

INTRODUCCIÓN A WORD Y SENTENCES EMBEDDINGS Y SUS APLICA

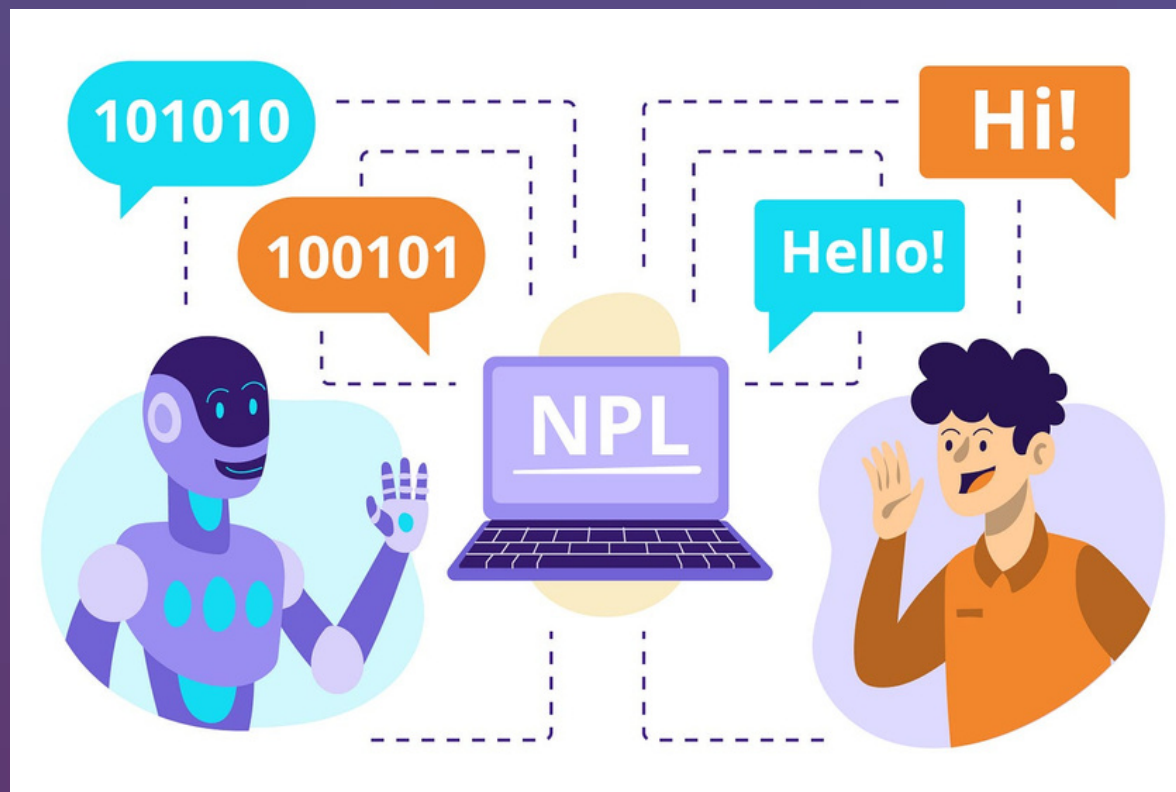
Gabriel Iturra-Bocaz, Msc Computer Science

<https://giturra.cl/>

Objetivos

- ¿Qué es el procesamiento de lenguaje natural? (NLP)
- Aplicaciones
- Como entienden los Computadores 🤖
- Representaciones Vectoriales 📋

¿Qué es el procesamiento de lenguaje natural?



Es una disciplina de la inteligencia artificial que busca que las máquinas entiendan e interpreten el lenguaje humano.

Principalmente se basa en técnicas de Estadística, Machine Learning y Deep Learning.

Pero... ¿Qué es el Lenguaje Natural?



El lenguaje natural se refiere al lenguaje que los seres humanos usan para comunicarse entre sí, ya sea hablado o escrito.

A diferencia de los lenguajes de programación o códigos formales, el lenguaje natural es complejo, ambiguo y evoluciona con el tiempo.



Pero... porque se dice que es ambiguo...

Condiseremos la frase: “Él no sabe si volar es seguro”



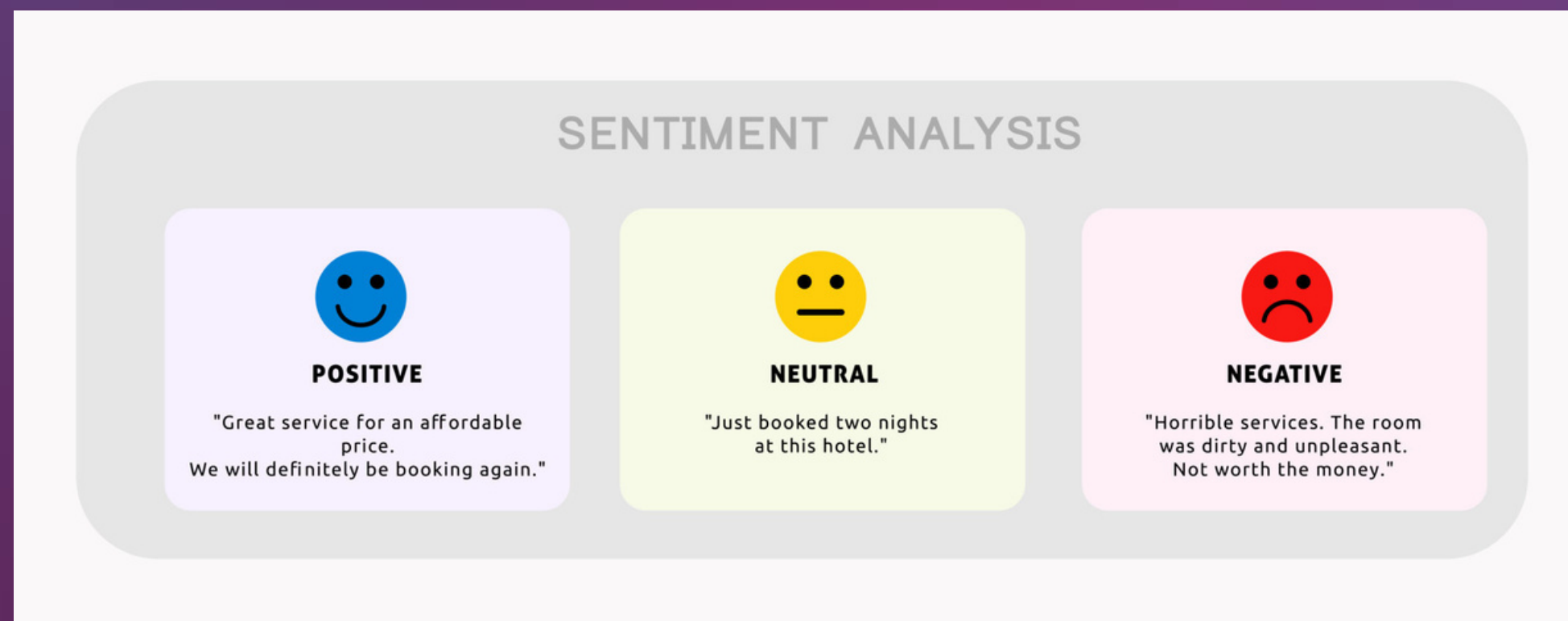
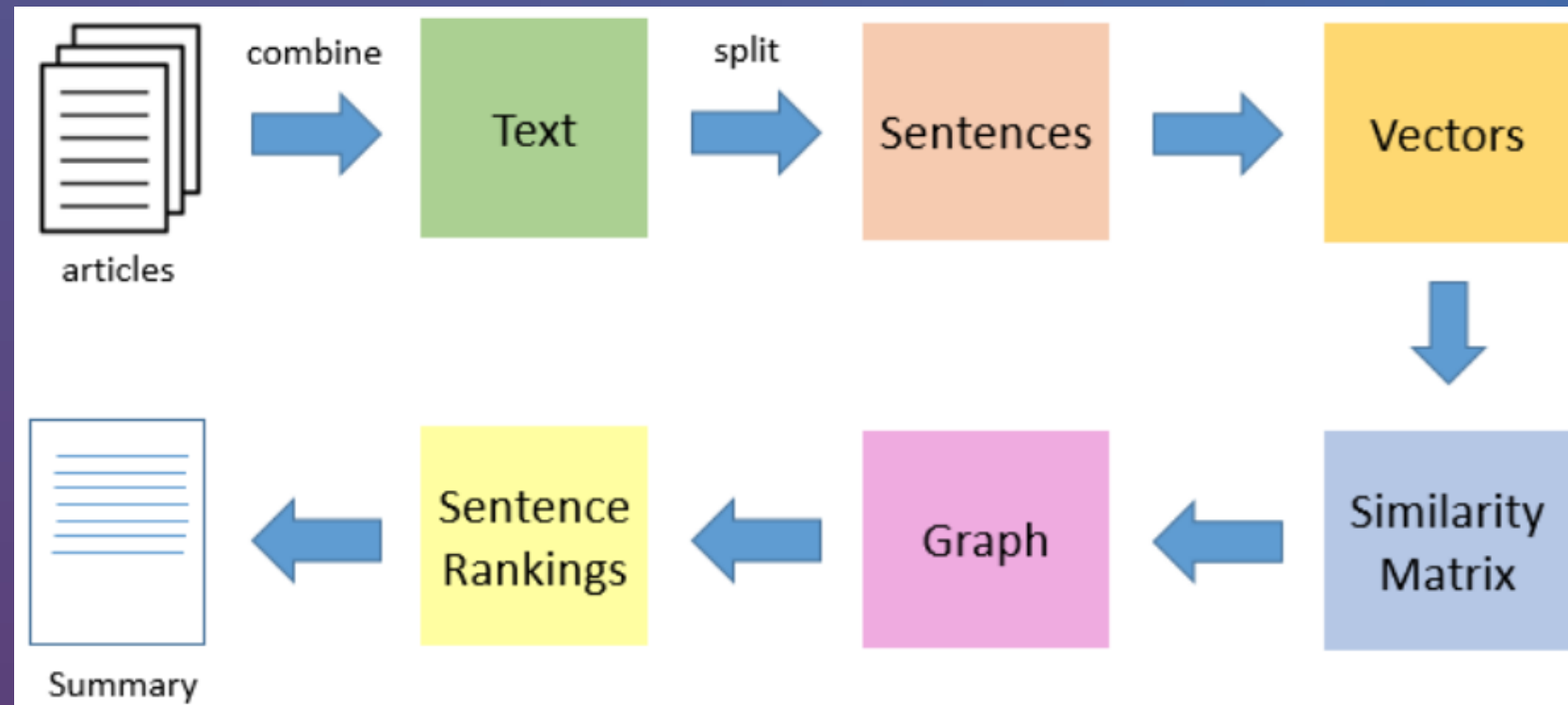
Dado que muchas frases dependen del contexto...



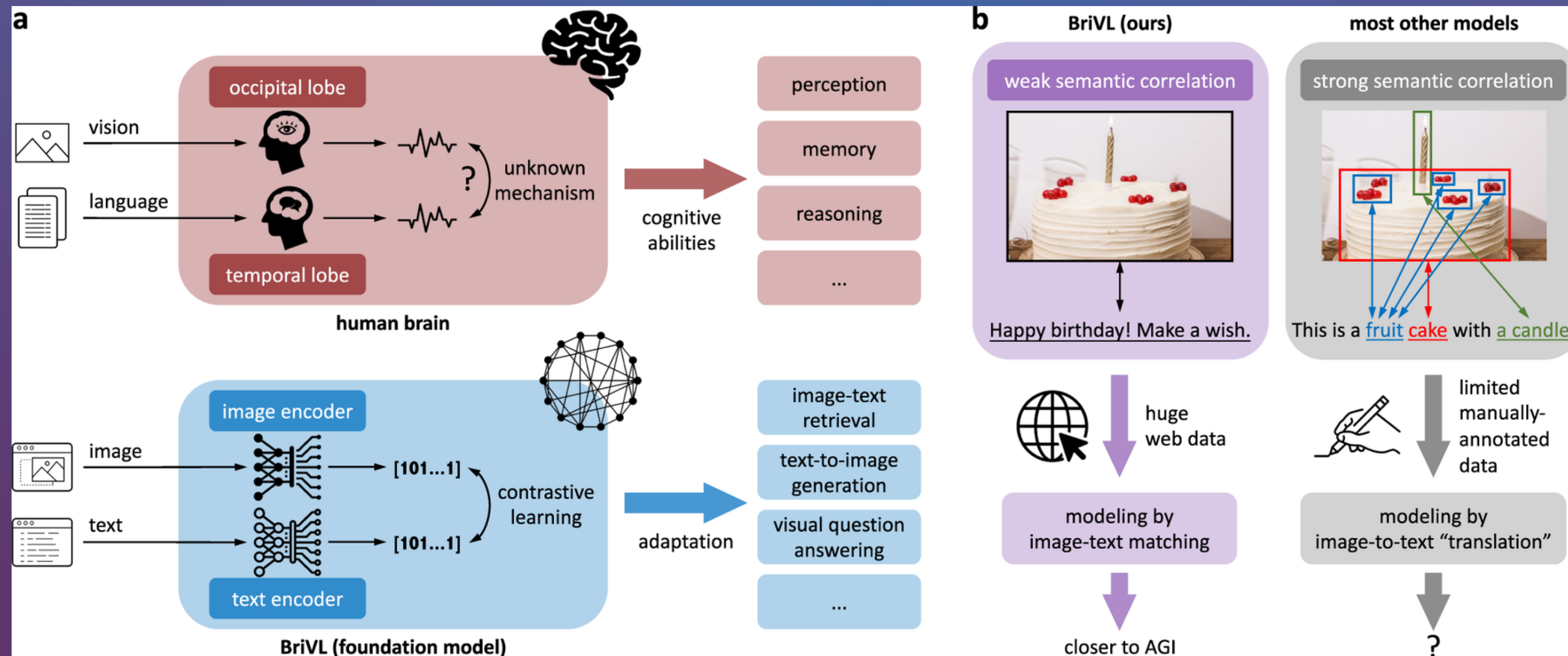
La interpretación de las palabras cambia de acuerdo a su orden, temporalidad y contexto.

Esto provoca que la resolución de los Problemas de NLP sean muy complejos de resolver para los computadores.

Aplicaciones



Aplicaciones

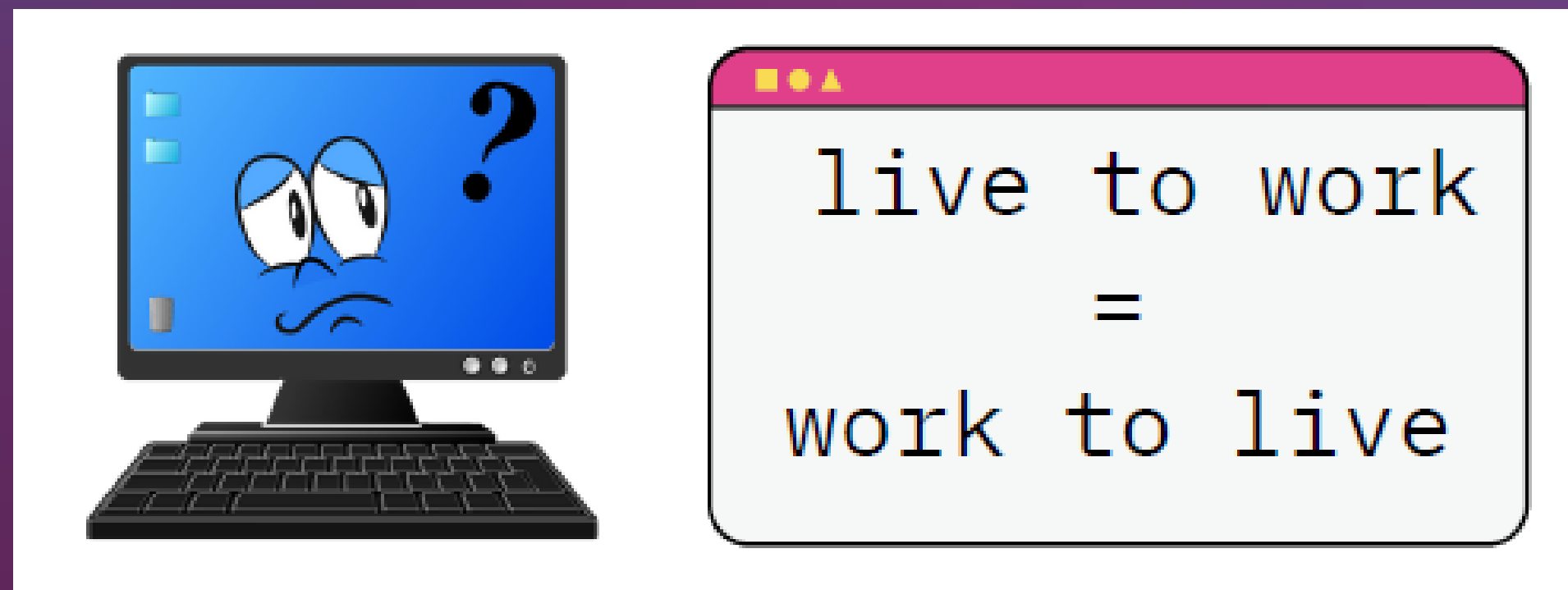


Multimodal models

¿Como funcionan los computadores?

Los computadores codifican los caracteres en bytes, como por ejemplo ASCII o utf-8.

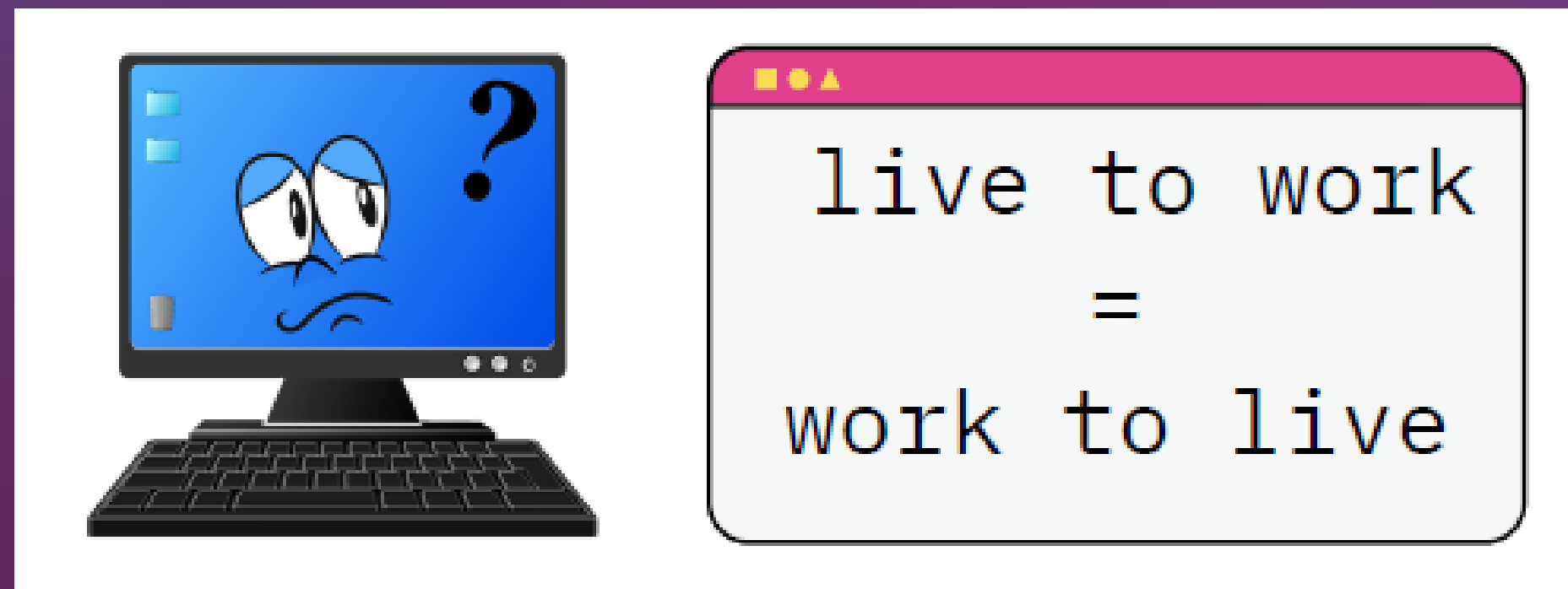
A vista del computador, cada palabra y oración son simplemente composiciones de bytes, lo que en si no poseen sentido alguno.



¿Como funcionan los computadores?

El orden de esta oración no implica una diferencia (en bytes) para el computador ya que no es capaz de identificar de forma natural el lenguaje y sus complejidades.

¿Entonces, como hacemos para que el computador entienda el lenguaje?...



Representaciones Vectoriales

Las representaciones vectoriales del lenguaje convierten palabras o frases en vectores numéricos que capturan su significado y relaciones semánticas.

		cat	mat	on	sat	the
the	=>	0	0	0	0	1
cat	=>	1	0	0	0	0
sat	=>	0	0	0	1	0

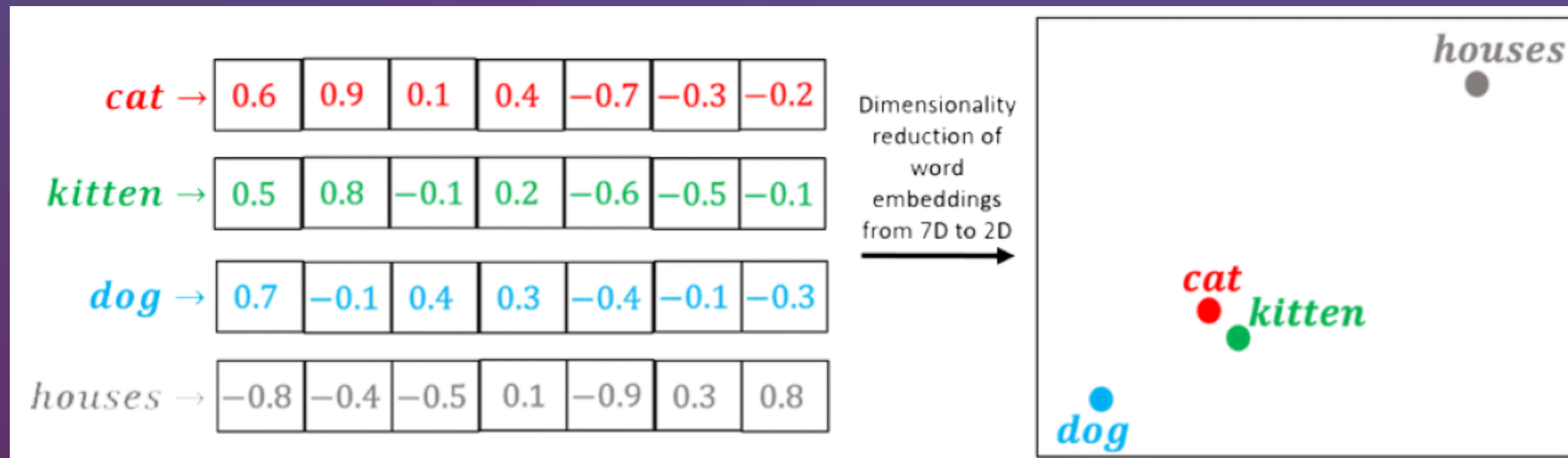
Representaciones Vectoriales: Bag of Words

Es un modelo que representa un texto como un conjunto de palabras, ignorando el orden y la gramática. Cuenta la frecuencia de cada palabra, generando un vector donde cada elemento corresponde a una palabra única del vocabulario.

	the	red	dog	cat	eats	food
1. the red dog →	1	1	1	0	0	0
2. cat eats dog →	0	0	1	1	1	0
3. dog eats food →	0	0	1	0	1	1
4. red cat eats →	0	1	0	1	1	0

Representaciones Vectoriales: Word Embeddings

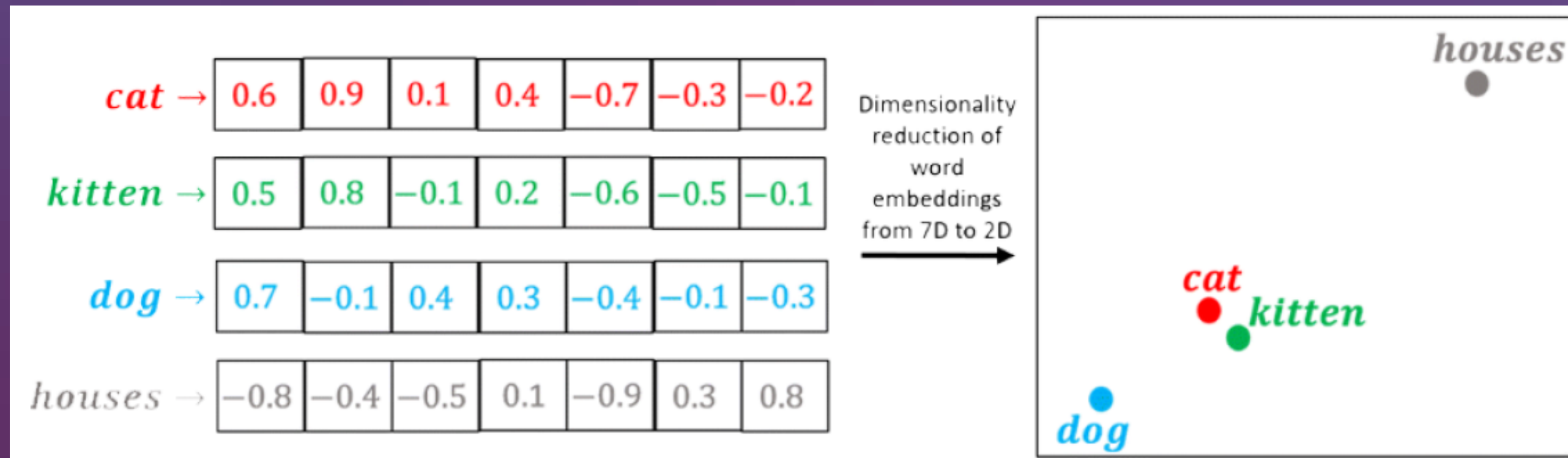
Son representaciones vectoriales de palabras que capturan relaciones semánticas y contextuales en un espacio de alta dimensión. Estos vectores se obtienen a través de modelos como Word2Vec o GloVe.



Representaciones Vectoriales: Word Embeddings

Se obtienen utilizando redes neuronales superficiales a través de una tarea auxiliar, un ejemplo es el modelo Skip-Gram.

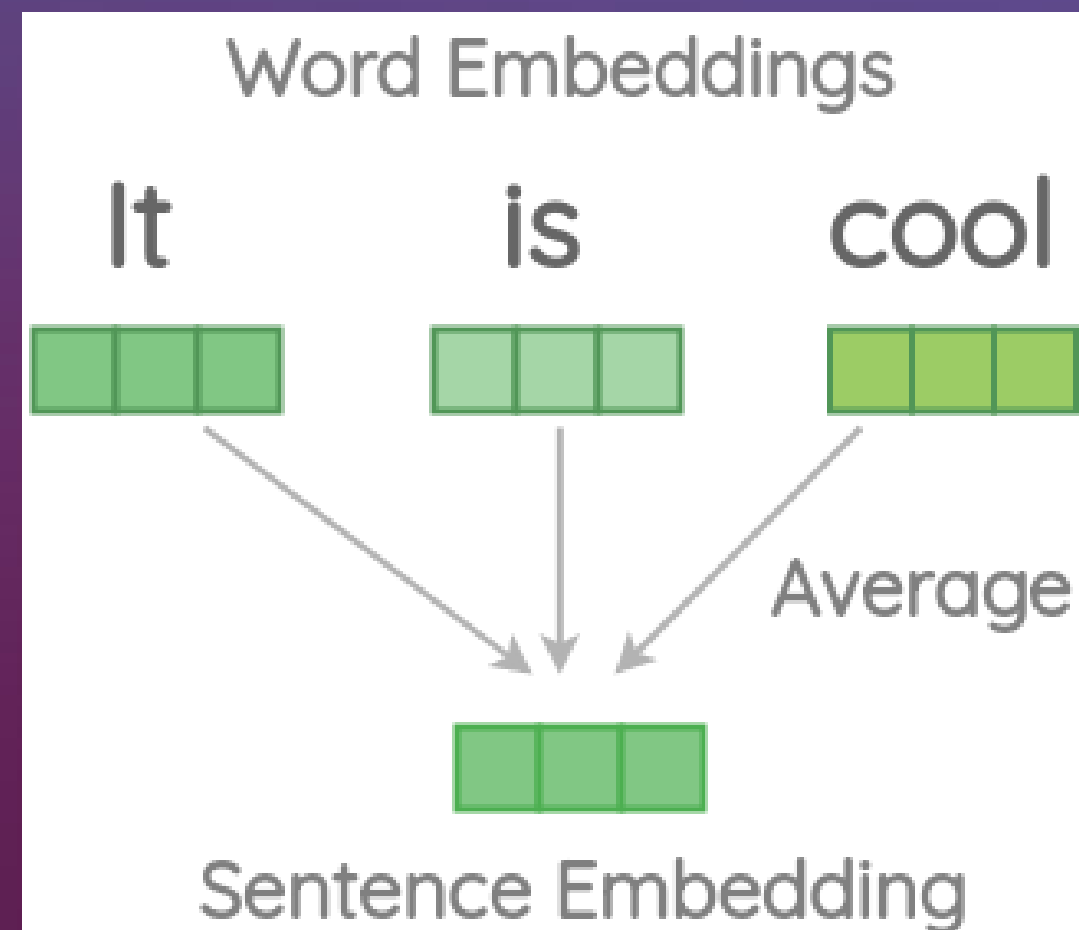
Dada una oración, se selecciona una palabra de esta y luego se intenta que la red prediga la palabras que la rodean.



Representaciones vectoriales: Sentence Embeddings

Representaciones de oraciones basados en distintas técnicas:

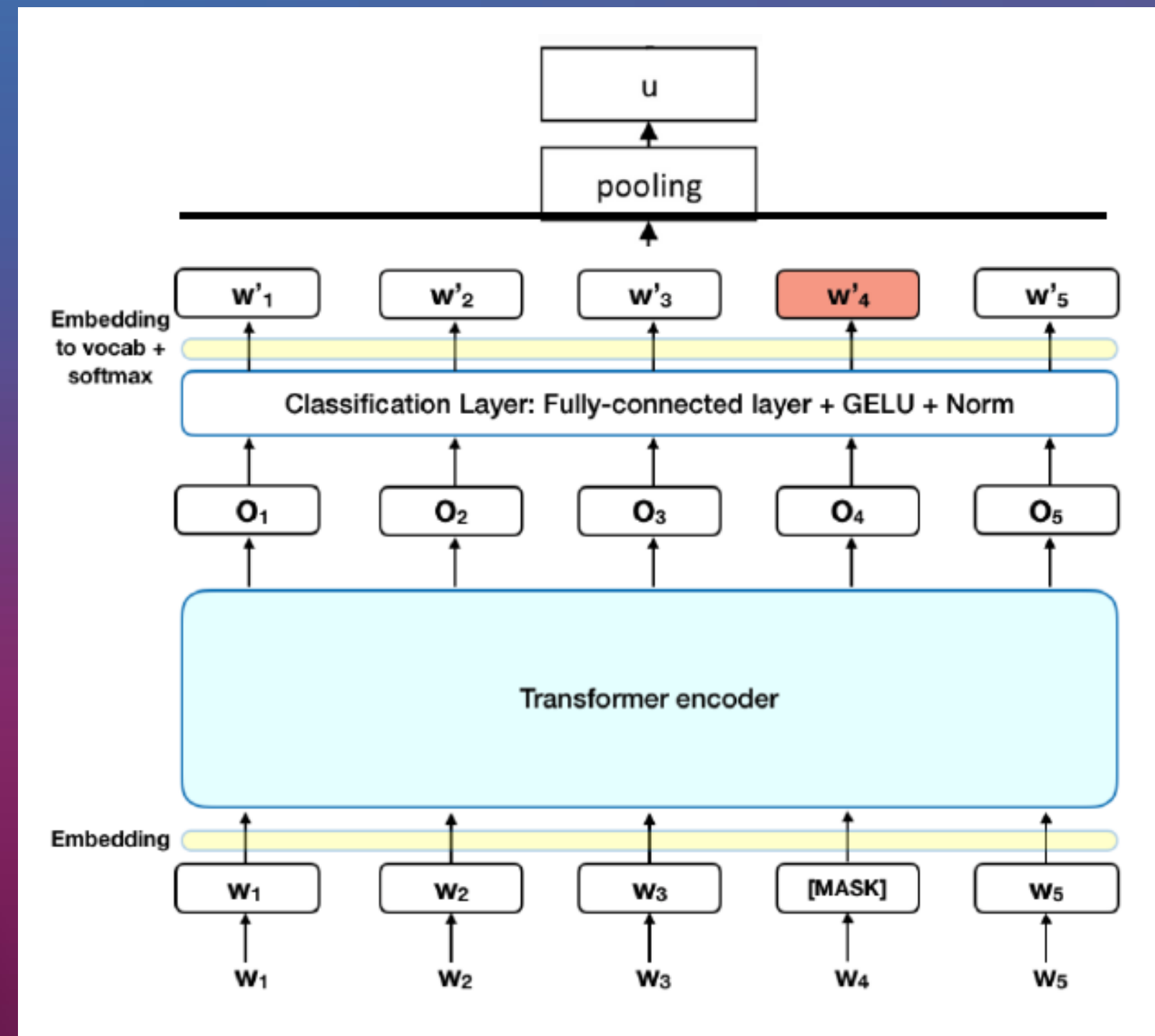
- Bag of Words.
- Suma, promedio o máximo de los Word Embeddings que componen una oración.



Representaciones vectoriales: Sentence Embeddings

Representaciones de oraciones basados en distintas técnicas:

- Bag of Words.
- Suma, promedio o máximo de los Word Embeddings que componen una oración.
- Modelos de Lenguaje basados en Transformers



Descargar material

Link repositorio: <https://github.com/giturra/JCC2023-WE/>
Short Link: <https://shorturl.at/dfqZ4>

INTRODUCCIÓN A WORD Y SENTENCES EMBEDDINGS Y SUS APLICA

Gabriel Iturra-Bocaz, Msc Computer Science
<https://giturra.cl/>