

Comunidad MyFuture-AI

# Introducción al Procesamiento de Lenguaje Natural (NLP)

By Laura López

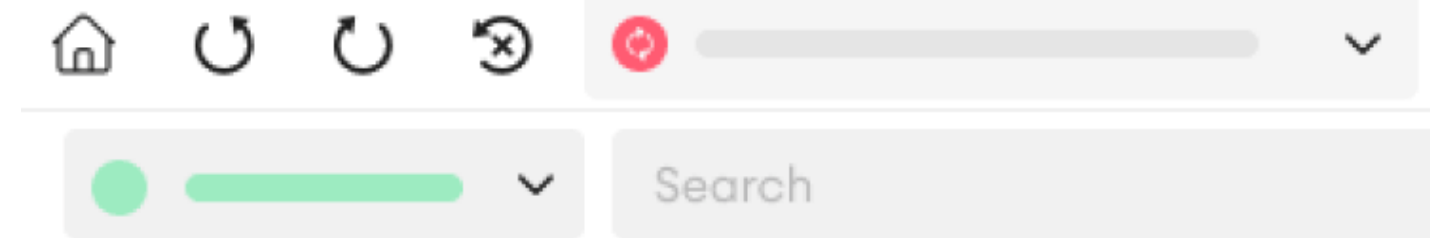
# Agenda

Esto es lo que abarcaremos:

- Embeddings
- Preprocesamiento de texto
- Análisis NLP
- Práctica

# Procesamiento de Lenguaje Natural (NLP)

“Es una tecnología de machine learning que brinda a las computadoras la capacidad de interpretar, manipular y comprender el lenguaje humano.” - Amazon



Despite the rapid advances in AI, computers are still challenged in matching the precision of human language. The training data here is as important as the accurate the input data annotation, and the model prediction.

How do we annotate data, though? There are many ways to go with this one, but it all depends on the purposes of this article, we'll take a look at some of the boxes as one of the most extensively used techniques. Moving forward, we'll work on the following:

¿Cómo representas  
un texto con  
números?

Asignando a cada letra un valor



Asignando a cada palabra un  
valor

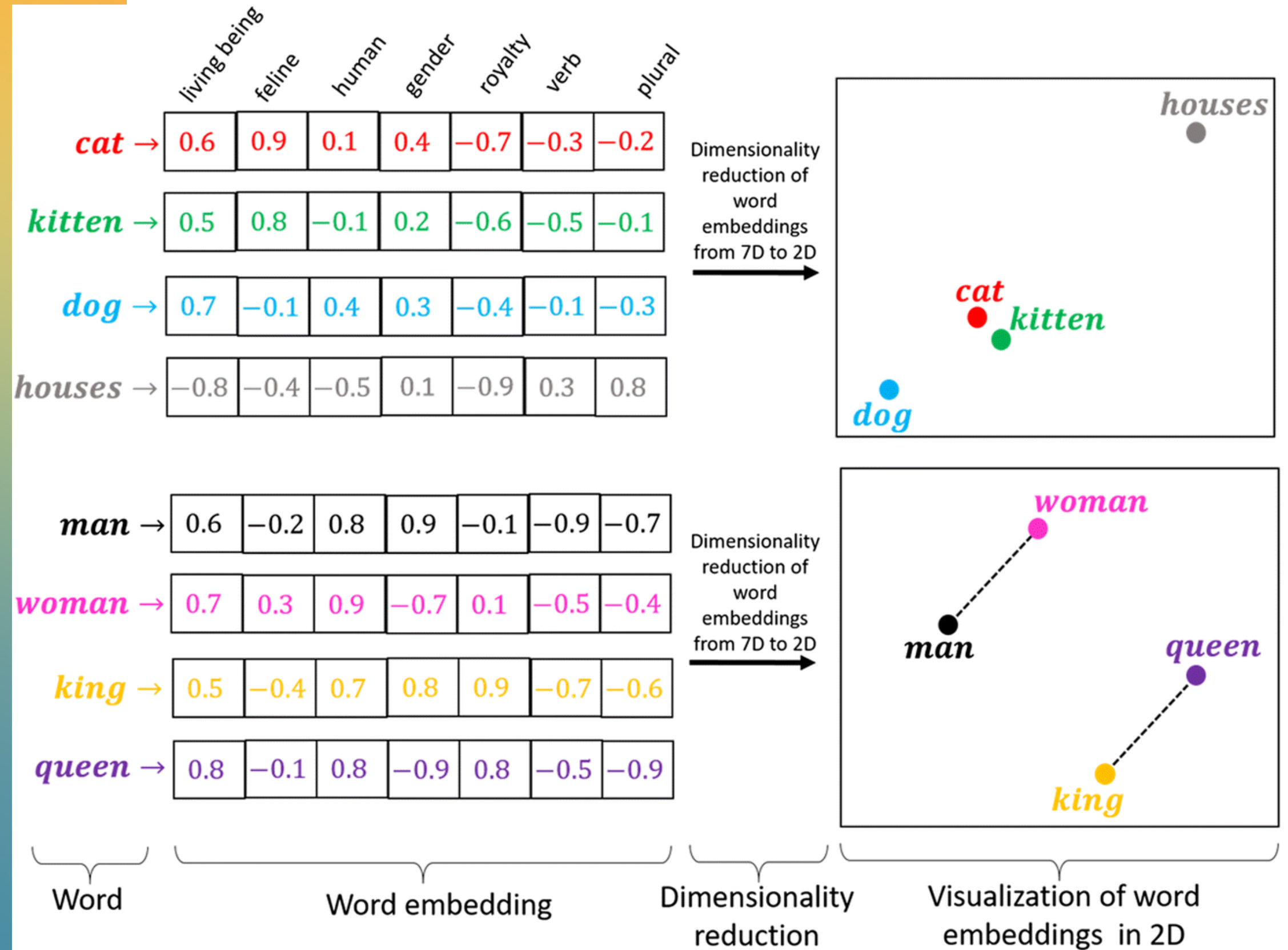


Asociación de palabras en  
contextos similares





# Embeddings



# Preprocesamiento

“Hey Amazon - my package never arrived  
[https://www.amazon.com/gp/css/order-history?ref\\_=nav\\_orders\\_first](https://www.amazon.com/gp/css/order-history?ref_=nav_orders_first)  
PLEASE FIX ASAP! @amazonhelp”

**Normalizar**

“hey amazon my package  
never arrived please fix asap”

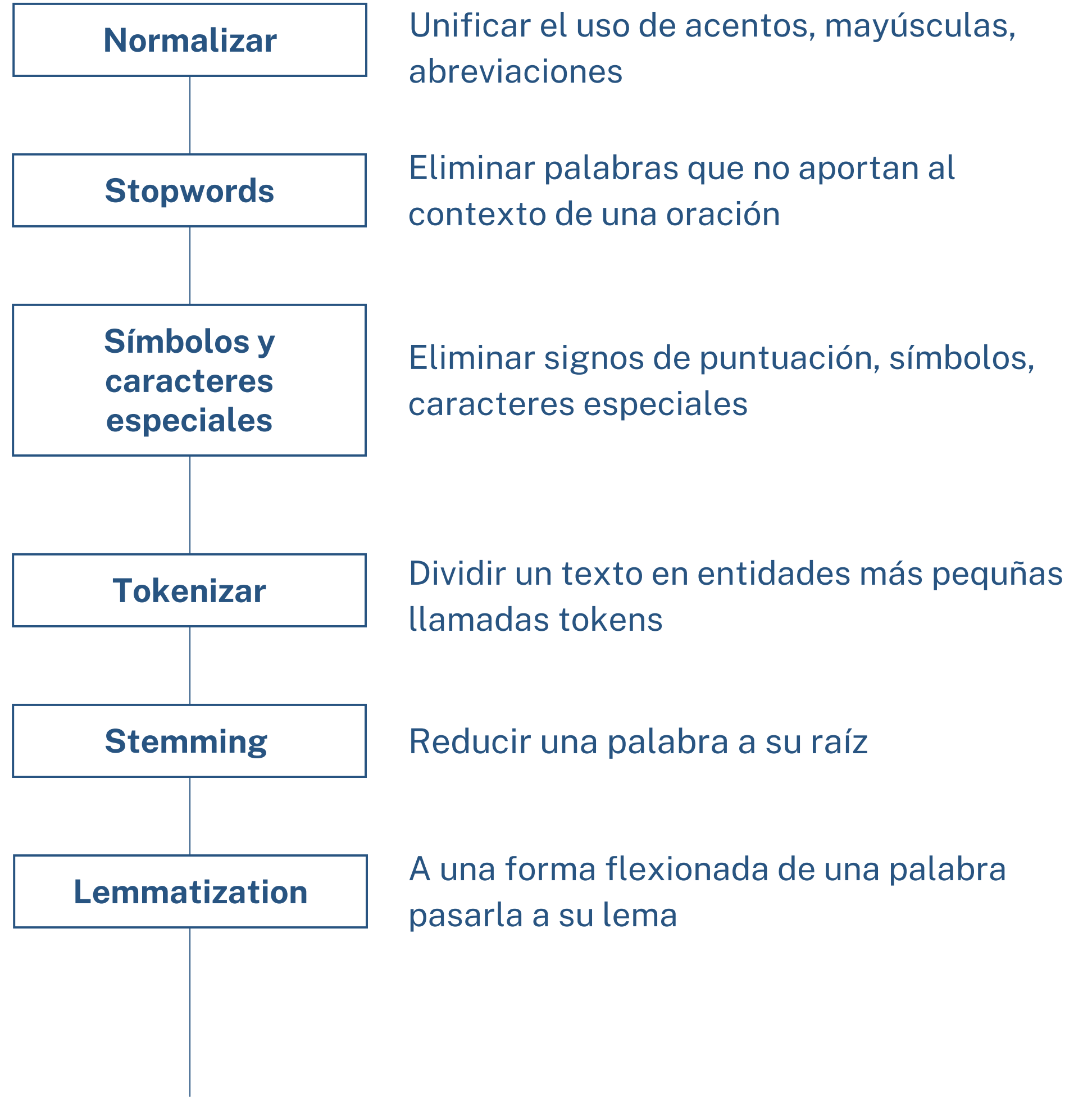
**Eliminar stopwords**

“amazon package never arrived fix  
asap”

**Tokenizar**

["amazon", "package",  
"never", "arrived", "fix",  
"asap"]

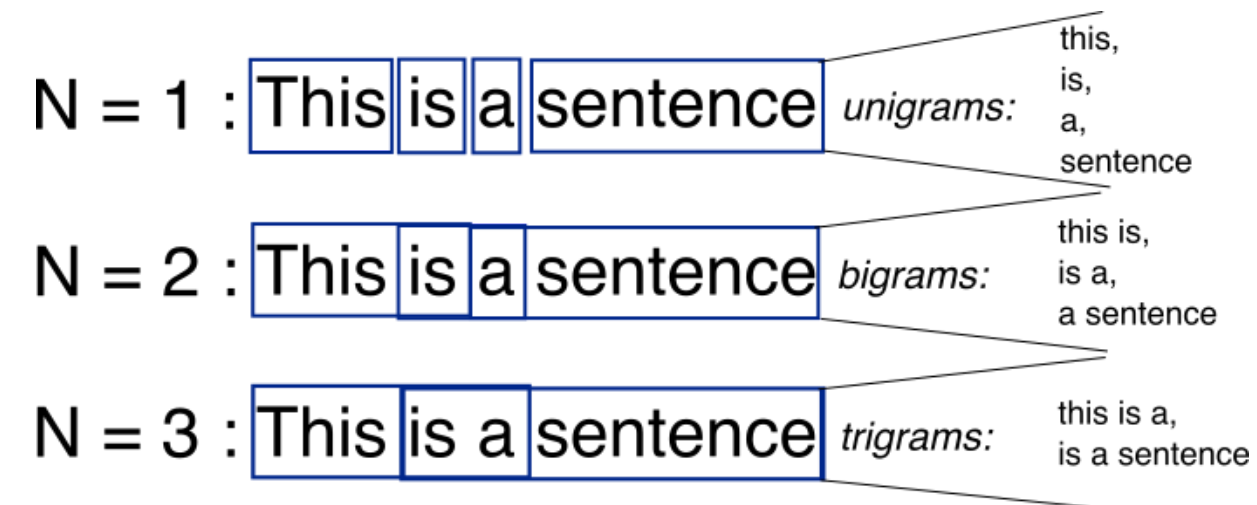
# Preprocesamiento



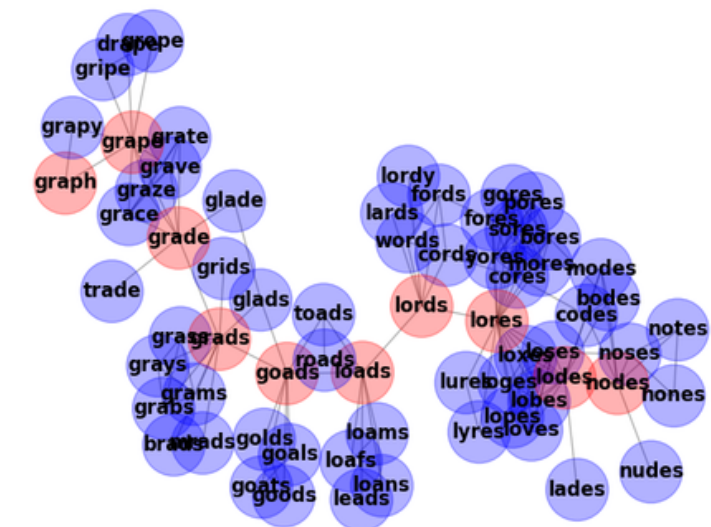
# ¿Qué podemos hacer con NLP?



## Wordcloud



## N-Grams



## Grafos de palabras

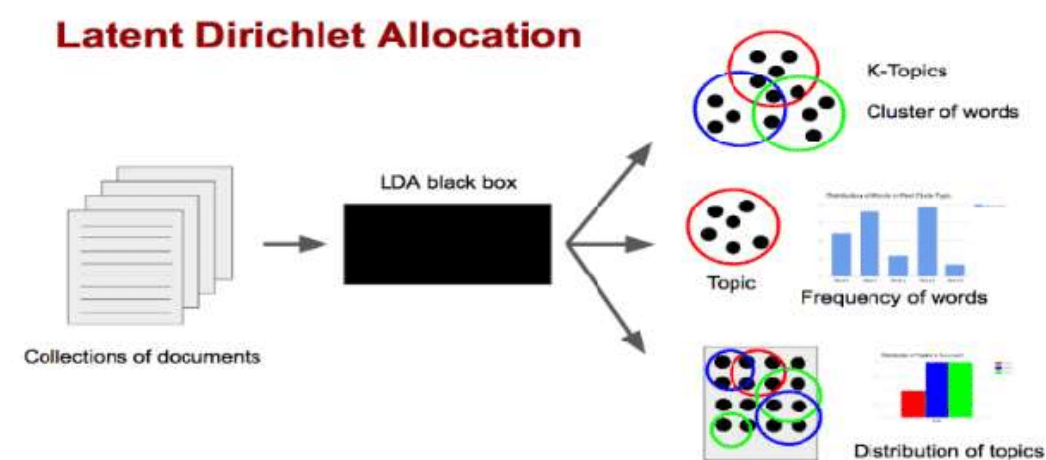


# ¿Qué podemos hacer con NLP?

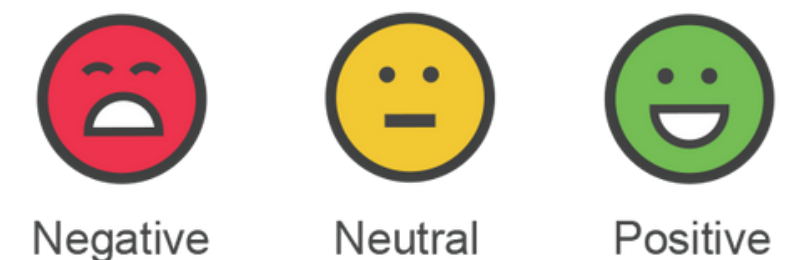
When **Sebastian Thrun** PERSON started at **Google** ORG in **2007** DATE, few people outside of the company took him seriously. "I can tell you very senior CEOs of major **American** NORP car companies would shake my hand and turn away because I wasn't worth talking to," said **Thrun** PERSON, now the co-founder and CEO of online higher education startup Udacity, in an interview with **Recode** ORG **earlier this week** DATE.

A little **less than a decade later** DATE, dozens of self-driving startups have cropped up while automakers around the world clamor, wallet in hand, to secure their place in the fast-moving world of fully automated transportation.

Named Entity Recognition (NER) y Part Of Speech Tagging (POS)



Latent Dirichlet Allocation (LDA)



Análisis de sentimientos y emociones

Vamos al código con  
algo de PySpark



# ¿Tienes alguna pregunta?

[github.com/lauralpezb/NLP\\_MyFutureAI](https://github.com/lauralpezb/NLP_MyFutureAI)  
<http://bit.ly/3RSzYAy>

@lauralpezb