

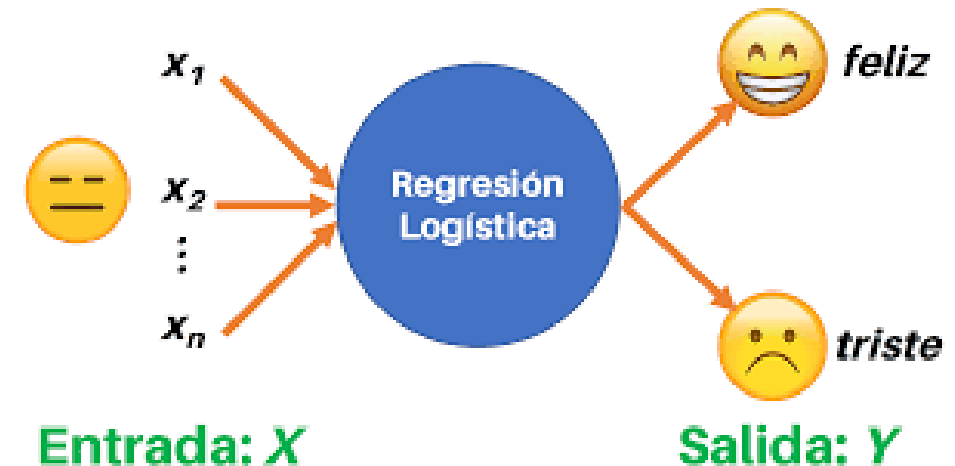
Modelos discriminativos vs generativos

Aplicaciones de Lenguaje Natural
Escuela Superior de Cómputo



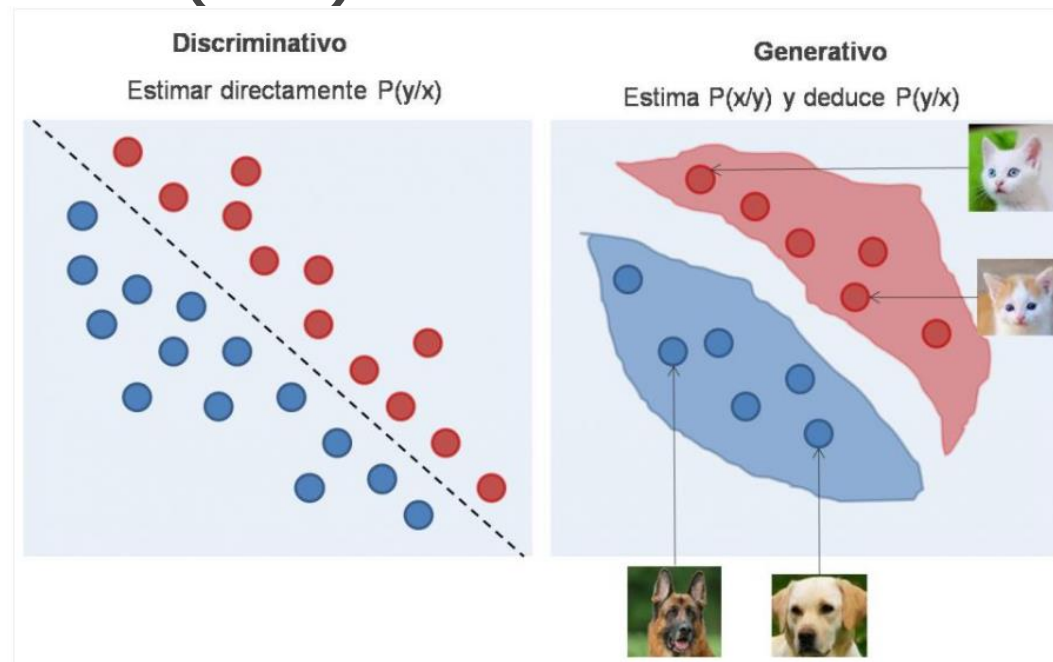
Algoritmos Discriminativos

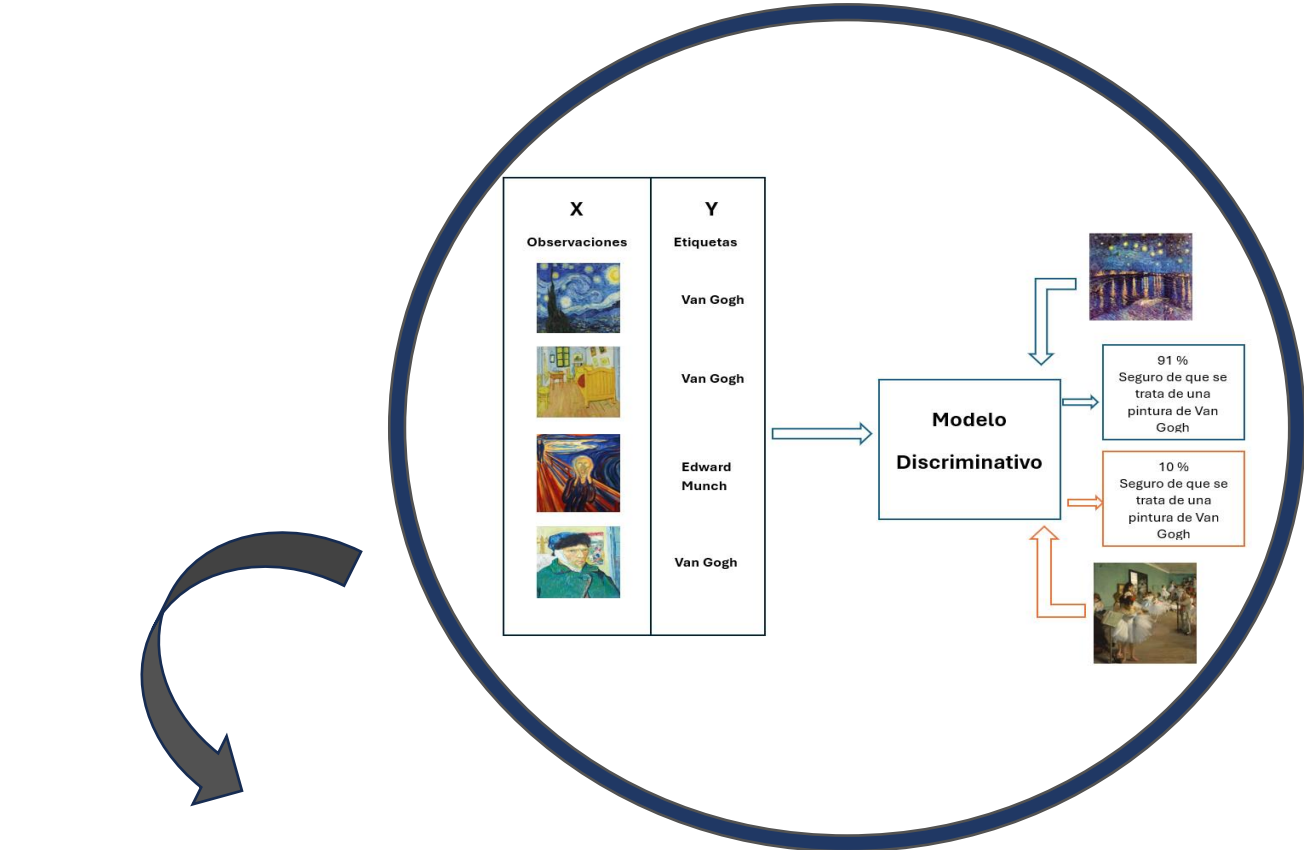
- Dadas las características de una instancia de datos, predicen una etiqueta o categoría a la que pertenecen esos datos.
 - Regresión logística
 - Redes neuronales de varias capas
- Se aprende un **mapeo o función directa desde las variables de entrada a las clases o variables objetivo.**
- Intentan clasificar los datos de entrada.



Algoritmo generativo

- Proporcionan un modelo de **cómo los datos se generan**
- Modela la distribución **$P(X/Y)$** y **$P(Y)$** .
- Ejemplo. Si la variable **Y** indica clases de animales, entonces la distribución **$P(X/Y)$** modela como se distribuyen las características de cada tipo de animal.

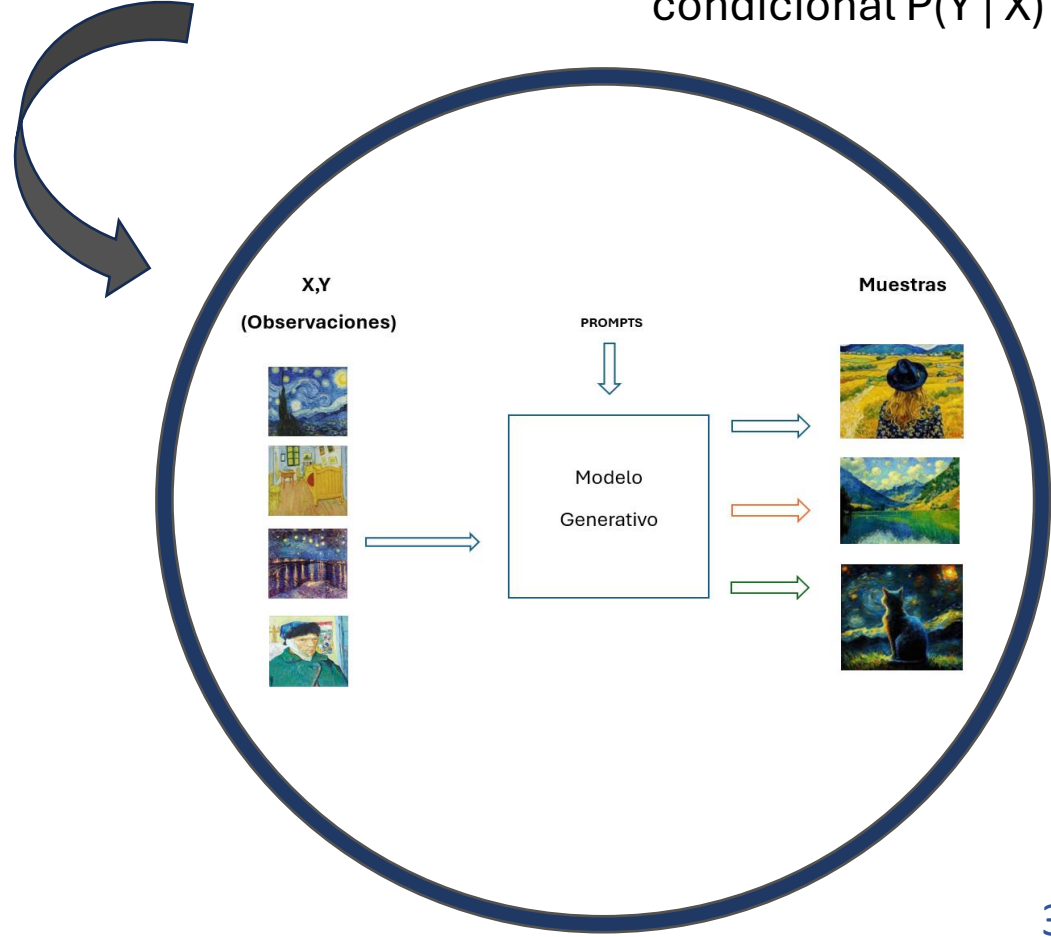




MODELOS DE IA DISCRIMINATIVOS

Predicen un valor Y dado un valor X, con base en el aprendizaje de la distribución de probabilidad condicional $P(Y | X)$,

MODELOS DE IA GENERATIVOS
Aprenden la distribución conjunta de probabilidad $P(X, Y)$ y luego predicen la probabilidad condicional $P(Y | X)$



INTRODUCCIÓN - Gen ai



El **ratoncito** comió **queso** porque estaba _____

El **ratoncito** comió **queso** porque estaba **hambriento**

El **ratoncito** comió **queso** porque estaba **delicioso**

ratoncito - - - - - estaba - - - - - **queso**

hambriento

ratoncito - - - - - estaba - - - - - **queso**

delicioso

ratoncito - - - - - estaba - - - - - **queso**

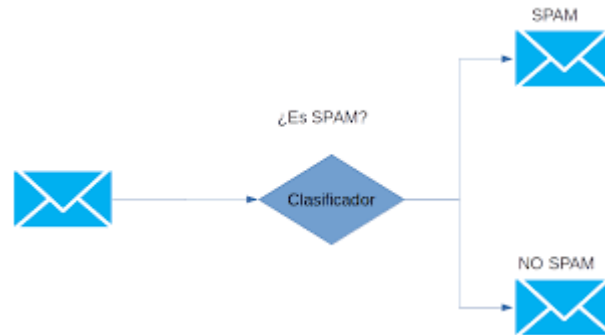
INTRODUCCIÓN

Tabla 1. Tareas para aplicación de IA Generativa

Tarea	Casos de uso	Plataformas y modelos	Técnicas y modelos
Creación de contenido	Creación de contenido textual, como artículos, historias, y código	GPT-4 (Open AI), LLAMA (Meta), Gemini (Google), BERT (Google) DistilBERT (HuggingFace), RoBERTa (Meta), LLaMA (Meta), BLOOM (BigScience)	Transformers, Procesamiento de lenguaje natural
Generación de imágenes y arte	Creación de imágenes y obras de arte, realistas o abstractas	DALL-E (OpenAI), Midjourney (Leap Motion), Imagine 3D (Luma AI)	Difusión, Visión por computadora, Procesamiento de lenguaje natural
Música y audio	Creación de música y efectos de sonido	Jukebox (OpenAI)	Transformers autorregresivos
Videojuegos y Simulaciones	Creación de escenarios, personajes y misiones	Scenario, Stable Diffusion (Stability AI)	Difusión

Ejemplo – Algoritmo Discriminativo

- Dadas todas las palabras en un correo electrónico, predecir si el mensaje es **spam** o no **not spam**
- Datos de entrada: La bolsa de palabras recopiladas del correo electrónico.
- "la probabilidad de que un correo electrónico sea spam dadas las palabras que contiene".



Ejemplo - Algoritmo Generativo

- Dadas todas las palabras en un correo electrónico, predecir si el mensaje es **spam** o no **not spam**
- Predecir características dadas una determinada etiqueta.
- “Suponiendo que este correo electrónico sea spam, ¿qué posibilidades hay de que sean estas características?”
- Mientras que los modelos discriminativos se preocupan por la relación entre **y** y **x**, los modelos generativos se preocupan por "cómo se obtiene x"

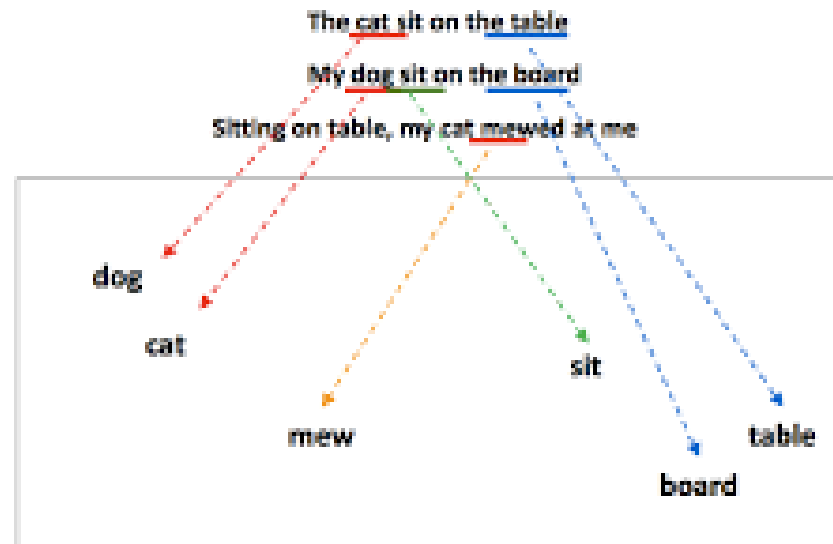


Modelos Funcionales

- Modelos de lenguaje de gran tamaño (LLM) no etiquetados
- Se centran específicamente en tareas basadas en el lenguaje, como el resumen, la generación de texto, la clasificación, las preguntas y respuestas abiertas y la extracción de información.

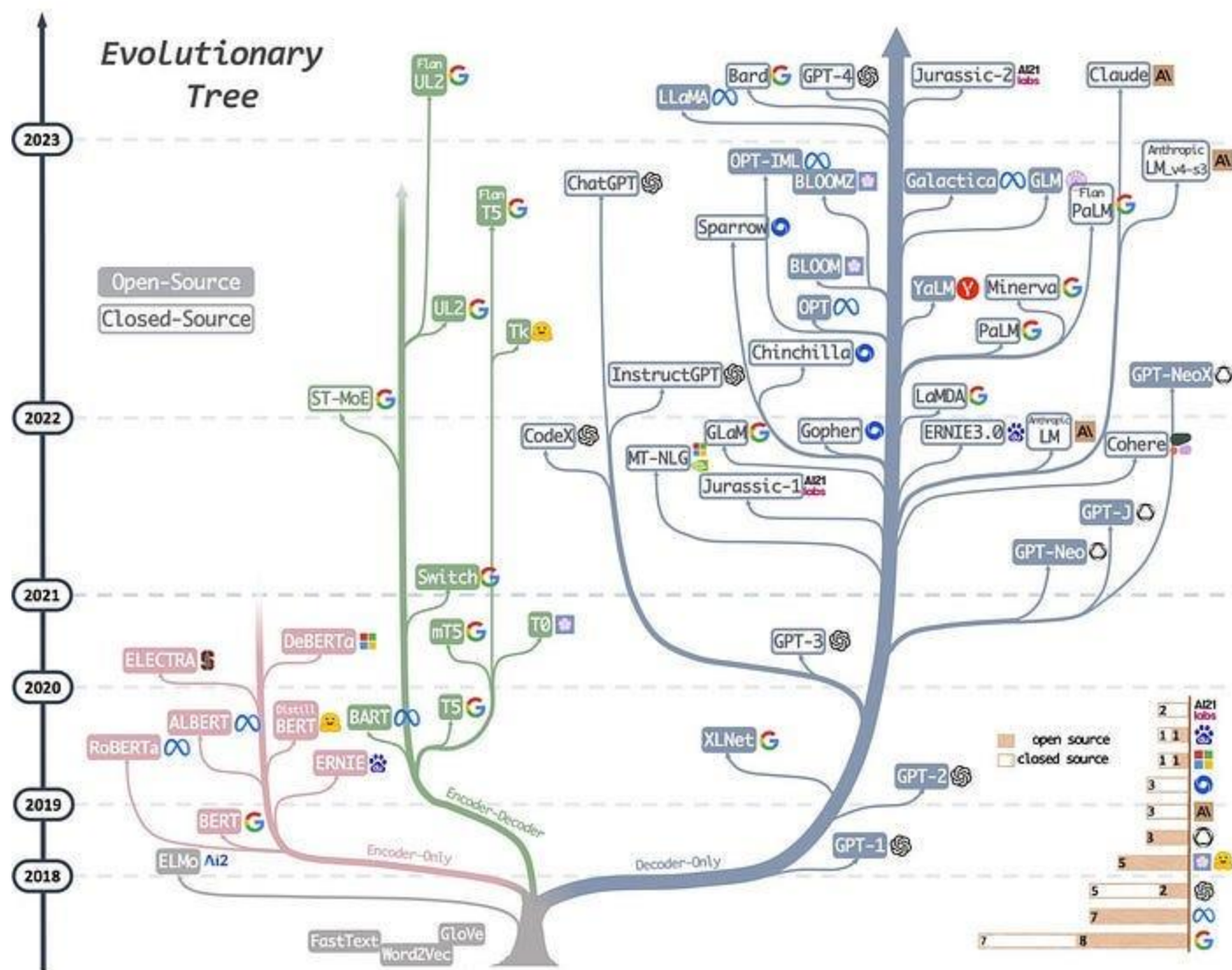
Modelos

- Small Models – Word Vectors – Mismo vector para el mismo Token
- Large Models – Context Sensitive Vectors – Vector del Token determinado por su contexto (tokens que le rodean)



LLMs

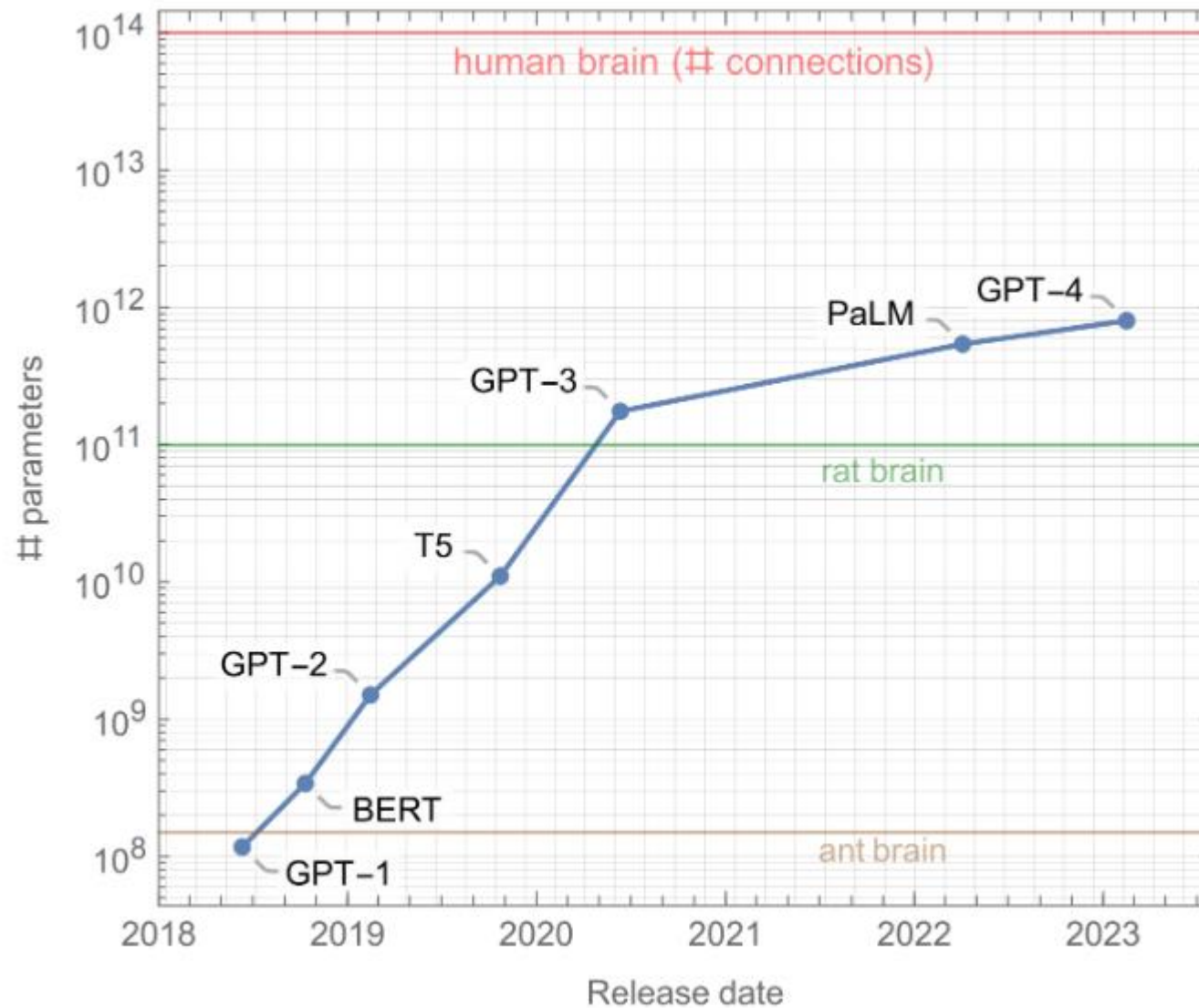
- Sistemas de IA que han sido entrenados para comprender y generar texto en lenguaje natural.
- Modelos de aprendizaje profundo muy grandes que se preentrenan con grandes cantidades de datos
- Estos modelos pueden interpretar y producir textos en diferentes idiomas y contextos
- Traducción automática, generación de resúmenes y creación de contenido.



LLMs

- LLM de generación de texto basados en transformadores
- OpenAI
 - GPT-1 (2018) - 100 millones de parámetros
 - GPT-2 (2019) - 1.500 millones de parámetros 40 GB.
 - GPT-3 (2020) - 175 mil millones de parámetros 700 GB.
 - GPT-4 (2020) – 1 Billón de parámetros 700 GB.
- Meta - LLaMA

LLMs notables



- La invención de los transformadores marcó el comienzo de la era de los grandes modelos de lenguaje modernos.



Transformers

- A diferencia de las redes neuronales recurrentes (RNN) anteriores que procesaban las entradas de forma secuencial, los transformadores procesan secuencias enteras en paralelo.
- Permite a los modelos aprender relaciones complejas entre palabras y oraciones, lo que resulta en una mejor comprensión y generación de texto en lenguaje natural.
- Consta de un codificador y un decodificador con capacidades de autoatención. El codificador y el decodificador extraen significados de una secuencia de texto y comprenden las relaciones entre las palabras y las frases que contiene.
- Capacidad para prestar atención a diferentes partes del texto de manera selectiva mediante un mecanismo llamado atención auto-regresiva, que permite a la máquina determinar qué palabras son relevantes para la generación de una respuesta coherente.

Análisis de Sentimientos basado en aspectos ABSA

Aplicaciones de Lenguaje Natural
Escuela Superior de Cómputo





Análisis de Sentimientos

El análisis de sentimiento aplica NLP, lingüística computacional y machine learning para identificar el tono emocional de un texto digital.

Análisis de sentimientos – NLP, Machine Learning, IA, Minería de Datos

- Subcategoría del NLP (reconocer entidades con nombre, responder a preguntas, resumir texto, identificar el idioma y generar lenguaje natural).
- Machine Learning. Aprender y realizar el análisis de un texto dado.
- IA. Imitar las capacidades de aprendizaje y resolución de problemas humanos (identificar sentimientos).
- Minería de Datos. Extrae y analiza grandes sets de datos a fin de descubrir varios tipos de información y patrones (identificación de palabras textuales que denotan un sentimiento).



Tipos de Análisis de Sentimientos

- Análisis detallado
- Detección de emoción
- Análisis basado en intención
- Análisis basado en aspectos



Análisis de sentimiento detallado

- Análisis de sentimiento graduado
- Permite estudiar las valoraciones contenidas en las reseñas.
- Refina las polaridades en las categorías muy positiva, positiva, neutral, negativa o muy negativa.
- Por ejemplo: una reseña de 1 estrella se considerará muy negativa, una reseña de 3 estrellas, neutral, y una reseña de 5 estrellas, muy positiva.



Análisis de sentimiento de detección de emoción

- Identifica los sentimientos como felicidad, tristeza o enojo.
- Este tipo de análisis puede usar léxico para evaluar el lenguaje subjetivo:
 - Las palabras como "horrible" y "deplorable" sugieren enojo.
 - "Miserable" y "devastador" pueden indicar tristeza.
 - "Emocionante" o "súper" pueden sugerir felicidad.
 - "atascado" y "frustrante" significan una emoción negativa
 - "generoso" es positivo.
- Este sentimiento está lleno de matices, y la emoción es difícil de clasificar.



Análisis de sentimiento basado en intención

- Permite identificar la intención del cliente y los niveles de interés.
- Los distintos tipos de intención incluyen: compra, actualización, actualización a un nivel inferior, cancelación o anulación de una suscripción.
- El análisis basado en intención requiere entrenamiento para clasificación con texto relevante, como consultas o correos electrónicos de clientes.
- Ejemplos:
 - "me quedé sin espacio de almacenamiento, ¿qué hago?" - oportunidad de actualización.
 - "no me gustan las muestras que recibo, no necesito más delineadores" - cancelación, pero también oportunidad de mejora del servicio.
- Este tipo de análisis ayuda a las empresas a gestionar y mantener su base de clientes y a maximizar las oportunidades de venta.



Análisis de sentimiento basado en aspecto ABSA

- Permite captar sentimiento respecto a un solo aspecto o parte específica de un servicio o producto.
- Extrae los aspectos individuales del documento introducido y clasifica cada uno de los aspectos en una de las clases de polaridad: positiva, negativa, mixta y neutra.
- Proporciona una puntuación de confianza para cada una de las clases y sus desfases correspondientes en el texto introducido.
- Las puntuaciones de confianza más cercanas a 1 indican una mayor confianza en la clasificación de la etiqueta, mientras que las puntuaciones más bajas indican una menor puntuación de confianza. El rango de la puntuación de confianza para cada clase está entre 0 y 1, y las puntuaciones acumuladas de las cuatro clases suman 1.



Análisis de sentimiento basado en aspecto ABSA

Ejemplo:

Una empresa lanza unos nuevos audífonos inalámbricos.

Aspectos: la conectividad, el diseño estético y la calidad del sonido.

- Reseñas:
 - "Estos nuevos auris son sexy" - sentimiento respecto al diseño estético de los auriculares.
 - "Me gusta cómo se ven, pero el control del volumen es un problema" - podría alertar a una empresa respecto a un defecto de diseño práctico.

Ejemplo (una opinión de un restaurante)

"La comida es minimalista, pero el servicio es tan malo."

- Sentimiento positivo hacia el aspecto de la comida.
- Sentimiento negativo hacia el aspecto de servicio.
- Una clasificación del sentimiento general etiquetaría esta reseña como negativa y pasaría por alto el hecho de que la comida era buena.
- ABSA aborda este problema haciendo referencia a un aspecto como atributo (o componente) de una entidad.

Ejemplo ABSA

"He tenido un buen día en el trabajo hoy"

- 100% de sentimientos positivos
- 0% neutros
- 0% mixtos
- 0% negativos

Casos de Uso

- Supervisión de la marca
- Supervisión de la investigación de mercado
- Análisis de comentarios de los empleados
- Análisis de revisiones y correos electrónicos de clientes
- Encuestas sobre productos

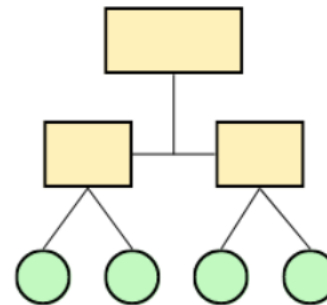
Proceso

Data Textual

Modelo
Jerárquico
de reglas

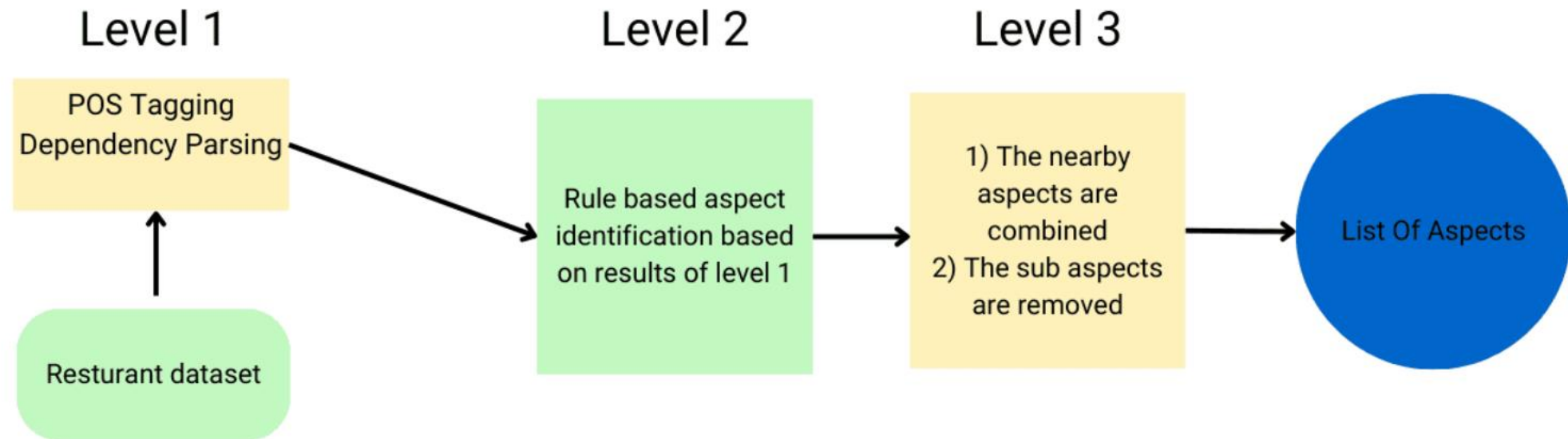
Clasificación
del sentimiento
por Polaridad

El comportamiento
del usuario puede
identificarse por
cada categoría (aspecto)
particular



Identificación de aspectos en un texto

- Level 1. Identificación de POS TAGS y dependencias
- Level 2. Identificación de aspectos basados en reglas
- Level 3. Nueva capa a partir de la combinación de los resultados



Ejemplo

```
import aspect_based_sentiment_analysis as absa
```

```
nlp = absa.load()
```

```
text = ("We are great fans of Slack, but we wish the subscriptions "  
        "were more accessible to small startups.")
```

```
slack, price = nlp(text, aspects=['slack', 'price'])
```

```
if(price.sentiment == absa.Sentiment.negative)  
    print("Price Negative")
```

```
if(slack.sentiment == absa.Sentiment.positive)  
    print("Slack Positive")
```

Ejemplo (Lista de opiniones)

```
test_reviews = [  
    "Good, fast service.",  
    "The hostess was very pleasant.",  
    "The bread was stale, the salad was overpriced and empty.",  
    "The food we ordered was excellent, although I wouldn't say the  
    margaritas were anything to write home about.",  
    "This place has totally weird decor, stairs going up with mirrored  
    walls - I am surprised how no one yet broke their head or fall off the  
    stairs"  
]
```



Preprocesamiento de Aspectos

Convertir a minúsculas

```
test_reviews = [review.lower() for review in test_reviews]
```

Lista donde almacenaremos los aspectos

```
test_aspect_terms = []
```

Recorrido de lista de opiniones a través de “chunks” de nlp – Identificación POS

```
for review in nlp.pipe(test_reviews):
```

 # Extracción de Sustantivos

```
    chunks = [(chunk.root.text) for chunk in review.noun_chunks if chunk.root.pos_ == 'NOUN']
```

```
    test_aspect_terms.append(chunks)
```

Preprocesamiento de Sentimientos

Lista donde almacenaremos los sentimientos asociados a aspectos

```
test_sentiment_terms = []
```

Recorrido de lista de opiniones a través de “chunks” de nlp – Identificación POS

```
for review in nlp.pipe(test_reviews):
```

```
    if review.is_parsed:
```

Extracción de Adjetivos y Verbos

```
        test_sentiment_terms.append(' '.join([token.lemma_ for token in review if (not token.is_stop  
and not token.is_punct and (token.pos_ == "ADJ" or token.pos_ == "VERB"))]))
```

```
    else:
```

```
        test_sentiment_terms.append("")
```

Uso de Modelo

```
test_aspect_categories =  
aspect_categories_model.predict_classes(test_aspect_terms)  
  
test_sentiment =  
sentiment_model.predict_classes(test_sentiment_terms)
```

```
Review 1 is expressing a positive opinion about SERVICE#GENERAL  
Review 2 is expressing a positive opinion about SERVICE#GENERAL  
Review 3 is expressing a negative opinion about FOOD#QUALITY  
Review 4 is expressing a positive opinion about FOOD#QUALITY  
Review 5 is expressing a negative opinion about AMBIENCE#GENERAL
```



Análisis Sintáctico de Dependencias

Aplicaciones de Lenguaje Natural
Escuela Superior de Cómputo

Análisis sintáctico de dependencias

- Técnica para analizar la estructura gramatical de una frase.
- Ayuda a comprender cómo se relacionan entre sí las palabras de una frase identificando las dependencias entre las mismas
- Marcar cómo las palabras dependen unas de otras - significado.

Ventajas análisis sintáctico de dependencias

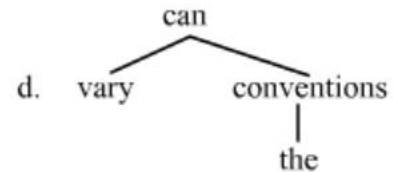
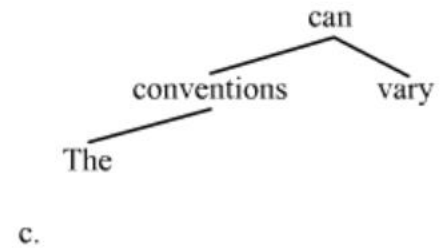
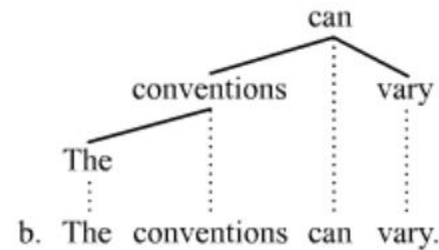
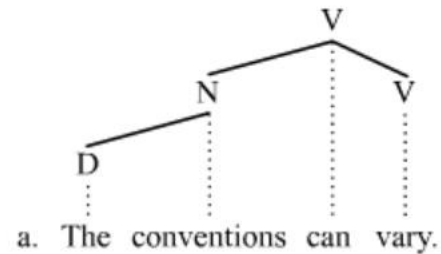
- Crucial para diversas tareas de PLN: traducción automática, extracción de información, respuesta a preguntas y análisis de sentimientos.
- Comprender la estructura sintáctica de las frases puede mejorar significativamente la precisión y la eficacia.
- Permite a los algoritmos captar el significado de las frases con mayor precisión al entender cómo están conectados los componentes de una frase (sujetos, predicados, objetos, etc.).

Árbol de dependencias

- Los nodos representan las palabras de una frase
- Las aristas representan las relaciones entre las palabras.
- Cada arista del árbol de dependencia se etiqueta con el tipo de relación gramatical que existe entre las palabras conectadas, como sujeto, objeto, modificador, etc.
- La raíz del árbol suele ser el verbo principal o la cláusula principal con la que se relacionan las demás palabras.

Algunas Abreviaturas y convenciones

S/O	para sentencia, (estructura de alto nivel).- Se divide en: sujeto de la sentencia y objeto de la sentencia.
NP/GN	noun phrase, sintagma nominal, sujeto
VP/GV	verb phrase, sintagma verbal, predicado
V	verbo.
D / Det	determinante.
N	sustantivo



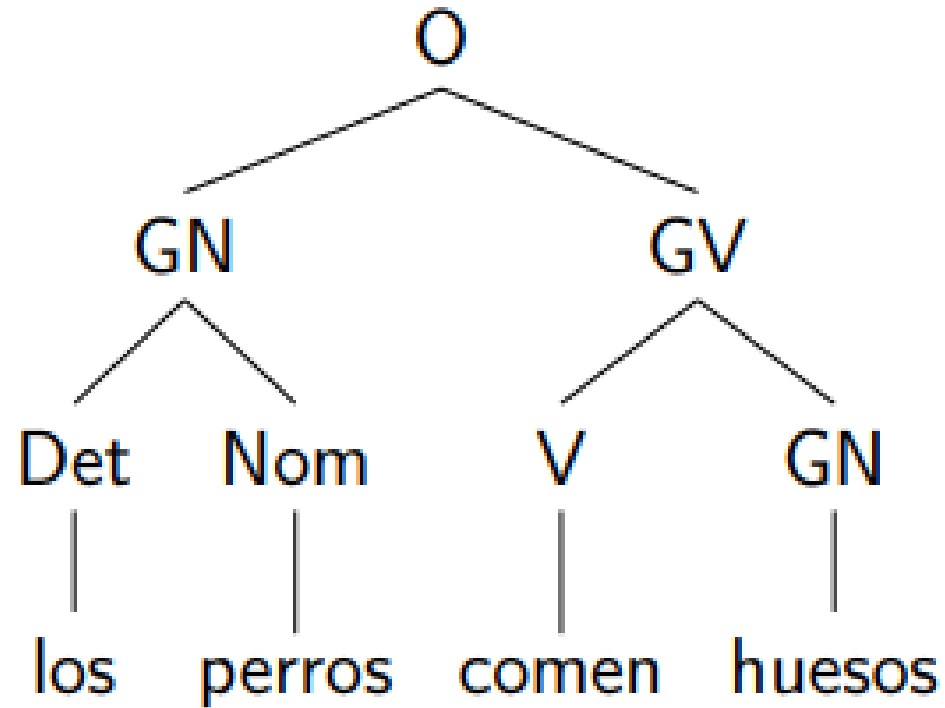
[S [N John] [VP [V hit] [NP [D the] [N ball]]]]

Árboles de análisis basados en dependencias

- Ven todos los nodos como terminales
- El verbo (finito) se toma como el centro estructural de la estructura de la cláusula.
- Todas las demás unidades sintácticas (palabras) están directa o indirectamente conectadas al verbo en términos de enlaces dirigidos, que se denominan *dependencias*.

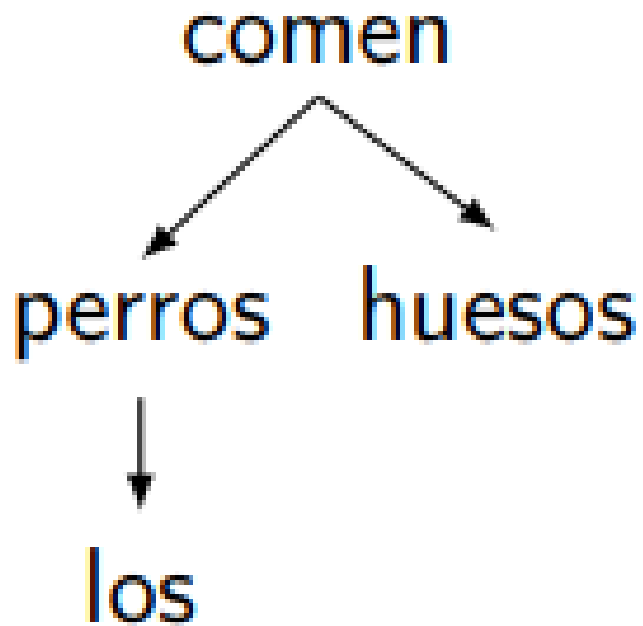
Árboles de análisis basados en dependencias

- Los perros comen huesos (Relaciones Biléxicas)



Árboles de análisis basados en dependencias

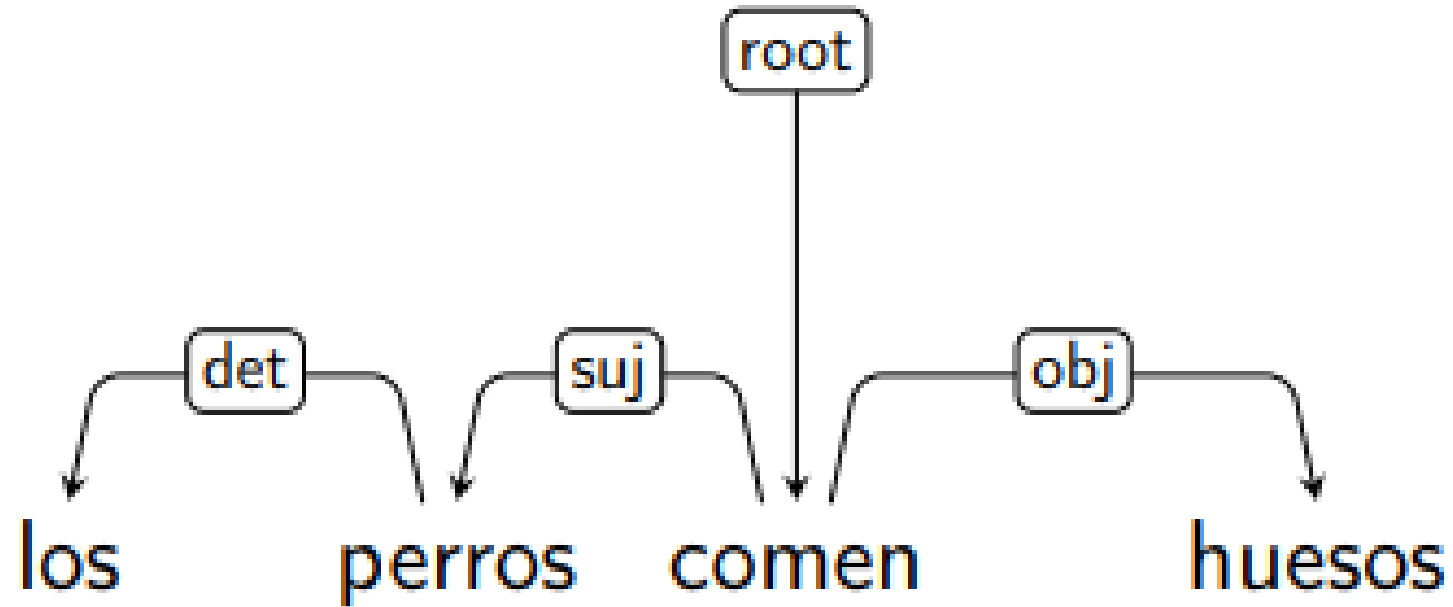
- Los perros comen huesos (Relaciones Asimétricas)



- ▶ "comen" gobierna a "perros" y a "huesos"
- ▶ "perros" gobierna a "los"

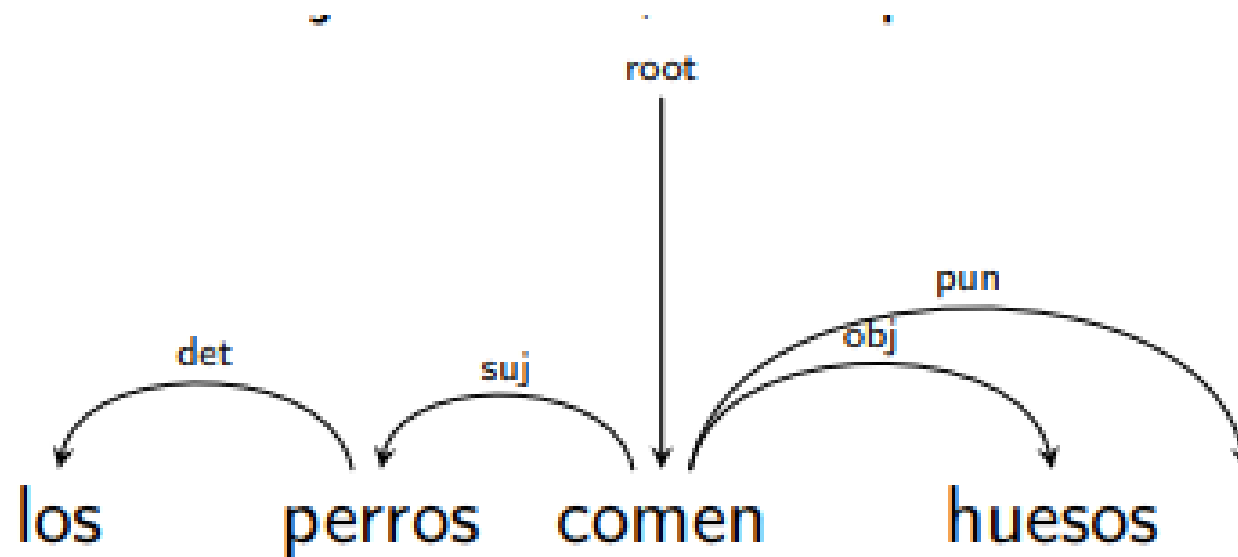
Árboles de análisis basados en dependencias

- Los perros comen huesos (Relaciones Etiquetadas)



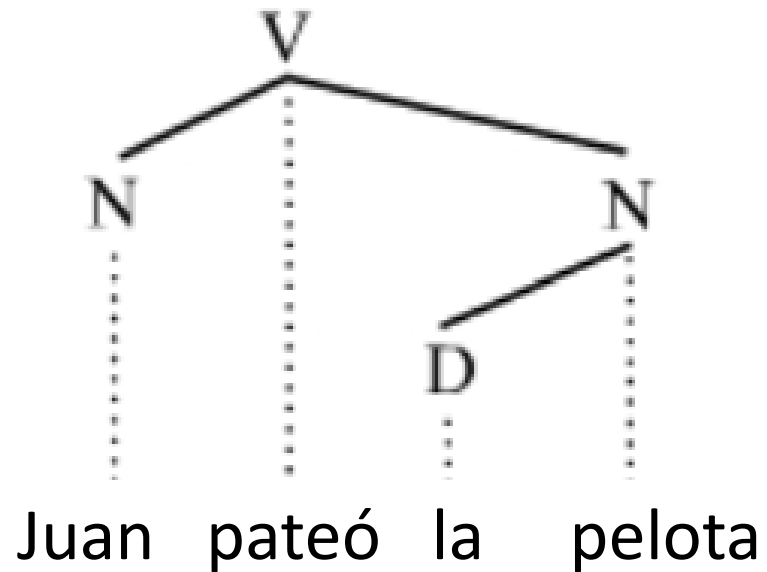
Árboles de análisis basados en dependencias

- Los perros comen huesos (proyección)



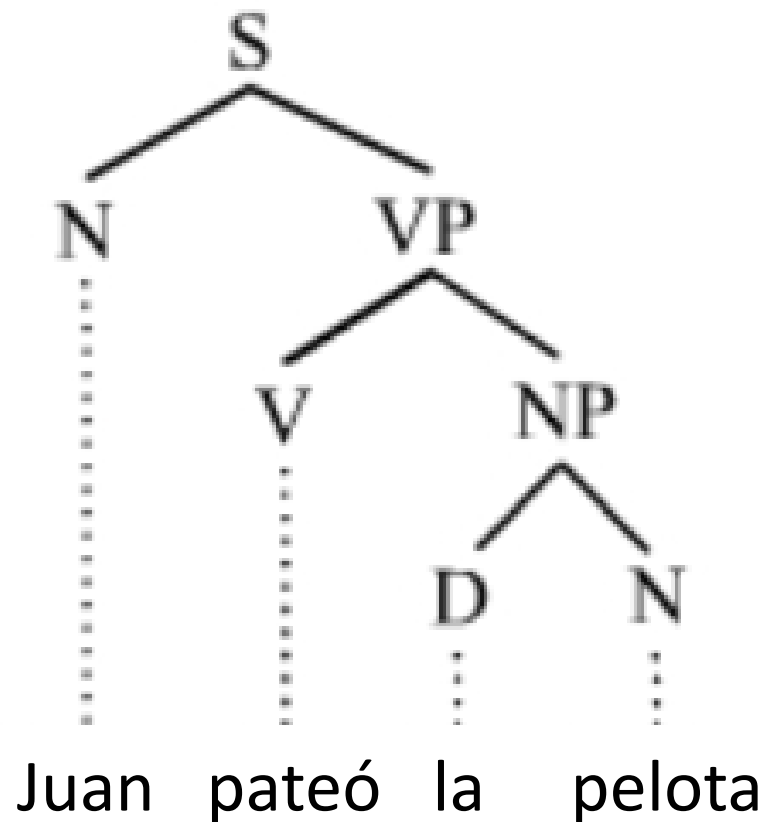
Árboles de análisis basados en dependencias

- Juan pateó la pelota (Relaciones Asimétricas)



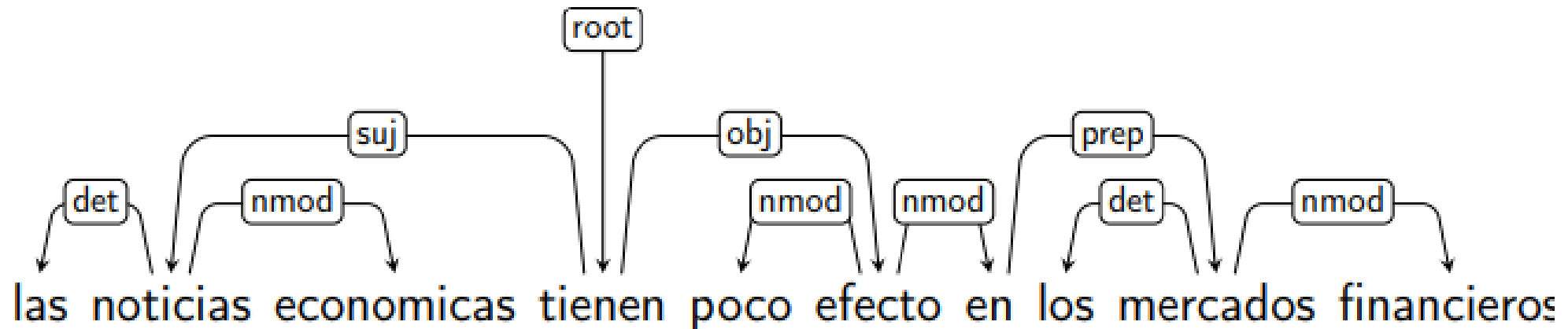
Árboles de análisis basados en dependencias

- Juan pateó la pelota (Relaciones Biléxicas)



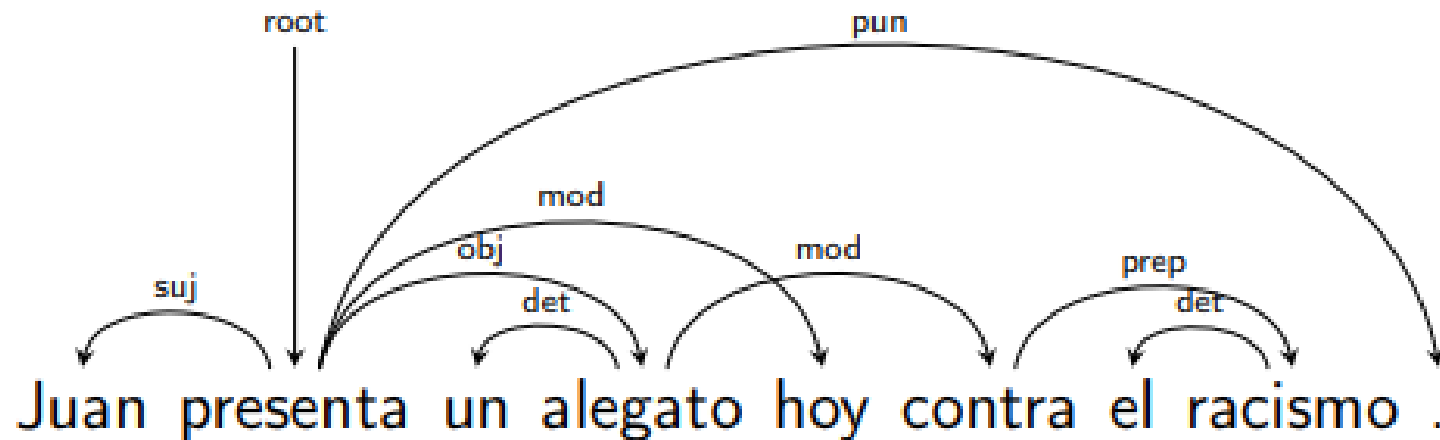
Árboles de análisis basados en dependencias

- Las noticias económicas tienen poco efecto en los mercados financieros (Relaciones Etiquetadas).



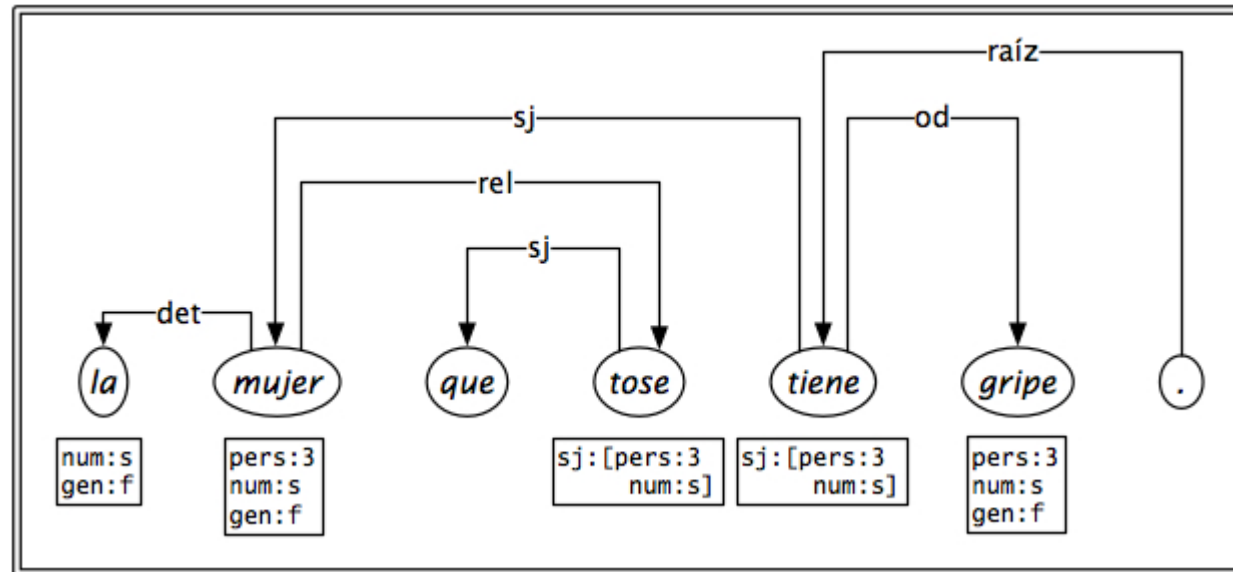
Árboles de análisis basados en dependencias

- Juan presenta un alegato hoy contra el racismo (Relaciones Etiquetadas).



Árboles de análisis basados en dependencias

- La mujer que tose tiene gripe (Relaciones Etiquetadas).





Aplicaciones de Lenguaje Natural

Escuela Superior de Cómputo

Atribución de autoría

¿Quién es el autor?

- “El que no sabe gozar de la ventura cuando le viene, no debe quejarse si se pasa.”
- “Ninguna ciencia, en cuanto a ciencia, engaña; el engaño está en quien no la sabe.”
- “La verdad adelgaza y no quiebra, y siempre anda sobre la mentira como el aceite sobre el agua.”
- "En esto, le vino al pensamiento cómo le haría, y fue que rasgó una gran tira de las faldas de la camisa, que andaban colgando, y diole once ñudos, el uno más gordo que los demás, y esto le sirvió de rosario el tiempo que allí estuvo, donde rezó un millón de avemarías. Y lo que le fatigaba mucho era no hallar por allí otro ermitaño que le confesase y con quien consolarse; y así, se entretenía paseándose por el pradecillo, escribiendo y grabando por las cortezas de los árboles y por la menuda arena muchos versos, todos acomodados a su tristeza,

¿Quién es el autor?

- "El mundo habrá acabado de joderse el día en que los hombres viajen en primera clase y la literatura en el vagón de carga".
- "Dice que se está muriendo por mí, como si yo fuera un cólico miserere".
- "Las cosas tienen vida propia, todo es cuestión de despertarles el ánima".
- "El secreto de una buena vejez no es otra cosa que un pacto honrado con la soledad".

Frases

- 1.En algún momento, tienes que olvidarte de los rencores, porque sólo hieren. ...
- 2.Al final todos nos sentimos atraídos por el amor y no podemos evitarlo. ...
- 3.Y solo mírame con esos ojito' lindo', que, con eso, yo estoy bien, hoy he vuelto a nacer
- 4.Con tu alma es la que yo conecto, tranquila, no tiene que ser perfecto, no
5. Mi mayor dolor en la vida es que nunca podré verme a mí mismo tocar en vivo".
6. "Todavía creo que soy el mejor".

Elementos de texto

- Cada texto tiene marcadores lexicales de estilo:
 - riqueza del vocabulario
 - frecuencia de las palabras de función.
- Vocabulario: conjunto de palabras que conforman un idioma.
- Las palabras de función, por ejemplo, podrían ser simples preposiciones (a, contra, bajo, etc.).



Estilometría

- Estudio de estilo lingüístico que analiza ciertos rasgos del estilo del autor y los utiliza para comparar dos o más textos, con el objetivo de determinar la autoría de documentos anónimos o en disputa.
- El estilo es algo propio de cada autor, ya que se encuentra dentro de su subconsciente
- Aplicaciones: determinación de autoría de una obra, la autenticidad, la clasificación de textos, la medición de frecuencias de palabras, la identificación de lenguas, entre otras.



Problema general

Authorship attribution of texts (AAT)

- Identificar autores al responder la pregunta: ¿Quién es el autor de un determinado texto?
- Determinar la probabilidad de que un escrito sea de la autoría de un individuo en particular, teniendo en consideración escritos anteriores del individuo.
- Agrupar los escritos de una misma autoría.



Categorías Fundamentales

Verificación de autoría.

- **“Un texto desconocido puede o no ser de determinado autor”**
- Se tiene una muestra de texto perteneciente a un autor determinado y se desea responder si cierto texto desconocido, que puede tener mayor o menor largo que el original, pertenece o no al autor determinado.
- Poder entregar qué tanto el algoritmo califica el texto desconocido dentro de la autoría.



Categorías Fundamentales

Caracterización del autor.

- Determinar ciertos atributos de un individuo, por ejemplo género, edad, entre otros o posibles combinaciones entre dichas características.
- Diversidad en la forma de escribir de un autor de acuerdo a un género literario frente a otro; así como también en el cambio del público objetivo, entre otros.



Categorías Fundamentales

Detección de estilos de escritura.

Author Masking – Obfuscation Evaluation

- Determinar a los individuos de un texto de múltiple autoría.
- Identificar un estilo de autor sin conocimiento de su identidad o perfil, logrando determinar si existen correlaciones validas dentro de textos o con otros documentos, relacionando sus similitudes y discriminando con base en esta información la autoría del individuo.



Estado del arte

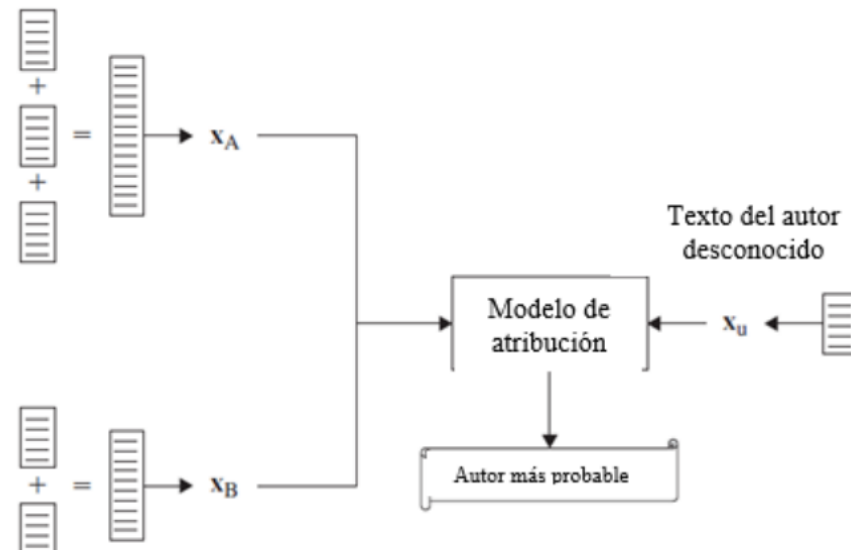
Análisis textual basado en características:

Tipo	Características
Léxico	Basado en tokens (largo de palabras, largo de oraciones, etc)
	Riqueza de vocabulario
	Frecuencia de palabras
	Frecuencia de N-gramas de palabras
	Errores
Caracteres	Tipo de caracteres (letras, dígitos, etc)
	N-gramas de caracteres (tamaño fijo)
	N-gramas de caracteres (tamaño variable)
	Métodos de compresión
Sintáctico	Parte del discurso
	Pedazos
	Estructura de frase y sentencia
	Reglas de frecuencia de re-escritura
Semántico	Mediciones de Sinónimos
	Dependencias semánticas
	Dependencias funcionales
Específicas de aplicación	Estructuras
	Contenido específico
	Lenguaje específico
	Frecuencia de palabras específicas.

Hasta ahora, algunos estudios han demostrado que las mediciones más efectivas son las de características léxicas y de carácter

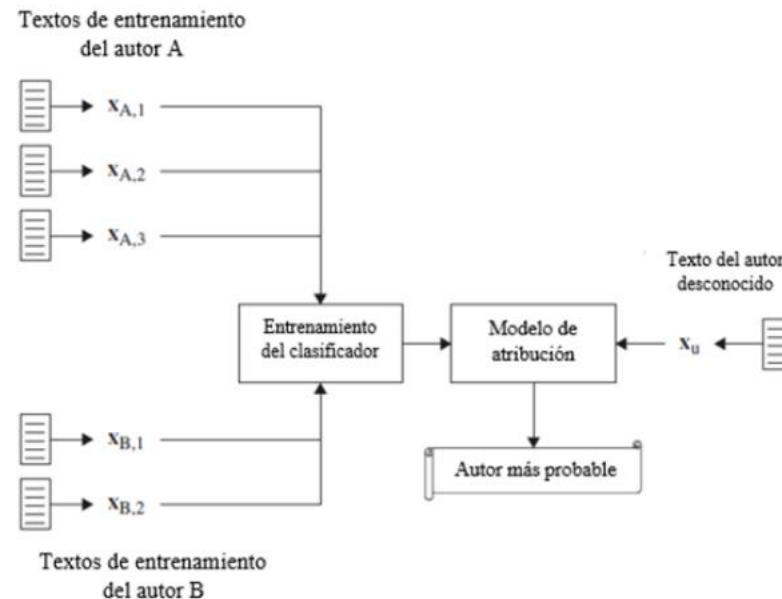
Enfoque basado en el perfil del autor

- Extracción de varias características de bajo nivel, tales como n-gramas de caracteres.
- Calcular la similitud entre los perfiles de autor generados y las características del nuevo documento



Enfoque basado en instancias


- Vectores con características variadas: caracteres, longitud de palabras, n-gramas de caracteres, n-gramas de palabras, y partes de la oración (POS).





Desambiguación del Sentido de las Palabras

Aplicaciones de Lenguaje Natural
Escuela Superior de Cómputo



Homonimia



- Dos o más palabras con la misma forma
- En las palabras homónimas, cada palabra tiene un significado completamente diferente, sin relación alguna entre sí.

Homonimia

- **Homonimia perfecta.** Dos palabras tienen la misma forma y la misma pronunciación, pero significados diferentes.
 - "banco" puede referirse tanto a una institución financiera como a un asiento largo para varias personas.
- **Homonimia imperfecta.** Dos palabras se escriben o se pronuncian de manera similar, pero tienen significados diferentes.
 - "Asta" se refiere a un mástil, mientras que la palabra "Hasta" es una preposición.



Ejemplos

- A (preposición) / Ha (forma del verbo “haber”).
- Banco (entidad bancaria o financiera) / Banco (mueble para sentarse) / Banco (forma del verbo “bancar”) / Banco (conjunto de peces).
- Vaya (forma del verbo “ir”) / Valla (cerca, vallado) / Baya (fruto).
- Rayar (hacer rayas) / Rallar (desmenuzar algo).
- Halla (forma del verbo “hallar”) / Haya (forma del verbo “haber”) / La Haya (ciudad de Holanda).
- Revelar (dar a conocer algo o hacer visible una fotografía) / Rebelar (sublevar).
- Varón (hombre) / Barón (título nobiliario).
- Bote (embarcación) / Bote (forma del verbo “botar”) / Vote (forma del verbo “votar”).
- Vello (pelo corporal) / Bello (sinónimo de lindo, hermoso).



Polisemia

- Una palabra es POLISÉMICA cuando tiene dos o más significados.
- Cualidad que se refiere a la presencia de múltiples sentidos en un mismo término.
- Los diferentes significados que tiene una misma palabra están relacionados entre sí, ya sea por su origen etimológico, su contexto de uso o por convención.



Ejemplos



Palabra polisémica	Significados
Cura	1. Sacerdote. 2. Sanación de alguna herida o dolor.
Café	1. Bebida. 2. Establecimiento donde se vende café y otros artículos de consumo. 3. Color. 4. Semilla del cafeto.
Acento	1. Forma de hablar de una persona en función de su región. 2. Fuerza de voz que se aplica en una sílaba de una palabra determinada.

Pico	1. Parte saliente en la cabeza de las aves. 2. Herramienta utilizada para cavar. 3. Cima de una montaña. 4. Extremo puntiagudo de un objeto o elemento.
Chile	1. Pimiento de origen sudamericano. 2. País ubicado en Sudamérica debajo de Perú.
Blanco	1. Color. 2. Artículo usado para practicar la habilidad de tiro. 3. Objetivo al que se destina un disparo o acción. 4. Espacio sin llenar en un escrito.
Nave	1. Barco o embarcación. 2. Espacio entre los muros dentro de un templo.

Ejemplos



Arco	<ol style="list-style-type: none">1. Arma con la cual se dispara flechas.2. Portería en un partido de fútbol.3. Sección continua de una curva.
Portero	<ol style="list-style-type: none">1. Vigilante que resguarda la entrada a un edificio o establecimiento.2. Jugador que defiende la portería en el fútbol.
Cuarto	<ol style="list-style-type: none">1. Habitación.2. Una de las cuatro partes en que se puede dividir algo.3. Siguiente al tercero.
Bomba	<ol style="list-style-type: none">1. Objeto explosivo y riesgoso.2. Máquina que funciona para impulsar todo tipo de líquidos.

Gato	<ol style="list-style-type: none">1. Felino doméstico.2. Herramienta que sirve para alzar peso.3. Baile tradicional originario de Argentina.
Tienda	<ol style="list-style-type: none">1. Establecimiento comercial.2. Alojamiento hecho con telas y un armazón de palos.
Mono	<ol style="list-style-type: none">1. Animal.2. Persona atractiva.3. Prenda de vestir.
Ratón	<ol style="list-style-type: none">1. Roedor.2. Dispositivo utilizado en las computadoras.

En resumen

- Una palabra es POLISÉMICA cuando tiene dos o más significados o acepciones.
- Dos palabras son HOMÓNIMAS si se escriben o se pronuncian igual.

Coma	Verbo comer	
Coma	Signo de puntuación	
Coma	Pérdida de la conciencia	

Hola  - Hola , Sam. ¿Cómo estás? - Muy bien, ¿y tú?	Ola  Me gustan las olas grandes para surfear
Bello  ¡Qué bello amanecer!	Vello  Él tiene muchos vellos en el pecho

Ejemplo idioma inglés

- Más del 38% de las palabras son polisémicas.
- La palabra “set”, por ejemplo, que puede usarse como sustantivo, verbo y adjetivo, tiene docenas de definiciones únicas, lo que la convierte en una de las palabras más polisémicas registradas en el diccionario de inglés.

“please set the set of cutlery on the table”

- ¿Cómo se puede saber la diferencia entre los dos usos de “**set**”?

Contexto

- Conjunto de palabras que acompañan a una palabra
- Relaciones sintácticas
- Categorías semánticas.
- Medio más eficaz para identificar el sentido de una palabra polisémica.

Marcos vio un faisán volando entre las montañas. Sus plumas eran de color café, morado y azul. Estaba protegiendo a la hembra mientras empollaba sus huevos.

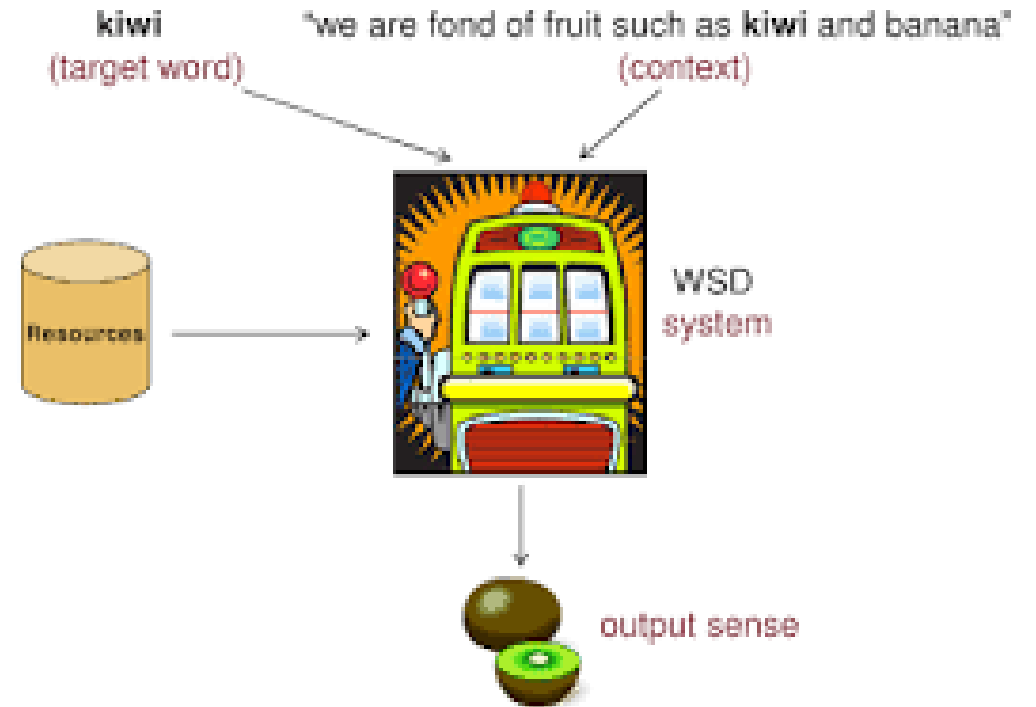


*El faisán es un ave. Las palabras que me ayudaron a encontrar el significado de faisán fueron: **volando, plumas y huevos.***

Desambiguación del Sentido de las Palabras

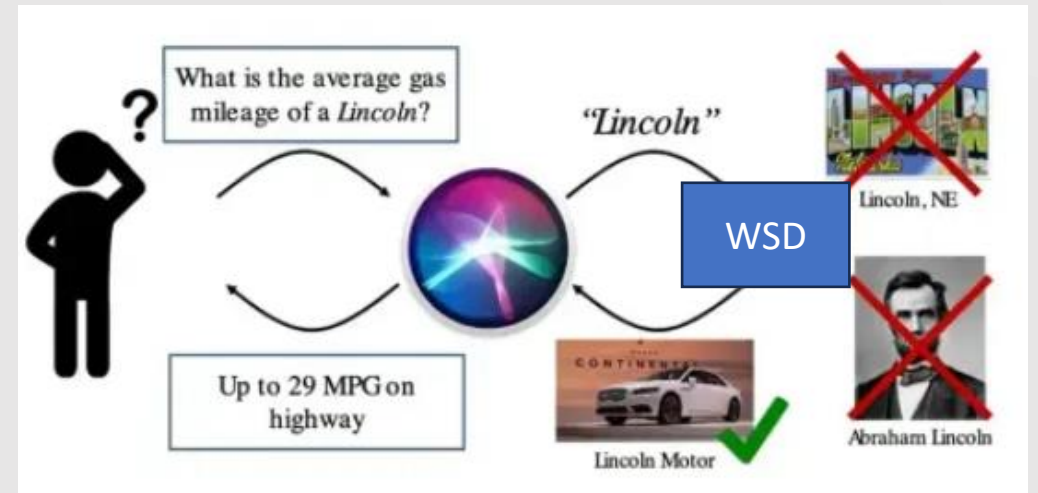
WSD: Word Sense Disambiguation

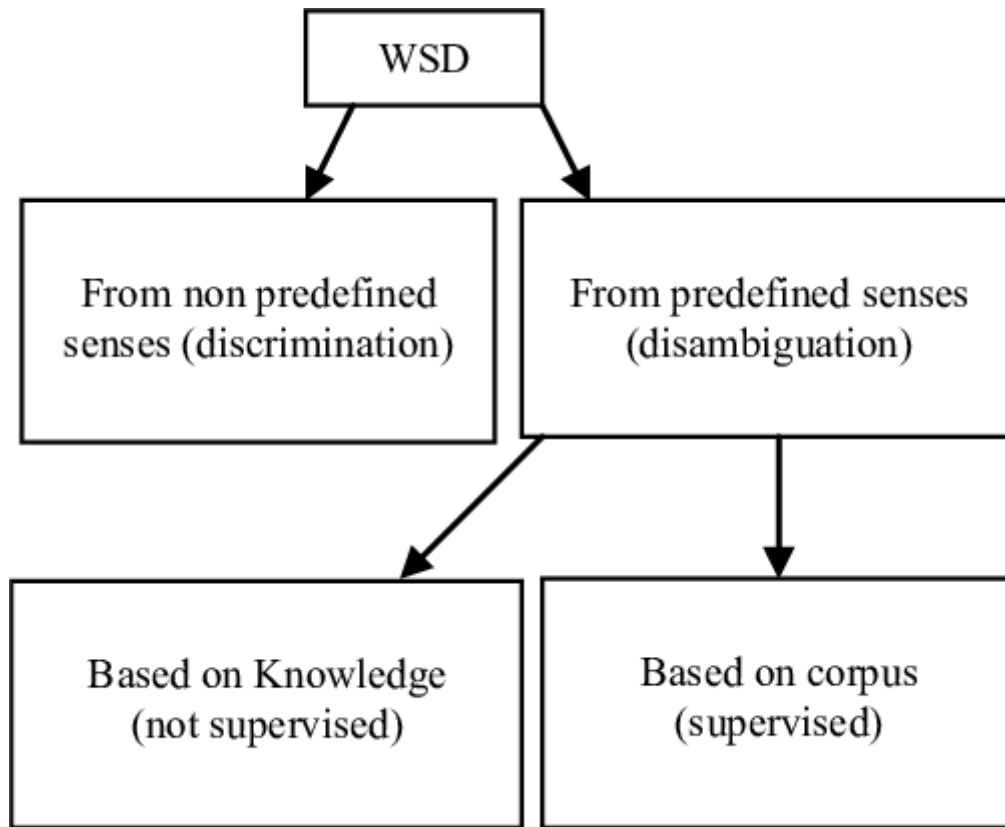
- Una de las tareas más difíciles en el PLN es la resolución de la ambigüedad de las palabras.
- Una palabra puede ser interpretada de diferentes formas, es decir, posee más de un significado o sentido (polisemia).
- Lo que persigue la WSD es la asignación automática de sentidos a las palabras de un texto, es decir, seleccionar el sentido de una palabra en su contexto.



Aplicaciones WSD

- Traducción Automática
- Recuperación de Información
- Minado de Texto

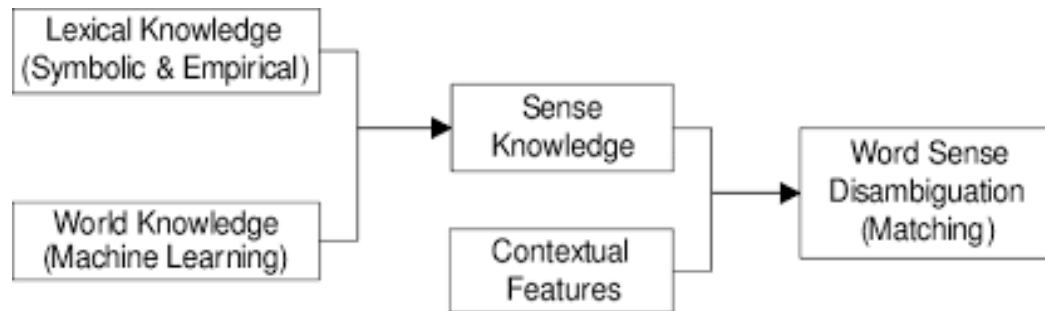




Métodos WSD

- Métodos Supervisados
 - Métodos basados en conocimiento
 - Métodos basados en tesauros/ontologías y corpus.
 - Métodos Basados en traducciones en otro lenguaje.
- Métodos No Supervisados
 - Métodos basados en contexto

Métodos Supervisados



- La palabra que se va a etiquetar con sentido siempre aparece en un contexto.
- El contexto se puede representar mediante un vector, llamado vector de contexto (palabra, características).
- Eliminar la ambigüedad del sentido de la palabra haciendo coincidir un vector de conocimiento de sentido y un vector de contexto.

Métricas

Medidas importantes de rendimiento para WSD: **Precision & Recall**

- Precision: Proporción de instancias clasificadas correctamente de las clasificadas
- Recall: Proporción de instancias clasificadas correctamente del total de instancias.

Métodos basados en conocimiento

LESK

Michael E. Lesk 1986

- Las palabras que aparecen juntas en el texto están relacionadas
- La relación y el contexto correspondiente de las palabras se pueden extraer a través de las definiciones de las palabras de interés y de las otras palabras a su alrededor.
- Utilizar la coherencia global del texto (total de sentidos de palabras relacionadas en el texto).
- Requerimientos: Un algoritmo de optimización para WSD y una medida de similitud para las definiciones de los sentidos.

Método Lesk

- Ventajas:
 - Fácil de implementar
 - Aplicable a una gran variedad de contextos
 - Generalizable
 - Sólo se necesita un diccionario de sentidos como recurso léxico.
- Desventajas:
 - Mientras se cuente con más palabras, más grande es el espacio de búsqueda
 - Para encontrar la combinación óptima de sentidos se necesita mucho tiempo.

Métodos tipo Lesk

Intentan obtener la combinación de sentidos óptima para un texto dado.

- Uso de diccionarios que proporcionen una lista de glosas (definición de sentido) para las palabras - WordNet.
- Elegir un sentido (de la lista) para cada palabra en un texto dado, tomando en cuenta el contexto en el que aparece.

Algoritmo de LESK

- Dada una palabra w , (ambigua) en un contexto c , y con sentidos s_1, \dots, s_k .
- Extraer del diccionario la bolsa de palabras correspondiente a cada sentido s_k .
- Comparar cada bolsa de palabras con las palabras del contexto (Contar número de palabras en la vecindad de la palabra como en la definición del diccionario de ese sentido)
- Seleccionar el sentido con el mayor traslape.
- No se toman en cuenta stopwords
- No es sintáctico

PINE

1. kinds of evergreen tree with needle-shaped leaves
2. waste away through sorrow or illness

CONE

1. solid body which narrows to a point
2. something of this shape whether solid or hollow
3. fruit of certain evergreen trees

Ejemplo: *My father deposits his money in a bank account*".

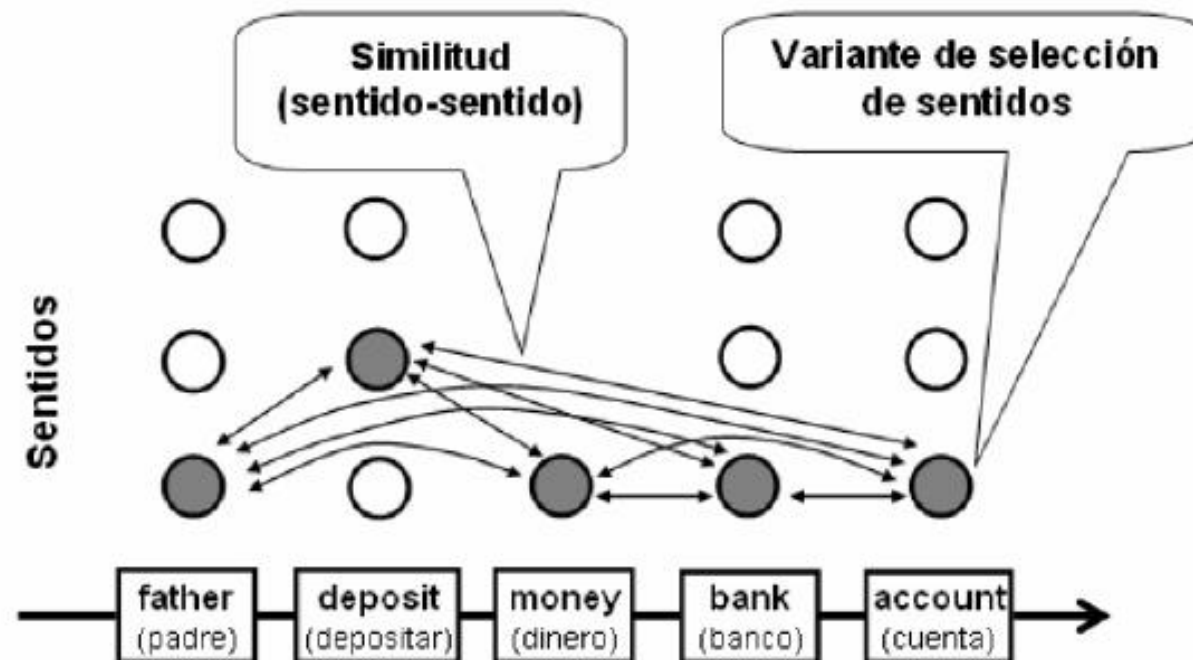
1. Extracción de los Sentidos de las palabras (máximo tres) de WordNet

Palabra	Sentidos
Father	1: a male parent (also used as a term of address to your father); "his father was born in Atlanta". 2: 'Father' is a term of address for priests in some churches (especially the Roman Catholic Church or the Orthodox Catholic Church); "'Padre' is frequently used in the military". 3: a person who holds an important or distinguished position in some organization; "the tennis fathers ruled in her favor"; "the city fathers endorsed the proposal".
Deposit	1: fix, force, or implant; "lodge a bullet in the table". 2: put into a bank account; "She deposits her paycheck every month". 3: put (something somewhere) firmly; "She posited her hand on his shoulder"; "deposit the suitcase on the bench"; "fix your eyes on this spot".
Money	1: the official currency issued by a government or national bank; "he changed his money into francs".

Bank	1: 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". 2: sloping land (especially the slope beside a body of water); "they pulled the canoe up on the bank"; "he sat on the bank of the river and watched the currents". 3: a supply or stock held in reserve for future use (especially in emergencies)
Account	1: a formal contractual relationship established to provide for regular banking or brokerage or business services; "he asked to see the executive who handled his account". 2: the act of informing by verbal report; "he heard reports that they were causing trouble"; "by all accounts they were a happy couple". 3: a record or narrative description of past events; "a history of France"; "he gave an inaccurate account of the plot to kill the president"; "the story of exposure to lead".

2. Traslape (Conteo del número de palabras que tienen en común).

- “deposit” y “bank” Comparar las tres definiciones de “deposit” contra las tres definiciones de “bank”.



A stylized background graphic featuring a grey document with several horizontal black bars representing text. A magnifying glass with a black handle and frame is positioned over the right side of the document, with its lens focusing on one of the text bars. The entire graphic is set against a solid dark red background.

Enfoques para el Resumen de textos automático

Aplicaciones de Lenguaje Natural

Escuela Superior de Cómputo

Enfoques

- **Resumen basado en extracción (términos, frases, taxonomía)**
- **Resumen basado en abstracción (conceptos)**



Resumen basado en extracción

- Implica extraer frases clave del documento de origen y combinarlas para hacer un resumen.
- La extracción se realiza de acuerdo con la métrica definida sin realizar ningún cambio en los textos.



Resumen basado en la abstracción

- Parafrasear y acortar partes del documento original.
- Puede superar las inconsistencias gramaticales del método extractivo.
- Crear nuevas frases y oraciones que transmiten la información más útil del texto original, al igual que los humanos do.
- Funciona mejor que la extracción.
- Los algoritmos de resumen de texto en Machine Learning requeridos para hacer abstracción son más difíciles de desarrollar.



Ejemplo de aplicación de ambos Enfoques

Texto fuente: *Jose y María cabalgaron en un burro hacia asistir el evento anual en **Jerusalén** . En la ciudad, **Mary** dio **nacimiento** a un niño llamado **Jesús** .*



Resumen de extracción: *José y María asistir evento Jerusalén. María, nacimiento Jesús.*

Resumen abstracto: José y María vinieron a Jerusalén, donde nació Jesús.

Métodos Extractivos

Posicional		Seleccionar las primeras y últimas oraciones
Frecuencia de palabras clave	TF	Seleccionar oraciones con palabras clave importantes
Palabras clave de señal	TF-IDF / LSA / LSI	Similar al método de frecuencia de palabras clave, pero en este caso se utilizan palabras clave específicas relacionadas con el tema del documento.
Plantillas	N-Grams / POS Tag / NER	Utiliza una plantilla predefinida para seleccionar las oraciones más relevantes del texto. Esta plantilla contiene una serie de slots que representan diferentes partes de la

Material de Apoyo

ABSA

<https://colab.research.google.com/drive/1ItS0-ZPXGcEeVmRmHaneX3w8eq6Vhdde?usp=sharing>

Desambiguación (LESK)

<https://colab.research.google.com/drive/1y6yJvCtGCqp815EVbCTIYU0fym1Zh98m#scrollTo=gwUDabDJuK1d>

Resumen automático

<https://colab.research.google.com/drive/1ejlXFelG3CZSWei3EQbeokZ-pD32Qp93?usp=sharing#scrollTo=OLu-YsOXMeqm>

Transformadores

<https://www.datacamp.com/es/tutorial/how-transformers-work>

https://youtu.be/aL-EmKuB078?si=cRIE86cRNV60m_7c

https://youtu.be/xi94v_jl26U?si=v49fs8-Zcjwtup9Tm

- Implementación Python Transformers

<https://colab.research.google.com/github/jaygala24/pytorch-implementations/blob/master/Attention%20Is%20All%20You%20Need.ipynb#scrollTo=zxtnjdKur3qe>