

# Asimetrías lingüísticas en la Inteligencia Artificial

*“La herramienta básica para la manipulación de la realidad  
es la manipulación de las palabras”  
Philip K. Dick*

Jennifer N. GONZÁLEZ  
[jgonzalez@unm.edu.ar](mailto:jgonzalez@unm.edu.ar)  
Estudiante de la Licenciatura  
en Comunicación Social - UNM

## Introducción

El objetivo de este artículo es analizar las asimetrías lingüísticas en los Modelos de Lenguaje de Gran Tamaño (LLM, por sus siglas en inglés). Un LLM es un tipo de programa de Inteligencia Artificial (IA) que puede reconocer y generar texto, entre otras tareas. Entendemos por asimetrías lingüísticas a las desigualdades que existen entre las lenguas, es decir, que algunas lenguas sean más prestigiosas que otras, el reconocimiento de la opacidad del lenguaje, de la diversidad lingüística y de las relaciones de poder entre hablantes y variedades lingüísticas.

A partir de un estudio realizado por el Instituto Federal Suizo de Tecnología (2024) se busca analizar las implicancias que las inteligencias artificiales tienen para comprender y procesar conceptos en varios idiomas, donde primero traducen estos conceptos al inglés, procesan la información y luego traducen las respuestas al idioma de origen. Siguiendo el concepto de *imperialismo lingüístico* de Phillipson (2007), el dominio del inglés se reafirma y se mantiene en el establecimiento y la reconstrucción continua de desigualdades estructurales y culturales entre el inglés y las demás lenguas.

Por otro lado, en un estudio realizado a estudiantes ingresantes de la Universidad Nacional de Moreno (UNM), el 12.6% afirma que alguien en su casa habla esta lengua (Abelenda et al., 2024). En este sentido, se buscará reflexionar acerca de las implicancias de la preponderancia del inglés sobre otras lenguas en la educación superior.

Este artículo es escrito en el marco del proyecto de investigación que dirige Paola Pereira<sup>1</sup> llamado “Ideologías lingüísticas y prácticas de lectu-

ra y escritura 2: Asimetrías lingüísticas”. Inscribimos nuestras reflexiones en la perspectiva disciplinar de la glotopolítica entendida ampliamente como el estudio de las intervenciones políticas en el espacio público del lenguaje. Estas intervenciones activan ideologías lingüísticas que asocian cuestiones lingüísticas con posiciones sociales. La Glotopolítica indaga, entonces, en los modos en que las ideologías lingüísticas participan en la instauración, reproducción o transformación de entidades políticas, relaciones sociales y estructuras de poder tanto en el ámbito local o nacional como regional o planetario (Guespin y Marcellesi, 1986, citado por Arnoux, 2008, p. 12). En consecuencia, en este trabajo nos proponemos reflexionar sobre las intervenciones glotopolíticas vehiculizadas a partir de la Inteligencia Artificial y el uso del inglés.

## Inteligencia Artificial y Chat GPT

En términos generales, podríamos afirmar que las computadoras eran capaces solo de realizar aquello para lo que las programáramos. Con el advenimiento del aprendizaje automático eso cambió, ahora las computadoras pueden, a partir de una gran cantidad de datos, construir conclusiones e inferir una respuesta sin ser programadas específicamente. Los LLM se basan en el aprendizaje automático: en concreto, en un tipo de red neuronal llamado modelo transformador. Previamente, los expertos indicaban y construían el sistema que las computadoras debían llevar a cabo. En la actualidad, con la construcción de algoritmos, las computadoras pueden, siguiendo patrones específicos, inferir reglas, a partir de la enorme cantidad de datos que se le disponen. En este punto, la IA se basa en una enorme cantidad de datos contruidos por la inteligencia humana. Y los modelos de lenguaje, específicamente, en palabras.

1. Docente e investigadora del Departamento de Humanidades y Ciencias Sociales de la UNM.

La capacidad de las computadoras de generar automáticamente texto o imágenes con la aparente facilidad de un ser humano fue anunciada a fines del 2022 a partir del lanzamiento de Chat GPT de la empresa estadounidense OpenAI. Un año después, un estudio realizado por un equipo de investigadores de la Universidad Politécnica de Madrid comprobó que el Chat GPT desconoce el 20% del idioma español y que, con respecto al 80% que sí conoce, interpreta significados equivocados en un 5% de las ocasiones. A partir del análisis de 90 mil palabras incluidas en el *Diccionario de la Real Academia Española* (casi la totalidad disponible), se reveló que el Chat GPT desconoce 18 mil.

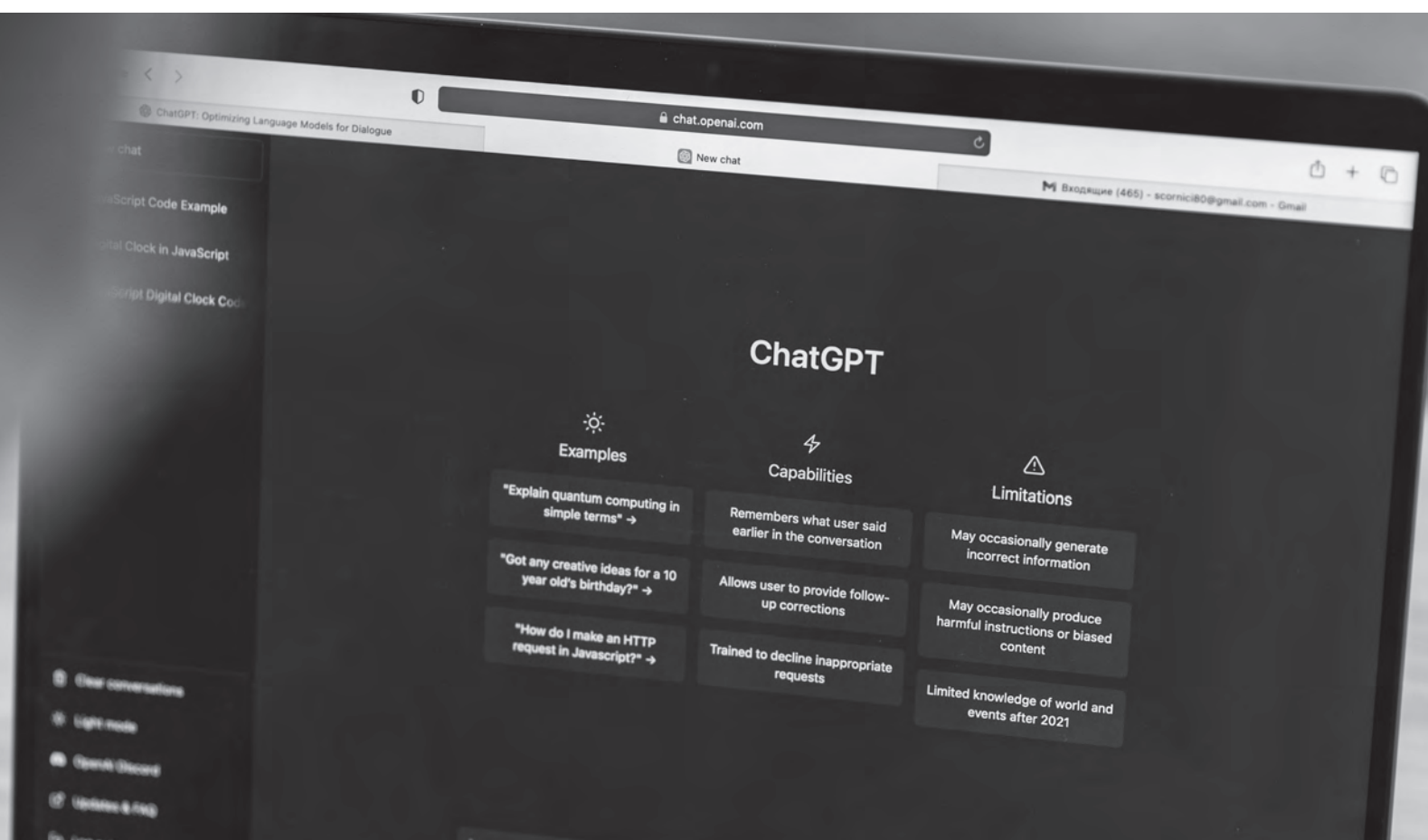
Por otro lado, la doctora en Filología Hispánica Silvia Gumiel Molina (2023) afirma que estas IAs no son capaces de entender ni lo que dicen ni lo que se les consulta. A continuación, se presenta la pregunta que la autora utiliza de ejemplo en el artículo:

¿Puedes decirme cual es la diferencia entre un 'joven viejo' y un 'viejo joven'? ¿O son sinónimas?

ChatGPT: “Un 'joven viejo' y 'un viejo joven' son expresiones que se utilizan para describir a una persona de una manera un tanto peculiar. Aunque no son sinónimas, ambas sugieren una contradicción en la edad aparente y el comportamiento de la persona:

“Un 'joven viejo' se refiere a alguien que, a pesar de tener una edad avanzada, actúa o se comporta de una manera más joven de lo que cabría esperar. Es decir, es un anciano que mantiene una actitud juvenil o intereses propios de personas más jóvenes.

Un 'viejo joven' se refiere a alguien que, a pesar de ser joven en términos de edad cronológica, se comporta o tiene actitudes



típicas de personas mayores. Es decir, es un joven que actúa como si fuera más viejo de lo que es. (Gumiel Molina, 2023, p. 3)

En resumen, la diferencia está en la edad cronológica y el comportamiento. Un “joven viejo” es mayor en edad pero actúa joven, mientras que un “viejo joven” es joven de edad pero actúa viejo (Gumiel Molina, 2023).

En este ejemplo el Chat GPT invierte las diferencias. Un joven viejo es un sintagma nominal cuyo núcleo es “joven” (por consiguiente, un joven viejo es un tipo de “joven”) mientras que en un viejo joven el núcleo es “viejo” y, por tanto, nos referimos a un tipo de “viejo”. La confusión parece obedecer a la gramática del inglés que coloca el adjetivo adelante del sustantivo.

Asimismo, realicé la misma pregunta al Chat GPT en el marco de esta investigación:

Yo: ¿Podés decirme la diferencia entre un joven viejo y un viejo joven?

ChatOn Bot: La diferencia entre un “joven viejo” y un “viejo joven” también puede ser interpretada de varias maneras, dependiendo del contexto:

1.Joven viejo: Este término puede referirse a una persona joven en edad cronológica pero que tiene una mentalidad, perspectiva o comportamientos más propios de alguien mayor. Puede implicar madurez, sabiduría o experiencia que no se corresponde con su edad física. A menudo, estos individuos pueden ser vistos como “sabios para su edad”.

2.Viejo joven: En contraste, este término puede describir a una persona mayor que actúa, se siente o tiene una actitud más juvenil. A pesar de su edad avanzada, puede mantener un espíritu joven, estar activo, o tener intereses y comportamientos típicos de personas más jóvenes. Esto puede reflejar una mentalidad abierta o un deseo de mantenerse conectado con las nuevas generaciones.

En resumen, el “joven viejo” destaca la madurez en la juventud, mientras que el “viejo joven” resalta la juventud en la vejez. Ambos términos

juegan con la idea de que la edad no siempre se alinea con la mentalidad o el comportamiento de una persona. (Consulta de marzo de 2024)

En menos de ocho meses el Chat GPT ha corregido la confusión gramatical del ejemplo extraído. De la misma manera se puede esperar a que se carguen las 18 mil palabras en español no reconocidas. Las críticas realizadas desde la perspectiva de la Real Academia Española seguramente serán consideradas debido a ser un actor de poder legitimado. El español estándar es considerado “lo que está bien” y el uso de cronolectos, sociolectos o geolectos está mal visto. Que las IA utilicen un lenguaje más neutro parece darle mayor autoridad y conocimiento. Sin embargo, ¿qué va a pasar con las lenguas o variedades “deslegitimadas” o minorizadas?

## El sesgo inglés

En un reciente estudio publicado por el Instituto Federal Suizo de Tecnología, los investigadores Wendler, Veselovsky, Monea y West (2024) pusieron a prueba empíricamente si los modelos de Inteligencia Artificial (LLMs) “piensan en inglés”. Intuitivamente, era una noción aceptada entre investigadores y la prensa, ya que para que los LLMs funcionen se le han dado suficientes ejemplos para que sea capaz de reconocer e interpretar el lenguaje humano u otros tipos de datos complejos.

Sin embargo, los investigadores se preguntaron qué ocurre cuando las entradas son desde otra lengua:

“Diseñamos cuidadosamente indicaciones que nos permiten determinar si una lente logit decodificada token es semánticamente correcto y en qué idioma pertenece (por ejemplo, un mensaje que le pide al modelo que traducir “fleur” [“flor”] del francés al chino “花”; cf. Figura 1). Siguiendo las probabilidades lingüísticas en capas, observamos que no hay contextos apropiados. Los tokens se decodifican en la primera mitad de las capas, seguidos por un cambio repentino de masa de probabilidad hacia la versión en inglés (“flor”) de la siguiente palabra correcta token, y finalmente un cambio al siguiente token correcto en el idioma de destino (“花”)” (Wendler, et al., 2024, p.2).

La metodología del estudio se basó en instruir a las LLMs con tres tipos de instrucciones en chino, francés, alemán y ruso: una solicitaba repetir la palabra dada; otra pedía traducir de uno de esos idiomas a otro; y la tercera solicitaba completar un espacio en blanco en una oración. Al rastrear cómo los modelos procesaban cada instrucción, descubrieron que casi siempre la ruta de procesamiento a través de las capas pasaba por lo que denominaron el subespacio inglés. Esto indica fuertemente que el inglés es utilizado como un puente para ayudar a entender los conceptos: para que los LLMs comprendan y procesen estas nociones en varios idiomas, primero traducen estos conceptos al inglés, procesan la información y luego traducen las respuestas de vuelta al idioma de origen.

Esta es la primera evidencia empírica de que el inglés funciona como lenguaje pivote<sup>2</sup> interno, cuando las entradas son realizadas desde otra lengua. Los investigadores alertan que el sesgo inglés está presente, predisponen patrones anglocéntricos y no tienen en cuenta la diversidad cultural y lingüística de nuestras sociedades.

Por su parte, Abelenda, González y Ponce Mónaco (2024) señalan que la propagación del inglés a nivel mundial aumenta la desigualdad y la venta de los angloparlantes en todos los campos: el científico, el cultural, el educativo, entre otros. Se obliga a los interlocutores a hacer una inversión para aprender el idioma y luego actuar en terreno ajeno, es decir, más allá de los recursos invertidos, siempre se será hablante de lengua segunda con menos competencia en la lengua que un hablante nativo.

## En la educación superior

En el ambiente académico cuando un artículo de procesamiento del lenguaje natural trabaja solo con el español es muy difícil que sea aceptado en una conferencia de primer nivel. Se espera que sea un estudio multilingüe y que incluya inglés, italiano, francés y otros. Este requerimiento no se da para el inglés, donde las cantidades de texto son además

enormes. La gente que trabaja con ese idioma puede hacerlo solo en ese idioma y nadie reclama traducción del texto en diversas lenguas.

Al respecto, Arnoux (2014) se refiere al vínculo profundo entre el investigador y los instrumentos de razonamiento o de conceptualización propios de su lengua principal, es decir, aquella en la cual ha aprendido a descubrir y nombrar el mundo, aquella en que domina mejor los útiles de precisión argumentativa. Más precisamente:

“Esta lengua (la propia) posee sus propias libertades y sus propias restricciones cognitivas, y es utilizándola que el investigador puede más fácilmente innovar, pues con comodidad puede llegar al extremo de sus intuiciones. Por el contrario, una investigación que se exprese solo en una lengua con pretensiones de universalidad frena la innovación, obliga a aquel que no la tiene como principal a formulaciones prestadas y, al limitar por eso mismo sus capacidades de conceptualización, termina por imponer una ciencia conservadora” (Hagège, 2013, p.122, citado por Arnoux, 2014, p.295).

En *Ideologías lingüísticas en estudiantes ingresantes a la Universidad Nacional de Moreno* advertimos junto a Abelenda y Ponce Mónaco (2024) la fuerte presencia del inglés en las respuestas, dado que el 12.6% afirma que alguien en su casa habla esta lengua. Esto demuestra que tiene un lugar de prestigio en el imaginario de nuestra comunidad lingüística. Los sujetos que respondieron que hablaban inglés muchas veces acompañaban su respuesta afirmando que lo hacían en un contexto de estudio o ayuda a algún familiar que se encontraba estudiando el idioma en el colegio o de manera extracurricular.

En consecuencia, observamos que el ambiente académico tiene una fuerte influencia en la construcción de la hegemonía del inglés, tanto como los productos culturales importados de Estados Unidos que logran “convencer” a la población –mundial– de que el inglés es un idioma funcional y de prestigio. En este sentido, por un lado, se produce una asimetría ya que los investigadores no angloparlantes no pueden innovar intelectualmente del mismo modo en una lengua segunda que en su lengua propia. Por otro lado, esos mismos investigadores son sometidos a esa situación por las mismas instituciones académicas, situación que

2. En Informática, “lenguaje pivote” refiere a un idioma intermediario usado en la traducción automática y la localización. En lugar de traducir directamente de un idioma A a uno B, se traduce de A al “idioma pivote” y luego del idioma pivote a B. Esto permite una traducción más eficiente y estandarizada entre múltiples idiomas, ya que solo necesita crear una interfaz de traducción entre cada idioma y el idioma pivote, en lugar de crear una interfaz para cada par de idiomas.

se complejiza con la IA “pensando” en inglés.

Según un estudio realizado por el Instituto Universitario para el Desarrollo Productivo y Tecnológico Empresarial de la Argentina (IUDPT), 9 de cada 10 estudiantes usan el ChatGPT. El 92% de los estudiantes usa el ChatGPT para ayudarse con las tareas académicas, sin importar que la herramienta no sea capaz de verificar la información que proporciona ni evaluar su fiabilidad ni hablar su propia lengua. Será tarea de los especialistas en educación dictaminar hasta qué punto puede considerarse como aprendizaje su utilización.

Por nuestra parte, problematizamos el hecho de que el denominado “subespacio inglés” por el cual las entradas se filtran, construya textos que no representan la cosmovisión de los estudiantes o su forma de entender el mundo. A su vez, entendemos que eso niega la posibilidad de aprender a redactar textos que respeten su propia variedad lingüística y sobre todo, abraza y aumenta la autoestima lingüística: el poder de escribir con sus propias palabras.

## A modo de cierre

Como dijimos, la glotopolítica es el estudio de las intervenciones en el espacio público del lenguaje y de las políticas lingüísticas que rigen en determinada región o a determinado grupo. Lo interesante de pensar los modelos LLMs desde esta perspectiva radica en comprender los actores políticos, el valor y el sentido de las lenguas o variedades minoritarias que se pierden debido a la centralidad del idioma inglés. Que los modelos de IA “piensen en inglés” y determinen una sola forma de ver el mundo acotan las posibilidades de existencia de lenguas heterogéneas por sobre los valores del norte global.

## Referencias bibliográficas

Abelenda, P., González, J., Ponce Mónaco, E. (2024). Ideologías lingüísticas en estudiantes ingresantes a la Universidad Nacional de Moreno. *Anáfora, Volumen Especial 8*. Facultad de Filosofía y Letras, Universidad de Buenos Aires. 1-22

de Arnoux, E. (2008). Ámbitos para el español: recorridos desde una perspectiva glotopolítica. *REVERTE. Revista de Estudos e Reflexões Tecnológicas da Faculdade de Indaiatuba*. Vol. 6. 1-28.

de Arnoux, E. (2014). Minorización lingüística y diversidad: en torno al español y al portugués como lenguas científicas. En Carvalho García, Marcus Vinicius et al. (orgs.), *Seminário Ibero-americano de Diversidade Lingüística*. IPHAN. 290-306.

De Vita, V. (2023). Cada vez más estudiantes usan inteligencia artificial para hacer trabajos: impacto sobre sus habilidades y un desafío para los docentes. *Los Andes*. <https://www.losandes.com.ar/sociedad/cada-vez-mas-estudiantes-usan-inteligencia-artificial-para-hacer-trabajos-impacto-sobre-sus-habilidades-y-un-desafio-para-los-docentes>

Martínez, G; Conde, J.; Reviriego, P.; Merino-Gómez, E.; Hernández, J. y Lombardi, F. (2023) How many words does ChatGPT know? The answer is ChatWords. *ArXiv. Cornell University*. <https://arxiv.org/pdf/2309.16777>

Gumiel Molina, S. (2023). *De lenguaje, lenguas y ChatGPT*. Fundación Telefónica. <https://telos.fundaciontelefonica.com/de-lenguaje-lenguas-y-chat-gpt/>

Phillipson, R. (1992). *Linguistic Imperialism*. Oxford University Press. [Traducción de Ana Bello (2007), bajo el nombre *El imperialismo lingüístico: bases teóricas*, realizada para la cátedra de Sociología del Lenguaje de la Universidad de Buenos Aires]

Wendler, C.; Veselovsky, V.; Monea G. y West, R. (2024). Do LlamasWork in English? On the Latent Language of Multilingual Transformers. En *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics. 15366–15394.