la importancia de la ambigüedad en el lenguaje.

*The WordNet 3.0 release has 117,798 nouns, 11,529 verbs, 22,479 adjectives, and 4,481 adverbs. The average noun has 1.23 senses, and the average verb has 2.16 senses. (JyM, 3ed.)*

## Ambigüedad referencial o distribucional

La ambigüedad es la posible generación de varios significados gramaticales distintos para un mismo objeto lingüístico.

- ▶ Le compró flores.
- ▶ Pueden presentarse personas con ambos títulos

## Ambigüedad de constituyentes sintácticos

La ambigüedad viene dada porque existen varios modos de analizar sintácticamente una oración (en constituyentes o en dependencias).

- ▶ Juan observó al hombre en el parque con el telescopio.
- ▶ Juan comió arroz con queso / con cuchara.
- ▶ Compró los libros baratos.
- ▶ Arregló el camión rápido.

## Ambigüedad de constituyentes sintácticos

La ambigüedad viene dada porque existen varios modos de analizar sintácticamente una oración (en constituyentes o en dependencias).

- ▶ (Juan observó al (hombre (en el parque) (con el telescopio))).
- ▶ (Juan ( observó (al hombre (en el parque)) (con el telescopio))).
- ▶ (Arregló (el camión rápido))
- ▶ (Arregló (el camión) rápido)