

# Redes Neuronales para el análisis y la generación de texto

Grupo PLN  
InCo- Fing- UdelaR

ELI - IV Escuela Latinoamericana de Informática  
Octubre 2025 - Valparaíso

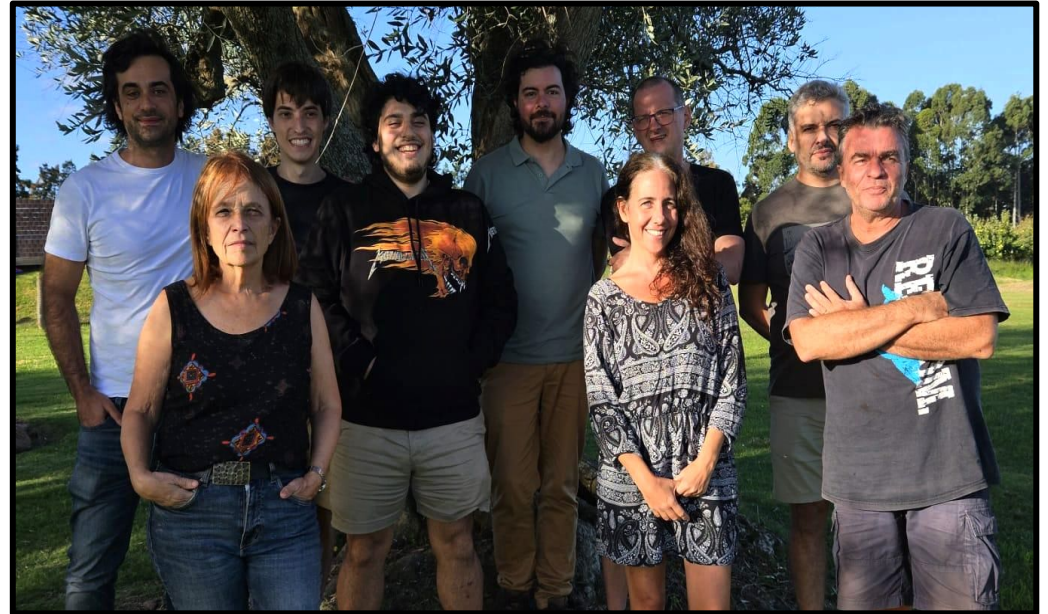
# Contenido

---

- 1.- Conceptos básicos de Procesamiento de Lenguaje Natural
- 2.- Fundamentos de Redes Neuronales
- 3.- Representaciones vectoriales de palabras
- 4.- Modelos de lenguaje: conceptos básicos, modelos basados en Transformers, grandes modelos de lenguaje (LLM)
- 5.- Usos de los LLM: entrenamiento, fine tuning, prompting, RAG, aplicaciones

# Grupo Procesamiento de Lenguaje Natural

¿Quiénes somos?



✉ [pln@fing.edu.uy](mailto:pln@fing.edu.uy)

✉ @PLN\_UdelaR

🌐 [www.fing.edu.uy/inco/grupos/pln](http://www.fing.edu.uy/inco/grupos/pln)



**Aiala  
Rosá**



**Luis  
Chiruzzo**



**Dina  
Wonsever**



**Guillermo  
Moncecchi**



**Juanjo  
Prada**



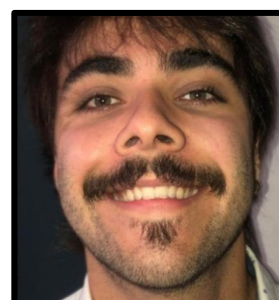
**Diego  
Garat**



**Ignacio  
Sastre**



**Santiago  
Góngora**



**Guillermo  
Rey**



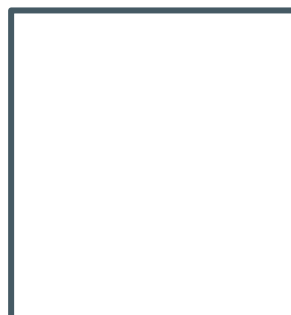
**Ignacio  
Remersaro**



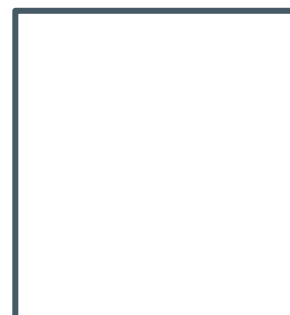
**Santiago  
Castro**



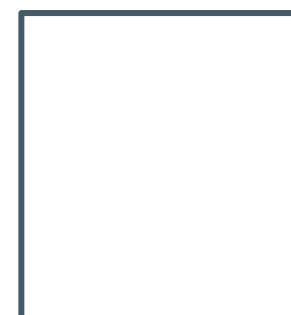
**Agustín  
Martínez**



**Facundo  
Díaz**



**Sofía  
Camacho**



**Juan Pablo  
Conde**

# ¿Qué nos interesa investigar?

Cómo las computadoras pueden **entender y escribir**  
las lenguas que hablan l@s human@s

Español

Inglés

Guaraní

Portugués

Traducción automática

Q&A

PLN en videojuegos

y cómo esos métodos son usados por los usuarios,  
qué resultados dan y cómo podrían ser mejorados (en general)

PLN para la educación

PLN para las lenguas de señas

# Introducción

---

¿Qué es el PLN?

Procesamiento del Lenguaje Natural

# Introducción

---

¿Qué es un lenguaje?

# Introducción

---

## ¿Qué es un lenguaje?

(1) Facultad del ser humano de expresarse y comunicarse con los demás a través del sonido articulado o de otros sistemas de signos

(2) Sistema de signos que utiliza una comunidad para comunicarse

(3) Conjunto de signos y reglas que permite la comunicación con una computadora u ordenador

...

(RAE: <https://www.rae.es/desen/lenguaje>)



# Introducción

- Alfabeto
- Reglas

```
1  ' Globales -----
2  Var Variable0:Booleano
3  Var Variable1:Cadena
4  ' Fin Globales -----
5  Proc Procedimiento ' <- Procedimiento sin retorno.
6      Var Variable2:Entero ' Locales
7      Var Variable3:Real
8
9      Si Variable0 = Falso Entonces ' Condición "If"
10         Contar Variable2 = 0 a 9 ' Bucle "For"
11         Variable1 = Variable1 + "1"
12         Seguir ' "End For"
13     FinSi ' "End If"
14
15     Variable3 = 5.13
16 FinProc
```

```
<program> → begin <stmt_list> end
<stmt_list> → <stmt>
               | <stmt> ; <stmt_list>
<stmt> → <var> = <expression>
<var> → A | B | C
<expression> → <var> + <var>
               | <var> - <var>
               | <var>
```

$0 \rightarrow SN \ SV$

$SN \rightarrow DET \ N$

$SV \rightarrow V \ SN$

...

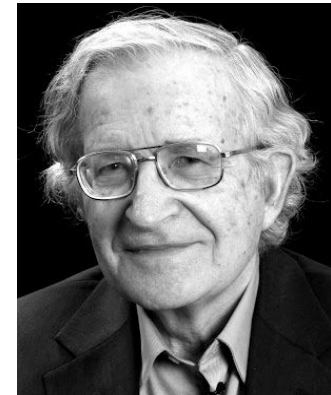
*Los perros comen un hueso*

# <> isn't actually a valid comparison operator in Python. It's here for the # sake of a `__future__` import described in PEP 401 (which really works :-)   
comp\_op: '<|>'=='|>'>='|<='|<>'!='|in'|not' 'in'|is'|is' 'not'   
star\_expr: '\*' expr   
expr: xor\_expr ('|' xor\_expr)\*   
xor\_expr: and\_expr ('^' and\_expr)\*   
and\_expr: shift\_expr ('&' shift\_expr)\*   
shift\_expr: arith\_expr (('<<'>>') arith\_expr)\*   
arith\_expr: term (('+'|'-') term)\*   
term: factor (('\*'|'/'|'%'|'//') factor)\*   
factor: ('+'|'-'|'~') factor | power

# Introducción

## ¿Qué es un lenguaje?

*Conjunto finito o infinito de oraciones, cada una de las cuales posee una extensión finita, construida a partir de un conjunto finito de elementos. (Chomsky 1957)*



*El lenguaje natural es **discreto** en cuanto a sus unidades, pero **infinito** en cuanto a las combinaciones que pueden hacerse de esas unidades*

# Introducción

---

¿Qué es ser natural?

*Adjetivo que refiere a la naturaleza*

Lenguaje Natural

- *es la lengua o idioma hablado o escrito por humanos para propósitos generales de comunicación*
- *complejo y diverso*

# Introducción

---

¿Qué es el PLN?

# Introducción

---

## ¿Qué es el PLN?

*El Procesamiento de Lenguaje Natural (PLN) es una subdisciplina de la **Inteligencia Artificial** que intenta resolver con computadoras tareas vinculadas al lenguaje humano, permitiendo la comunicación entre el humano y la computadora a través del lenguaje natural o resolviendo diferentes tareas que implican algún tipo de procesamiento de texto o habla.*  
(Jurafsky & Martin, 2008)

# Introducción

---

## ¿Qué es el PLN?

- *subdisciplina de la Inteligencia Artificial*
- *conjunto de métodos y técnicas eficientes desde un punto de vista computacional para la **comprensión / análisis** y **generación** de lenguaje natural*

# Introducción

---

¿PLN = Lingüística Computacional?

# Introducción

---

## Lingüística Computacional:

- campo multidisciplinario de la **lingüística** y de la **computación**.
- en constante evolución
- desarrollo de formalismos descriptivos del funcionamiento del lenguaje natural, que puedan ser transformados en programas ejecutables
- involucra a lingüistas, informáticos, lógicos, psicólogos cognitivos...

*El PLN puede verse como la rama ingenieril de la LC*



# Introducción

---

## Lingüística Computacional

busca fundamentación teórica en los modelos abstractos y métodos computacionales

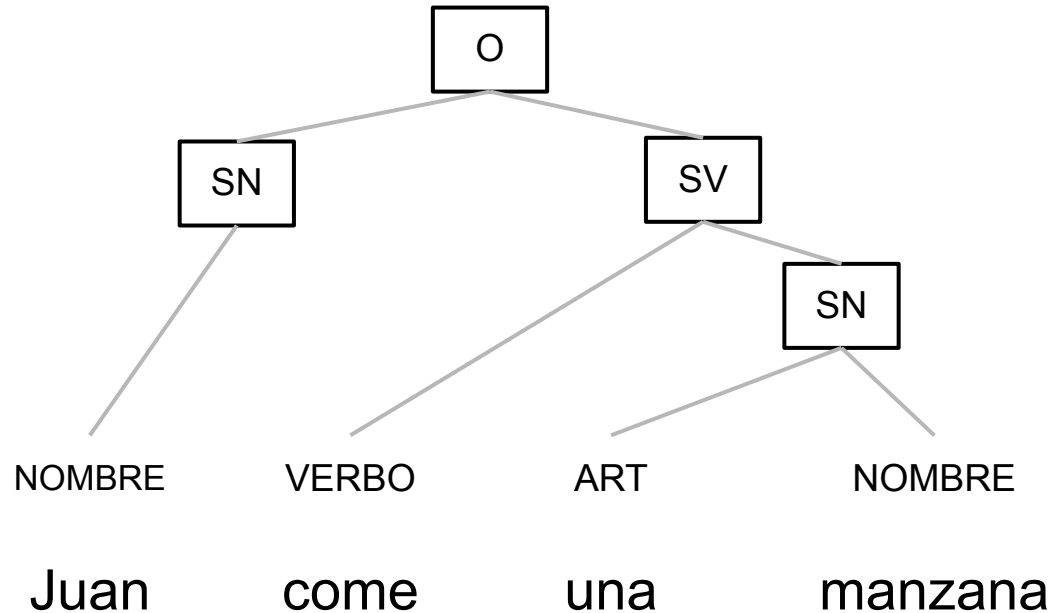
2 ramas → teórica: desarrollar teorías lingüísticas computables, compresión formal de los procesos y basa su trabajo en la ciencia cognitiva (psicolingüística computacional)

→ aplicada: orientación más tecnológica, centrada en aspectos prácticos (ingeniería lingüística)

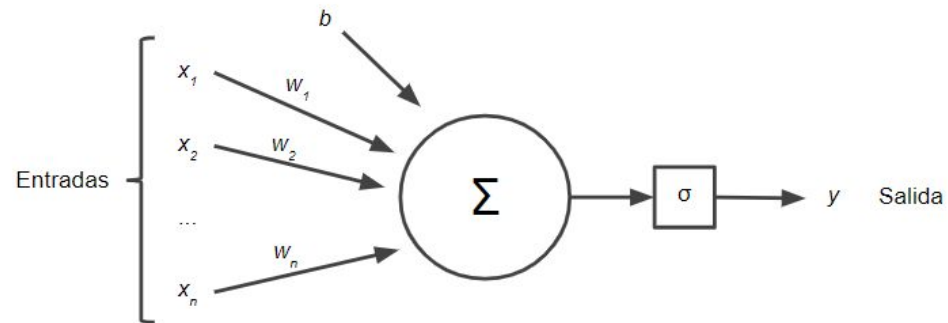
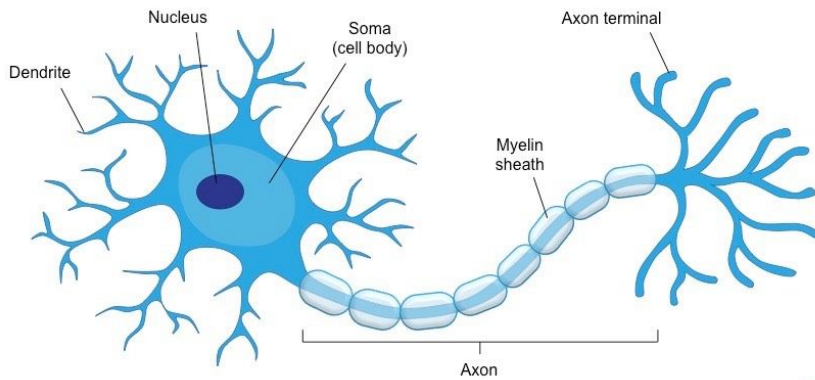
que se complementan

# Introducción

- Parsing
- POS Tagging
- Morfología
- y mucho más ...



# Introducción



$$y = \sigma\left(\sum_i x_i w_i + b\right)$$

Neurona de  
McCulloch-Pitts,  
1943

Cuéntame la historia de las redes neuronales desde su inicio hasta hoy.



La historia de las redes neuronales es fascinante y abarca varias décadas de investigación y avances tecnológicos. Aquí tienes un resumen de los momentos clave en el desarrollo de las redes neuronales desde su inicio hasta el presente:

### 1. Década de 1940-1950: Neurona Artificial

- El concepto de una "neurona artificial" fue introducido por Warren McCulloch y Walter Pitts en 1943. Propusieron una simplificación matemática de las neuronas biológicas y cómo podrían interactuar para realizar cálculos lógicos.

### 2. Década de 1950-1960: Perceptrón

El perceptrón fue desarrollado por Frank Rosenblatt en 1957. Era un modelo de red neuronal simple que podía aprender a reconocer patrones visuales.

Ahora escríbelo como si fuera una payada.



En la historia de las redes neuronales me voy a aventurar,  
Una payada les traigo, para todos narrar.  
Comenzó en los cuarenta, con McCulloch y Pitts,  
Una neurona artificial, ¡vaya ingenio sin fin!

Llegó el cincuenta, y el perceptrón apareció,  
Rosenblatt lo creó, en el campo se impuso.  
Clasificando patrones, brillaba con pasión,  
Pero estancado quedó, retos en la misión.

Los sesenta y setenta trajeron su pesar,  
Mindelay Report, cuestionaron sin cesar.

# Introducción

## 2 aspectos clave:

1. comprensión
2. generación

*Vi a un conejo con un telescopio*

*Vi [ a un conejo ] [ con un telescopio ]*

*Vi [ a un conejo [ con un telescopio ] ] X*

*Vi a un conejo con una zanahoria*

*Vi [ a un conejo [ con una zanahoria ] ]*

*Vi [ a un conejo ] [ con una zanahoria ] X*

# Introducción

## 2 aspectos clave:

1. comprensión
2. generación

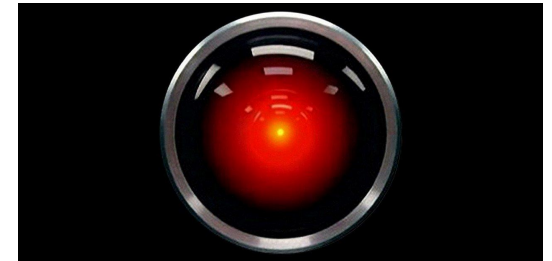
HAL 9000  
(1968)



# Introducción

## 2 aspectos clave:

1. comprensión
2. generación



- *Dave:* Open the pod bay doors, HAL.
- *HAL:* I'm sorry Dave. I'm afraid I can't do that.
- *Dave:* Abre las compuertas, HAL.
- *HAL:* Lo siento, Dave. Me temo que no puedo hacerlo.

# HAL 9000

---

## Habilidades de HAL

- comprensión de humanos vía:
  - reconocimiento del habla
  - comprensión de lenguaje natural
- comunicación con humanos vía:
  - generación de lenguaje natural
  - síntesis del habla
- pero también...
  - juega al ajedrez
  - toma decisiones
  - ...



# HAL 9000

---

Señal sonora  Secuencia de palabras

Reconocer/Generar

## Conocimientos de:

- **Fonética:** naturaleza física de los sonidos
- **Fonología:** cómo los sonidos funcionan en una lengua

# HAL 9000

---

- Debe saber, por ejemplo:
  - que los sustantivos tienen género y número:
    - Perr-o, Perr-o-s, Perr-a, Perr-a-s.
    - Pero:
      - Cas-a no es el femenino de Cas-o.
      - Ni Luz-s ni Luz-es son plurales de Luz.

# HAL 9000

- Debe saber, por ejemplo:
  - que los sustantivos tienen género y número:
    - Perr-o, Perr-o-s, Perr-a, Perr-a-s.
    - Pero:
      - Cas-a no es el femenino de Cas-o.
      - Ni Luz-s ni Luz-es son plurales de Luz.
  - que se pueden formar palabras agregando prefijos y sufijos a palabras existentes:
    - in-creíble (*in-* denota negación)
    - calmada-mente (*-mente* transforma adjetivo en adverbio)
- Conocimientos de **Morfología**: estudio de la estructura interna de las palabras

# HAL 9000

- Debe conocer el orden correcto en el que las palabras deben decirse para que la respuesta tenga sentido.

- HAL dice: *Lo siento, Dave. Me temo que no puedo hacerlo.*

Incluso podría decir: *Dave, lo siento. Que no puedo hacerlo me temo.*

Pero ... (\*) *Lo puedo Dave siento que no temo me hacerlo.*

- Conocimientos de **Sintaxis**: estudio de la estructuración (orden y agrupamiento) de las palabras en unidades mayores.

# HAL 9000

---

- La sintaxis no es suficiente:
  - Abre las compuertas, HAL. (*VC + ART + SUST + SP + SUST*)
  - Saca las persianas, HAL.
  - Baja los dados, HAL.
- Es necesario comprender el **significado** de lo que Dave está diciendo:
  - significado de cada palabra: **Semántica Léxica**
  - significado de la combinación de palabras para obtener: **Semántica Composicional**.

# HAL 9000

- Adicionalmente, HAL presenta una utilización “educada” del lenguaje:

*Lo siento, Dave. Me temo que no puedo hacerlo.*

- Significa en realidad:
  - (1) no lo siente
  - (2) puede abrir las compuertas
- Conocimientos de:
  - **Discurso:** estudio de las unidades mayores a la oración.
  - **Pragmática:** estudio del modo en el que el contexto influye en la interpretación del significado. Cómo el lenguaje se utiliza para ciertos fines.

# Etapas clásicas en PLN

---

- ***Fonética y Fonología***: estudio de los sonidos lingüísticos (usados para la comunicación humana)
- ***Morfología***: estudio de la estructura interna de las palabras
- ***Sintaxis***: estudio de la estructuración (orden y agrupamiento) de las palabras en unidades mayores
- ***Semántica***: estudio del significado
- ***Discurso***: estudio de las unidades mayores a la oración
- ***Pragmática***: estudio en la relación entre la lengua, los hablantes y el contexto de la comunicación

**Un poco de historia...**



# Década del '50

## Traducción Automática

- En particular del Ruso al Inglés
  - Guerra Fría
  - Experimento *Georgetown* (1954)
    - tenía 6 reglas gramaticales
    - 250 palabras
    - + de 60 oraciones
  - En tres años la traducción estará resuelta....
  - Informe ALPAC (1966)



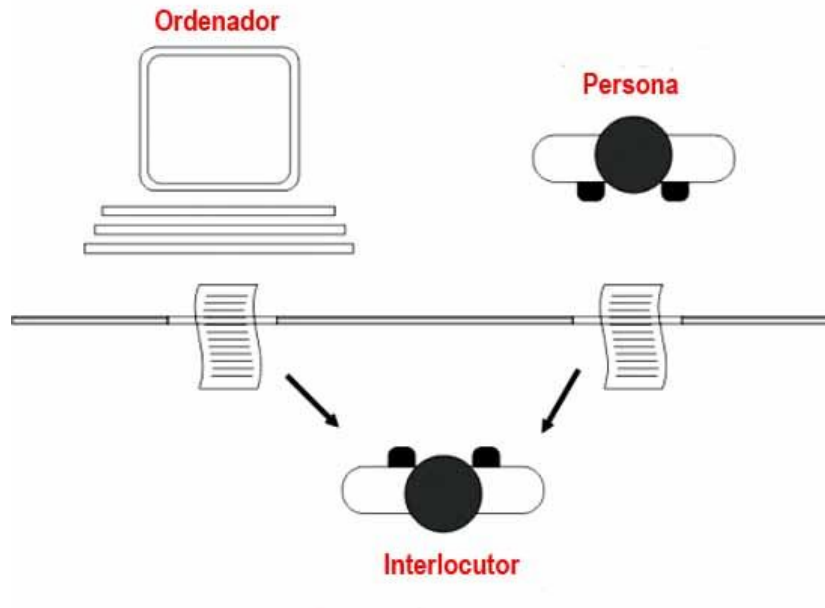
**Warren Weaver**

# Década del '50

## Alan Turing: "Computing Machinery and Intelligence"

(I propose to consider the question, "*Can machines think?*")

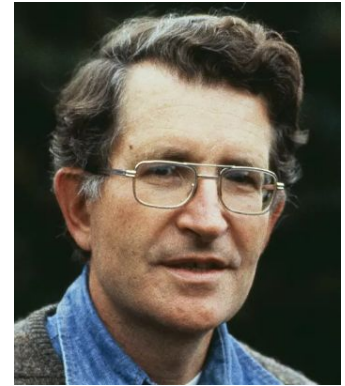
### *Test de Turing*



# Década del '50

## Noam Chomsky: "Syntactic Structures" (1957)

La gramática es un sistema combinatorio discreto que permite construir infinitas frases a partir de un número finito de elementos mediante un conjunto de reglas que pueden formalizarse



*Colorless green ideas sleep furiously* – 1955

(Las ideas verdes incoloras duermen furiosamente)

*"probabilistic models give no insight into the basic problems of syntactic structure"*

Gramática Universal → *todos los idiomas que usamos los seres humanos tienen unas características o principios comunes en su propia estructura*

# Década del '60

- **Cocke – Kasami – Younger (1965)**

(*Parser GLC*)



- **Earley (1968)**

(*Parser GLC*)



- **Kay (1967)**

(*Chart Parser*)



El foco está puesto en el Análisis Sintáctico

# Década del '70



**Gerard Salton**



**Karen Spärck Jones**

- Recuperación de información: obtener documentos más relevantes dada una consulta
- Modelo Vectorial (1968)
- ***TF-IDF***: medida de importancia de un documento

# Década del '70

**Richard Montague**



**“English as a Formal Language”**

- Fue pionero en el enfoque lógico de la semántica del lenguaje natural
- La gramática de Montague está basada en lógica formal:
  - alto orden
  - lambda cálculo
  - gramática categorial de unificación y combinatoria

# Década del '70

---

**Alain Colmerauer**



- Lenguaje de Programación basado en Lógica (PROLOG)
- Pensado originalmente para Procesamiento de Lenguaje Natural

# Década del '70

## Sistemas Expertos

- **Expresiones Regulares**

Paciente: Men are all alike

**ELIZA: IN WHAT WAY**

Paciente: They're always bugging us about something or other

**ELIZA: CAN YOU THINK OF A SPECIFIC EXAMPLE**

Paciente: Well, my boyfriend made me come here

**ELIZA: YOUR BOYFRIEND MADE YOU COME HERE**

Paciente: He says I'm depressed much of the time

**ELIZA: I AM SORRY TO HEAR YOU ARE DEPRESSED**

(Weizenbaum1966)

s/. \* I AM (depressed|sad) .\*/ I AM SORRY TO HEAR YOU ARE \1 /

s/. \* MY (mother | father | boyfriend | girlfriend | friend) MADE ME COME HERE .\*/  
WHY DO YOU THINK YOUR \1 YOU COME HERE /

s/. \* all .\*/ IN WHAT WAY /

s/. \* always .\*/ CAN YOU THINK OF A SPECIFIC EXAMPLE /

- **Razonamiento basado en reglas**

Enfoque basado en reglas "si-entonces" para representar el conocimiento y permitir al sistema inferir nuevas conclusiones a partir de la información dada



# Década del '80

---

- Se comienzan a construir sistemas de laboratorio con reglas hechas a mano
- Sistemas de interrogación de BDs relacionales
- Mayor poder de cómputo

***Surgen los sistemas basados en aprendizaje automático***

# Década del '90

## Frederik Jelinek



- Modelos IBM de traducción automática y reconocimiento de voz
- A partir del corpus se infieren las reglas
- Mayor uso de algoritmos no supervisados y semi-supervisados
- EL PLN se mueve hacia modelos estadísticos y métodos basados en datos
- Se introduce el concepto de *perplejidad* (capacidad de predecir la siguiente palabra)

# Los 2000...

## Vladimir Vapnik



- Primeros modelos de lenguaje capaces de predecir siguiente palabra con técnicas estadísticas
- Pero... poca capacidad “predictiva”
- Support Vector Machines
- AT&T, NEC, Facebook IA Research

# A partir del 2000...

## Geoffrey Hinton, Yoshua Bengio, Yann LeCun

(The Canadian Mafia)

- Red FeedForward
- Deep Learning
- Word Embeddings



---

### Attention Is All You Need

---

Ashish Vaswani\*  
Google Brain  
avaswani@google.com

Noam Shazeer\*  
Google Brain  
noam@google.com

Niki Parmar\*  
Google Research  
nikip@google.com

Jakob Uszkoreit\*  
Google Research  
usz@google.com

Llion Jones\*  
Google Research  
llion@google.com

Aidan N. Gomez\* †  
University of Toronto  
aidan@cs.toronto.edu

Lukasz Kaiser\*  
Google Brain  
lukaszkaizer@google.com

Illia Polosukhin\* ‡  
illia.polosukhin@gmail.com

# La investigación en PLN hoy

---

- Las compañías de Internet se llevan a los grandes investigadores (casi todos)
- Se publica primero en Arxiv, luego en conferencias, luego en (pocos) journals
- Mucho es *Open Source*
- Casi completamente *data-driven*
- Las redes neuronales son las estrellas al haber cada vez más datos
  - Hoy: Large Language Models (LLM)
- Pero ojo....los métodos simbólicos aún están vigentes

# Algunas aplicaciones ...

- Traducción automática
- Resumen automático
- Recuperación de información
- Extracción de información
- Análisis de sentimientos
- Respuestas a preguntas
- Generación de texto

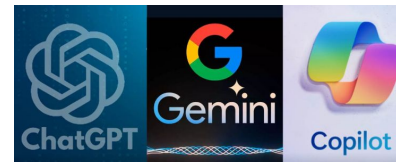
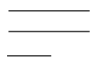
receta de  
merengue  
italiano → 意大利蛋白  
酥皮食譜

Google

Busca en Google o escribe una URL



+



- Proyectos Grupo PLN (<http://www.fing.edu.uy/inco/grupos/pln>)
- ...

# Lenguajes

---

## ➤ Formales

- Definidos por reglas pre-establecidas

## ➤ Naturales

- Evolucionan con el tiempo
- Utilizados para la comunicación humana
- Las reglas “se desarrollan” después que sucede el hecho

¿Qué tiene el lenguaje  
natural que no tienen los  
lenguajes formales?



# Ambigüedad



The image shows a screenshot of the SER website. The header features the SER logo in the center, a menu icon on the left, and a language selection dropdown on the right set to 'Cadena SER'. Below the header is a navigation bar with links for INICIO, DEPORTES, HUMOR, OCIO Y CULTURA, OPINIÓN, PROGRAMAS, and PODCASTS (which is highlighted). To the right of the navigation bar, it shows the date '05 ago 2025', the time 'Actualizado 16:21', an email icon, and a notification icon with the number '8'. Below the navigation bar, there is a breadcrumb trail: 'A vivir que son dos días' (highlighted in yellow), 'Prensa Ibérica', 'Sociedad | Actualidad', and 'Prensa Ibérica'. The main headline reads: 'Fallece por segundo día consecutivo una mujer de 103 años'.

Menu icon

SER

Selecciona tu emisora

Cadena SER

INICIO DEPORTES HUMOR OCIO Y CULTURA OPINIÓN PROGRAMAS **PODCASTS**

05 ago 2025 Actualizado 16:21

A vivir que son dos días | Prensa Ibérica | Sociedad | Actualidad | Prensa Ibérica

## Fallece por segundo día consecutivo una mujer de 103 años

# Ambigüedad

---



# Ambigüedad



# Fuentes de ambigüedad

---

- Ambiguo: que admite distintas interpretaciones
- Homonimia: dos palabras con misma forma que tienen distinto significado
  - Homografía: capital, banco
  - Homofonía: Ola/Hola, As/Has, Cocer/Coser
- Polisemia: una palabra con múltiples significados pero que de alguna manera “tienen que ver”

*El hombre **desciende** del mono y el mono **desciende** del árbol*

# Ambigüedad fonética

---

## Ejemplos de calambures:

- Ató dos palos. / A todos palos.
- Yo loco, loco, y ella loquita. / Yo lo coloco y ella lo quita.
- Mi madre estaba riendo. / Mi madre está barriendo.
- El dulce lamentar de los pastores. / El dulce lamen tarde los pastores. (Garcilaso de la Vega)
- ***Entre el clavel blanco y la rosa roja, su majestad escoja.***  
(Quevedo)

# Ambigüedad a nivel morfológico

---

*Nosotros **plantamos** papas.*

¿El verbo plantar está conjugado en pasado o en presente?

# Ambigüedad sintáctica

---

Pedro vio a Juan con el telescopio.

- a) Pedro vio [a Juan] con el telescopio.
- b) Pedro vio [a Juan con el telescopio].

Los hombres y las mujeres que hayan cumplido 60 años pueden solicitar una pensión.

- a) [Los hombres y las mujeres que hayan cumplido 60 años] pueden solicitar una pensión.
- b) [Los hombres] y [las mujeres que hayan cumplido 60 años] pueden solicitar una pensión.

# Ambigüedad semántica

---

Cuantificadores:

Todos los hombres aman a una mujer.

Todos los estudiantes leyeron un libro.

a) Es la misma mujer/libro para todos.

b) Para cada hombre/estudiante existe una mujer/un libro



# Ambigüedad semántica

---

La perra de mi vecina me ladró.

a) mi vecina realmente tiene una perra

b) no tengo un buen trato con mi vecina

# Ambigüedad a nivel pragmático

---

Llego a las ocho. Esperame.

- ¿A qué hora llegarás?
- Llego a las ocho. Esperame. (**Previsión**)
- Nunca llegás en hora.
- Llego a las ocho. Esperame (**Promesa**)
- Eso me lo vas a tener que decir cara a cara.
- Llego a las ocho. Esperame. (**Amenaza**)

# Ambigüedad a nivel de discurso

---

Tomé el alfajor del escritorio y lo comí.

a) Tomé el alfajor que estaba en el escritorio y comí el alfajor.

b) Tomé el alfajor que estaba en el escritorio y comí el escritorio.

# ¿Se puede resolver la ambigüedad?

---

Juan mató al carpincho con una escopeta.

- No puede ser el carpincho quien lleve la escopeta.

Puse la camisa en la lavadora y la lavé.

- Las lavadoras lavan. La ropa se lava.

*Se requiere conocimiento del mundo*

# El PLN es difícil porque:

---

- Alta ambigüedad en todos los niveles
- Complejo y sutil
- Jergas, modismos
- Ironía, sarcasmo, doble sentido
- Involucra razonar y tener conocimiento acerca del mundo
- ...