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

ENSAYO 14

Pre-entrenamiento: la sobremesa de Betty Shannon

Claudia Marsico · Hernán Inverso

DOI 10.5281/zenodo.23575274

¿Por qué es tan importante adivinar qué viene después?



F. Porter, The Living Room (1952)

El 23 de enero de 1913, ante la Academia Imperial de Ciencias de San Petersburgo, un matemático de cincuenta y seis años, célebre por su rigor y su mal carácter, presentó los resultados de un experimento extravagante. Andréi Márkov había tomado las primeras veinte mil letras de *Eugenio Oneguin* de Pushkin, les había borrado los espacios y la puntuación y las había acomodado en doscientas grillas de diez por diez. Después, contando a mano, determinó cuántas eran vocales y consonantes y cuántas veces una vocal seguía a otra vocal, una consonante a otra consonante o ambas se alternaban. Aquella

faena tenía detrás una pelea bastante menos literaria. El matemático Pável Nekrásov había sostenido que la ley de los grandes números, según la cual las frecuencias tienden a estabilizarse cuando aumenta enormemente el número de observaciones, requería sucesos independientes y había llegado a vincular esa independencia con el libre albedrío. Márkov llevaba años insistiendo en que se equivocaba. También una secuencia de acontecimientos dependientes unos de otros puede mostrar regularidades estadísticas estables. Pushkin le ofrecía un laboratorio perfecto, ya que cada letra cambia la probabilidad de lo que viene después y, sin embargo, sobre miles de letras, las frecuencias se estabilizan. Las cadenas de Márkov nacieron en ese clima. La inteligencia artificial de hoy desciende de aquellas grillas garabateadas en San Petersburgo, porque la tarea en preentrenamiento de modelos, simple y misteriosa a la vez, es adivinar la palabra que sigue.

El vocabulario de la anticipación nos puede orientar. *Predecir* es "decir antes", en general. Puede implicar "adivinar" (*divinare*), que es hacer lo que hace un dios, arrogándose el conocimiento de lo que viene. La adivinación tiene versiones inspiradas y otras casi técnicas. El augur romano que trazaba con su bastón un rectángulo en el cielo y leía el porvenir en cómo lo cruzaban los pájaros usaba la *conjetura*, "arrojar junto", como quien junta los signos para interpretar lo que sigue en ese conjunto más bien caótico. Cuando Cicerón escribió su tratado contra la adivinación, salvó en parte esta variante que se limita a inferir frecuencias a partir de señales. También un modelo de lenguaje "pre-dice" sin acceso a lo que viene ni iluminación divina, conjeturando a partir de lo arrojado, es decir, el contexto y el corpus, para apostar al símbolo siguiente. El modelo es un augur estadístico que observa pasar las palabras como pájaros.

Treinta y ocho años después de Márkov, la escena se repitió en una casa de Nueva Jersey, esta vez como juego de sobremesa. Claude Shannon, que acababa de fundar la teoría de la información, quería medir cuánta redundancia tiene el inglés, es decir, cuánto de un texto está, por así decir, ya dicho por el texto anterior. Su esposa se convirtió en instrumento de medición. Betty Shannon recibía un pasaje de una novela tapado y tenía que ir adivinando letra por letra. Si acertaba, avanzaban; si no, Claude le decía la letra y seguían. Contando cuántos intentos hacían falta en cada posición, Shannon calculó que el inglés impreso lleva cerca de un bit de información por letra. Eso quiere decir que en un alfabeto de veintisiete símbolos, la letra siguiente está tan condicionada por las anteriores que adivinarla es mucho más fácil de lo que parece. La conclusión, publicada en 1951 con el título "Predicción y entropía del inglés impreso", es el puente entre Pushkin y los chatbots. Predecir un texto y comprimirlo son la misma operación mirada de dos lados y el grado de predicción acertada permite medir cuánto se sabe de la lengua. Jugando después de la cena, Betty estaba ejecutando a mano el trabajo que hoy hacen los centros de datos.

F. Porter, The Living Room (1952)

Shannon había hecho, además, el experimento inverso. En su artículo fundacional de 1948 fabricó series de "aproximaciones al inglés" generadas eligiendo cada símbolo según las estadísticas del anterior. Primero letras sueltas al azar, después pares, después palabras encadenadas por sus frecuencias. La serie arranca en sopa tipográfica y se va ordenando. En la aproximación por pares de

palabras aparecen frases como “the head and in frontal attack on an english writer that the character of this point is therefore another method”, que no dicen nada y sin embargo *suenan* a inglés, como una radio mal sintonizada en la que podemos reconocer más o menos un idioma pero no el mensaje. En 1948 Shannon generaba texto con una cadena de Márkov. Entre esa radio mal sintonizada y la prosa de un chatbot hay setenta años de mejora.

Para entrenar una máquina hacen falta ejemplos corregidos. Fotos de perros etiquetadas “perro” por alguien, radiografías anotadas por médicos, traducciones validadas por traductores y textos que un hablante nativo confirma que suenan bien. Ese etiquetado humano es lento y costoso. Sin embargo, la predicción de la próxima palabra esquiva el problema porque el texto trae su propia hoja de respuestas. Al tomar cualquier oración escrita, por ejemplo, “El gato se subió al tejado”, se tapa la última palabra, se le pide al modelo su conjetura y la corrección está en la palabra tapada sin que nadie tenga que anotar nada. Este procedimiento se llama autosupervisión y convierte el corpus de billones de palabras en un gigantesco cuaderno de ejercicios. En algún sentido, el entrenamiento y su tapar, conjeturar, comparar y ajustar los pesos una y otra vez parece monótono, pero no lo es tanto. Para completar “el gato se subió al...” alcanza con estadística aplicada a vecindarios para elegir entre tejado, árbol y sillón, pero al completar “la capital de Argentina es...” hay una referencia al mundo, al completar “ciento diecisiete por tres es...” hay que calcular, y cerrar la frase final de una novela policial, al estilo de “el asesino es...”, requiere haber seguido coartadas y motivos, es decir, haber realmente leído.

La tarea en apariencia humilde de predecir bien la próxima palabra de cualquier texto posible exigiría modelar aquello de lo que hablan los textos, incluyendo gramática, hechos, regularidades del razonamiento, estilos y un largo etcétera, si es que el modelo va a mejorar en la adivinanza. La vereda opuesta es escéptica. ¿Y si solo son loros estocásticos, es decir, profieren cosas al azar? Combinar piezas según probabilidades, al fin y al cabo, no es comprender. La discusión sigue abierta porque toca cuestiones que discutimos desde siempre sin consenso. ¿Qué es la comprensión? ¿Puede darse sin cuerpo, sin mundo y sin intención comunicativa? Ya nos ocuparemos de esto, pero ahora solo registremos la tensión entre la predicción como una forma de comprensión y como imitación cuya fluidez simula esa comprensión. Es un viejísimo problema y mientras tanto el modelo sigue mejorando su conjetura de la palabra que viene.

Cuando el modelo termina de entrenarse, la predicción se convierte en generación. En lugar de conjeturar solamente la próxima palabra, la escribe y la incorpora como parte del contexto desde el cual se conjetura la siguiente. Palabra a palabra, la profecía se autoalimenta. En la jerga técnica se llama autorregresión, pero no dista de los juegos de las vanguardias surrealistas. En 1925, en una casa de la rue du Château de París, un colorido grupo donde estaban Breton, Prévert y Tanguy escribía frases en cadena. Cada uno anotaba su parte viendo apenas el final de lo anterior, doblaba el papel y lo pasaba. La primera frase se volvió el nombre del juego que hurgaba en el inconsciente colectivo: “El cadáver exquisito beberá el vino nuevo”. El mismo mecanismo de continuar lo apenas visto entretiene a un jugador solitario que se pasa el papel a sí mismo a enorme velocidad. Si la bolsa de Tzara se parece a la máquina sin orden, el cadáver exquisito es la máquina en funcionamiento, una escritura donde cada

paso conjetura desde lo que alcanza a ver y sin embargo sale prosa con sentido y código que compila. El modelo escribe hacia adelante y sin borrador conjeturando palabra a palabra, atado por las previas, como un improvisador encadenado que empieza una frase larga en voz alta y solo puede seguir adelante.

Entre todas las definiciones de inteligencia que ensayaron la filosofía y la ciencia, muchas elucubran sobre resolver problemas, adaptarse, abstraer o crear, pero ninguna mencionó adivinar la palabra siguiente. Eso a lo sumo era un juego de salón de esposa paciente después de la cena o de matemático terco y resultó ser la clave. Era entrenable, con respuestas gratis en cantidad infinita y posibilidad de absorber todo lo que además vive en un texto. Es muy curioso que Márkov contara vocales para ganar una disputa sobre el libre albedrío y un siglo después los nietos de sus cadenas discutan de libre albedrío con quien quiera hacerlo.

Mientras la ingeniería descubría cuánto rinde la conjetura, una corriente influyente de las neurociencias venía proponiendo que el cerebro humano es, él también y ante todo, una máquina de predicción: un órgano que anticipa sin descanso la próxima señal. La próxima sílaba del interlocutor, el próximo peldaño de la escalera, ¿no implican aprender la diferencia entre lo esperado y lo llegado? La hipótesis del procesamiento predictivo, que Andy Clark resumió en un título perfecto, *Surfing Uncertainty*, es discutible, pero invita a pensar qué hay detrás de la anticipación como mecanismo asociado con la inteligencia.

◊

CÓMO CITAR

Marsico, Claudia y Hernán Inverso. 2026. “Pre-entrenamiento: la sobremesa de Betty Shannon”. *Mapas del laberinto*, ensayo 14. Phantom Maze — AI & Language Lab. https://doi.org/10.5281/zenodo.23575274

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).

PHANTOM MAZE

Leer este ensayo en línea · Todos los ensayos · El libro

Fuentes. Claude E. Shannon, “Prediction and Entropy of Printed English”, *The Bell System Technical Journal* 30, n.º 1 (1951), pp. 50–64; Claude E. Shannon, “A Mathematical Theory of Communication”, *The Bell System Technical Journal* 27, n.º 3 (1948), pp. 379–423, y 27, n.º 4 (1948), pp. 623–656; Andy Clark, *Surfing Uncertainty: Prediction, Action, and the Embodied Mind*, Oxford University Press, 2016; Alec Radford, Karthik Narasimhan, Tim Salimans e Ilya Sutskever, “Improving Language Understanding by Generative Pre-Training”, OpenAI, 2018; Yoshua Bengio, Réjean Ducharme, Pascal Vincent y Christian Jauvin, “A Neural Probabilistic Language Model”, *Journal of Machine Learning Research* 3 (2003), pp.

1137–1155; Emily M. Bender, Timnit Gebru, Angelina McMillan-Major y Margaret Mitchell, "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?", en *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21)*, Association for Computing Machinery, 2021, pp. 610–623.