MAPAS DEL LABERINTO
UNA SERIE DE PHANTOM MAZE — AI & LANGUAGE LAB

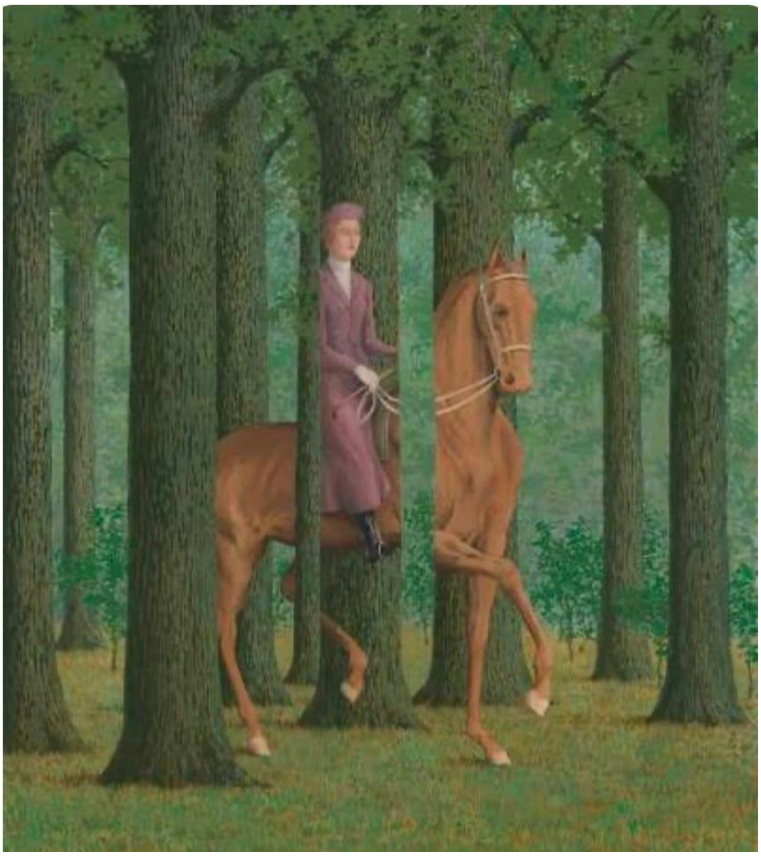
ENSAYO 19

La extraordinaria bolsa de Tristan Tzara

Claudia Marsico · Hernán Inverso

DOI 10.5281/zenodo.23575221

La codificación posicional en los modelos de IA o hasta dónde importa el orden de las palabras



R. Magritte, Le Blanc-Seing (1965)

En 1920, en pleno escándalo dadaísta, Tristan Tzara publicó primero en *Littérature* y a los meses en la recopilación *Manifiesto sobre el amor débil y el amor amargo* una receta para fabricar poemas:

Tome un periódico.
Tome unas tijeras.
Elija en el periódico un artículo de la longitud que quiera darle a su poema.

Recorte el artículo.
Recorte en seguida con cuidado cada una de las palabras que forman el artículo y métalas en una bolsa.
Agítela suavemente.
Ahora saque cada recorte uno tras otro.
Copie concienzudamente
en el orden en que hayan salido de la bolsa.
El poema se parecerá a usted.
Y es usted un escritor infinitamente original y de una sensibilidad hechizante, aunque incomprendida del vulgo.

Esta burla del antiarte contra la solemnidad literaria roza un punto curioso. Si les sacamos el orden a las palabras, el sentido no sobrevive. La bolsa de Tzara contiene exactamente las mismas palabras que el artículo del diario y no dice nada parecido. Miremos ahora el transformer, la arquitectura de todos los modelos de lenguaje actuales. A su modo es un lector de bolsas de Tzara, ya que su diseño no le deja ver el orden. Al procesar tokens, esos fragmentos con que opera, todos los términos de la frase se procesan a la vez, cada uno consultando a los demás y por tanto sin orden. Si a la máquina le damos "el perro mordió al cartero" y "al cartero mordió el perro", desordenadas, sus cuentas internas producen lo mismo. Los matemáticos dicen con elegancia que la atención es invariante ante permutaciones. Tzara podría haber dicho que el modelo lee lo que sale de la bolsa, de modo que necesita que alguien le devuelva el orden, porque no pretende ser un dadaísta. ¿Cómo se le enseña orden a quien no lo tiene?

R. Magritte, Le Blanc-Seing (1965)

Poner "orden" era en el origen acomodar los hilos del tejido, urdir la trama, lo cual lo vuelve de la misma familia que "texto" y su referencia a tejer. Pero "orden" saltó del telar al mundo entero organizándolo como si fuera una tela hecha de órdenes religiosas y de caballería, ordenaciones de sacerdotes, números ordinales, órdenes del día y ordenadores. El orden de las palabras es el bastidor del texto, de modo que la bolsa de Tzara desteje sin cortar los hilos.

En rigor, hay lenguas que vivieron con la bolsa medio abierta. El latín y el griego clásicos tenían un orden de palabras muy libre. La *Eneida* arranca *arma virumque cano*, "las armas y al varón canto". Ningún romano se perdía. La poesía lo extremó en el hipérbaton, la figura que disloca el orden esperable. El récord de exceso glorioso lo tiene Góngora, que podía escribir en las *Soledades* "pasos de un peregrino son errante" en lugar de "son pasos de un peregrino errante", desparramando el sintagma por el verso, confiado en que el lector lo iba a poder rearmar. "Paciencia debes tener", "mucho que aprender todavía te queda" dice el maestro Yoda en la saga de *La guerra de las galaxias* con hipérbaton puro.

El orden es más alterable cuanto más depende el sentido de la morfología y no de la posición. El latín declina, es decir, le da a cada palabra una terminación que dice qué papel juega en la frase. *Canis virum mordet* y *virum canis mordet* dicen exactamente lo mismo (el perro muerde al hombre) porque "canis" lleva puesta la marca de sujeto y "virum" la de objeto, vaya cada uno donde vaya. Las lenguas

modernas, del inglés al chino, prefirieron debilitar o abolir las declinaciones y entregarse al orden y por eso la bolsa de Tzara es letal para un diario actual y habría sido casi inofensiva para Virgilio, ya que cada recorte latino sale de la bolsa con su función tatuada.

Si miramos desde la filología, lo que hicieron los autores del transformer en 2017 es asombrosamente clásico: convirtieron una lengua de orden en una lengua de declinación. Como su máquina no podía leer en fila, le estamparon a cada palabra, en el cuerpo mismo de la palabra, la marca de su posición. Antes de entrar al juego, cada token viene convertido en un vector, esa lista de números que es su punto en el espacio del sentido. La codificación posicional le suma a ese vector otra lista de números que codifica "soy la palabra número siete" y así el token adquiere la marca de posición. No importa el desorden de la bolsa de Tzara porque el orden está adentro de cada palabra y ese hilo ya no va a enredarse.

Cada posición recibe su firma hecha de ondas, en una colección de senos y cosenos de frecuencias distintas, unas que oscilan rápido y otras lentísimas. Imaginemos un reloj con muchísimas agujas donde ninguna sola dice la hora, pero todas juntas identifican el instante sin ambigüedad. Dos instantes cercanos tienen fotos parecidas, dos lejanos, fotos muy distintas, con lo cual la distancia entre palabras queda también insinuada. La firma posicional se suma, número a número, al vector del significado, de modo que posición y sentido terminan mezclados en las mismas coordenadas, como si el acusativo no se agregara a la palabra sino que se disolviera en ella. Que una suma tan sencilla alcance para conservar a la vez identidad y posición fue uno de esos hallazgos en los que la ingeniería mostró primero que el mecanismo funcionaba y la teoría tuvo después que explicar por qué funcionaba tan bien.

La copia de libros ya había enfrentado un problema parecido que revisamos a propósito de la arquitectura transformer y la atención simultánea y resulta significativa también aquí por la remisión de una parte a otra. Un códice medieval se copiaba por cuadernillos sueltos y luego esos pliegos independientes llegaban al encuadernador. Como no había leído el texto y debía coserlos en el orden correcto, se estampaban marcas de orden en cada pieza por medio de *signaturas*, letras y números al pie de los cuadernillos, y el *reclamo*, que indicaba la primera palabra del cuadernillo siguiente. La codificación posicional es el reclamo del token, es decir, la marca que le permite a un procesamiento ciego y simultáneo reconstruir una secuencia que nunca leyó de ese modo.

La firma de ondas de 2017 fue sólo el primer modo de resolver el problema. Poco después, modelos como GPT-2 y BERT empezaron a aprender también la posición. En lugar de recibir una firma matemática ya diseñada para "posición 1", "posición 2" o "posición 3", el modelo aprende un vector distinto para cada lugar de la secuencia. Eso le da más libertad, con la desventaja de que conoce bien las posiciones para las que fue entrenado y no sabe qué hacer cuando el texto va más allá de eso. ¿Qué hace un encuadernador que aprendió a numerar hasta cierto número de páginas cuando de pronto recibe un volumen muchísimo más largo?

En 2018 surgió la idea de enfocarse en la relación entre las palabras más que en su número absoluto. No interesa tanto que "perro" esté en la posición siete como que esté tres lugares después de "el" o dos

antes de “ladra”. La red podía aprender cuánto separaba unas palabras de otras. En rigor, la gramática se trata más de vecindades y distancias que de números fijos. De esa idea salió RoPE, “Rotary Position Embedding”, propuesta en 2021, donde la posición deja de sumarse como una segunda etiqueta numérica y se incorpora haciendo girar la representación de cada palabra de acuerdo con el lugar que ocupa. La palabra sigue representada por un conjunto de números, pero ahora rotados de una manera ligeramente distinta según aparezca primera, quinta o centésima. La diferencia entre esas rotaciones permite recuperar información sobre cuánto separa determinadas palabras, como si cada palabra llevara una aguja de reloj que permite saber dónde está respecto de las demás.

Una solución todavía más austera fue ALiBi, presentada en 2022, que en lugar de estampar la posición sobre las palabras, actúa directamente sobre la atención entre ellas: cuanto más lejos se encuentra una de otra, más se penaliza la relación, de modo que a igualdad de otras condiciones, las palabras próximas conservan más influencia que las remotas sin necesidad de aprender una marca específica para cada posición. La simplicidad del procedimiento tenía además una ventaja considerable, resumida en el título del paper, “Train Short, Test Long”: un modelo entrenado con secuencias relativamente breves podía extenderse después a textos más largos sin perderse en un sistema de posiciones que no había aprendido. Este detalle terminó convertido en un campo de batalla. La manera en que el modelo representa la posición condiciona hasta dónde puede seguir el hilo de un texto antes de quedarse sin referencias útiles. De eso depende, en buena medida, la “ventana de contexto” que las empresas anuncian como si fuera la cilindrada de un motor y aumenta con cada mejora.

La receta de Tzara terminaba con una promesa irónica: después de recortar las palabras, meterlas en una bolsa y sacarlas al azar, “el poema se parecerá a usted”. El modelo de lenguaje crea una sensación parecida. ¿Cuánta gente siente que el modelo los conoce? Está entrenado para responder de maneras que nos resulten atractivas y cuanto mejor lo hace, más fácil es confundir esa adaptación con semejanza. La respuesta puede adoptar nuestro estilo y anticipar nuestras expectativas y debajo de esa familiaridad sigue estando la bolsa de Tzara con sus piezas discretas, posiciones, probabilidades y relaciones numéricas. El parecido del resultado no debería hacernos olvidar las diferencias.

◊

CÓMO CITAR

Marsico, Claudia y Hernán Inverso. 2026. “La extraordinaria bolsa de Tristan Tzara”. *Mapas del laberinto*, ensayo 19. Phantom Maze — AI & Language Lab. https://doi.org/10.5281/zenodo.23575221

Mapas del laberinto es una serie de Phantom Maze, AI & Language Lab. La colección vive en **phantommaze.cloud**; los ensayos aparecen también en phantommazeailabes.substack.com. La suscripción es gratuita.

Algunas de las ideas que enmarcan este texto se desarrollan en *El fantasma sin ego. Un mapa en el laberinto de la inteligencia artificial*, de Hernán Inverso y Claudia Marsico (Phantom Maze Press, 2026; ISBN 978-987-3729-10-2).

Fuentes. Ashish Vaswani y col., “Attention Is All You Need” (NeurIPS, 2017), sección 3.5; Jacob Devlin y col., “BERT: Pre-training of Deep Bidirectional Transformers” (2018); Alec Radford y col., “Language Models are Unsupervised Multitask Learners” (2019); Peter Shaw, Jakob Uszkoreit y Ashish Vaswani, “Self-Attention with Relative Position Representations” (NAACL, 2018); Jianlin Su y col., “RoFormer: Enhanced Transformer with Rotary Position Embedding” (2021); Ofir Press, Noah A. Smith y Mike Lewis, “Train Short, Test Long: Attention with Linear Biases Enables Input Length Generalization” (ICLR, 2022).