

Procesamiento de Lenguaje Natural

Tópicos Avanzados en Analítica

Maestría en Analítica para la Inteligencia de Negocios

Sergio Alberto Mora Pardo - H2 2023

Procesamiento de Lenguaje Natural (NLP)

Clase 1 - Introducción y aplicaciones

Introducción:

1. ¿Qué es procesamiento de lenguaje Natural?
2. ¿Por qué el entendimiento del lenguaje natural es una tarea compleja?

Tareas del procesamiento de lenguaje natural.

Evolución: de reglas a aprendizaje de maquina.

Introducción

**“A language is not just words. It’s a culture, a tradition,
a unification of a community,
a whole history that creates what a community is.
It’s all embodied in a language.”**

Noam Chomsky

“The art of creating machines that perform functions that require intelligence when performed by people”

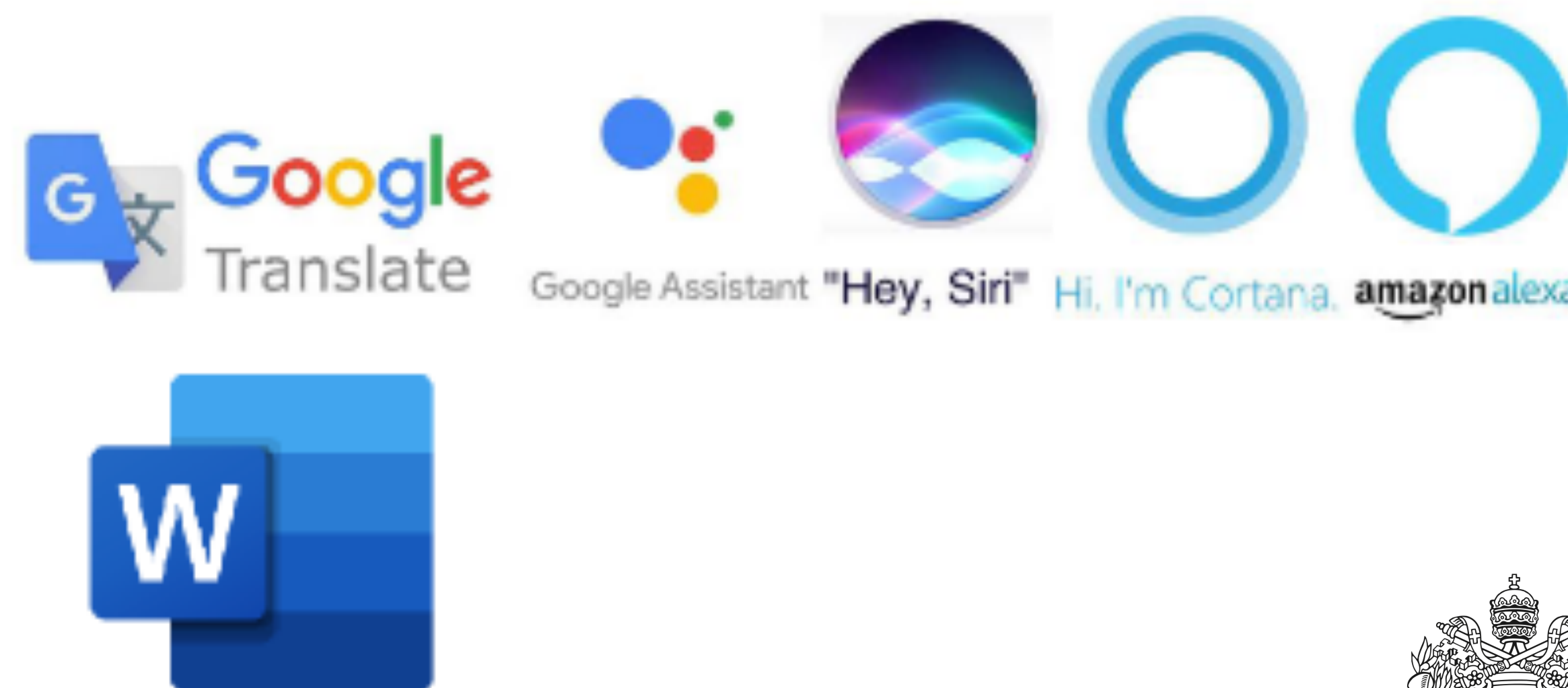
Raymond Kurzweil - 1990

Procesamiento de Lenguaje Natural

El procesamiento del lenguaje natural es el área que busca dotar a las computadoras con la capacidad de comprender el lenguaje natural del ser humano (escrito y hablado)



By:Seobility





2011 - IBM Watson vence a Ken Jennings y Brad Rutter en el show de TV Jeopardy.

Fuente: IBM research

<https://www.techrepublic.com/article/ibm-watson-the-inside-story-of-how-the-jeopardy-winning-supercomputer-was-born-and-what-it-wants-to-do-next/>



Buscadores

Recuperación de información de corpus gigantescos...

nature

Explore content ▾ About the journal ▾ Publish with us ▾ Subscribe

nature > news feature > article

NEWS FEATURE | 03 March 2021

Robo-writers: the rise and risks of language-generating AI

A remarkable AI can write like humans – but with no understanding of what it's saying.

Fuente: <https://www.nature.com/articles/d41586-021-00530-0>

Generación de texto usando modelos de lenguaje...

¿Del conjunto de palabras de mi vocabulario cuál es la palabra más probable?

Buenos Aires es la capital de _____

Bogotá

Universidad

Argentina

Generación de texto usando modelos de lenguaje...

¿Del conjunto de palabras de mi vocabulario cuál es la palabra más probable?

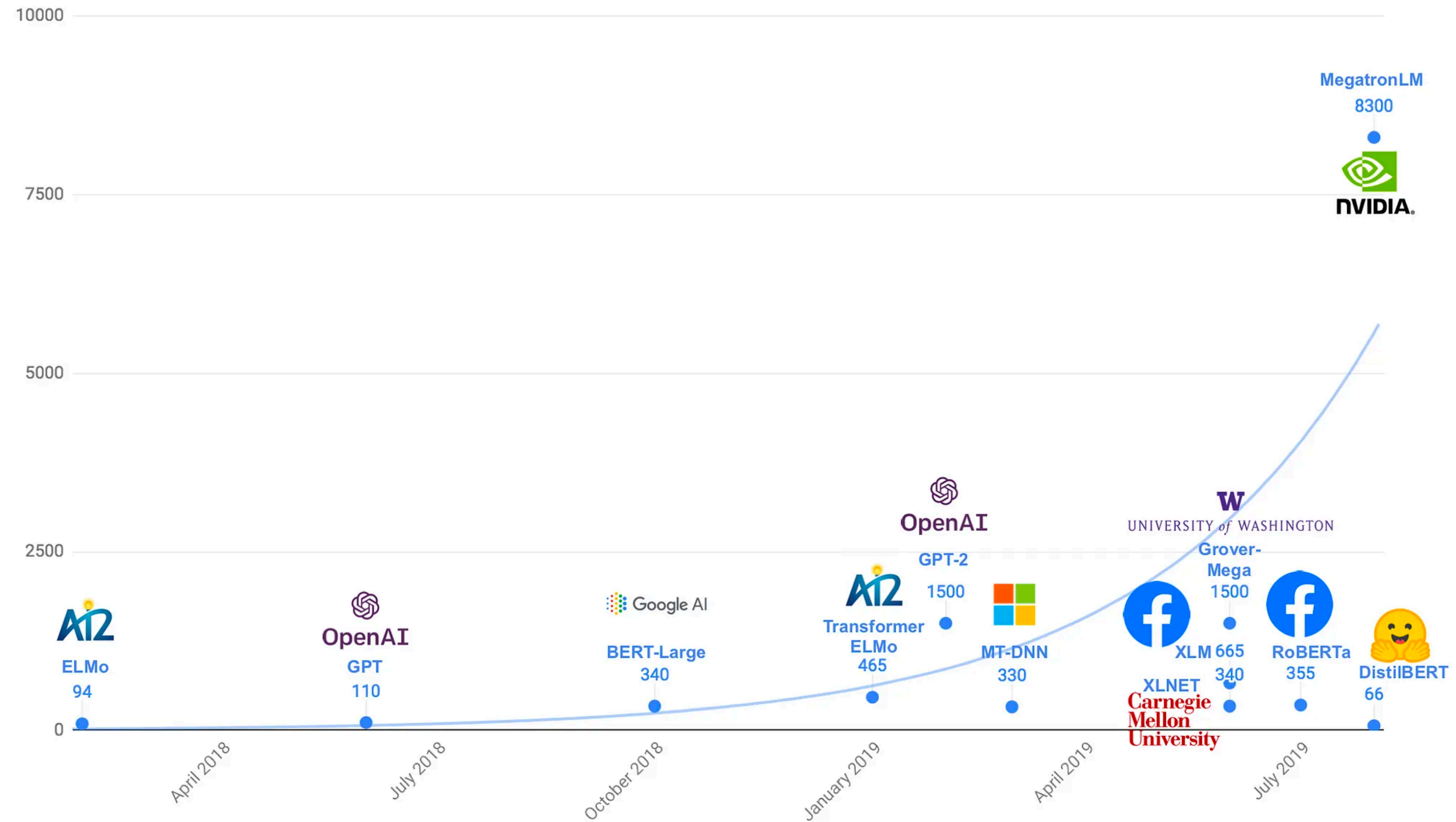


Chat GPT

Chat Generative Pre-Trained Transformer

Generación de texto usando modelos de lenguaje...

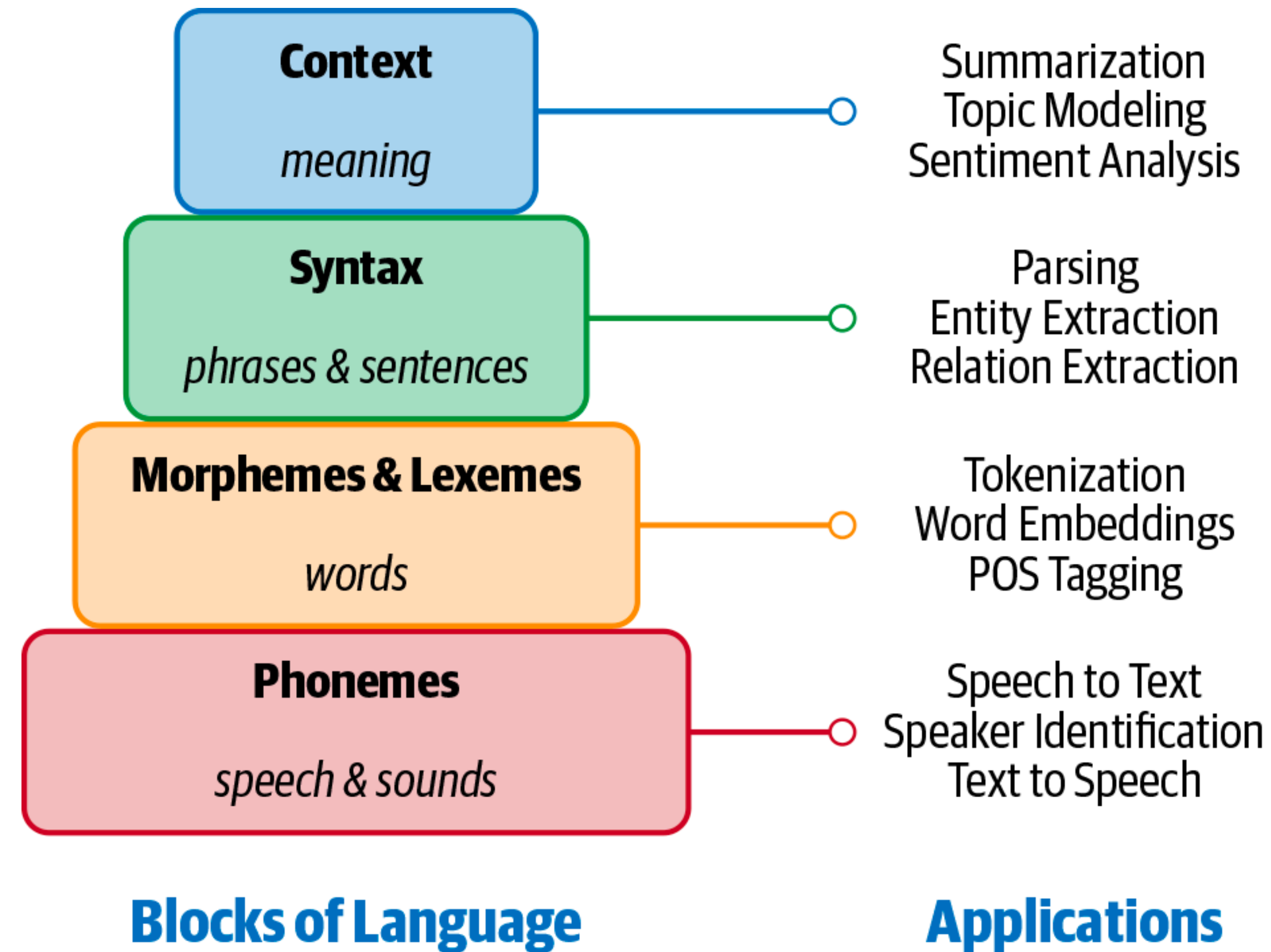
Generación de texto usando modelos de lenguaje...



Here's are some of the latest large models and their **size in millions of parameters**.

Linguistics

Fonemas, morfemas y lexemas, sintaxis y contexto.



Some concepts of linguistics

Sonido del lenguaje (fonemas)

Consonant phonemes, with sample words		Vowel phonemes, with sample words	
1. /b/ - bat	13. /s/ - sun	1. /a/ - ant	13. /oi/ - coin
2. /k/ - cat	14. /t/ - tap	2. /e/ - egg	14. /ar/ - farm
3. /d/ - dog	15. /v/ - van	3. /i/ - in	15. /or/ - for
4. /f/ - fan	16. /w/ - wig	4. /o/ - on	16. /ur/ - hurt
5. /g/ - go	17. /y/ - yes	5. /u/ - up	17. /air/ - fair
6. /h/ - hen	18. /z/ - zip	6. /ai/ - rain	18. /ear/ - dear
7. /j/ - jet	19. /sh/ - shop	7. /ee/ - feet	19. /ure/ ⁴ - sure
8. /l/ - leg	20. /ch/ - chip	8. /igh/ - night	20. /ə/ - corner (the 'schwa' - an unstressed vowel sound which is close to /u/)
9. /m/ - map	21. /th/ - thin	9. /oa/ - boat	
10. /n/ - net	22. / th / - then	10. / oo / - boot	
11. /p/ - pen	23. /ng/ - ring	11. /oo/ - look	
12. /r/ - rat	24. /zh/ ³ - vision	12. /ow/ - cow	

Some concepts of linguistics

Lexemas y morfemas

unbreakable
un + break + able

cats
cat + s

tumbling
tumble + ing

unreliability
un + rely + able + ity

Lexema

CULP-

culpa

culpar

culpable

Morfema

Inexplicablemente

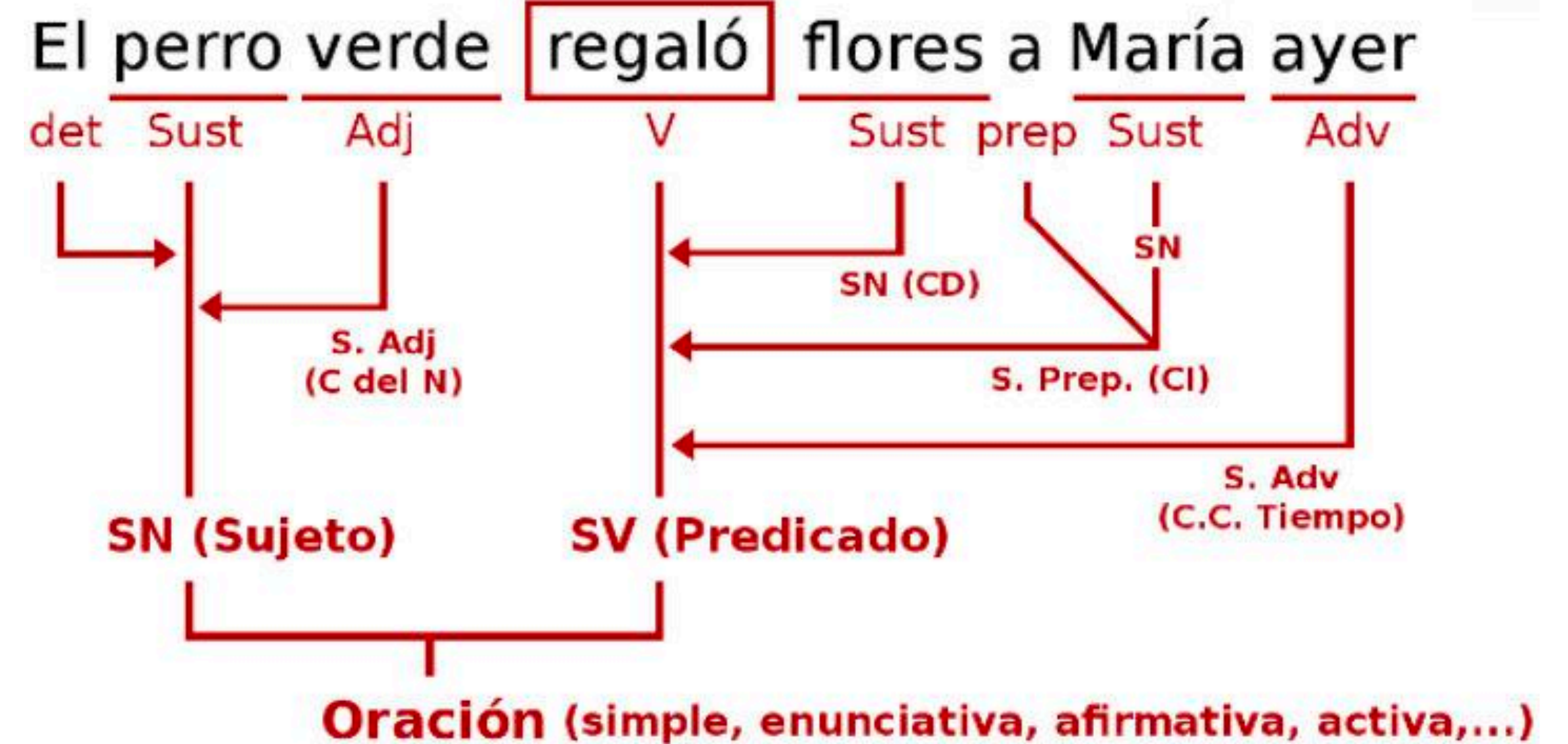
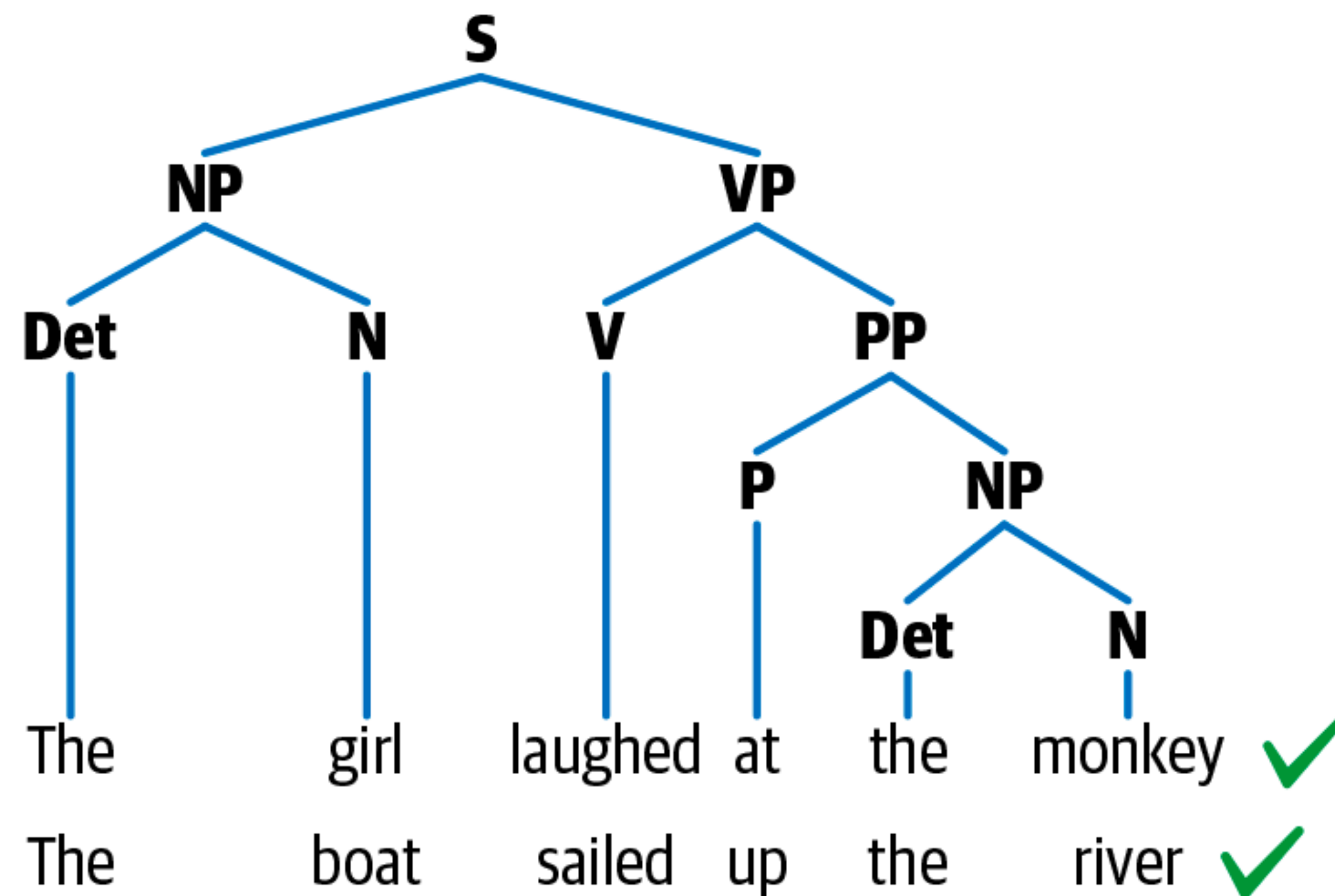
-in

-mente

El lexema es la raíz de una palabra que lleva su significado principal, mientras que el morfema es la unidad más pequeña de significado en una palabra y puede ser un lexema o un afijo que modifica ese significado.

Some concepts of linguistics

Sintaxis



Some concepts of linguistics

Contexto

Semántica

Pragmatica

Significado de las palabras y oraciones sin contexto
externo

Agrega el conocimiento del mundo y contexto
externo

Significado directo

Significado implícito

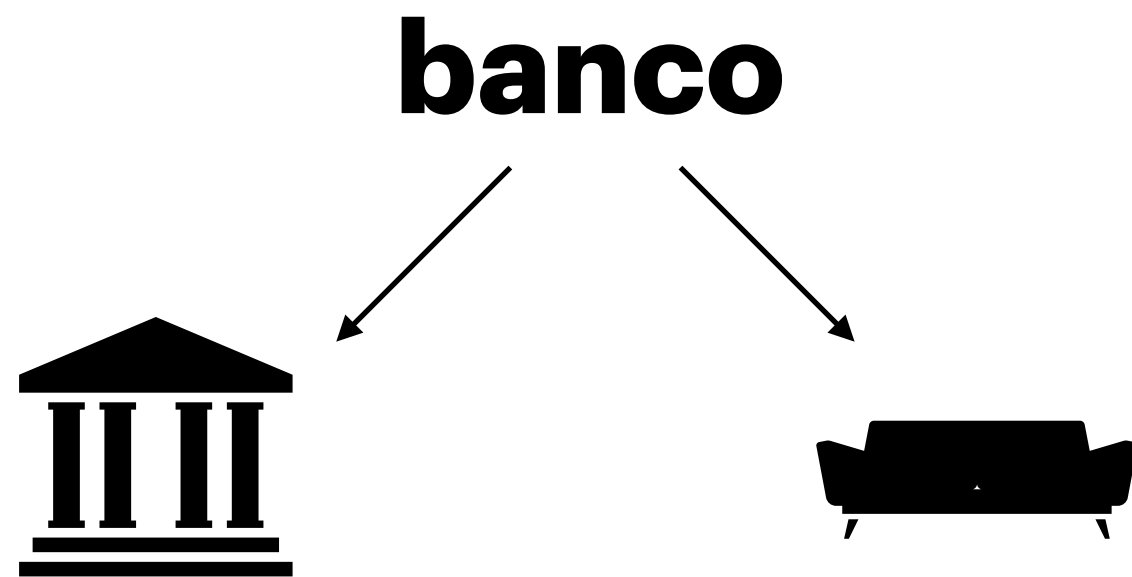
Ejemplos en NLP: detección del sarcasmo, Resumen, Modelación de temas

¿Porque el entendimiento del lenguaje es una tarea compleja?

Ambigüedad

Léxico-semántica (Polisemia)

La cita es en el **banco** donde les tomaron la fotografía



Sintáctica (Anfibología)

Vimos a Pedro tomando clase.

Sig. 1. Cuando tomábamos la clase, vimos a Pedro.

Sig. 2. Vimos a Pedro, que estaba tomando la clase.

Fonética

¿**Me diste** la ventana?

Sig. 1. Mediste- de medir

Sig. 2. Me diste- de dar

¿Porque el entendimiento del lenguaje es una tarea compleja?

Correferencia



La **lámpara** no cabe por la **puerta** porque **esta** es muy *angosta*.

¿A que hace referencia?

- A. La lámpara
- B. La puerta

¿Porque el entendimiento del lenguaje es una tarea compleja?

Correferencia



La **lámpara** no cabe por la **puerta** porque **esta** es muy *alta*.

¿A que hace referencia?

- A. La lámpara
- B. La puerta

¿Porque el entendimiento del lenguaje es una tarea compleja?

Sarcasmo/Ironía

- No estarás muy cansado, ¿no?
- Me gusta su perfume, ¿cuánto tiempo ha pasado sumergido en él?



flipkart.com Search for a product, category or brand

MOST HELPFUL REVIEWS (5 OF 43)
View all Sennheiser HD 800 Headphone reviews

★★★★★
Frateek Dhupey
12 December 13

must Buy!!!! just 90k.
After selling my left kidney i got the right product. Its worthy.
There are other organs also but i preferred Kidney.
Awsome Experience.

98% of 333 users found this review helpful.

Tienen que comprarlo!!! Solo 90k.

Después de vender mi riñón izquierdo, obtuve el producto correcto. Vale la pena.

También hay otros órganos, pero prefiero el riñón. Increíble experiencia.

Permalink Report Abuse Share

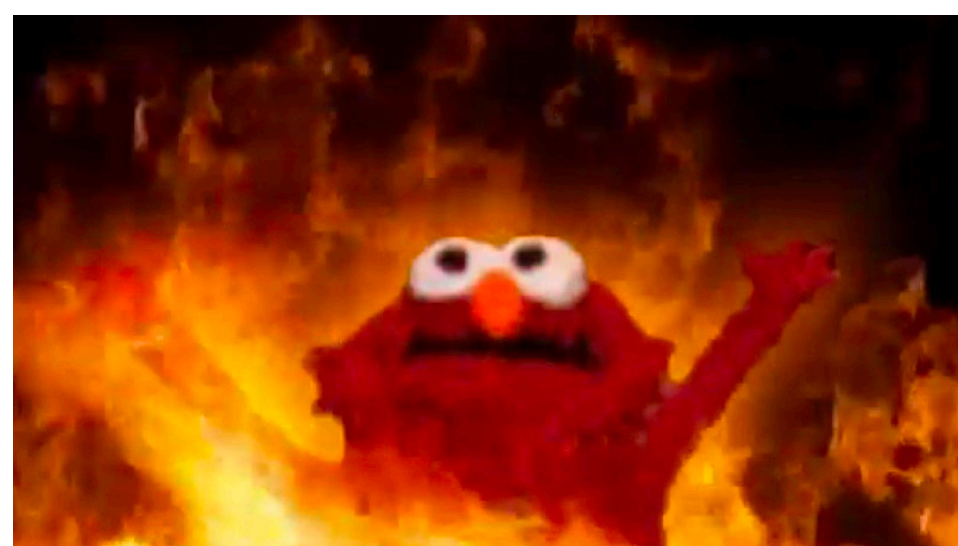
Fuente: <https://flipkart.com/>

¿Porque el entendimiento del lenguaje es una tarea compleja?

Otros

Lenguaje no estándar (redes sociales)

RT @ Pere4L: Guy PLZ d/I the letter
I've written 2 Jeff Gaspin, he's THE man
who can giv us #Heroes S5 <http://tinyurl.com/y9pcaj7> #Heroes100 :)



Modismos

Chanda, seba, guayabo,
camello

Neologismos

Textear, Hipertexto, clickear,
escanear

Nombre de entidades

Musica ligera fue grabado en...

Luisito comunica no ha dado
declaraciones...

Problemas abordados

Clasificación de texto

Ej. 1: Spam/No spam

La reunión quedó programa para el ... (**NO SPAM**)

Gana dinero fácil, invierte en P1pOpl3 ... (**SPAM**)

Ej. 2: Análisis de polaridad

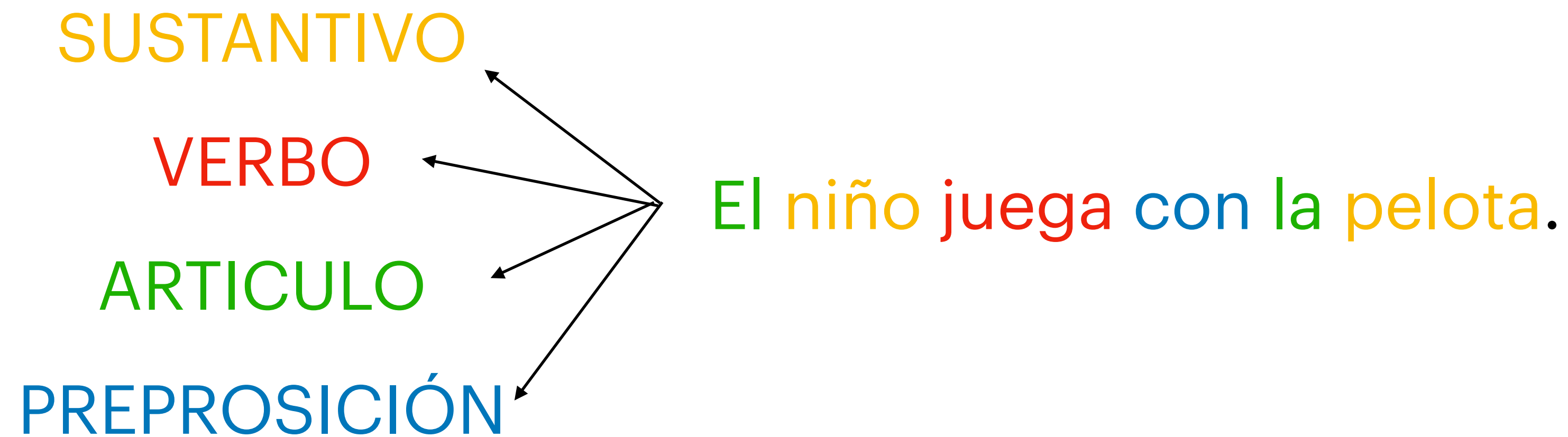
Mejor PC gamer de la historia (+)

La pantalla se rompió a la semana, no la compren. (-)

Problemas abordados

Identificación de partes de la oración (POS)

El niño juega con la pelota.



**Comprensión sintáctica
(Parsing)**

Corrección de texto

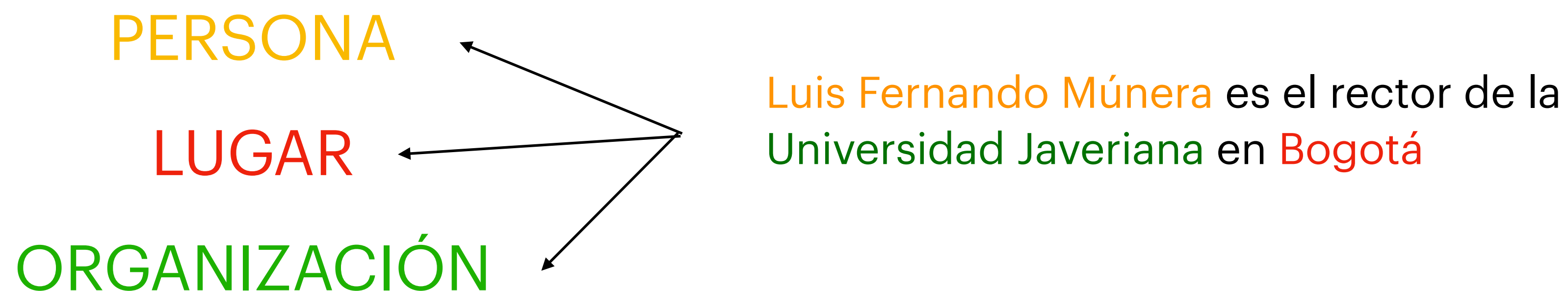
Problemas abordados

Identificación de entidades nombradas (NER)

Luis Fernando Múnera es el rector de la
universidad Javeriana en Bogotá

NER:

clasificar en el texto entidades,
personas, organizaciones,
lugares, y expresiones de texto.

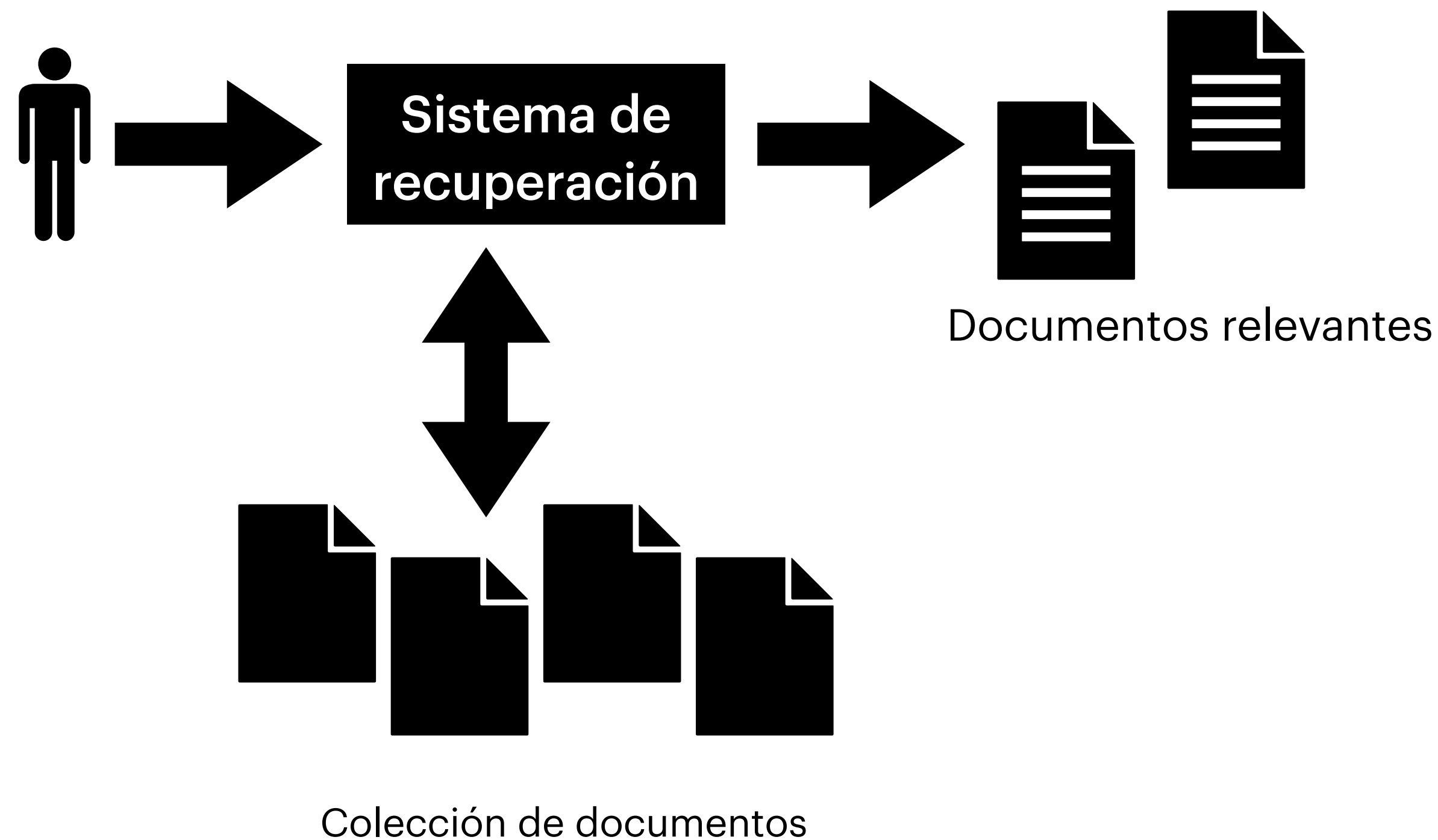


Problemas abordados

Recuperación de Información

Sistema de recuperación

Acceder y recuperar la información textual (documentos) más apropiada a en base a una consulta (query).



Problemas abordados

Correferencia

Def.

Relación entre dos o más elementos lingüísticos cuyo referente es común.

Ej.

*El profesor le dice al estudiante que **él** debe renunciar a sus malos hábitos de estudio. A pesar de esto sus malos hábitos **éste** ha logrado superar la moría de los exámenes.*

Comentario:

En el texto anterior, ¿las palabras **él** y **éste** hacen referencia al profesor o al estudiante?

En textos largos los pronombres sigue siendo un problema desafiante.

Problemas abordados

Desafíos

Preguntas a Respuestas.

Los sistemas de Q&A requieren un conocimiento “no genérico” que debe establecerse mediante análisis de intenciones o modelos de conocimiento.

Sistema de dialogo (chatbot)

Human: What is your favorite animal?

GPT-3: My favorite animal is a dog.

Human: Why?

GPT-3: Because dogs are loyal and friendly

Resúmenes Automáticos.

Sintetizar o reducir el texto preservando la idea o ideas más importantes.

NLP Tasks

Principales tareas que aparecen con frecuencia en proyectos de NLP

- Modelación de Lenguaje:
 - A. Esta es la tarea de predecir cuál será la siguiente palabra en una oración en función de la historia de las palabras anteriores.
 - B. El objetivo de esta tarea es aprender la probabilidad de que una secuencia de palabras aparezca en un idioma determinado.

NLP Tasks

Principales tareas que aparecen con frecuencia en proyectos de NLP

- Modelación de Lenguaje
- Clasificación de Texto
- Extracción de Información
- Recuperación de Información
- Agente Conversacional
- Resumen de Texto
- Question answering
- Machine translation
- Topic modeling

NLP Tasks

Principales tareas que aparecen con frecuencia en proyectos de NLP

