

# Introducción a PLN

---



**iimas**

Dra. Helena Gómez Adorno  
[helena.gomez@iimas.unam.mx](mailto:helena.gomez@iimas.unam.mx)

Dr. Orlando Ramos  
[orlando.ramos@aries.iimas.unam.mx](mailto:orlando.ramos@aries.iimas.unam.mx)

Correo del curso:  
[pln.cienciadedatos@gmail.com](mailto:pln.cienciadedatos@gmail.com)

# ¿Qué es lenguaje natural?



**LENGUAJE  
NATURAL**

**LENGUAJE  
FORMAL**

**LENGUAJE  
ARTIFICIAL**

# Dimensiones del lenguaje natural



Conocimiento mundo

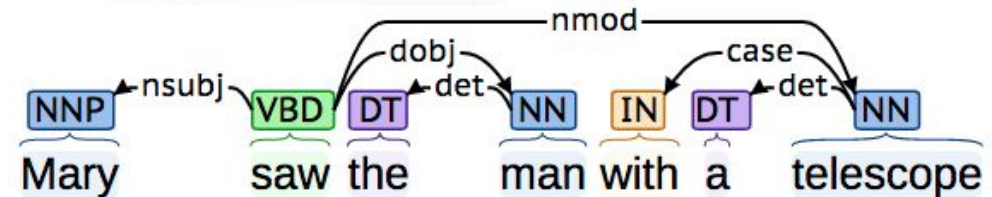
Pragmática

Semántica

Sintaxis

Morfología

Fonética



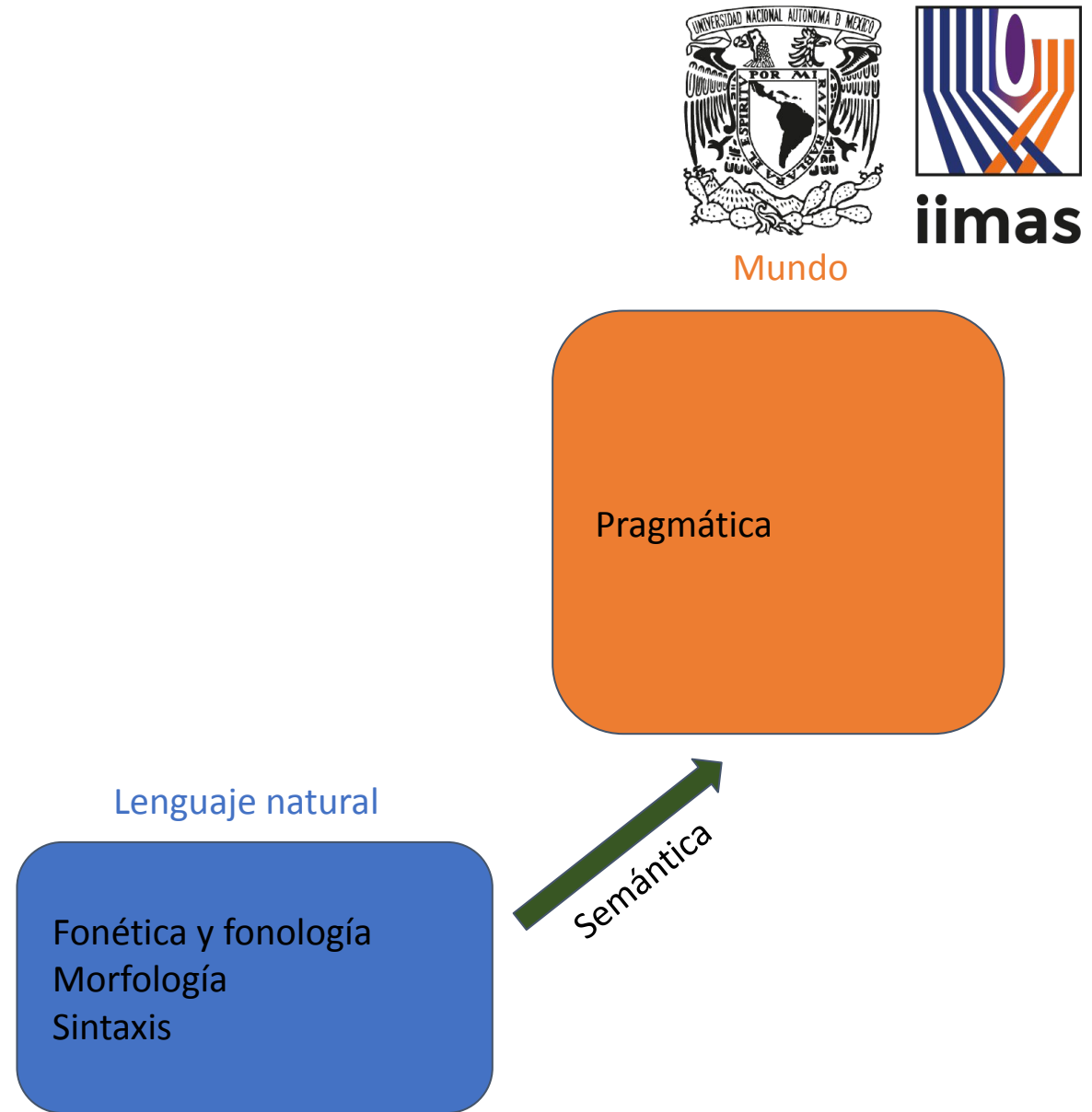
PP (1s) see (past)...

əj sɔ e mən wið e teləskop

I saw a man with a telescope

# Partes de la lingüística

- **Fonética y fonología:** el estudio de los sonidos
- **Morfología:** el estudio de los componentes lingüísticos mínimos con significado.
- **Sintaxis:** el estudio de las relaciones estructurales entre palabras.
- **Semántica:** el estudio de la relación del lenguaje con el mundo
- **Pragmática:** el estudio del lenguaje desde los emisores y receptores
- **Discurso:** el estudio de unidades lingüísticas más amplias que un simple enunciado



# Características (problemas) del LN



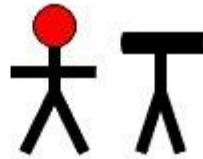
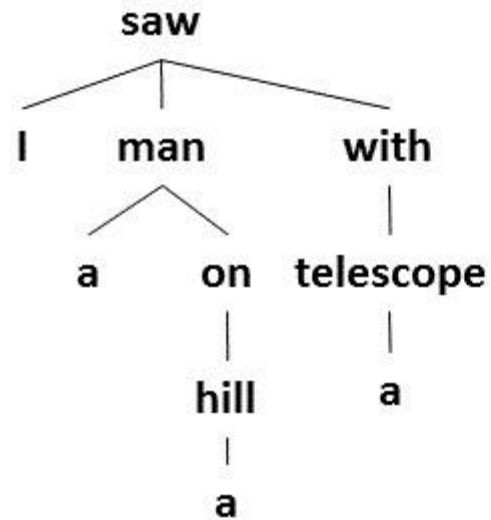
**iimas**

- Ambigüedad
  - Sintáctica
  - Semántica
  - Pragmática
  - ...
- Multiplicidad de variantes
- Evolución
- Oscuridad, slang, etc

# Ambigüedad sintáctica



- Ambigüedad

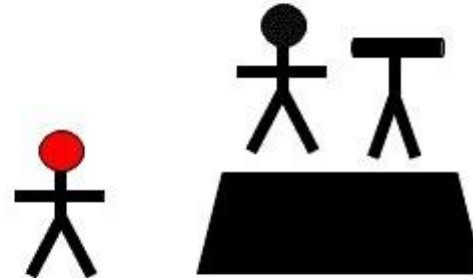
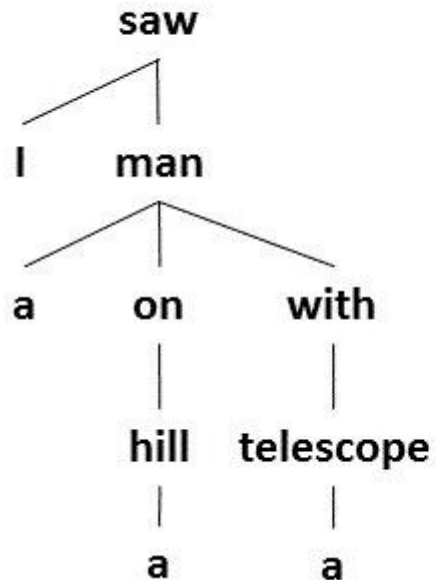


I saw a man with a telescope

# Ambigüedad sintáctica



- Ambigüedad

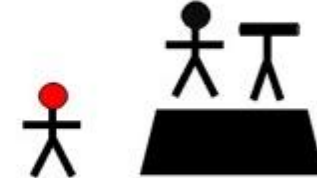
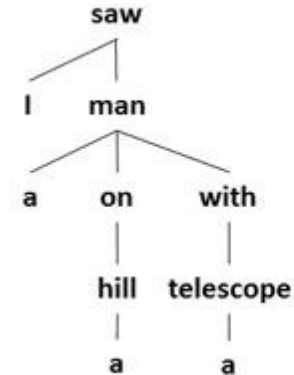


I saw a man with a telescope

# Ambigüedad sintáctica



- Ambigüedad



I saw a man with a telescope



# Ambigüedad léxico-semántica



iimas

- ¿Se quedará a dormir?
- Sí.
- Quizá debería saber que la casa está encantada.
- Ah, pues díglele que a mí también me hace ilusión quedarme.

- ¿Por qué los de Lepe tiran a sus hijos a un pozo?
- Porque saben que en el fondo son buenos.

El capitán dijo: “¡Bajen las velas!” Y los de arriba se quedaron sin luz.

- ¡Qué fresca está la mañana!
- Normal, es de hoy.

- ¿Qué pasa si un elefante se queda de pie encima de una pata?
- a- Que se cae.
- b- Que el pato se queda viudo.
- c- Que aplasta a su domador.

# Ambigüedad fonológica / morfosintáctica

Dígame su nombre.

- Peter O'Brian
- Decídase por favor.

- ¡Acusado! ¡Hable ahora o calle para siempre!

- Elijo calle.

Mi marido se ha ido de ca[S]a

1. de casa
2. de caza

¿Qué esconde?

¿Que es conde?

# Ambigüedad pragmática



iimas

- ¡Camarero! ¿Se puede saber qué está haciendo esta mosca en mi sopa?

- Mmm, yo creo que está nadando a braza, señor.”

- ¿Cómo estás?

- Han perdido los Pumas

- Golpeó el armario con un palo y lo rompió.

MARÍA SIMARRO VÁZQUEZ, Humor verbal basado en la ambigüedad léxica y competencia léxico-semántica, Pragmalingüística 25 (2017): 618-636

# Multiples variantes

- Existen unas 7000 lenguas en el mundo.
- Con diferentes palabras, estructuras sintácticas, reglas morfológicas, sistemas fonéticos y escrituras.
- El intercambio-traducción entre unas y otras no es obvio.

<https://wals.info/>



iimas



# Múltiples variantes



<https://universalcustomer.com/2016/08/26/traductor-automatico-mejores-peores-traduccion/>

<https://www.strommeninc.com/funny+bad+translations+part+1/attachment/bad-english-translation/>



# Múltiples variantes



iimas

a b c d

latino

α β γ δ

griego

Ⲁ Ⲃ Ⲅ Ⲇ

copto

а б г д

cirílico

ⴀ ⴁ ⴂ ⴃ

mjedruli

Ա Բ Գ Դ

armenio

ⵍ ⵍⵎ ⵍⵏ ⵍⵔ

tifinagh

አ ቡ ጊ ዳ

geez

א ב ג ד

hebreo

أ ب ج د

árabe

ܐ ܒ ܓ ܕ

siriaco

ᐱ ᐳ ᐵ ᐷ

mandeo

DIFERENTES ALFABETOS

# Múltiples variantes



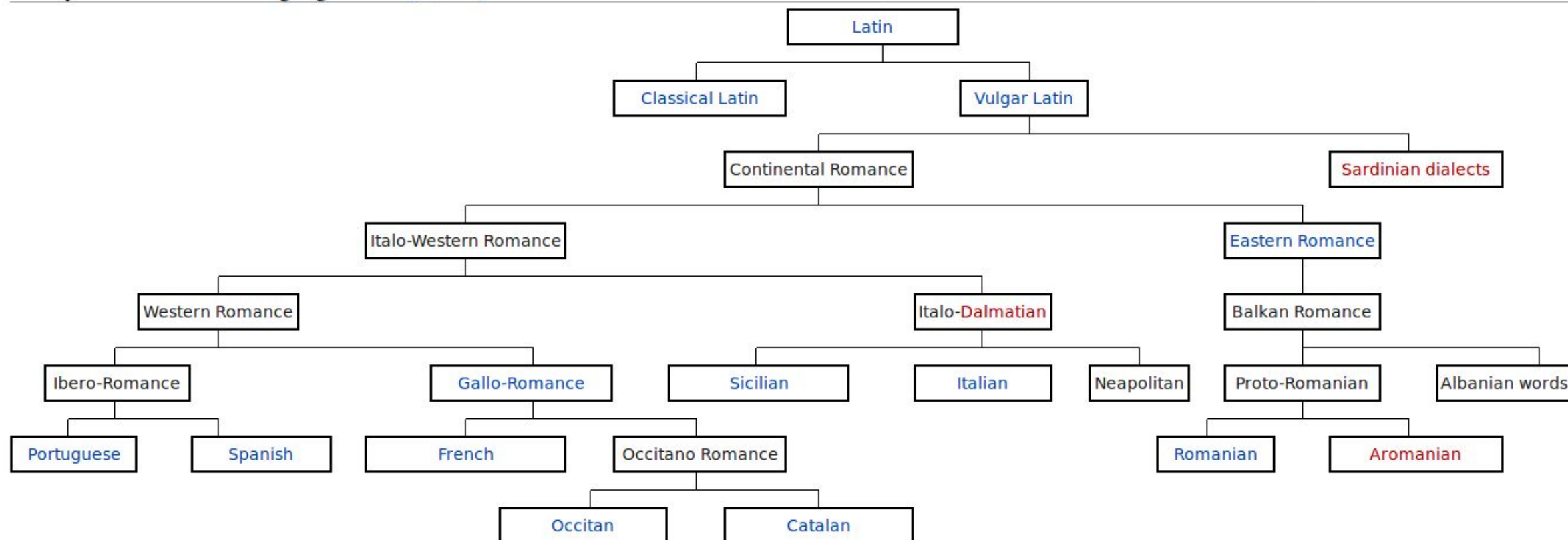
ESCRITURA NO ALFABÉTICA

# Evolución y cambio



iimas

Family tree of Romance languages [ [change](#) | [change source](#) ]



Fuente: Wikipedia



# Evolución y cambio



- Hay palabras que se incorporan a la lengua: tuit, celular, selfie
- Hay palabras que desaparecen: doncel, jumento, vuestra merced, vosotros
- Otras cambian de sentido: hasta, celular, ratón
- Algunas estructuras sintácticas cambian: SOV (latín) -> SVO (español)
- Algunos fonemas (sonidos) desaparecen y aparecen otros nuevos: caballo [kabaɫo] > [kabaio] > [kabaʒo] / [kavaɫo]
- Muchas lenguas desaparecen. Otras se mezclan (pidgin y criollos). Otras resucitan (hebreo), y de otras solo quedan testimonios escritos (etrusco).

# Los humanos a veces, no son claros



iimas

- <https://www.poemas-del-alma.com/pablo-neruda-oda-al-amor.htm>

eminentemente presidencialista, a su Presidente, el imputado Miguel Blesa. Y ello, es lógico si se parte de que el Regulador no conoce, ni puede conocer, la profunda y concreta estructura de negocio de las diferentes Entidades bajo su control.

**TERCERO.-** Por todo ello, desprendiéndose de las actuaciones indicios racionales de responsabilidad criminal del imputado en relación, o bien, con la posible comisión de un delito societario de administración desleal; o bien, de un delito de apropiación indebida; en posible concurso con un delito de falsedad en documento público, resulta proporcionado acordar la medida cautelar instada de prisión provisional del imputado; pues, junto a tales elementos objetivos, se aprecia un innegable riesgo de fuga, fundamentalmente, porque el pronóstico de pena correspondiente a los hechos imputados permite sustentar, en tal sentido, la adopción de la medida que se recurre; respecto de lo cual, por lo demás, no puede descartarse la posible conexidad con los hechos relacionados y antes descritos en relación con los créditos concedidos al otro imputado, Díaz Ferrán; lo que, desde luego, no fundamenta

direccionamente esta medida; pero, permite establecer una cierta continuidad por parte del imputado Miguel Blesa, en términos tales que permiten concluir que no se trata de un proceder esporádico, aislado de conductas incriminatorias previas, o de hechos extremos o circunstancias que pueden ser muy relevantes para recabar evidencias a nivel de instrucción, en virtud de la relación o connivencia del imputado con otros miembros integrantes de los cuadros directivos de CAJA MADRID al más alto nivel, de tal modo que la libertad del imputado pudiera incurrir en ocultación, alteración o destrucción de las fuentes de prueba relevantes para el enjuiciamiento, bajo peligro fundado y concreto, relacionado con las personas que específicamente pudieron colaborar con el mismo, o con comunicaciones de correo electrónico o de similar naturaleza que pudieran alterar el ritmo o culminación de la instrucción, porque desaparezcan o porque se alteren o sesguen manipulatoriamente a favor del imputado (artículo 503.3.b de la Ley de Enjuiciamiento Criminal).

Por lo demás, como es sabido, el riesgo de fuga no debe valorarse y pronosticarse, exclusivamente, desde la perspectiva del empadronamiento o ubicación residencial, más o menos estable o previsible del imputado; pues, tal premisa no



Madrid



# Inteligencia Artificial

## Lingüística computacional

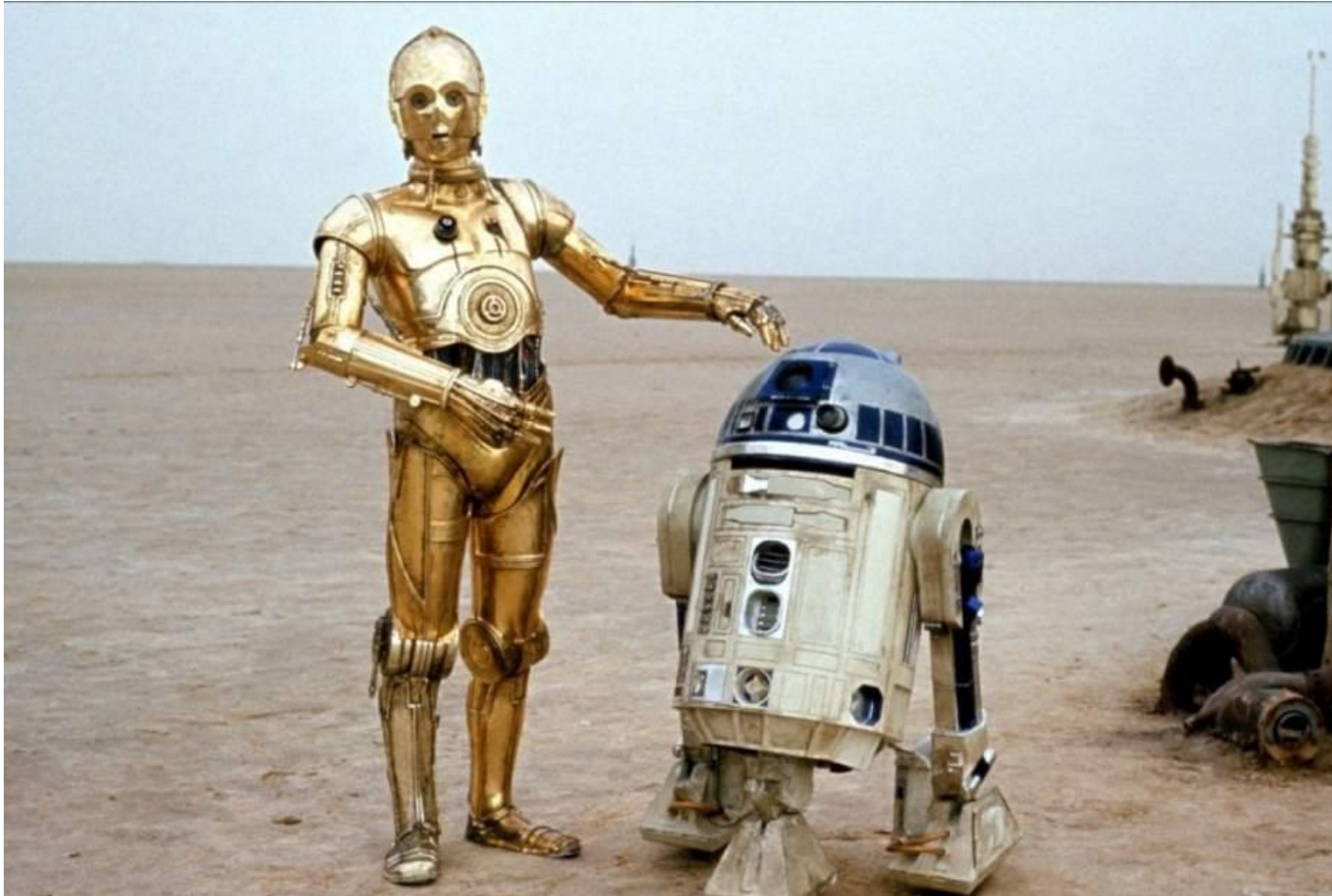
### Procesamiento del LN



# Inteligencia artificial



**iimas**



Los humanos deseamos crear inteligencia de forma artificial.

¿Puede haber inteligencia sin lenguaje?

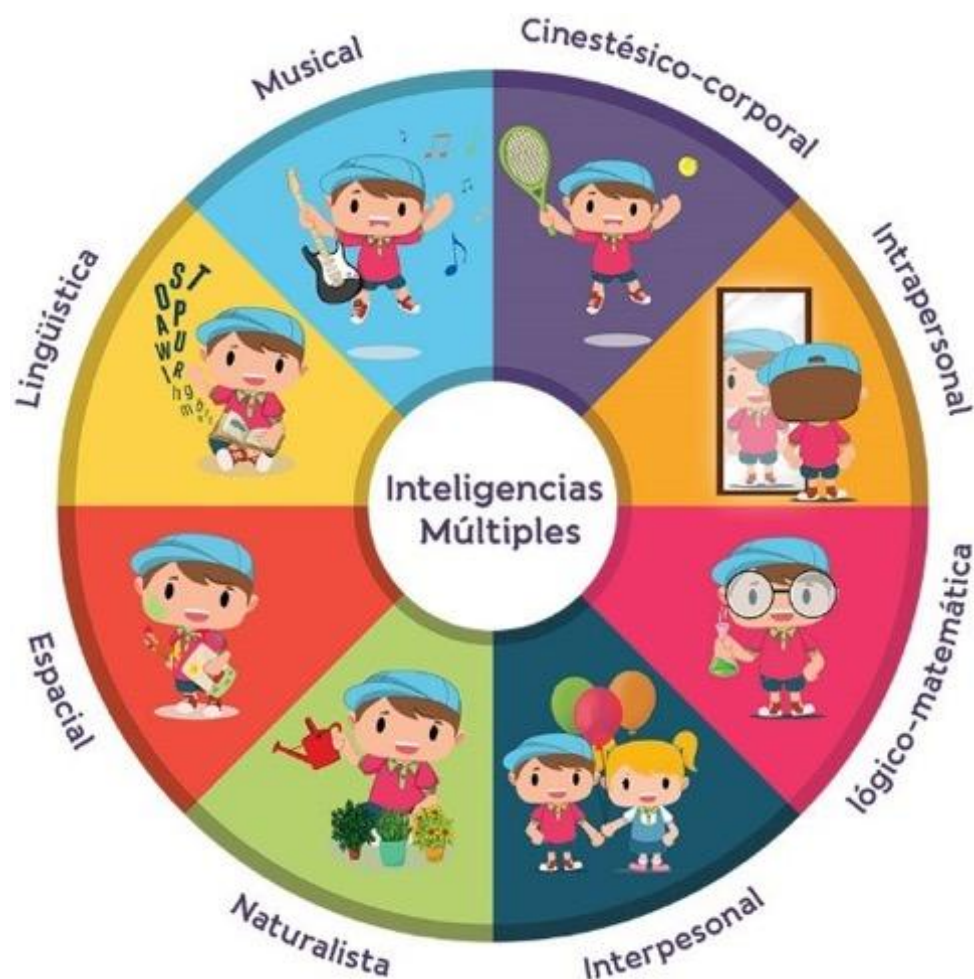
¿Se puede desarrollar una teoría del lenguaje sin conectarla con otras funciones cognitivas?

¿Se puede desarrollar un componente lingüístico si no existe conocimiento contextual y enciclopédico?

# Inteligencia artificial



iimas



- La inteligencia tiene muchos componentes.
- Howard Gardner desarrolló la teoría de las inteligencias múltiples.
- Las capacidades lingüísticas son un componente fundamental de la inteligencia.

Fuente:

<http://materialdidacticolittlesherlocks.com/aprendiendo-con-miss-yiyi/inteligencias-multiples-en-la-ensenanza/>

# Lingüística computacional

- Entre la lingüística y la computación.
- Busca formular modelos que ayuden a diseñar una inteligencia lingüística.
- Pretende conocer como funciona la comprensión y generación del lenguaje mediante las computadoras.



Nuevo Conocimiento

# Procesamiento del lenguaje natural



- Realiza tareas concretas de la lingüística computacional, con la ayuda de las tecnologías del lenguaje.
- Cubre necesidades sociales actuales, que desarrollan tecnologías que ayuden a vivir en la sociedad de la información.



# Procesamiento del lenguaje natural



**Lenguaje hablado**

**Lenguaje escrito**

## NATURAL LANGUAGE UNDERSTANDING

- Reconocimiento de voz -> Escuchar

- Análisis de texto -> Lectura

- Síntesis de habla -> Hablar

- Generación de texto -> Escritura

## NATURAL LANGUAGE GENERATION



# Análisis del habla



- El reconocimiento de voz consiste en traducir el lenguaje hablado a texto.
- Algunas de las aplicaciones de reconocimiento de voz son:
  - Interfaces de sistemas de habla (sistemas de dictado, sistemas GPS y teléfonos móviles).
  - Enseñanza de un segundo idioma.
  - Ayuda a personas con discapacidades.



# Generación de habla



- La síntesis del habla es la producción artificial de habla humana a partir de texto o datos.
- Algunas aplicaciones de síntesis de voz son:
  - Ayuda a personas con discapacidades como *Stephen Hawking*.
  - Interacción con dispositivos móviles a través de interfaces de procesamiento de lenguaje natural.



# Análisis de textos



- El análisis de texto es el proceso de extraer información de alta calidad del texto. Tal información se usa para una amplia gama de tareas.
- Algunas aplicaciones de ejemplo de la vida real son:
  - Aplicaciones de búsqueda web (Google, Bing, etc.)
  - Errores gramaticales (Ofimática, editores de texto, etc.)
  - Minería de Opiniones (sistema de recomendación de Netflix)
  - Colocación automatizada de anuncios
  - Monitoreo de redes sociales
  - Aplicaciones de seguridad

# Generación de textos



iimas

- La generación de texto es la tarea de generar un texto legible por humanos a partir de los datos.
- Se pueden generar textos a partir de otros textos (ej. Generación de fake news).
- O se pueden generar textos originales
  - Sistemas de diálogo (humano-humano, humano-máquina)





**iimas**

# ANÁLISIS DE TEXTOS

# Pre-requisitos para analizar un texto



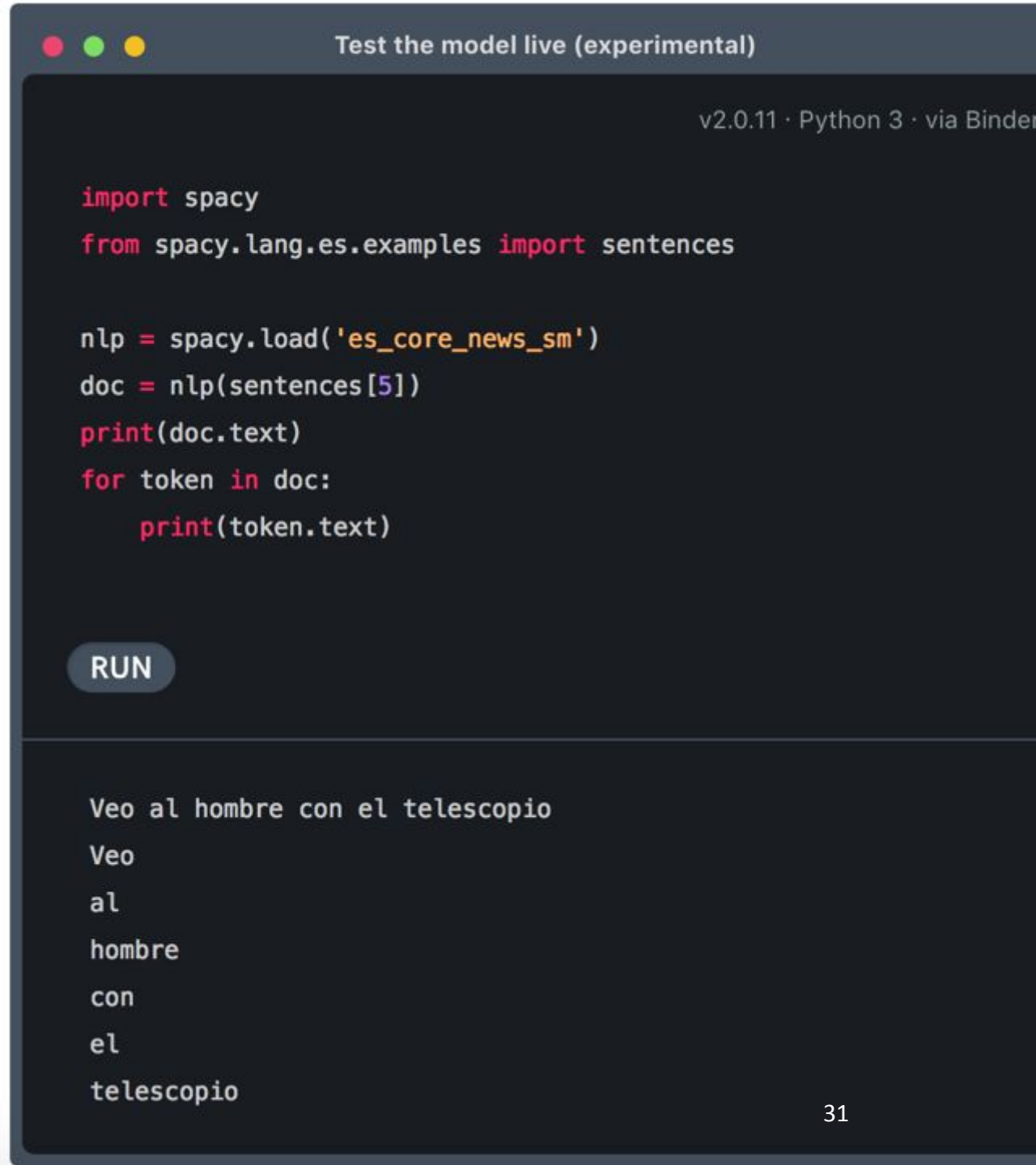
- Ser capaz de dividir el texto por frases.
- Ser capaz de encontrar las palabras.
- DEFINICIÓN DE PALABRA: Todo aquello que se encuentra entre dos espacios en blanco o espacio en blanco y signo de puntuación.



# Tokenizador

- El tokenizador simplemente separa texto en una lista de tokens usando algunos caracteres como referencia para dividir.
- Los tokens son generalmente palabras y símbolos de puntuación.
- La tokenización no implica ningún nivel de análisis.
- La tokenización se realiza muy rápido.

<https://spacy.io/models/es>



The screenshot shows a Jupyter Notebook window titled "Test the model live (experimental)". The code in the notebook is as follows:

```
import spacy
from spacy.lang.es.examples import sentences

nlp = spacy.load('es_core_news_sm')
doc = nlp(sentences[5])
print(doc.text)
for token in doc:
    print(token.text)
```

Below the code is a "RUN" button. The output of the code is displayed in the cell below:

```
Veo al hombre con el telescopio
Veo
al
hombre
con
el
telescopio
```

# Dimensiones del lenguaje natural



Conocimiento mundo

Pragmática

Semántica

Sintaxis

Morfología

Fonética



NN

Mary

saw the

man with

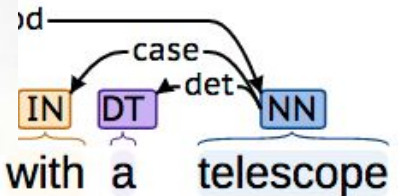
IN

DT

a

NN

telescope



PP (1s) see (past)...

aj sɔ e mæn wið e teləskop

I saw a man with a telescope



# Análisis de textos



## Niveles

- Fonética-fonología (sonidos) – Corresponde al análisis de habla
- Morfología (clases de palabras y segmentación)
- Sintaxis (oraciones, sintagmas y orden de las palabras)
- Semántica (significados)
- Pragmática (interacciones, uso y contexto)
- Discurso (expresiones correferenciales, estructura retórica)

# Morfología



- ¿Cuáles son las clases de palabras y por qué importa saberlo?
- ¿Qué partes tienen las palabras? ¿Cómo se segmentan?

# ¿Cuáles son las palabras?



iimas

Pago por \$475 347.50 M.N. (cuatrocientos setenta y cinco mil trescientos cuarenta y siete pesos 50/100 M.N.) del Restaurant Pomme de terre D'Opera por haberlo traspasado a Mary-Carmen da Cunha en San Luis Potosí.

7/02/2017

# Clases de palabras

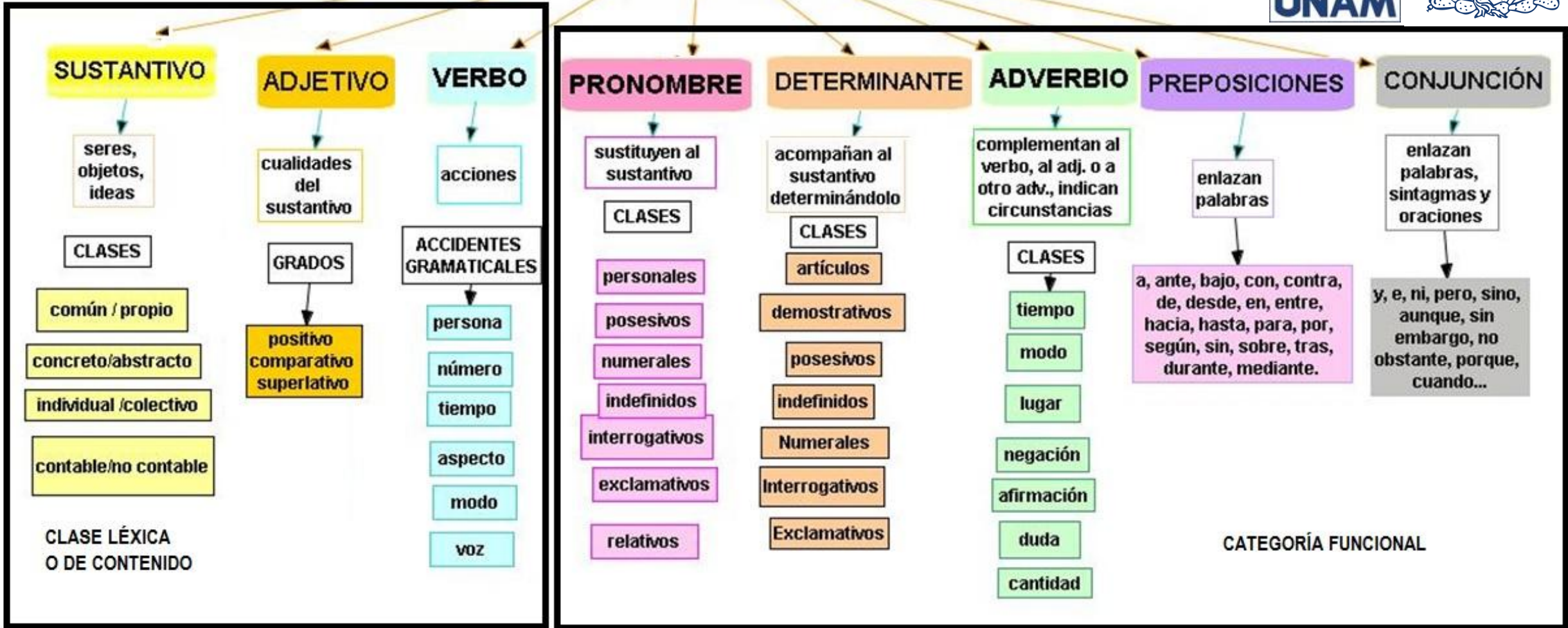


iimas

Part-of-speech (POS), clases morfológicas, categorías gramaticales



# CLASES DE PALABRAS





# Etiquetado Penn Treebank



| Number | Tag  | Description                              |
|--------|------|--|
| 1.     | CC   | Coordinating conjunction                 |
| 2.     | CD   | Cardinal number                          |
| 3.     | DT   | Determiner                               |
| 4.     | EX   | Existential <i>there</i>                 |
| 5.     | FW   | Foreign word                             |
| 6.     | IN   | Preposition or subordinating conjunction |
| 7.     | JJ   | Adjective                                |
| 8.     | JJR  | Adjective, comparative                   |
| 9.     | JJS  | Adjective, superlative                   |
| 10.    | LS   | List item marker                         |
| 11.    | MD   | Modal                                    |
| 12.    | NN   | Noun, singular or mass                   |
| 13.    | NNS  | Noun, plural                             |
| 14.    | NNP  | Proper noun, singular                    |
| 15.    | NNPS | Proper noun, plural                      |
| 16.    | PDT  | Predeterminer                            |
| 17.    | POS  | Possessive ending                        |
| 18.    | PRP  | Personal pronoun                         |

|     |       |                                       |
|-----|-------|---------------------------------------|
| 19. | PRP\$ | Possessive pronoun                    |
| 20. | RB    | Adverb                                |
| 21. | RBR   | Adverb, comparative                   |
| 22. | RBS   | Adverb, superlative                   |
| 23. | RP    | Particle                              |
| 24. | SYM   | Symbol                                |
| 25. | TO    | <i>to</i>                             |
| 26. | UH    | Interjection                          |
| 27. | VB    | Verb, base form                       |
| 28. | VBD   | Verb, past tense                      |
| 29. | VBG   | Verb, gerund or present participle    |
| 30. | VBN   | Verb, past participle                 |
| 31. | VBP   | Verb, non-3rd person singular present |
| 32. | VBZ   | Verb, 3rd person singular present     |
| 33. | WDT   | Wh-determiner                         |
| 34. | WP    | Wh-pronoun                            |
| 35. | WP\$  | Possessive wh-pronoun                 |
| 36. | WRB   | Wh-adverb                             |

# Etiquetas EAGLES

<http://www.lsi.upc.es/~nlp/tools/parole-sp.html>

1. [ADJETIVOS](#)
2. [ADVERBIOS](#)
3. [ARTÍCULOS](#)
4. [DETERMINANTES](#)
5. [NOMBRES](#)
6. [VERBOS](#)
7. [PRONOMBRES](#)
8. [CONJUNCIONES](#)
9. [NUMERALES](#)
10. [INTERJECCIONES](#)
11. [ABREVIATURAS](#)
12. [PREPOSICIONES](#)
13. [SIGNOS DE PUNTUACIÓN](#)

| NOMBRES |                  |             |        |
|---------|------------------|-------------|--------|
| Pos.    | Atributo         | Valor       | Código |
| 1       | Categoría        | Nombre      | N      |
| 2       | Tipo             | Común       | C      |
|         |                  | Propio      | P      |
| 3       | Género           | Masculino   | M      |
|         |                  | Femenino    | F      |
|         |                  | Común       | C      |
| 4       | Número           | Singular    | S      |
|         |                  | Plural      | P      |
|         |                  | Invariable  | N      |
| 5       | Caso             | -           | 0      |
| 6       | Género Semántico | -           | 0      |
| 7       | Grado            | Apreciativo | A      |

| VERBOS |           |             |        |
|--------|-----------|-------------|--------|
| Pos.   | Atributo  | Valor       | Código |
| 1      | Categoría | Verbo       | V      |
| 2      | Tipo      | Principal   | M      |
|        |           | Auxiliar    | A      |
| 3      | Modo      | Indicativo  | I      |
|        |           | Subjuntivo  | S      |
|        |           | Imperativo  | M      |
|        |           | Condicional | C      |
|        |           | Infinitivo  | N      |
|        |           | Gerundio    | G      |
|        |           | Participio  | P      |
| 4      | Tiempo    | Presente    | P      |
|        |           | Imperfecto  | I      |
|        |           | Futuro      | F      |
|        |           | Pasado      | S      |
| 5      | Persona   | Primera     | 1      |
|        |           | Segunda     | 2      |
|        |           | Tercera     | 3      |
| 6      | Número    | Singular    | S      |
|        |           | Plural      | P      |
| 7      | Género    | Masculino   | M      |
|        |           | Femenino    | F      |

# POS Tagging: ... problema

Las palabras, tomadas en forma aislada,  
son ambiguas respecto a su categoría



Yo bajo con el hombre bajo a

|    |    |    |    |    |    |    |
|----|----|----|----|----|----|----|
| PP | VM | SP | TD | NC | VM | NC |
|    | VM |    |    |    | VM | SP |
|    | AQ |    |    |    | AQ |    |
|    | NC |    |    |    | NC |    |
|    | SP |    |    |    | SP |    |

tocar el bajo bajo la escalera .

|    |    |    |    |    |    |    |
|----|----|----|----|----|----|----|
| VM | TD | VM | VM | TD | NC | FP |
| VM |    | VM | VM | NC |    |    |
|    |    | AQ | AQ | PP |    |    |
|    |    | NC | NC |    |    |    |
|    |    | SP | SP |    |    |    |



# pero ...

La categoría de la mayoría de las palabras no es ambigua dentro de un contexto



Yo bajo con el hombre bajo a

|    |    |    |    |    |    |    |
|----|----|----|----|----|----|----|
| PP | VM | SP | TD | NC | VM | NC |
|    | VM |    |    |    | VM | SP |
|    | AQ |    |    |    | AQ |    |
|    | NC |    |    |    | NC |    |
|    | SP |    |    |    | SP |    |

tocar el bajo bajo la escalera .

|    |    |    |    |    |    |    |
|----|----|----|----|----|----|----|
| VM | TD | VM | VM | TD | NC | FP |
| VM |    | VM | VM | NC |    |    |
|    |    | AQ | AQ | PP |    |    |
|    |    | NC | NC |    |    |    |
|    |    | SP | SP |    |    |    |

# solución



desambiguador morfosintáctico (Pos tagger)

El objetivo de un desambiguador (tagger) es el de asignar a cada palabra la categoría más "apropiada", dentro de un contexto

- Basados en reglas
- Estadísticos
- Basados en transformación

# Herramientas en línea



**iimas**

- <https://linguakit.com/es/etiquetador-morfosintactico>
- <http://protos.dis.ulpgc.es/investigacion/desambigua/morfosintactico.htm>
- <http://nlp.stanford.edu:8080/parser/index.jsp>

# Etiquetador grammatical (POS)

- El etiquetado de parte del discurso (POS) es el proceso de marcar una palabra en un texto con una parte particular del discurso, en función de su definición y contexto.
- Los etiquetadores POS generalmente requieren un corpus marcado manualmente.
- El etiquetado POS es una forma de análisis morfo-sintáctico.

Test the model live (experimental)

v2.0.11 · Python 3 · via Binder

```
import spacy
from spacy.lang.es.examples import sentences

nlp = spacy.load('es_core_news_sm')
doc = nlp(sentences[5])
print(doc.text)
for token in doc:
    #print(token.text, token.pos_, token.dep_)
    print(token.text, token.pos_)
```

RUN

Veo al hombre con el telescopio  
Veo VERB  
al ADP  
hombre NOUN  
con ADP  
el DET  
telescopio NOUN

# Segmentación



iimas

- Algunas palabras son indivisibles: que, no, ya, y...
- Pero en general las palabras tienen partes:
  - Raíz o raíces
  - Afijos
    - Flexivos: singular/plural, femenino/masculino, tiempos verbales.
    - Derivativos: prefijos, interfijos, sufijos.

CANT-O

CANT-ABA-MOS

CANT-O-S

GAT-A

GAT-IT-O

GAT-IT-OS

AMOR

EN-AMOR-AR

DES-EN-AMOR-AR

CONSTITU-IR

CONSTITU-CIÓN

CONSTITU-CION-AL

CONSTITU-CION-AL-IZ-AR

CONSTITU-CION-AL-IZ-A-CIÓN

ABRE-LATA-S

CORRE-CAMINO-S

# Segmentación



- A LA GATA LE GUSTA TRAER GATAS QUE TRAEN A MÁS GATOS Y A SUS GATITOS
- ¿Cuántas palabras hay en este texto?
- ¿Cuántas palabras diferentes?
- Hay dos tareas diferentes que ayudan a buscar las palabras y sus raíces: lematización y stematización.



# Stemmer



iimas

- Los algoritmos de *Stemming* intentan reducir las palabras flexionadas y derivadas en su forma raíz
- El *Stemming* es una forma de análisis morfológico
- Los algoritmos *Stemming* no usan ningún recurso y generalmente son muy rápidos

veo al hombre con el telescopio en la terraza del edificio

spanish

⬆

Stem!

**veo al hombre con el telescopio en la terraza del edifici**

<http://proiot.ru/jssnowball/>

# Lematizador

- La lematización consiste en mapear una palabra en una entrada del diccionario conocida como lema.
- La lematización requiere etiquetado POS.
- La lematización requiere un diccionario como WordNet o Wikipedia.
- La lematización es un proceso computacionalmente costoso, en comparación con el stemming.

Test the model live (experimental)

v2.0.11 · Python 3 · via Binder

```
import spacy
from spacy.lang.es.examples import sentences

nlp = spacy.load('es_core_news_sm')
doc = nlp(sentences[4])
print(doc.text)
for token in doc:
    #print(token.text, token.pos_, token.dep_)
    print(token.text, token.lemma_)
```

RUN

El gato come pescado  
El El  
gato gato  
come comer  
pescado pescar

# Sintaxis



iimas

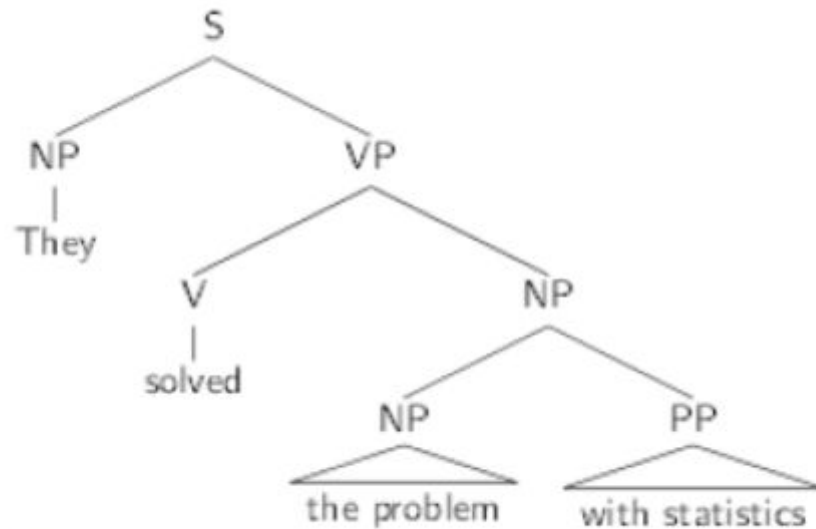
La tarea sintáctica más importante en procesamiento del lenguaje natural es el **PARSING** o análisis oracional, reconoce unidades de estructura de nivel superior que nos permiten comprimir nuestra descripción de una oración.

Objetivo del análisis sintáctico (**parsing**):

- Detectar si una oración es correcta
- Proporcionar una estructura sintáctica de una oración.
- El **parsing** es la tarea de descubrir la estructura sintáctica del lenguaje y, a menudo, se considera un requisito previo importante para construir sistemas capaces de comprender el lenguaje.
- La estructura sintáctica es necesaria como primer paso hacia la interpretación semántica, para detectar fragmentos compuestos para indexar en un sistema IR ...

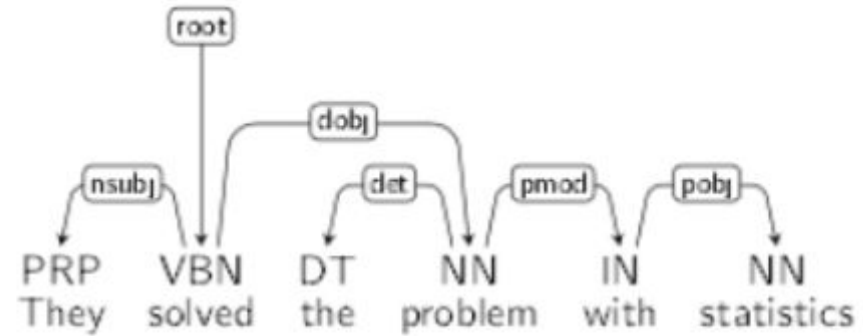
# Teorías de estructura sintáctica

## Árbol de constituyentes



- Main element: constituents (or phrases, or bracketings)
- Constituents abstract linguistic units
- Results in nested trees

## Árbol de dependencias



- Main element: dependency
- Focus on relations between words
- Handles **free word order** nicely.  
Exercise: parse **He saw the dog yesterday that barked all night**

# Factores para el análisis sintáctico

- Expresividad de la Gramática
- Cobertura
- Fuentes de conocimiento involucradas
- Estrategia de análisis
- Dirección de análisis
- Orden de solicitud de producción
- Gestión de la ambigüedad.



**iimas**

# Analizadores hoy

- Gramáticas libres de contexto - CFG (extendido o no)
  - Tabular
  - Gráficas
- Basado en la unificación
- Estadística
- Análisis de dependencia
- Análisis robusto (superficial, fragmentario, fragmentado, observador)





# Análisis de constituyentes



Book that flight

$S \rightarrow NP VP$

$S \rightarrow Aux NP VP$

$S \rightarrow VP$

$NP \rightarrow Det Nominal$

$Nominal \rightarrow Noun$

$Nominal \rightarrow Noun Nominal$

$NP \rightarrow Proper-Noun$

$VP \rightarrow Verb$

$VP \rightarrow Verb NP$

$Det \rightarrow that \mid this \mid a$

$Noun \rightarrow book \mid flight \mid meal \mid money$

$Verb \rightarrow book \mid include \mid prefer$

$Aux \rightarrow does$

$Prep \rightarrow from \mid to \mid on$

$Proper-Noun \rightarrow Houston \mid TWA$

$Nominal \rightarrow Nominal PP$

# Análisis de dependencias

- El análisis de dependencia extrae un árbol sintáctico de un texto dado.
- Los árboles representan las relaciones sintácticas entre las palabras de una oración.
- El análisis utiliza un corpus marcado manualmente.
- El análisis requiere más recursos computacionales que los anteriores.



Test the model live (experimental)

v2.0.11 · Python 3 · via Binder

```
import spacy
from spacy.lang.es.examples import sentences

nlp = spacy.load('es_core_news_sm')
doc = nlp(sentences[5])
print(doc.text)
for token in doc:
    print(token.text, token.dep_)
```

RUN

Veo al hombre con el telescopio  
Veo ROOT  
al case  
hombre obj  
con case  
el det  
telescopio obl

# Tarea: Language Design - Noam Chomsky



Ver una breve entrevista a Noam Chomsky, creador de los lenguajes formales. En este tema tendrás que comprender la jerarquía de Chomsky y la tipología de cada clase de lenguaje. El gran lingüista y computólogo nos ayuda a comprender cuáles son los elementos fundamentales del diseño de un lenguaje formal.

<http://serious-science.org/language-design-679>