

Procesamiento de lenguaje natural

Materia: Procesamiento de lenguaje natural
Blanca Vázquez

Introducción al procesamiento de lenguaje natural

¿Qué es el lenguaje?

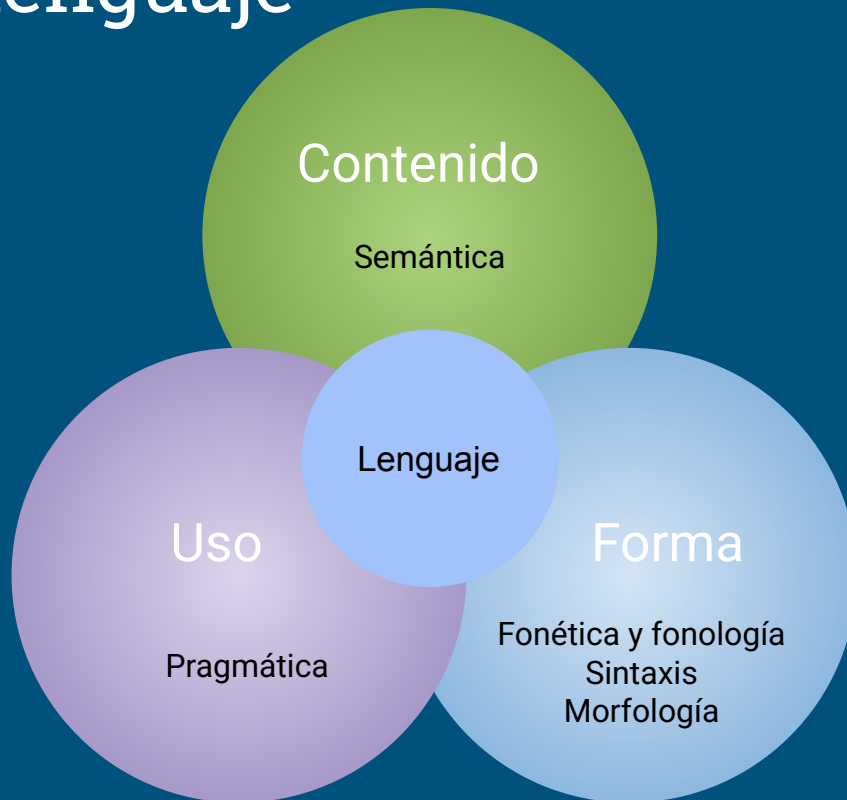
¿Qué es el lenguaje?

- Lo usamos todos los días para comunicarnos
- Pocas veces nos detenemos a conocer “cómo funciona”
- Su estudio se ha desarrollado a partir de la rama humanística y de las ciencias sociales.

Componentes del lenguaje

- Fonética
 - Sonidos del habla
- Fonología
 - Organización lingüística de los sonidos
- Morfología
 - Estructura de las palabras
- Sintaxis
 - Orden de las palabras y frases
- Semántica
 - Significado literal de frases y enunciados
- Pragmática
 - Significado en el contexto del discurso

Aspectos del lenguaje



Características del lenguaje natural

- Ambigüedad
 - Sintáctica
 - Semántica
 - Pragmática
- Variantes
- Evolución

Ambigüedad sintáctica

- Vi al hombre en el parque con binoculares.
 - La casa del perro grande.
 - Eligió un coche rápido.

Ambigüedad léxica

- Te esperaré en el **banco**
- Me gustan los **entremeses**
- Es una persona **noble**

Ambigüedad léxica - semántica

- El capitán dijo: “¡Bajen las velas!” Y los de arriba se quedaron sin luz.
- ¿Qué pasa si un elefante se queda de pie encima de una pata?
- Estamos buscando al asesino en tres estados: reduzcan la búsqueda a un solo estado, el sólido es el más probable.

Ambigüedad fonológica

- vaca / baca
- ora / hora
- tuvo / tubo
- grave / grabe

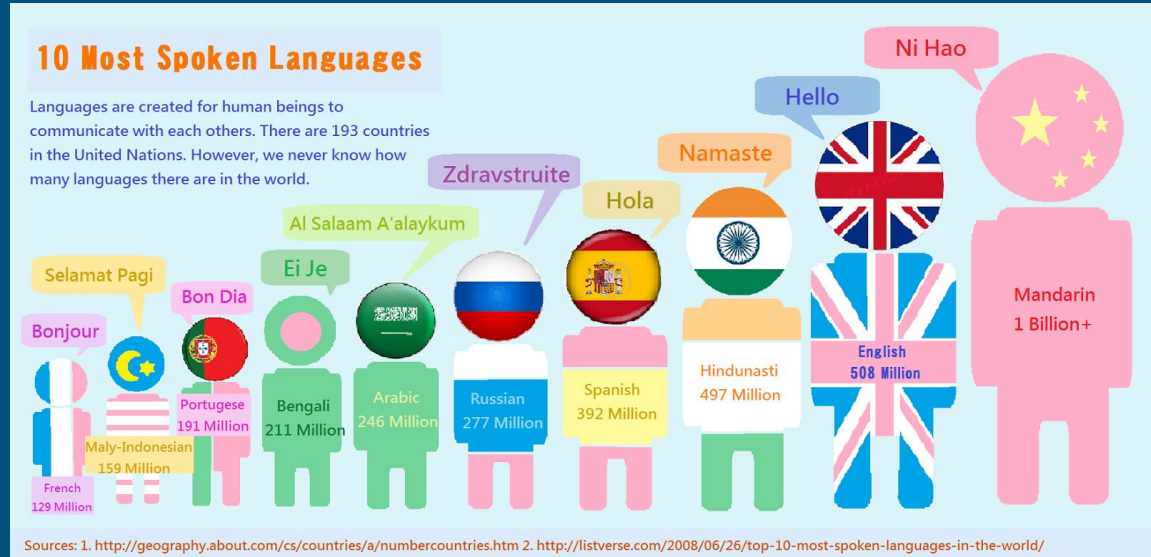
Ambigüedad pragmática

- Iba a comprar un café, pero me faltan 9 personas:
AVISO: sólo se atenderá a grupos de 10 personas.
- ¡Qué bonito día!

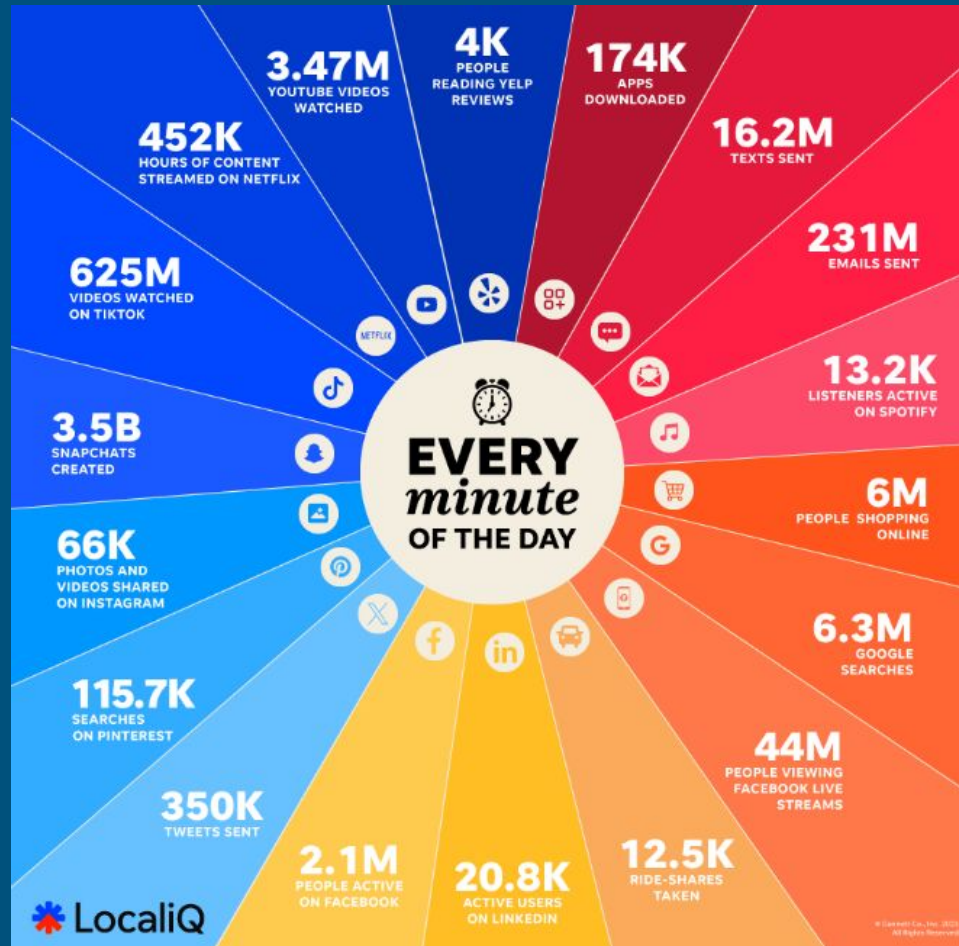
¿Sabías que en México
hay 68 lenguas
indígenas, además del
español?



¿Sabías que en el mundo existen 7,159 lenguajes activos?



<https://www.ethnologue.com/insights/how-many-languages/>



Temario

01	Procesamiento de lenguaje natural
02	Word embeddings
03	Modelos basados en recurrencias
04	Modelos de lenguaje

- Introducción al PLN
- Expresiones regulares
- Etiquetado y procesamiento de texto

Temario

01	Procesamiento de lenguaje natural
02	Word embeddings
03	Modelos basados en recurrencias
04	Modelos de lenguaje

- Introducción al PLN
- Expresiones regulares
- Etiquetado y procesamiento de texto

- Introducción a los word embeddings
- Word2Vec, GloVe, FastText
- Similitud coseno, La ley de Zipf

Temario

01	Procesamiento de lenguaje natural	<ul style="list-style-type: none">• Introducción al PLN• Expresiones regulares• Etiquetado y procesamiento de texto
02	Word embeddings	<ul style="list-style-type: none">• Introducción a los word embeddings• Word2Vec, GloVe, FastText• Similitud coseno, La ley de Zipf
03	Modelos basados en recurrencias	<ul style="list-style-type: none">• RNN• LSTM, GRU• Aplicaciones: análisis de sentimientos y generación de música
04	Modelos de lenguaje	

Temario

01	Procesamiento de lenguaje natural	<ul style="list-style-type: none">• Introducción al PLN• Expresiones regulares• Etiquetado y procesamiento de texto
02	Word embeddings	<ul style="list-style-type: none">• Introducción a los word embeddings• Word2Vec, GloVe, FastText• Similitud coseno, La ley de Zipf
03	Modelos basados en recurrencias	<ul style="list-style-type: none">• RNN• LSTM, GRU• Aplicaciones: análisis de sentimientos y generación de música
04	Modelos de lenguaje	<ul style="list-style-type: none">• Arquitectura encoder – decoder• Mecanismos de atención y transformers• Embeddings contextualizados: BERT• Large language models

Horario

- Lunes 11:30 am a 13:00 pm
- Miércoles 11:30 am a 13:00 pm

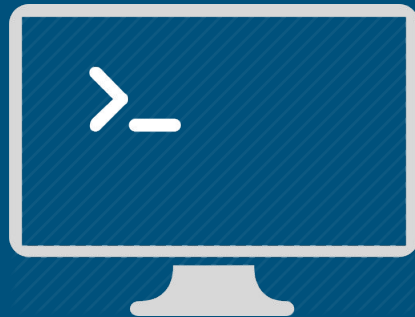


Ambientes de programación

Google Colab

Python

Jupyter notebook



Criterios de evaluación

- Tareas (40%)
- Participación (10%)
- Proyecto final (50%)



Repositorio y contacto

- Github: <https://github.com/blancavazquez/PLN>
- Correo: blancavazquez2013@gmail.com