

## Cierre de la materia

Fernando Carranza  
fernandocarranza86@gmail.com

Clase 15  
Sábado 28/06/2025

- Unidad 1: Introducción
- Unidad 2: Los algoritmos supervisados
- Unidad 3: Anotación morfológica y de clase de palabra
- Unidad 4: Anotación sintáctica
- Unidad 5: Anotación para propósitos específicos
  - i) el análisis de sentimiento....
  - b. Reconocimiento de entidades nombradas (NER)...
  - c. Implementación de identificador de similitud semántica...
  - **d. Entendimiento del lenguaje natural (Natural Language Understanding) en la era de los grandes modelos de lenguaje.**

## Bibliografía

- Bender, Emily y Alexander Koller. 2020. "Climbing towards NLU: On Meaning, Form and Understanding in the Age of Data". *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp. 5185-5198.

Este seminario se ocupó básicamente de algoritmos de clasificación, es decir, algoritmos que toman datos y les asignan una determinada clase.

Al comienzo de esta cursada vimos que existen dos grandes tipos de programación:

- Programación clásica (aplicado a una tarea de clasificación, por ejemplo, la creación de un árbol de decisión)
- Aprendizaje automático o *Machine learning*

En el área del aprendizaje automático existen distintos tipos de algoritmos:

- Algoritmos supervisados
- Algoritmos no supervisados
- Aprendizaje por refuerzo
- Algoritmos autosupervisados

En esta cursada vimos tres métodos para entrenar algoritmos supervisados:

- Bayesiano ingenuo
- Máquina de soporte vectorial
- Regresión logística

En todos los casos, el entrenamiento da como resultado un modelo que permite predecir las etiquetas de datos nuevos. Hay algunas librerías que disponen de modelos entrenados disponibles, como por ejemplo, NLTK, Spacy y Stanza, así como repositorios como Kaggle y Huggingface.



En años recientes, el panorama del procesamiento del lenguaje natural se ha visto fuertemente impactado por la emergencia de los modelos de lenguaje.

The insight of large language modeling is that many practical NLP tasks can be cast as word prediction, and that a powerful-enough language model can solve them with a high degree of accuracy. For example, we can cast sentiment analysis as language modeling by giving a language model a context like:

The sentiment of the sentence “I like Jackie Chan” is:

and comparing the following conditional probability of the words “positive” and the word “negative” to see which is higher:

$P(\text{positive} | \text{The sentiment of the sentence “I like Jackie Chan” is:})$

$P(\text{negative} | \text{The sentiment of the sentence “I like Jackie Chan” is:})$

If the word “positive” is more probable, we say the sentiment of the sentence is positive, otherwise we say the sentiment is negative

Jurafsky y Martin. “Large Language Models”

Esta misma idea se puede aplicar a otras tareas de procesamiento del lenguaje natural, como *question answering*:

Q: Who wrote the book "The Origin of Species"? A:

$P(w|Q$ : Who wrote the book ' The Origin of Species"? A:)

Lo mismo se puede hacer para otras tareas, como *summarization*, etc.

- La tarea de decidir una palabra a generar dado un contexto se denomina *decoding*.
- Hacerlo repetitivamente se denomina *autoregressive generation* o *causal LM generation*.

La generación autorregresiva es, en esencia, una tarea de clasificación, en la que cada input lingüístico es el dato a clasificar y la palabra a elegir es la clase.

Esta tarea tiene la ventaja de que se puede entrenar mediante aprendizaje autosupervisado, lo que hace que no sea necesaria la anotación manual.

En la primera clase presentamos el test de Turing:

- **Juego de imitación:** una persona y una computadora interactúan con un evaluador. Este evaluador debe decidir, solo a partir de las respuestas que recibe, quién es la persona y quién es la computadora. Si la computadora logra engañar al evaluador, podrá decirse que su comportamiento es en alguna medida indistinguible del de la persona.



**Natural Language Understanding:** Disciplina que estudia la posibilidad de que un sistema de inteligencia artificial de entender el lenguaje.

Según Bender y Koller (2020), la explosión de los modelos de lenguaje ha generado una enorme exageración de sus habilidades para comprender:

- (1) a. In order to train a model that understands sentence relationships, we pre-train for a binarized next sentence prediction task. (Devlin et al., 2019)
- b. Using BERT, a pretraining language model, has been successful for single-turn machine comprehension... (Ohsugi et al., 2019)

Sin embargo, no terminamos de saber qué saben estos modelos. Esto llevó al surgimiento de lo que Bender y Koller (2020) llaman *bertología*, el estudio de qué tipo de conocimiento lingüístico representan los modelos de lenguaje.

Según Bender y Koller (2020), la evidencia apunta a que los modelos no tienen una real comprensión lingüística. Los autores asumen que, de hecho no pueden tenerla.

- Forma: cualquier realización observable del lenguaje.
- Significado: relación entre la forma y algo externo al lenguaje.

La noción de semántica que subyace a estos modelos es la de la semántica distribucional, que se sigue a su vez de la teoría wittgensteniana del significado.

- **Teoría wittgensteniana del significado:** El significado es el uso.
- **Semántica distribucional.** El significado es la distribución de las palabras en corpus.

## **Hipótesis Distribucional:**

Las palabras que aparecen en contextos similares tienden a tener significados similares.

(Joos 1950, Firth 1957, Harris 1954, ver también Sahlgren 2008)

Siguiendo a Gamut (1991), podemos identificar tres tipos de teorías sobre el significado:

- ① **Conceptualismo**: El significado es una relación entre símbolos y contenidos de conciencia.
- ② **Platonismo**: Los conceptos y proposiciones no son entidades mentales, sino cosas reales, solo que no pertenecen al mundo de los fenómenos observables, sino al mundo de las ideas.
- ③ **Realismo**: Las entidades con las que los símbolos lingüísticos guardan la relación de significado pertenecen a la realidad concreta y observable que nos rodea.



Siguiendo a Gamut (1991), los tres tipos presuponen que los significados se corresponden con alguna clase de entidad, ya sea mental, platónica o real. Cualquier teoría que cumpla con esta asunción se denomina **teoría referencial del significado**.

Según sostiene (Lewis, 1976, p. 1-2), esos referentes que constituyen los significados lingüísticos no pueden ser de naturaleza lingüística, porque, si así lo fueran, interpretar sería traducir de una lengua a otra, y en ese caso, esa otra lengua también debería poder ser traducida en los términos de otra y así hasta el infinito.

Para Bender y Koller (2020), cuando los humanos usan el lenguaje, lo hacen para cumplir con alguna intención comunicativa  $i$ . La relación de significado se formaliza entonces del siguiente modo, en donde  $E$  es el conjunto de expresiones e  $I$  es el conjunto de intenciones comunicativas:

$$(2) \quad M \subseteq E \times I$$

El significado convencional (el que se mantiene estable en todos los contextos de uso posibles) se formaliza del siguiente modo, en donde  $S$  es el conjunto de todos los significados posibles:

$$(3) \quad C \subseteq E \times S$$

Los hablantes utilizan C como uno de los medios disponibles para llegar a M.

Construir un sistema que realmente entienda el lenguaje es construir uno que sea capaz de vincular cada elemento de E con I o con S. Bender y Koller sostienen que un modelo entrenado puramente con E es incapaz de hacer eso.

Asumir entendimiento de comportamiento lingüístico ya fue criticado en la metáfora del cuarto chino de Searle:

- **El cuarto chino (Searle 1980).** Una persona que no sepa chino encerrado en un cuarto podría simular entenderlo si al recibir estímulos en chino desde afuera, pudiera responder en base a libros y un conjunto de reglas predefinidas.

En términos de Harnad (1990), el hablante del cuarto chino adolecería del *symbol grounding problem*.



Bender y Koller (2020) proponen, entonces, el test del pulpo (*octopus test*).

- Dos individuos A y B quedaron varados en dos islas inhabitadas y descubrieron que pueden hablarse a través de un aparato morse conectado por un cable que pasa por debajo del océano.
- Un pulpo O que se siente solo y que es una máquina experta en estadísticas descubrió que puede sentir los impulsos eléctricos del cable.
- Un día, O decide hacerse pasar por B y empezar a conversar con A. Para eso, utiliza patrones estadísticos para reproducir el tipo de respuestas que A y B se hacían frente a los estímulos de A.

- Un día, A inventa una catapulta de cocos y le cuenta a B con el objetivo de que B le dé consejos. O, que se quiere hacer pasar por B no podría ser capaz de eso, porque no tiene acceso al conocimiento de mundo necesario para poder ayudar a A.
- Otro día, A descubre que en su isla hay un oso y tiene que defenderse. Le pide a B ayuda para construir una defensa. O no podría ser capaz de darse una idea de qué necesita A.

Pensar que un modelo de lenguaje entiende, es como entrenar un modelo con todo el código en Java disponible en github y pedirle, en función de eso, que compile un documento programado en java.

Para que un modelo realmente entienda, es necesario entrenarlo con datos lingüísticos asociados a sus representaciones semánticas. Algunas opciones para esto es entrenar multimodalmente, para lo cual es necesario contar con datos lingüísticos anotados.

Las anotaciones siguen siendo relevantes y tienen todavía un rol que jugar en el desarrollo futuro de la inteligencia artificial y de la NLU.

# Bibliografía I

- Bender, E. M. y Koller, A. (2020). Climbing towards NLU: On meaning, form, and understanding in the age of data. En *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp. 5185–5198.
- Firth, J. (1957). *A synopsis of linguistic theory, 1930-1955*, pp. 1930–1955. Philological Society.
- Gamut, L. T. F. (1982/1991). *Lógica, Lenguaje y Significado. Vol. 2: Lógica Intensional y Gramática Lógica*. Eudeba, Buenos Aires. 2009.
- Harris, Z. S. (1954). Distributional structure. *Word*, 10:146–162.
- Joos, M. (1950). Description of language design. *The Journal of the Acoustical Society of America*, 22(6):701–707.
- Lewis, D. (1976). General semantics. En Partee, B., editor, *Montague Grammar*, pp. 1–50. Academic Press, New York.

## Bibliografía II

Sahlgren, M. (2008). The distributional hypothesis. *Italian Journal of linguistics*, 20:33–53.