

# Introducción al Procesado de Lenguaje Natural



# ¿Qué es el NLP?

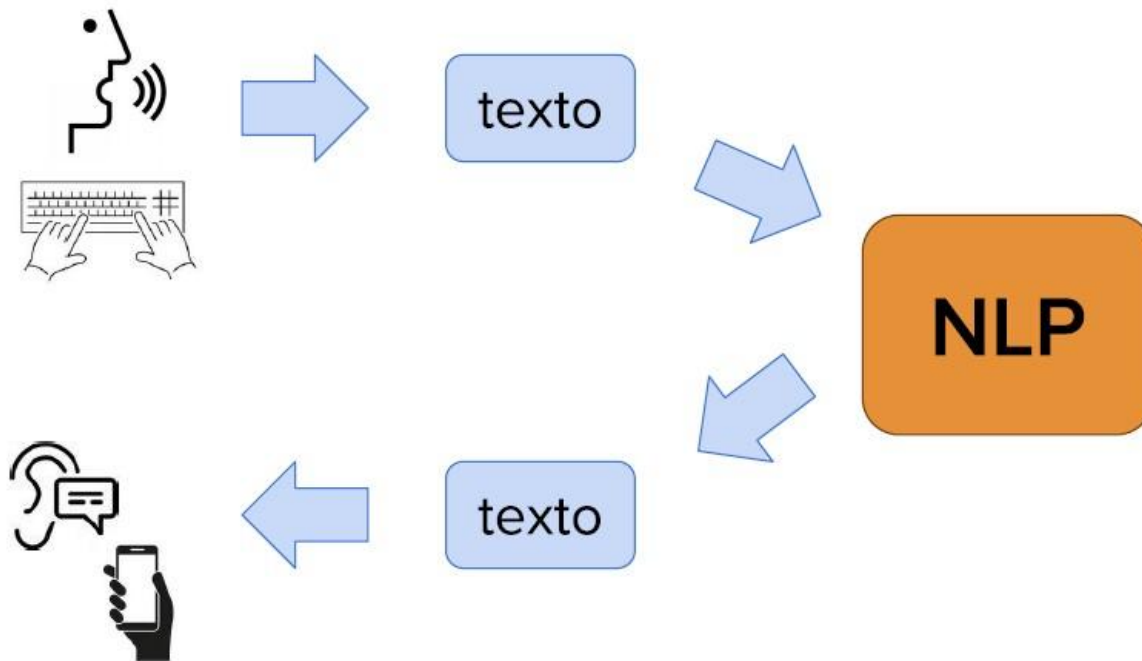
- Natural Language Processing (NLP)
- Intersección entre la IA, la lingüística y las ciencias de la computación
- Casi siempre se trabaja con texto, aunque también voz
- Es keyword matching, es normalización de textos, es ML, es clustering
- ...
- Entendimiento humano-máquina



# Perfiles en NLP

- Lingüistas
- Matemáticos
- Ingenieros
- Estadísticos,.....

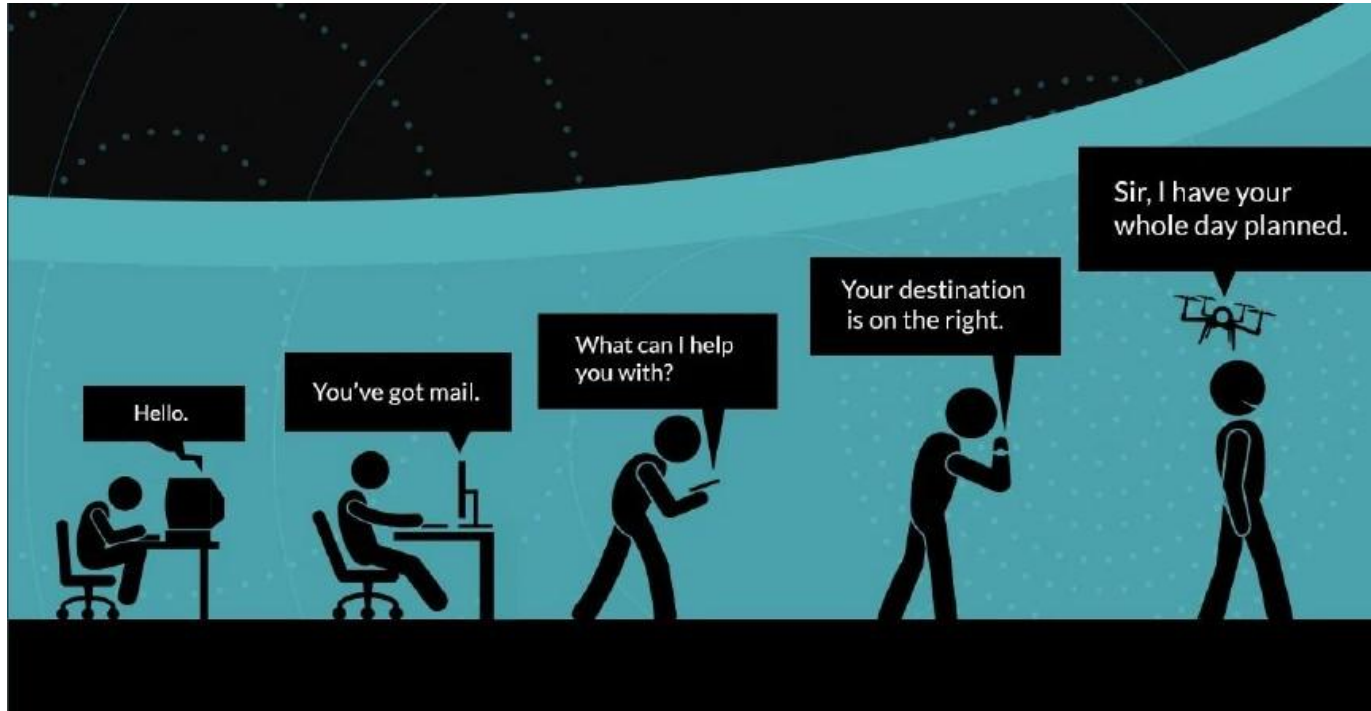
# Cadena típica de NLP



# Tareas en NLP

- Dominio concreto
  - Clasificación de documentos
  - Análisis de sentimiento en comentarios sobre marcas, productos, mercados, políticos,....
  - Búsqueda de información en un buscador
  - Detección de los temas de moda en redes sociales
- End-to End
  - Agentes conversacionales
  - Traducción bidireccional en tiempo real
  - Módulo de entendimiento de lenguaje (NLU) completo

# Estado del arte

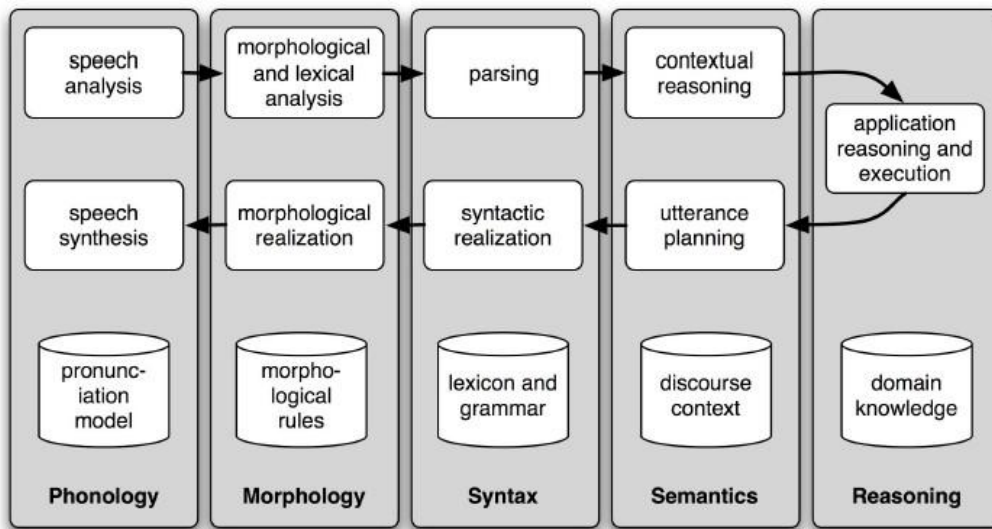


# Estado del arte

Prácticamente Resueltas	<ul style="list-style-type: none"><li>• Detección de spam</li><li>• PoS Tagging</li><li>• "Named Entity Recognition"</li></ul>
Progresando adecuadamente	<ul style="list-style-type: none"><li>• Sentiment Analysis</li><li>• Machine Translation</li><li>• Desambiguación</li><li>• Extracción de Información</li></ul>
Aún muy lejos	<ul style="list-style-type: none"><li>• Diálogo</li><li>• Parafrasear</li><li>• Preguntas y Respuestas</li><li>• Resumir texto</li><li>• Detección de sarcasmo</li></ul>

# Tareas en NLP

- Algunas tareas de NLP en un asistente virtual





# Estado del arte



# Estado del arte

Query a Alexa. El día a día ....



Alexa: remind me to feed the baby

# Procesado de Texto

## texto

Del lat. *textus*; propiamente 'trama', 'tejido'.

1. m. Enunciado o conjunto coherente de enunciados orales o escritos.

### coherencia

Del lat. *cohaerentia*.

1. f. Conexión, relación o unión de unas cosas con otras.
2. f. Actitud lógica y consecuente con los principios que se profesan.
3. f. *Fís.* **cohesión** (f. unión entre moléculas).
4. f. *Ling.* Estado de un sistema lingüístico o de un texto cuando sus componentes aparecen en conjuntos solidarios. *La coherencia del sistema de adverbios de lugar en español se manifiesta en tres grados.*

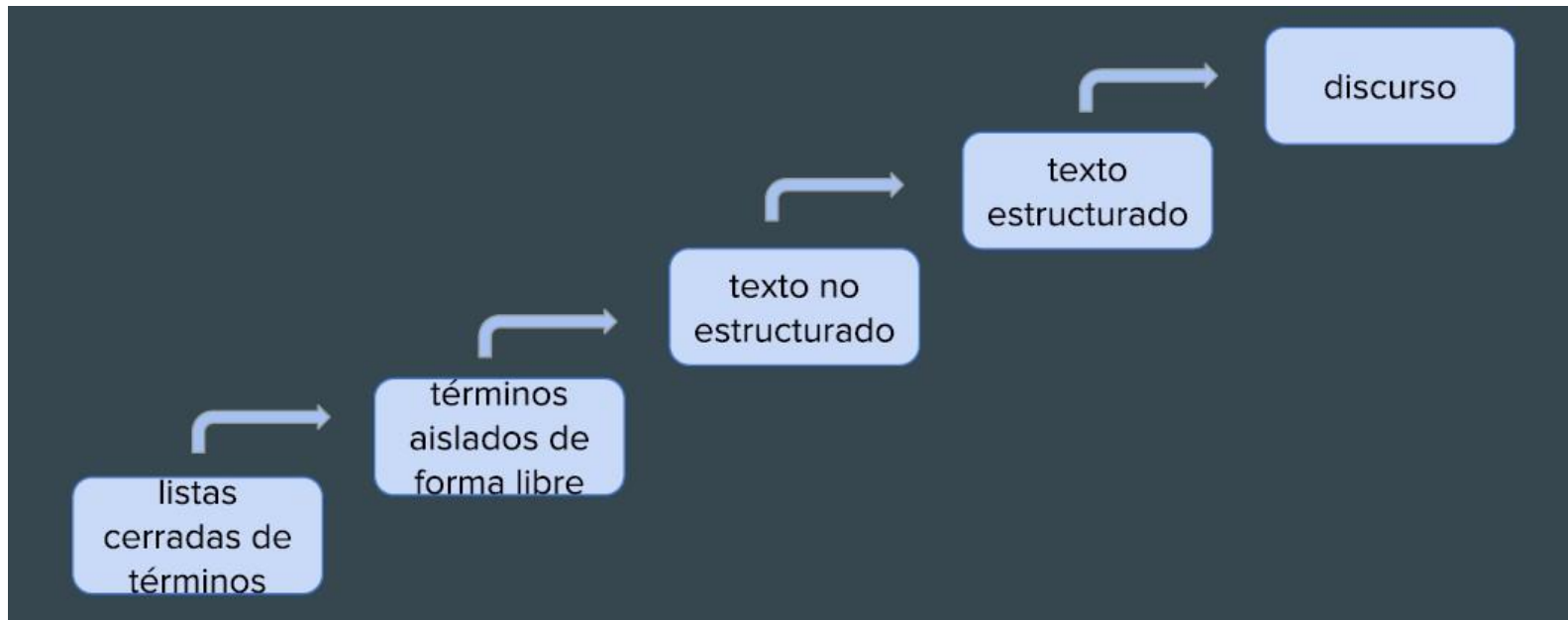
### enunciado +

De *enunciar*.

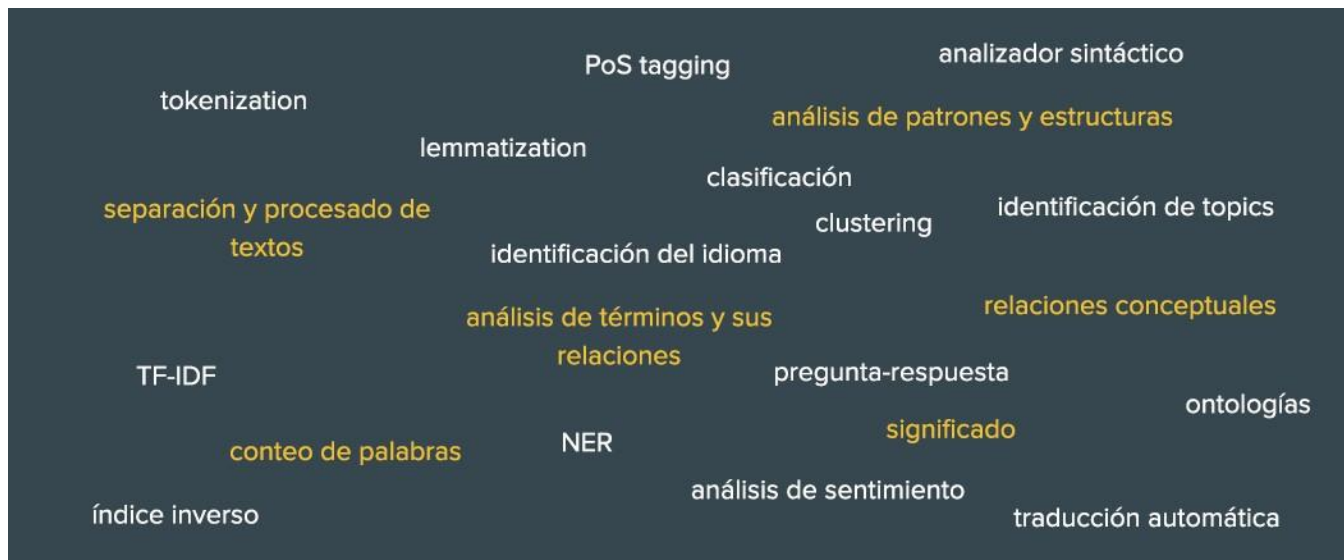
1. m. **enunciación**.
2. m. Secuencia de palabras delimitada por pausas muy marcadas, que puede estar constituida por una o varias oraciones.
3. m. *Ling.* Secuencia con valor comunicativo, sentido completo y entonación propia.

# Procesado de texto

¿Qué es un texto?



# Procesado de texto



# Algunos retos

Entre lo pienso,  
lo que quiero decir,  
lo que creo decir,  
lo que digo,  
lo que quieres oír,  
lo que oyes,  
lo que crees entender,  
lo que quieres entender,  
lo que entiendes

**Existen 9 posibilidades de no entenderse**

# Algunos retos

Regla de la comunicación del 7-38-55 % de Mehrabian



# Algunos retos

## Mapear palabras y conceptos

- Homonimia: misma forma, distinto significado, distinta etimología
  - Homografía: misma escritura
    - Nada (nadar) y nada, vela (velar) y vela, bota (calzado) y bota (vino)
  - Homofonía: misma pronunciación
    - Basta y vasta, sabia y savia, botar y votar
- Polisemia: misma palabra (misma etimología), distintos significados
  - Cabo (escalón militar) y cabo (accidente geográfico), gato (animal) y gato (herramienta)
- Sinonimia: relación de igualdad entre el significado de +2 palabras/enunciados
  - Barato y económico, demente y loco, idioma y lengua,.....



# Algunos retos

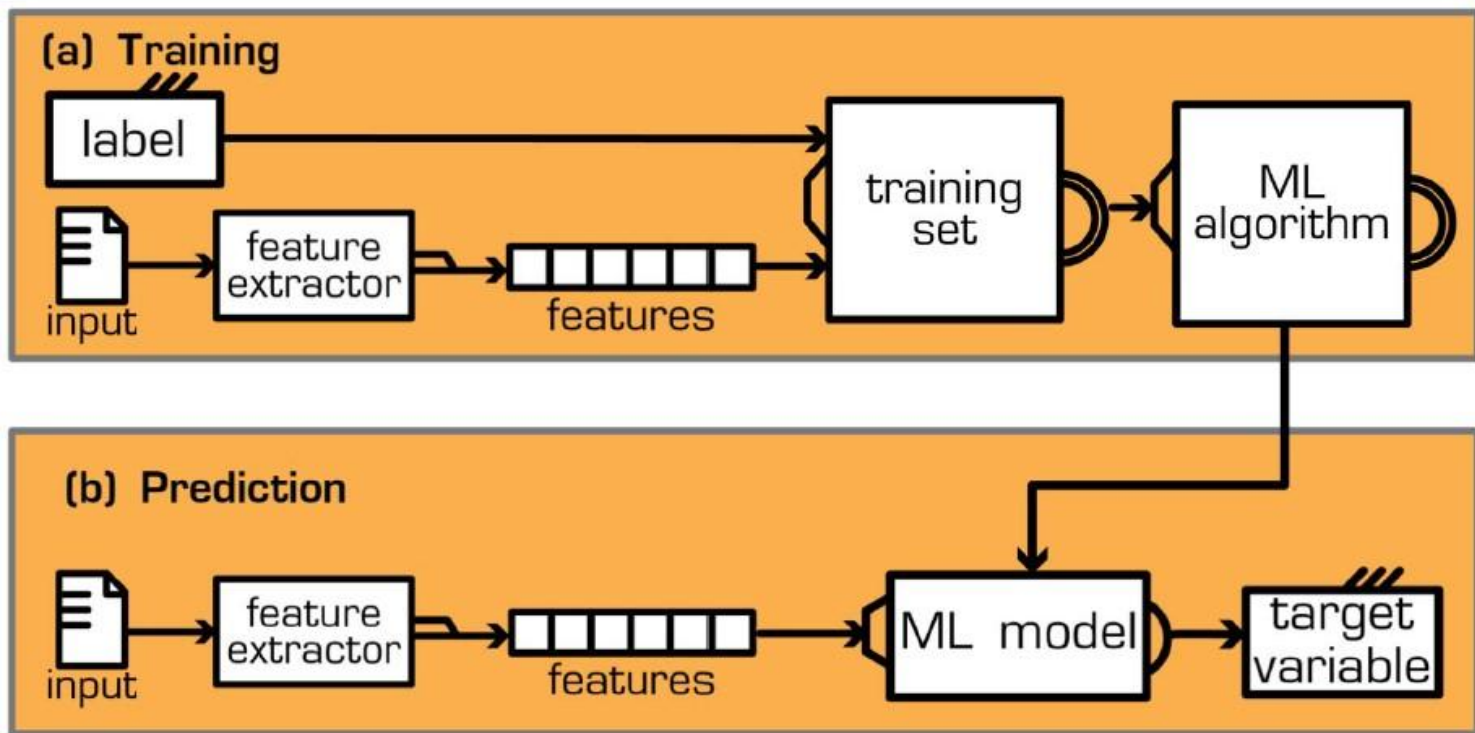
- Tipos
  - Ejemplos de respuestas en un formulario (¿dónde reside usted?): Illescas, illescas, ILLESCAS, ILL
- Ground truth
  - ¿Con qué entrenamos? ¿Con qué validamos?
- Tropos
  - Ironía, Sarcasmo, Metáfora, Alegoría, Oxímoron,...
- Contexto
  - P: ¿Qué temperatura hace en Roma?
  - R: 27 grados centígrados
  - P: ¿Y en Lima?
  - R: ....
- Idioma
- Puede llegar cualquier texto (sobre todo en asistentes virtuales)

# NLP vs Computer Vision

- En general, los modelos de CV performance mejor
- Modelos pre-entrenados
- Debate
  - Retos en modelos de Visión Artificial
  - Retos en modelo de NLP



# Pipeline de NLP





# KEEPCODING

## Tech School

Madrid | Barcelona | Bogotá

**Datos de contacto**