



Aplicaciones de Data Science

Unidad 2: Minería de Texto

Introducción al Procesamiento del Lenguaje Natural (NLP)

Unidad 2

Minería de Texto

Introducción al Procesamiento del Lenguaje Natural (NLP)



Contenido

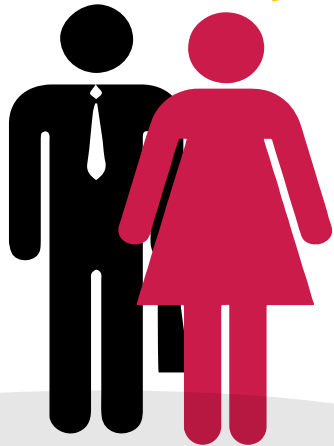
1. NLP: Perspectivas y Estado del Arte
2. Usos del NLP en la actualidad
3. Evolución y avances del NLP
4. Conceptos básicos de NLP
5. Software para NLP



Resumen

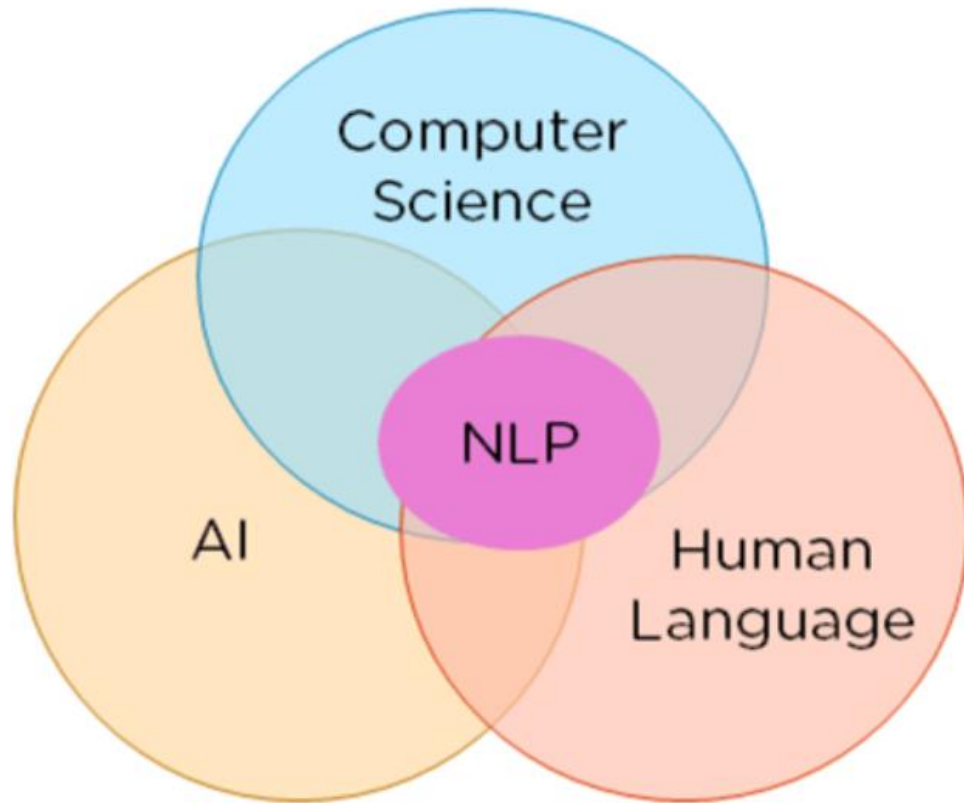
1. NLP: Perspectivas y Estado del Arte

¿Por qué es importante el NLP?



**Natural Language
Processing**

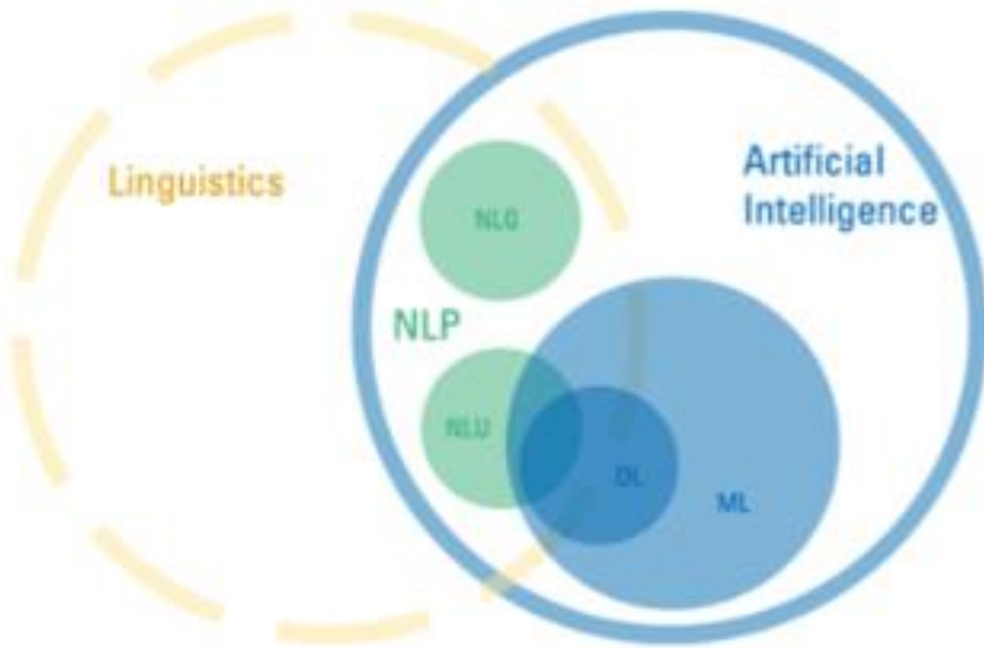
1. NLP: Perspectivas y Estado del Arte



- **NLP** es un área que intercepta tres grandes categorías: Lingüística, Ciencias de la Computación e Inteligencia Artificial
- Su objetivo es lograr la comunicación entre humanos y máquinas a través del lenguaje natural (lenguaje que utilizamos a diario para comunicarnos).

1. NLP: Perspectivas y Estado del Arte

Líneas de desarrollo de la IA



- **NLP** es un área que combina Lingüística, Ciencias de la Computación e Inteligencia Artificial
- Su objetivo es lograr la comunicación entre humanos y máquinas a través del lenguaje natural (lenguaje que utilizamos a diario para comunicarnos).

NLP es la intersección entre dos grandes categorías, la AI y lingüística.

1. NLP: Perspectivas y Estado del Arte

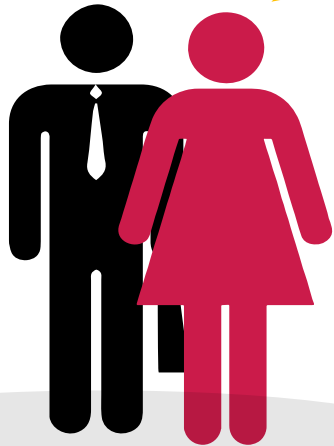
Tipología del Lenguaje Natural



1. NLP: Perspectivas y Estado del Arte

Natural Language Understanding

¿Qué es
el NLU?

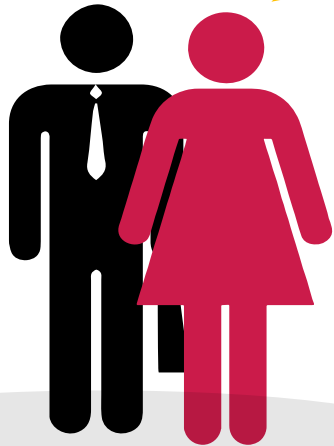


- El **NLU** es parte del NLP, encargada de tareas específicas que las máquinas puedan ejecutar en el proceso de comunicación con los seres humanos.
- Dichas tareas demuestran que la máquina no solo puede procesar nuestro lenguaje sino que también lo puede entender (sus respuestas demuestran dicha comprensión).

1. NLP: Perspectivas y Estado del Arte

Natural Language Generation

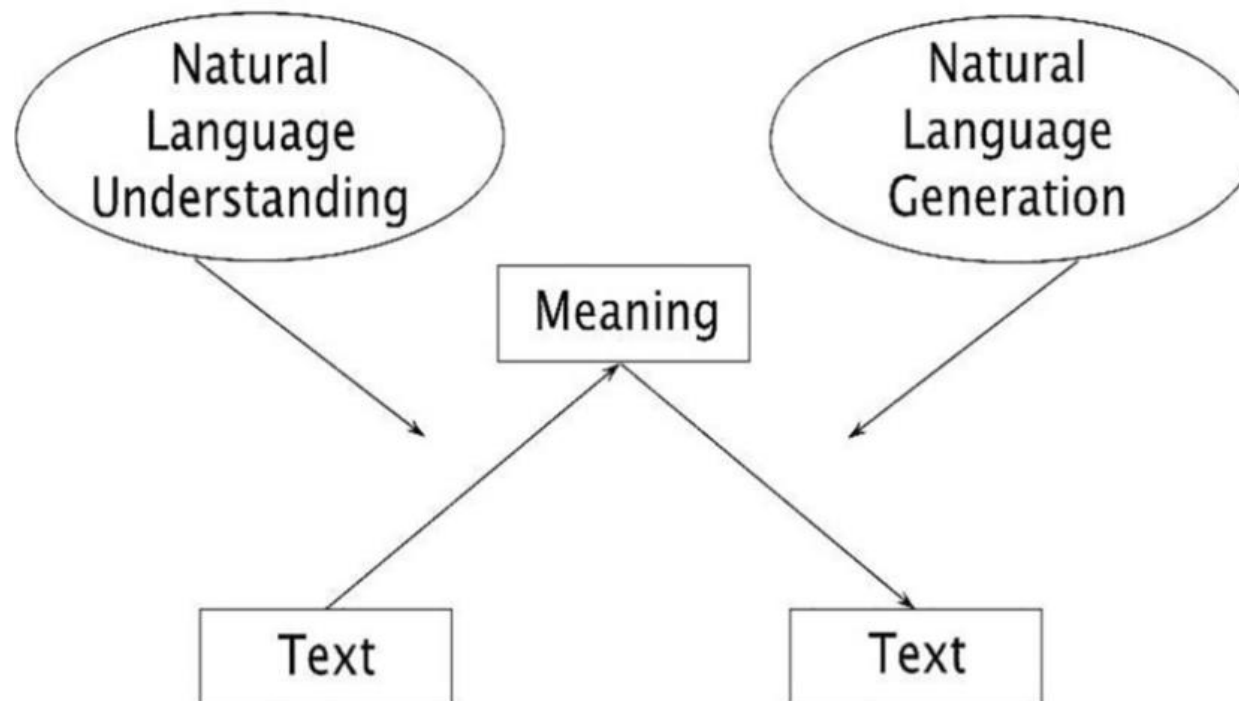
¿Qué es
el NLG?



- El **NLG** es parte del NLP, y comprende el uso de la programación de inteligencia artificial (IA) para producir narrativa escrita o hablada a partir de un conjunto de datos.
- Es utilizado en los bots para dar una respuesta exacta y clara para un mensaje de un usuario teniendo en cuenta el lenguaje abstracto y ambiguo de los humanos

1. NLP: Perspectivas y Estado del Arte

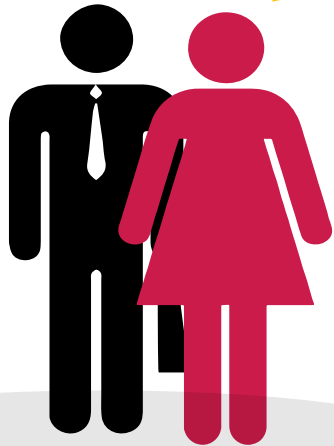
$$\text{NLP} = \text{NLU} + \text{NLG}$$



1. NLP: Perspectivas y Estado del Arte

Si, porque:

¿Es el NLP
difícil?



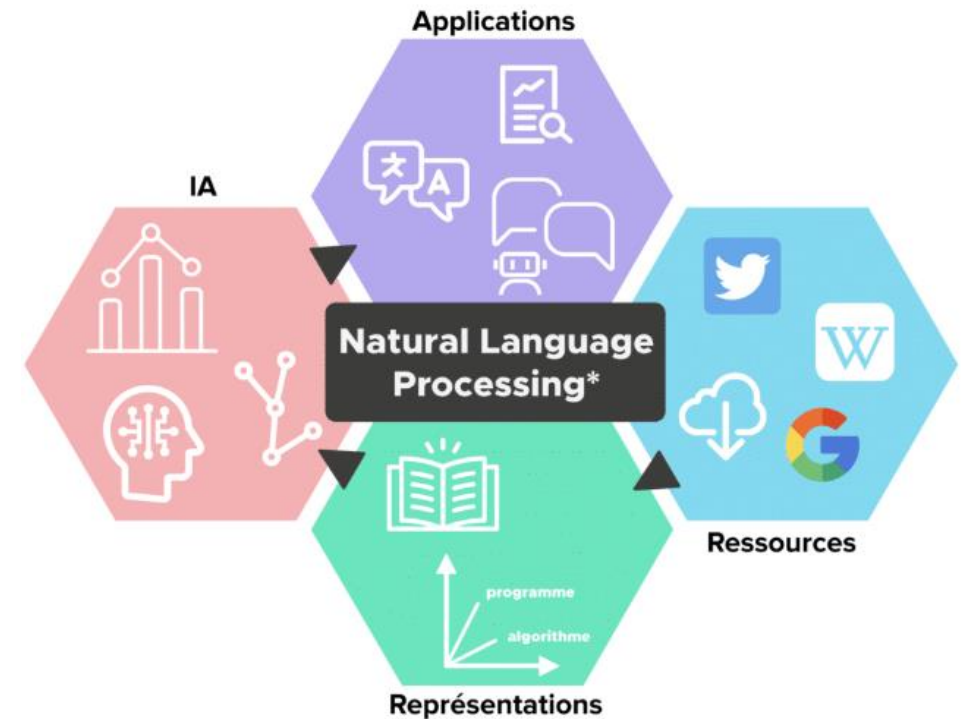
El lenguaje humano es **difuso**,
ambiguo y requiere **mucho**
contexto

1. NLP: Perspectivas y Estado del Arte

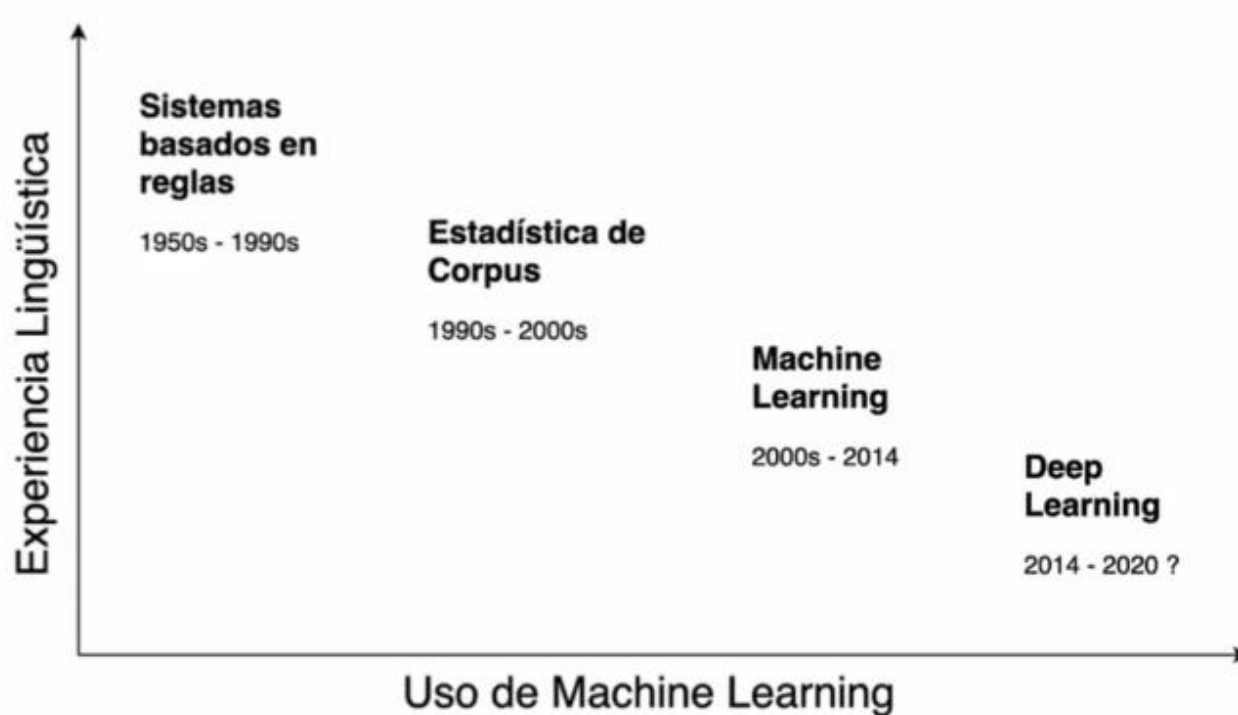


2. Usos del NLP en la actualidad

- Maquinas de búsqueda (ej. buscador de Google)
- Traducción de texto (ej. Google translate)
- Chatbots (interfases de comunicación natural) y asistentes virtuales
- Análisis de discurso (escrito)
- Análisis de Sentimientos
- Clasificación de texto
- Reconocimiento del habla, etc.



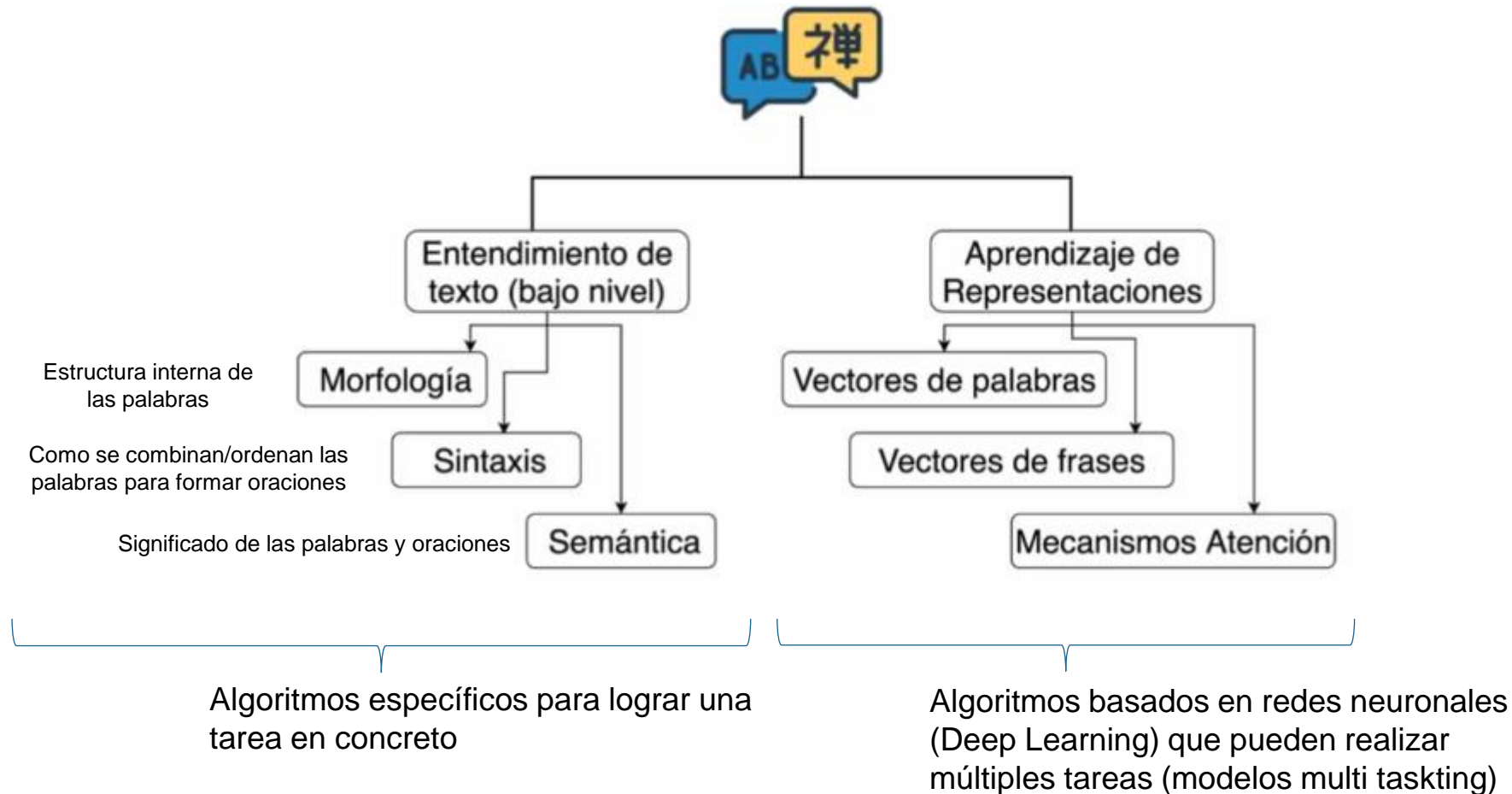
3. Evolución del NLP



- **1950 -1990** Se programan las reglas del lenguaje de forma tradicional
- **1990 -2000** Se incorpora la estadística a una colección de diferentes textos (corpus)
- **2000 -2014** Intenso uso de las GPU
- **2014 - Hoy** Tendencia para crear algoritmos de NLP y NLU más eficientes.

3. Evolución del NLP

Áreas desarrolladas en NLP



3. Evolución del NLP

Principales pasos para el entendimiento del Texto

(1) Etiquetado gramatical o Part-Of-Speech Tagging (POST)

Está relacionado con la morfología: determinar la función de cada palabra en forma individual.

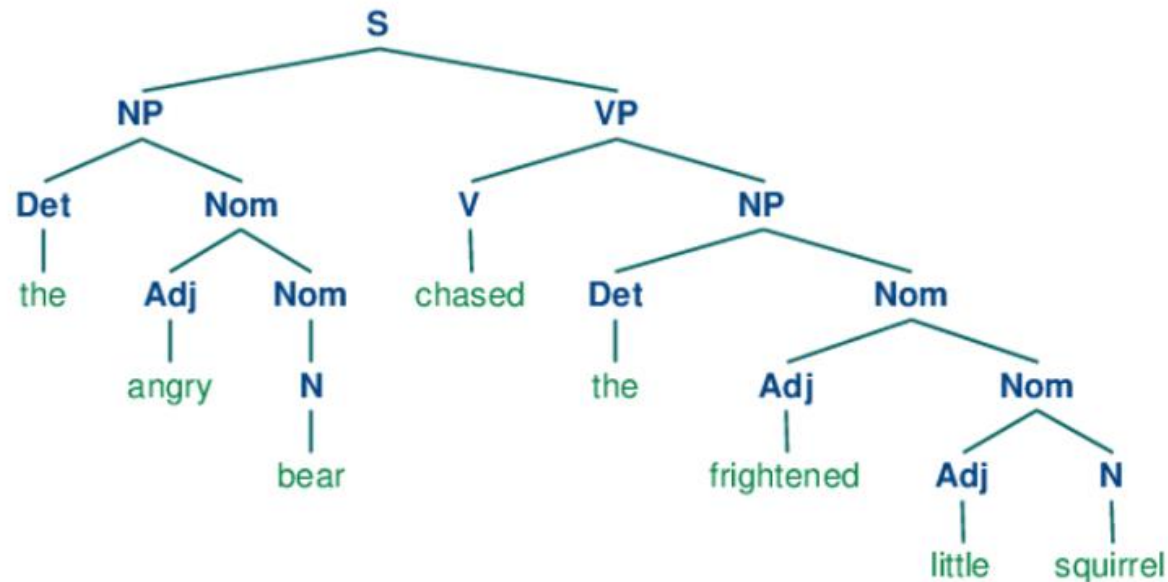
Why	not	tell	someone	?
adverb	adverb	verb	noun	punctuation mark, sentence closer

3. Evolución del NLP

Principales pasos para el entendimiento del Texto

(2) Parse trees

- Aquí se emplean los conocimientos extraídos de la sintaxis para entender la estructura de las oraciones.
- Se utilizan **diagramas de análisis sintáctico** con los que se pueden dividir las oraciones en sintagmas denominados parse trees

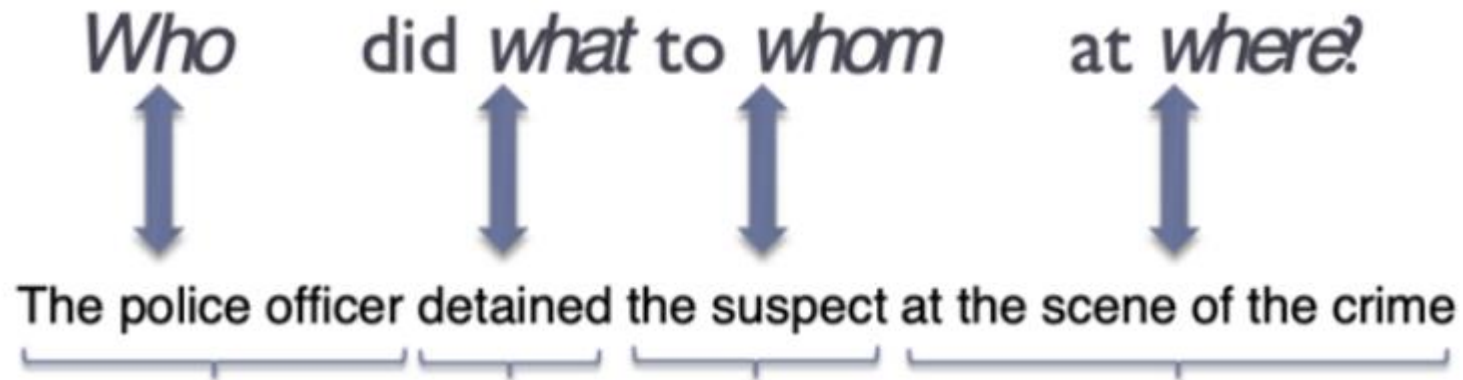


3. Evolución del NLP

Principales pasos para el entendimiento del Texto

(3) Semántica

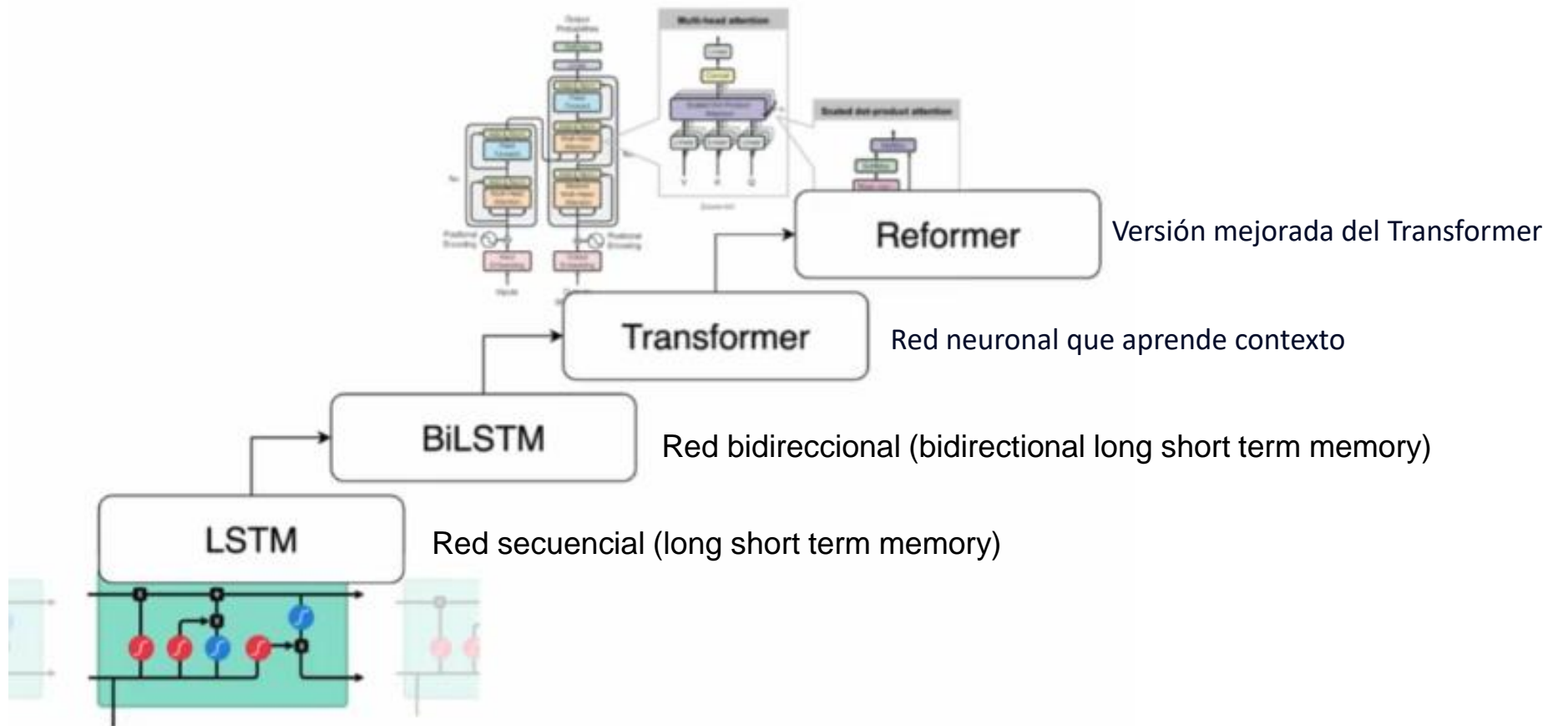
- Se intenta determinar el significado de una palabra con ayuda de las palabras que le preceden o le siguen.
- Estas diferencias se pueden aprender mediante el uso de corpus de textos, en los que el significado de cada palabra se reproduce su exactitud.



3. Evolución del NLP

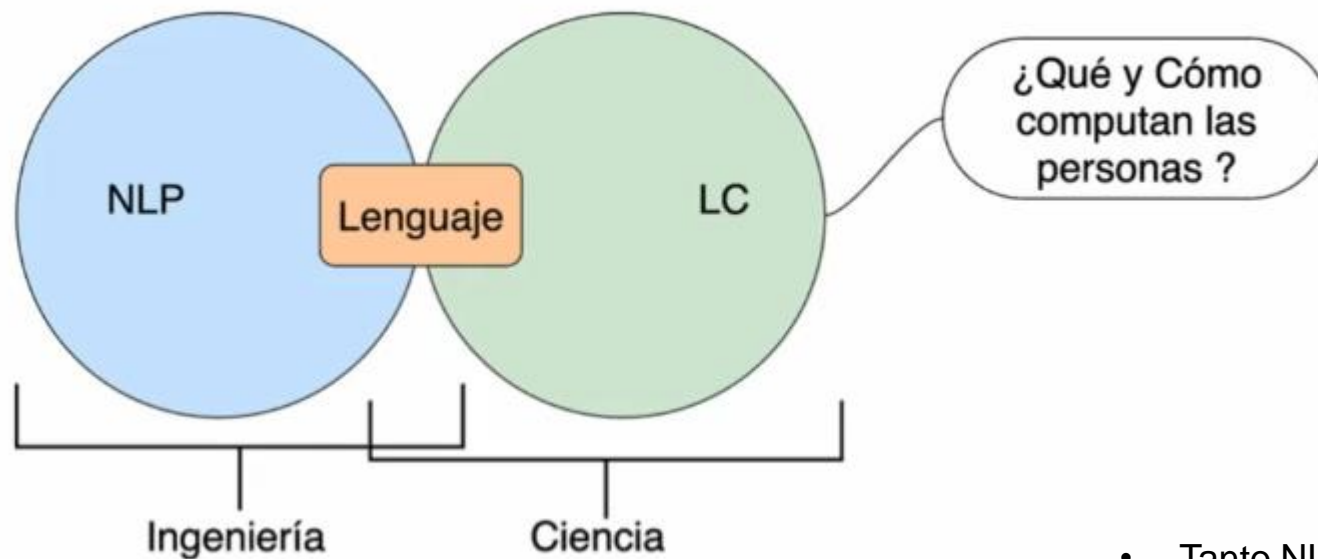
AVANCES DEL NLP

(dentro del Aprendizaje de Representaciones)



4. Conceptos básicos de NLP

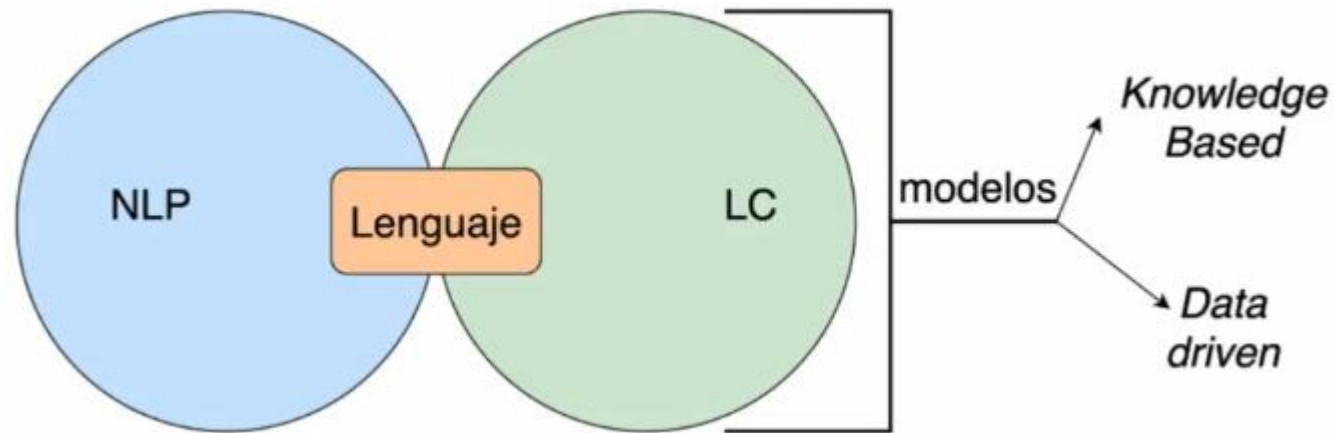
Lingüística Computacional (LC)



- Tanto NLP como la LC tienen como objeto de estudio el lenguaje pero desde perspectivas diferentes

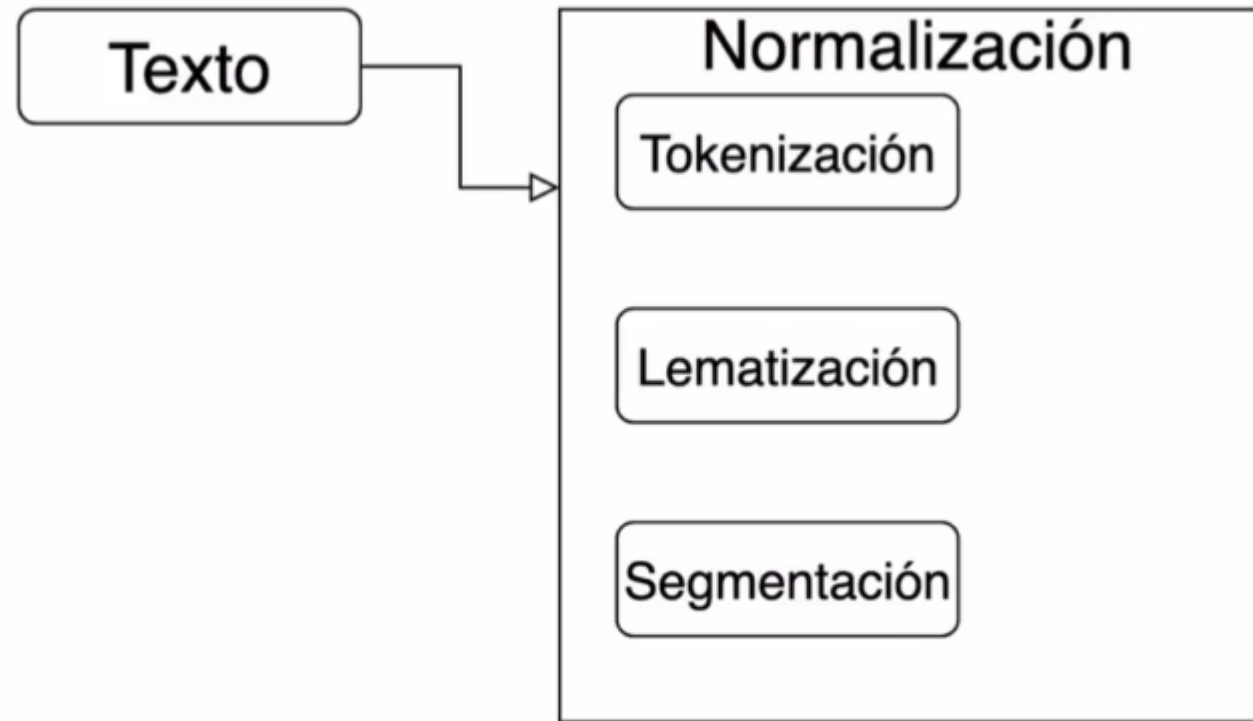
4. Conceptos básicos de NLP

Lingüística Computacional (LC)



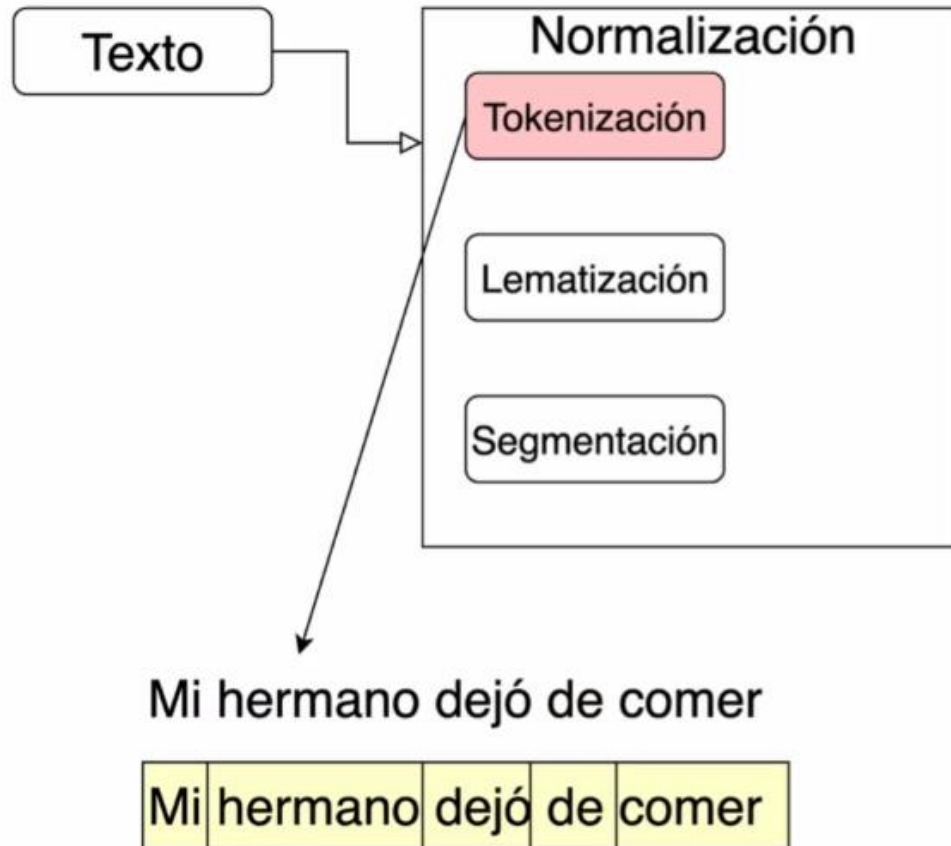
4. Conceptos básicos de NLP

Normalización del Texto



4. Conceptos básicos de NLP

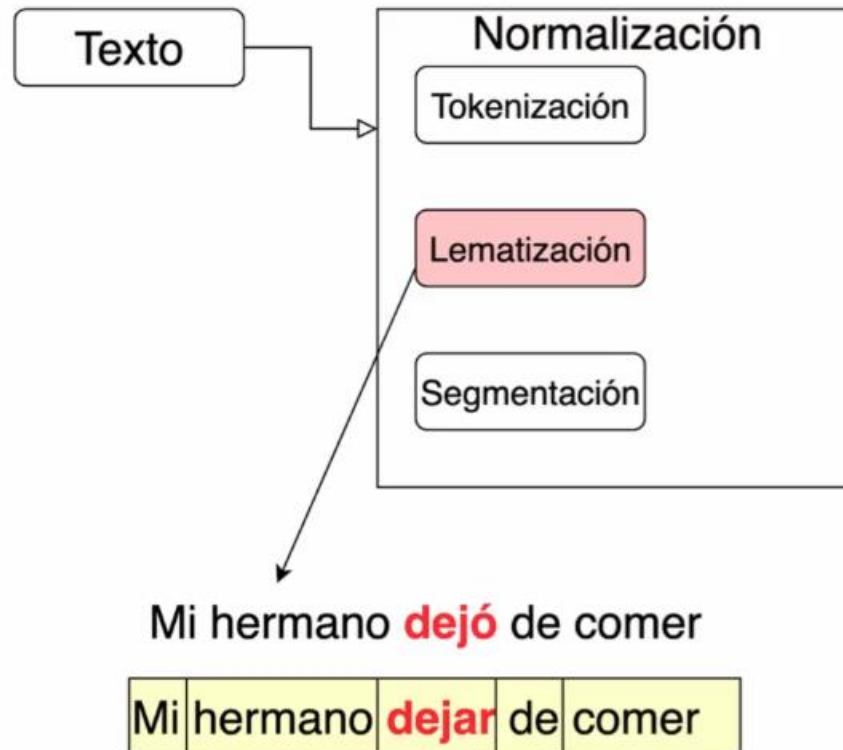
Normalización del Texto: **Tokenización**



- Consiste en separar una frase en tokens o unidades mínimas lingüísticas (palabras)

4. Conceptos básicos de NLP

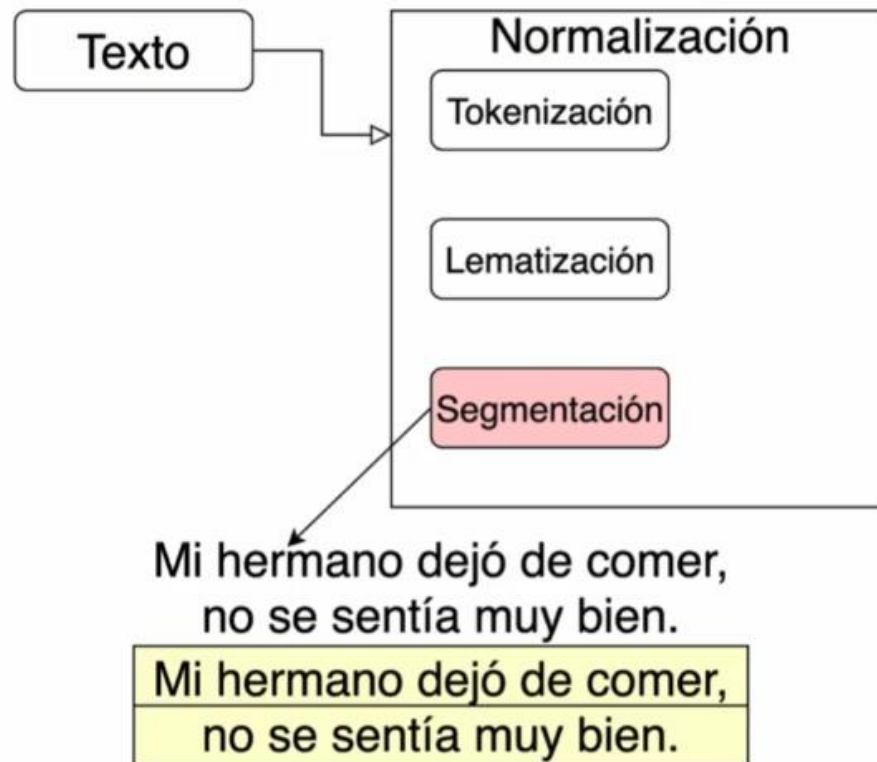
Normalización del Texto: **Lematización**



- Consiste en convertir cada una de las palabras de una frase a su raíz fundamental. En el caso del verbo en su conjugación “infinitiva”.

4. Conceptos básicos de NLP

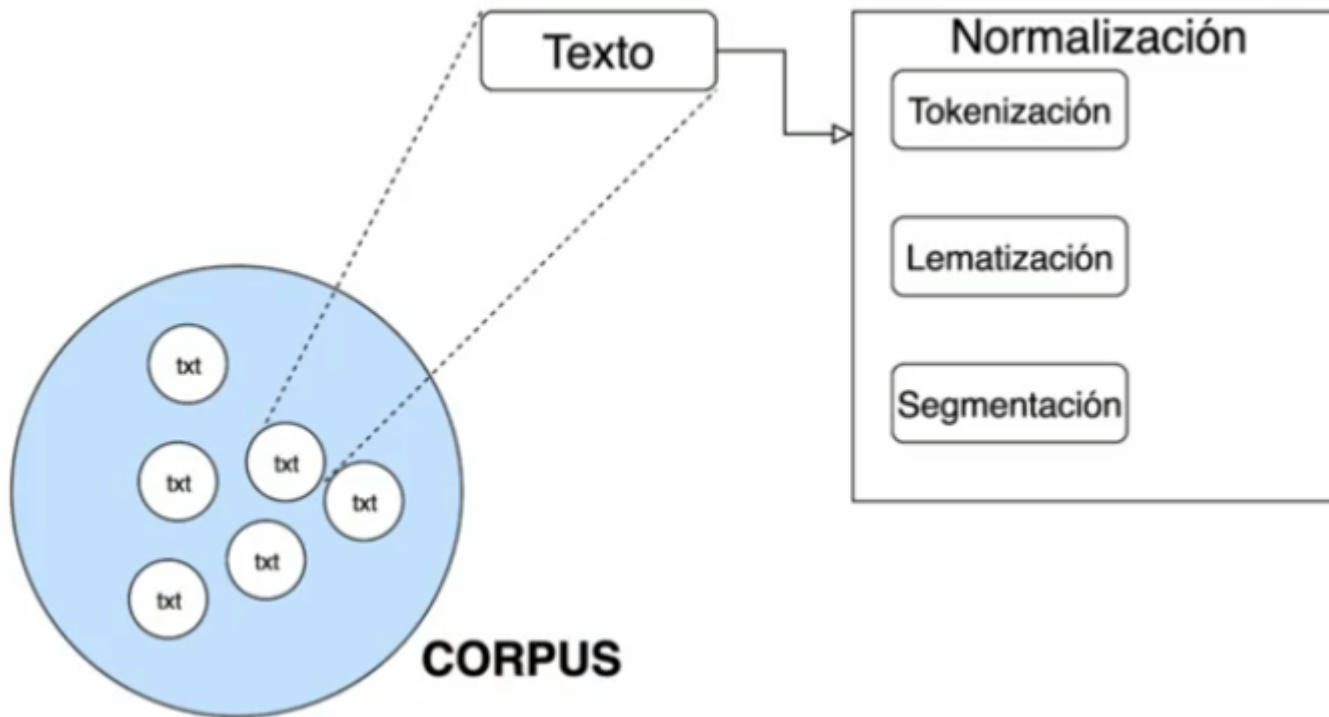
Normalización del Texto: **Segmentación**



- Consiste en segmentar o separar las frases del texto (generalmente las separadas por comas).

4. Conceptos básicos de NLP

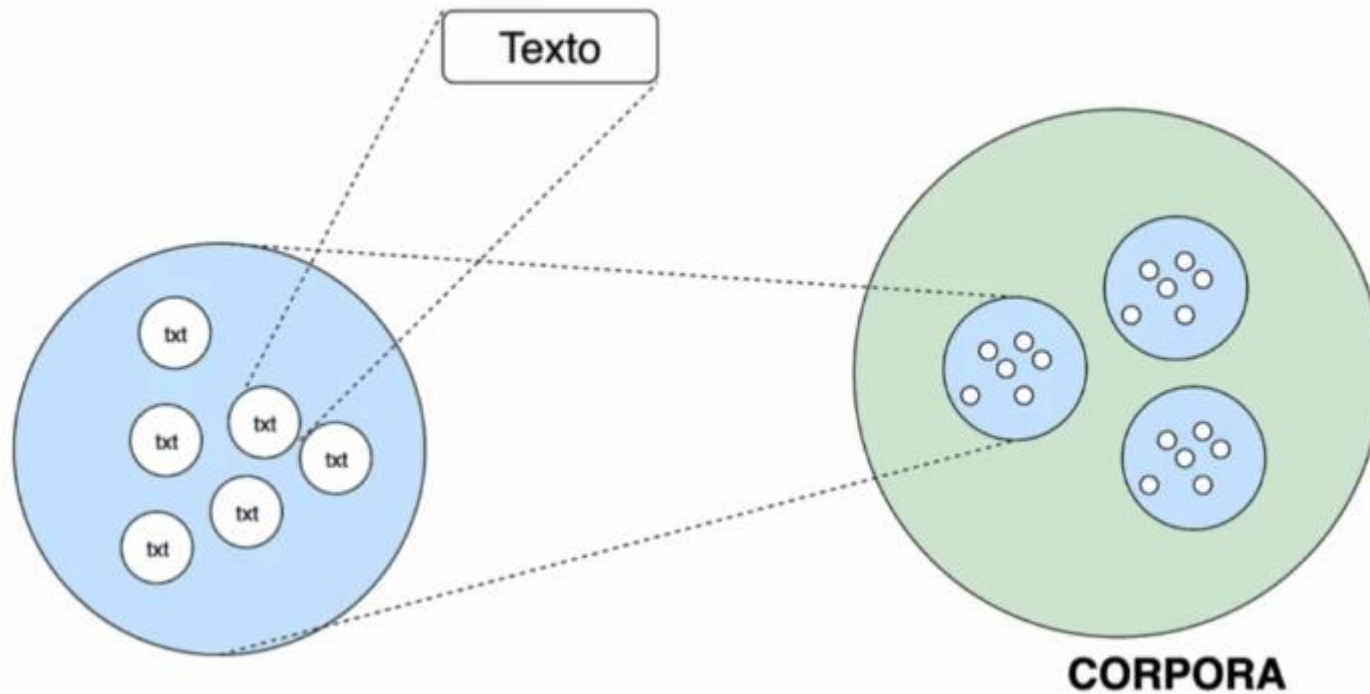
CORPUS



- Consiste en una colección de textos que componen un dataset que servirá para entrenar un algoritmo de aprendizaje de representaciones.

4. Conceptos básicos de NLP

CORPORA



- Consiste en una colección de colecciones de texto.

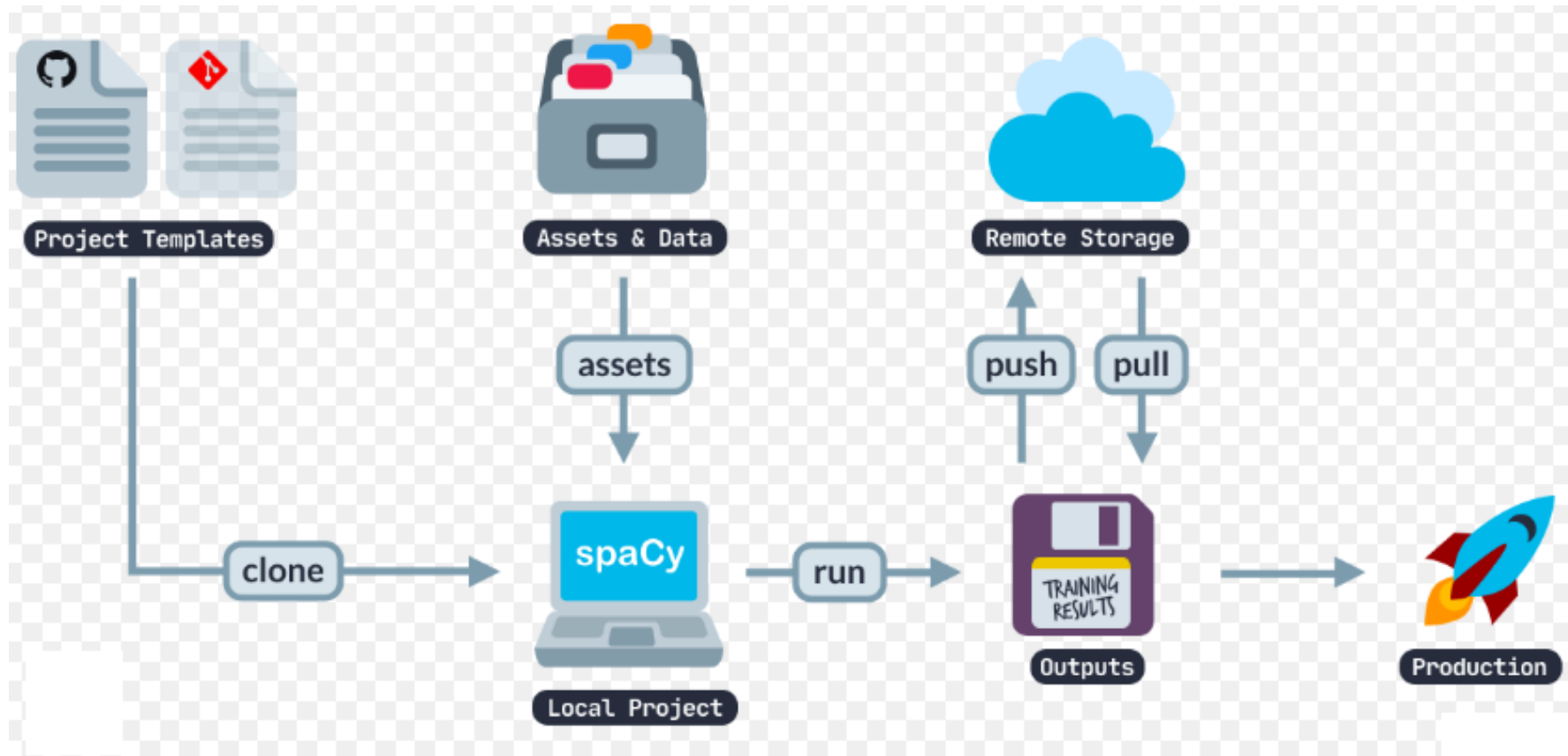
5. Software para NLP



Natural Language
Tool Kit (NLTK)

5. Software para NLP

SPACY



5. Software para NLP

Otro software:

MonkeyLearn

Aylien

IBM Watson

Google Cloud

Amazon Comprehend

Stanford Core NLP

TextBlob

GenSim

PREGUNTAS

Dudas y opiniones