

Modelos de lenguaje y normativa del español

**Mabel
Giammatteo**

ggiammat@gmail.com

UBA y USAL

**Fernando
Carranza**

fernandocarranza86@gmail.com

Instituto de Filología y Literaturas
Hispánicas "Dr. Amado Alonso",
UBA

El uso de la IA en la escuela

Buen uso de la IA en la escuela



Puede ser beneficioso para asistir en la producción de textos y como contenido en sí, en un mundo cada vez más atravesado por la IA.

Mal uso de la IA en la escuela



Errores afianzados

Puede reforzar errores ortográficos y gramaticales

Progreso estancado

Puede impedir el desarrollo de habilidades propias de la escritura.

Objetivo del trabajo



Estudiar la adecuación de los grandes modelos de procesamiento del lenguaje natural en cuanto a la normativa ortotipográfica y gramatical

Utilidad

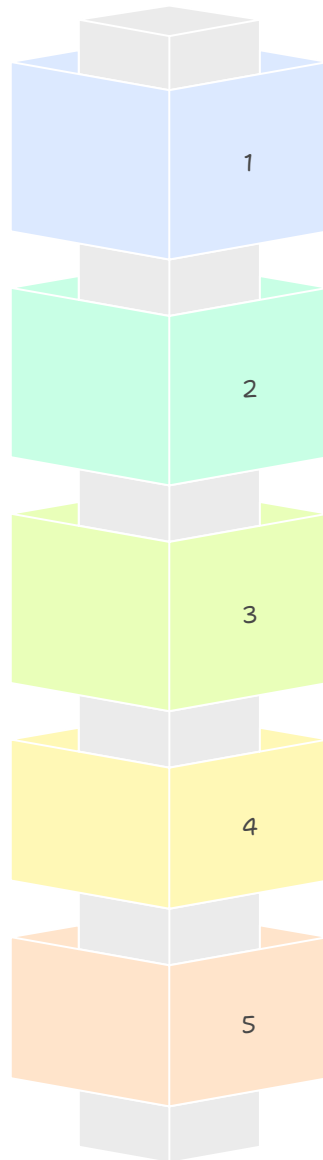


Tomar consciencia de los límites de estos modelos en cuanto a la producción textual adecuada a la normativa

Brindar herramientas para identificar textos escritos con IA

Eventualmente, dar herramientas para mejorar los modelos a futuro

Estructura del trabajo



Introducción a la IA

Proporciona una visión general de la IAG y de los modelos de procesamiento del lenguaje.



Metodología

Describe el enfoque y los métodos utilizados en el estudio.



Análisis del corpus

Presenta los rasgos de inadecuación normativa identificados.



Discusión

Analiza e interpreta los resultados del estudio.



Conclusiones

Resumen los hallazgos clave y las implicaciones del trabajo.

Evolución de los modelos de lenguaje



¿Cómo funciona un gran modelo de lenguaje (LLM en inglés)?

Preentrenamiento con grandes conjuntos de datos sobre tareas muy generales.

Fine tuning en cuanto a tareas más específicas.

Retroalimentación humana con aprendizaje por refuerzo

Entrenamiento multimodal incorpora imágenes, videos, audios

Otros aspectos del desarrollo

- Tokenización o descomposición en unidades.
- Captura de información posicional de los transformers a fin de construir modelos de aprendizaje profundo eficientes.
- Estrategias de atención para que el modelo se enfoque en tokens relevantes para aplicar su mecanismo predictivo probabilístico.
- Entrenamiento con grandes corpus de conocimiento científico para mejorar el razonamiento y la comprensión
- Modelos de diálogo centrados en ofrecer respuestas de alta calidad y seguras en interacciones conversacionales.

Metodología

Selección de modelos de lenguaje

- i) Gemini-2.5-flash-preview-09-2025 (Google)
- ii) Claude-3-7-sonnet-20250219 (Anthropic)
- iii) Qwen3-Max (Alibaba)
- iv) Deepseek-v3-0324
- v) GPT-5-Chat
- vi) Llama-3.3-70b-instruct (Meta, acceso abierto)
- vii) Chatgpt-4o-latest-20250326

Tarea de Producción Textual

Escribí un texto de aproximadamente 1000 palabras sobre las mariposas

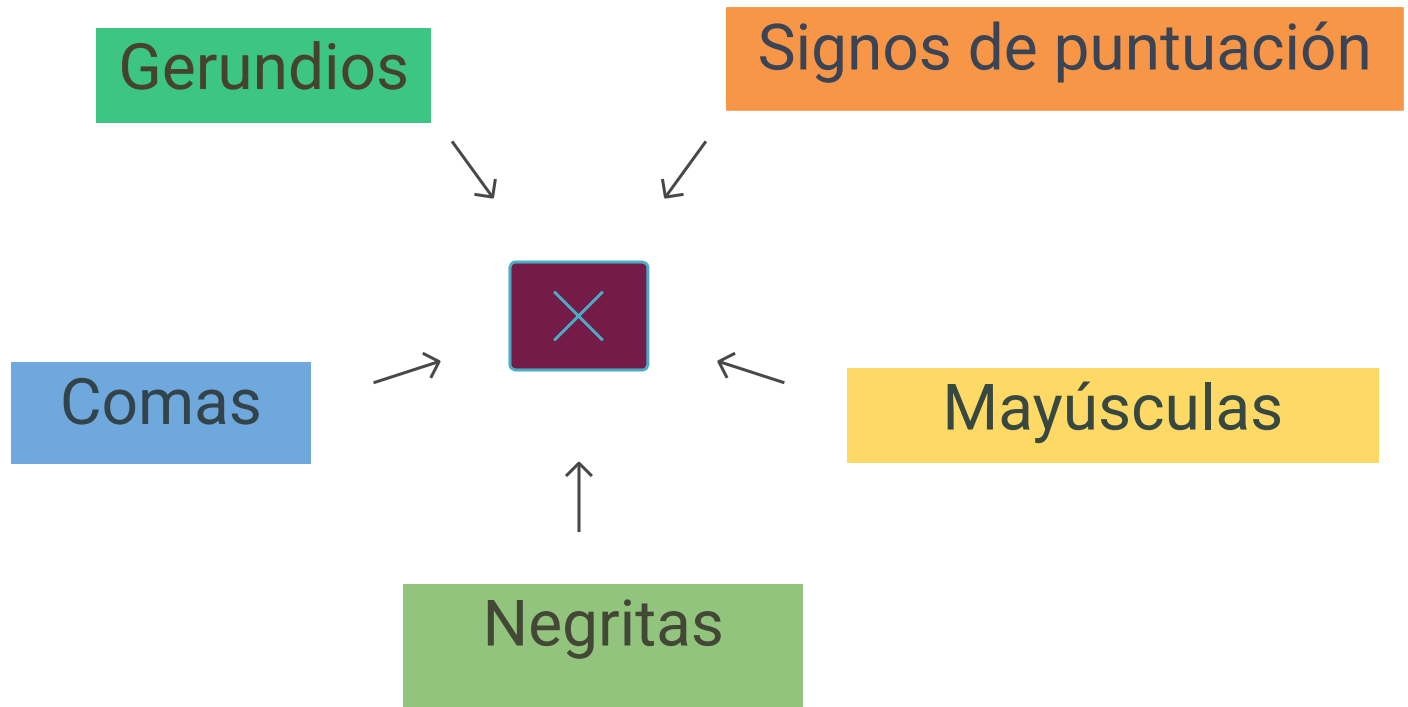


Chatbot Arena

Uso de <https://lmarena.ai/>
para interactuar con todos los
modelos

Recopilación de los textos e
identificación de
inadecuaciones normativas

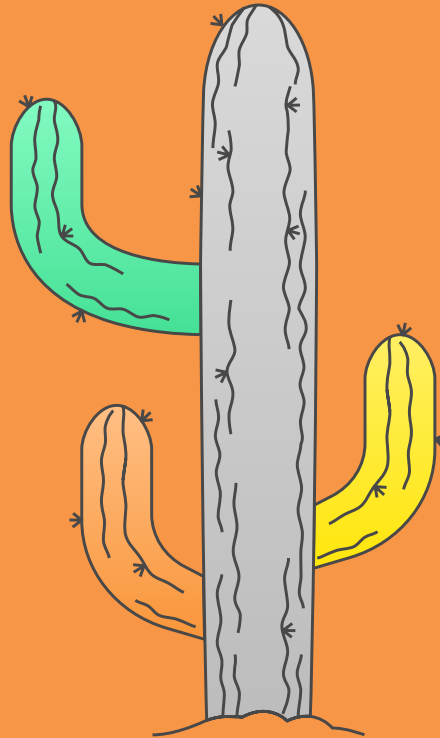
Algunas de las inadecuaciones normativas encontradas en los textos producidos con IAG



Desavenencias respecto del uso de los dos puntos

**Falta de elemento anticipador
en estructuras enumerativas**

**Punto en lugar de
dos puntos**



Normativa (RAE 2010, pp. 354-364)

Los usos de los dos puntos incluyen, entre otros, la presentación de enumeraciones y estructuras no enumerativas con un elemento anticipador (358-359), la introducción de discurso directo (359-360), la intercalación entre oraciones yuxtapuestas (360-361), la aparición luego de conectores como *pues bien* o *a saber* (361-362) y la separación del concepto general del contenido más específico en títulos o epígrafes (362).

Falta de elemento anticipador

1. Algunas especies notables incluyen:

Monarca (*Danaus plexippus*): Famosa por su migración masiva desde Canadá y EE.UU. hasta México, recorriendo miles de kilómetros.

Morpho Azul: Con alas de un azul metálico brillante, habita en los bosques de América Central y del Sur.

Atlas (*Attacus atlas*): Una de las mariposas más grandes del mundo, con una envergadura de hasta 30 cm.

Algunas adaptaciones sorprendentes son:

Mimetismo: Especies como la *Heliconius* imitan los colores de otras mariposas tóxicas para evitar depredadores.

Criptismo: Las alas de la *Kallima* parecen hojas secas, camuflándolas en su entorno.(Deepseek-v3-0324)

2. Lamentablemente, las mariposas están enfrentando numerosos peligros. Entre las amenazas más significativas se encuentran:

- **Pérdida de hábitat:** La deforestación, la urbanización descontrolada y el uso intensivo de pesticidas afectan directa e indirectamente a las poblaciones de mariposas.

- **Cambio climático:** Las modificaciones en las estaciones y temperaturas globales impactan los ciclos migratorios, la disponibilidad de plantas hospedadoras y el ritmo reproductivo de muchas especies.

- **Contaminación lumínica:** La luz artificial nocturna afecta el comportamiento y orientación de algunas mariposas nocturnas, dificultando su apareamiento y supervivencia.(Chatgpt-4o-latest-20250326)

En nuestro corpus encontramos algunos ejemplos de puntos en que habría correspondido más bien dos puntos:

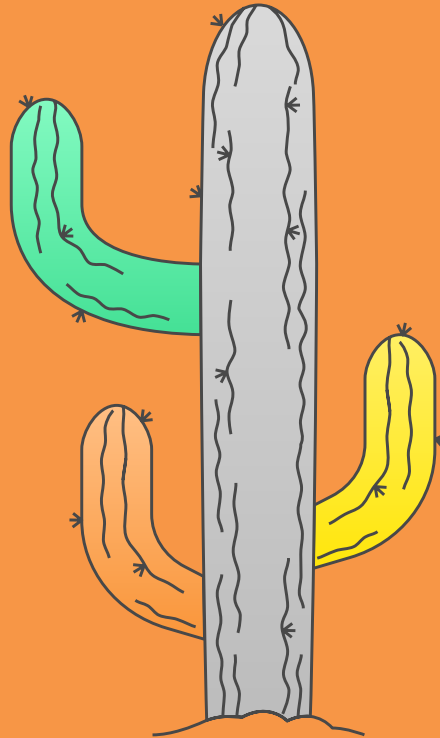
(3) La belleza de la mariposa reside, innegablemente, en sus alas, que son mucho más que membranas coloridas. Son estructuras biológicas altamente sofisticadas. (Gemini-2.5-flash-preview-09-2025) **Oposición**

(4) Proteger a las mariposas no es solo una cuestión de preservar la belleza; es una necesidad ecológica. (Gemini-2.5-flash-preview-09-2025) **Oposición**

(5) Cada especie de oruga tiene adaptaciones específicas; algunas poseen colores de advertencia (aposematismo), otras presentan camuflajes perfectos que las mimetizan con hojas o ramas. (GPT-5-Chat) **Explicación**

Desavenencias en el uso de la coma

Oxford comma



Normativa

No debe ir coma antes de
conjunción copulativa.

(6) Su declive es un indicador de la salud de nuestros propios ecosistemas, **y** su supervivencia depende de la protección de los hábitats de los que dependen en cada una de sus cuatro fases vitales. (Gemini-2.5-flash-preview-09-2025)

(7) Muchas especies muestran comportamientos sofisticados: migraciones estacionales (como la célebre monarca, *Danaus plexippus*, que recorre miles de kilómetros entre Canadá y México), cortejos elaborados donde los machos exhiben patrones luminosos o liberan feromonas, **y** mecanismos de defensa como la mimetización batesiana, en la que una especie inofensiva imita la apariencia de otra venenosa. (GPT-5-Chat)

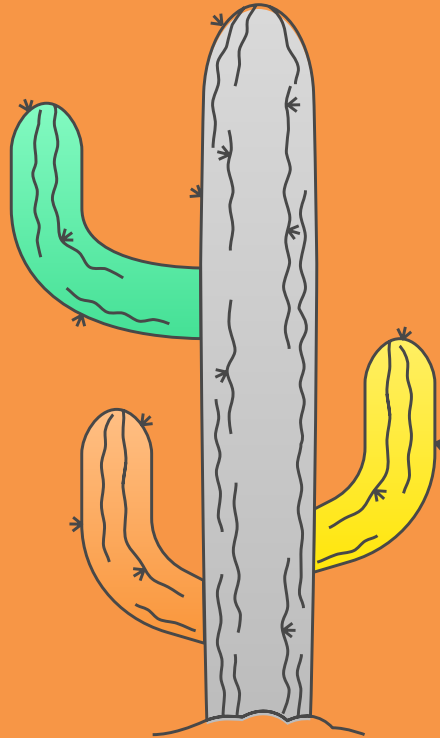
(8)

a. Esta migración es uno de los fenómenos naturales más impresionantes del planeta, **y** se cree que las mariposas utilizan la posición del sol y la temperatura para guiarse durante su viaje.

b. La destrucción de hábitats, la contaminación del aire y del agua, **y** el cambio climático están afectando a las poblaciones de mariposas en todo el mundo. Muchas especies de mariposas están en peligro de extinción, **y** se necesitan esfuerzos para proteger y conservar sus hábitats. (Llama-3.3-70b-instruct)

Abuso de negritas

No se lo recomienda especialmente



Normativa

En español, las negritas se utilizan principalmente para destacar títulos y subtítulos, ocasionalmente en combinación con la cursiva.

Su uso no está especialmente regulado en RAE (2010), pero en general se recomienda la moderación y su uso como método para destacar o resaltar algunas palabras. Se asocia típicamente a los textos periodísticos (García Negroni 2010, p. 171)

V. ej. (2). Diapositiva 10

(9) - **Polinización:** Aunque no son tan eficaces como las abejas, muchas mariposas ayudan a polinizar flores mientras se alimentan del néctar. Esta relación simbiótica beneficia tanto a las flores —que se reproducen— como a las mariposas que se nutren.

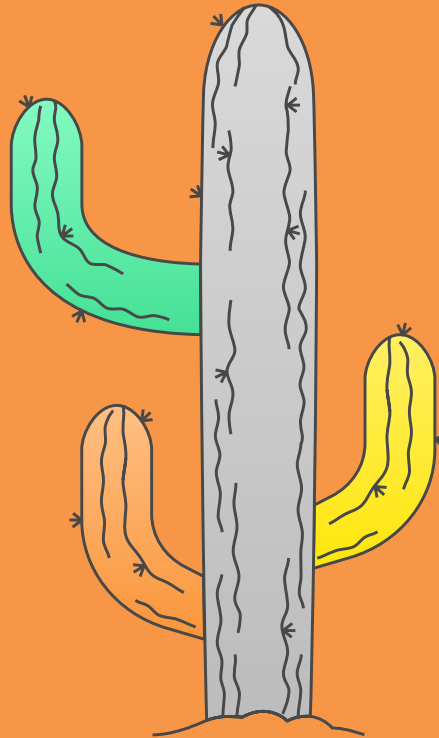
- **Indicadores ambientales:** Las mariposas son sensibles a los cambios en el clima, la calidad del aire y la pérdida de hábitats. Por eso, los científicos las estudian como indicadores biológicos del estado de los ecosistemas.

- **Fuente de alimento:** Las mariposas y sus larvas forman parte esencial de la cadena alimentaria, siendo presas de aves, reptiles, anfibios y otros insectos. (chatgpt-4o-latest-20250326)

Abuso de mayúsculas

Títulos

Mayúscula de relevancia



El uso incorrecto de mayúsculas abarca, por un lado, el uso recurrente en las palabras léxicas de títulos y subtítulos, algo que es típico de las convenciones de escritura del inglés, pero que no es aceptado para el español y, por otro lado, el uso, vinculado con el anterior, de la llamada mayúscula de relevancia (RAE 2010, p. 514).

(10)

- a. Las Mariposas: Joyas Aladas del Mundo Natural
- b. El Milagro de la Metamorfosis Completa
- c. Anatomía y Estructura Alar (Gemini-2.5-flash-preview-09-2025)

(11) Las Mariposas: Embajadoras aladas de la biodiversidad (GPT-5-Chat)

- Salvo Qwen3-Max y Chatgpt-4o-latest-20250326, el resto de los modelos hizo un uso extensivo de las mayúsculas en los subtítulos.
- El modelo Llama-3.3-70b-instruct no incluyó ni títulos ni subtítulos en la primera instrucción, por lo que este rasgo no pudo ser comprobado en la primera prueba. Sin embargo, al pedirle especialmente que agregue títulos y subtítulos en el texto, devolvió títulos que también abusan de las mayúsculas.

(12) Las Mariposas: Criaturas de la Naturaleza (Llama-3.3-70b-instruct)

Gerundios

Las incorrecciones que encontramos en los textos generados por IA también redundan en lo que Zorrilla (2004, p. 21-22) llama **norma morfosintáctica**. Esto incluye, por ejemplo, lo que tradicionalmente se consideran solecismos, es decir, incorrecciones normativas de naturaleza sintáctica antes que léxica, semántica, fónica u ortográfica. De entre este tipo cabe destacar el uso de los gerundios.

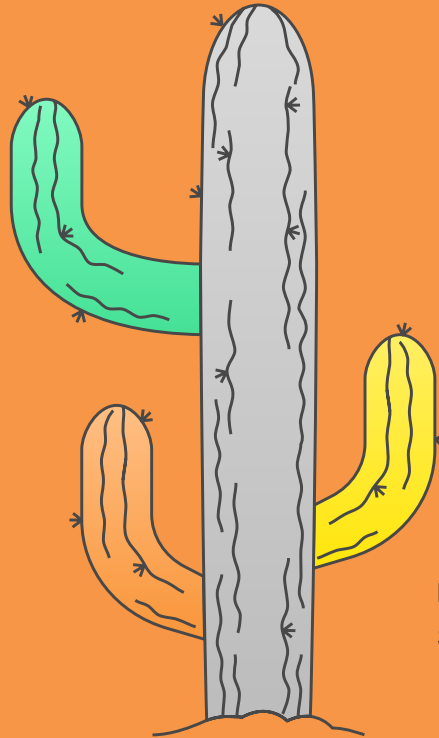
García Negroni (2010) reconoce, entre otros, los siguientes usos incorrectos del gerundio:

Gerundio de posterioridad

Estudió en Santiago, yéndose después a Bogotá

Gerundio especificativo

Vi un anciano teniendo ciento dos años.



Gerundio de consecuencia

Estudió durante el verano pudiendo recibirse a principios de marzo.

Gerundio como modificador de objetos indirectos o de circunstanciales

Les dieron una beca a unos estudiantes sabiendo chino
Le hablaba a un niño pidiendo limosna.

(13)

a. Su única misión es el crecimiento acelerado, **consumiendo** una cantidad de biomasa sorprendente.

b. La luz interactúa con estas estructuras a nivel nanométrico, **causando** refracción e interferencia, lo que produce los colores iridiscentes y metálicos que cambian según el ángulo de visión, como el azul resplandeciente de las mariposas Morfo. (Gemini-2.5-flash-preview-09-2025)

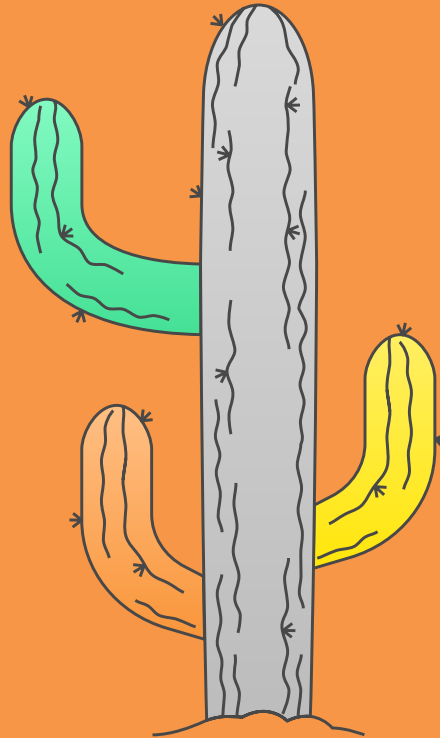
(14) Existen aproximadamente 160.000 especies descritas de lepidópteros, de las cuales unas 18.000 corresponden a mariposas diurnas (Rhopalocera), el resto **siendo** polillas (Heterocera)

(GPT-5-Chat)

Otros casos

Uso del paréntesis

Combinación de oraciones



Concordancia

Combinación de oraciones

(14) Esta es la fase de la quietud aparente y la transformación interna violenta

Mejor: Esta es la fase de la quietud aparente **en la que se produce** la transformación interna violenta.

(15) Las alteraciones en los patrones de temperatura y precipitación afectan la sincronización entre el momento en que eclosionan las mariposas y la disponibilidad de las plantas con flores que necesitan para alimentarse.

Mejor: entre... y el (momento) de disponibilidad de las plantas...

Concordancia

(16) **Las mariposas** desempeñan un papel ecológico crucial como **polinizadores**

(17) Durante este período, su objetivo principal es reproducirse y, en muchas especies, polinizar flores mientras se alimenta**(n)**.

(18) La pérdida de hábitat debido a la expansión agrícola, la urbanización y la deforestación es la principal causa de su declive.

Referencia

(19) El cuerpo de la larva se descompone en una "sopa" celular, pero a la vez, los "discos imaginales"—cúmulos de células especializadas—utilizan esa energía para construir las estructuras del adulto: alas, patas largas, ojos compuestos y la espiritrompa. **Es un crisol de destrucción y creación.**

Preposiciones

(20) La mariposa monarca, por ejemplo, ha experimentado un declive del 80% en las últimas décadas debido principalmente a la pérdida de plantas de algodoncillo y **(a)** la deforestación en sus sitios de hibernación.

Discusión

1) En primer lugar, una pregunta válida, sobre todo si se quiere utilizar las observaciones realizadas para identificar si un texto fue generado por un humano o por una inteligencia artificial, es si los hablantes cumplen esos rasgos de la normativa o no. Al respecto, debe destacarse que no siempre ese es el caso. De hecho, algunos errores, como el uso de los gerundios, el uso de las mayúsculas y los problemas a la hora de combinar oraciones, suelen aparecer en textos escritos por estudiantes o personas no formadas en corrección de textos.

2) Es posible que muchas de las inadecuaciones sean interferencias del inglés (e.g., *Oxford* comma, gerundios), lo cual se explicaría por el sesgo en la carga de datos de esos procesadores.

3) Las inadecuaciones registradas podrían usarse para trabajar en el aula mediante tareas de identificación y corrección de errores cometidos por la IA. Se estima que este tipo de tareas podría servir para que los estudiantes se habituaran a revisar sus textos, producidos autónomamente o con ayuda de la IA.

Conclusiones



Hicimos un relevamiento exploratorio de distintos modelos de lenguaje

Comprobamos experimentalmente que los modelos cometen algunos errores de normativa, por ejemplo, con respecto al uso de los dos puntos, las mayúsculas, las negritas, las comas, los gerundios, entre otros.



Esperamos que estas observaciones sirvan, sobre todo a los docentes

- **para tomar consciencia de las limitaciones que, al menos hasta la actualidad, presentan estos modelos a la hora de escribir textos adecuados a las normas de escritura del español.**
- **como posibles indicadores para poder identificar textos escritos con inteligencia artificial.**

Bibliografía

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, n, J. D. M. W. C., & Toutanova, L. K. (2019). Bert: Pre-training of deep bidirectional transformers for language understanding. En *Actas de NAACL-HLT* (Vol. 1, No. 2). <https://aclanthology.org/N19-1423.pdf>

Draxler, F., Buschek, D., Tavast, M., Hämäläinen, P., Schmidt, A., Kulshrestha, J., & Welsch, R. (2023). *Gender, age, and technology education influence the adoption and appropriation of LLMs*. arXiv preprint arXiv:2310.06556.

García Negroni, María Marta. (2010). *Escribir en español. Claves para una corrección de estilo*. Santiago Arcos.

Liang, Weixin, Yaohui Zhang, Mihai Codreanu, Jiayu Wang, Hancheng Cao y James Zou. (2025). *The Widespread Adoption of Large Language model-Assisted Writing Across Society*. arXiv. <https://arxiv.org/abs/2502.09747>.

Naveed, H., Khan, A. U., Qiu, S., Saqib, M., Anwar, S., Usman, M., Akhtar, N., Barnes, N., y Mian, A. (2023). *A comprehensive overview of large language models*. ArXiv. <https://arxiv.org/pdf/2307.06435>

Real Academia Española (RAE) y Asociación de Academias de la Lengua Española (ASALE). (2010). *Ortografía de la lengua española*. Espasa.

Real Academia Española (RAE). (2018). *Libro de estilo de la lengua española según la norma panhispánica*. Espasa.

Zorrilla, Alicia María. (2004). *Normativa lingüística española y corrección de textos*. Fundación Instituto Superior de Estudios Lingüísticos y Literarios Litterae.

Vaswani, A, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser. (2017). *Attention is all you need*. *Advances in Neural Information Processing Systems*. <https://proceedings.neurips.cc/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf>

Wang, F., Z. Zhang, X. Zhang, Z. Wu, T. Mo, Q. Lu, J. Xu, X. Tang, Q. He, Y. Ma, M. Huang & S. Wang. (2024). *A comprehensive survey of small language models in the era of large language models: Techniques, enhancements, applications, collaboration with llms, and trustworthiness*. ArXiv. <https://arxiv.org/pdf/2411.03350>

Muchas
GRACIAS