Lengua y Sociedad. Revista de Lingüística Teórica y Aplicada,
Vol. 23, n.º 2, julio-diciembre 2024, pp. 895-910,
ISSN-L 1729-9721; eISSN: 2413-2659
https://doi.org/10.15381/lengsoc.v23i2.28336

Análisis léxico de textos generados por modelos de lenguaje: reflejo de sus modelos de mundo

Lexical analysis of texts generated by language models: reflection of their world models

Análise lexical de textos gerados por modelos linguísticos: reflexão sobre os seus modelos de mundo

Gabriela Kotz
Universidad de Concepción, Concepción, Chile
gkotz@udec.cl
https://orcid.org/0000-0001-5300-7669

Pedro Salcedo
Universidad de Concepción, Concepción, Chile
psalcedo@udec.cl
https://orcid.org/0000-0002-1741-714X

Karina Fuentes
Universidad de Concepción, Concepción, Chile
kafuente@udec.cl
https://orcid.org/0000-0002-3644-3327

Resumen

La inteligencia artificial (IA) ha transformado numerosos campos, incluyendo la lingüística. Los Grandes Modelos de Lenguaje (LLM) han revolucionado la interacción con el texto al proporcionar respuestas que imitan el lenguaje humano. Estos modelos no solo generan texto, sino que también reflejan su interpretación del mundo. Sin embargo, la comprensión del mundo de estos modelos es limitada, lo que ha llevado a proponer el desarrollo de los Grandes Modelos del Mundo (LWM), que integran datos textuales, visuales y auditivos para una comprensión más completa. En este artículo, se analiza cómo los LLM articulan respuestas basadas en sus modelos de mundo mediante una perspectiva léxico-estadística. Mediante un diseño cuasiexperimental comparativo, se evaluaron seis LLM diferentes. La metodología se centró en medir la diversidad y densidad léxica de los textos generados por estos modelos. Los resultados mostraron que ChatGPT-4 tiene una alta densidad léxica y una diversidad léxica moderada, mientras que Copilot presenta la mayor diversidad léxica pero con una densidad léxica menor. Este análisis es crucial para entender las capacidades y limitaciones de los LLM, con implicaciones para aplicaciones en diversas áreas. Se presentan los conceptos y la metodología, se discuten los hallazgos y se concluye con reflexiones sobre futuras investigaciones y aplicaciones prácticas.

Palabras clave: modelo de mundo; modelo de lenguaje; inteligencia artificial; diversidad léxica; densidad léxica.

© Los autores. Este artículo es publicado por la revista ***Lengua y Sociedad*** de la Facultad de Letras y Ciencias Humanas, Universidad Nacional Mayor de San Marcos. Este es un artículo de acceso abierto, distribuido bajo los términos de la licencia Creative Commons Atribución 4.0 Internacional (CC BY 4.0) [https://creativecommons.org/licenses/by/4.0/deed.es] que permite el uso, distribución y reproducción en cualquier medio, siempre que la obra original sea debidamente citada de su fuente original.

Abstract

Artificial intelligence (AI) has transformed numerous fields, including linguistics. Large Language Models (LLMs) have revolutionized interaction with text by providing responses that mimic human language. These models not only generate text, but also reflect their interpretation of the world. However, these models' understanding of the world is limited, which has led to the the proposal of developing Large World Models (LWMs), which integrate textual, visual, and auditory data for a more complete understanding. This article employs a lexicostatistical perspective to analyze how LLMs articulate responses based on their world models. A comparative quasi-experimental design was utilized to evaluate six different LLMs. The methodology focused on measuring the diversity and lexical density of the texts generated by these models. The results demonstrated that ChatGPT-4 has high lexical density and moderate lexical diversity, while Copilot has the highest lexical diversity but lower lexical density. This analysis is of great importance for understanding the capabilities and limitations of LLMs, with implications for their applications in various areas. The concepts, and methodology are presented, the findings are discussed, and the paper concludes with reflections on future research and practical applications.

Keywords: world model; language model; artificial intelligence; lexical diversity; lexical density.

Resumo

A inteligência artificial (IA) transformou vários domínios, incluindo a linguística. Os grandes modelos de linguagem (LLM) revolucionaram a interação com o texto, fornecendo respostas que imitam a linguagem humana. Estes modelos não só geram texto, como também reflectem a sua interpretação do mundo. No entanto, a compreensão do mundo por parte destes modelos é limitada, o que levou à proposta de desenvolvimento de Large World Models (LWMs), que integram dados textuais, visuais e auditivos para uma compreensão mais completa. Este artigo utiliza uma perspetiva léxico-estatística para analisar a forma como os LLM articulam as respostas com base nos seus modelos do mundo. Utilizou-se uma conceção quase-experimental comparativa para avaliar seis LLMs diferentes. A metodologia centrou-se na medição da diversidade e da densidade lexical dos textos gerados por esses modelos. Os resultados demonstraram que o ChatGPT-4 tem uma densidade lexical elevada e uma diversidade lexical moderada, enquanto o Copilot tem a maior diversidade lexical mas uma densidade lexical mais baixa. Esta análise é de grande importância para a compreensão das capacidades e limitações dos LLMs, com implicações para suas aplicações em diversas áreas. Os conceitos, a metodologia e os resultados são apresentados e discutidos, concluindo-se com reflexões sobre futuras investigações e aplicações práticas.

Palavras-chave: modelo mundial; modelo linguístico; inteligência artificial; diversidade lexical; densidade lexical.

Recibido: 20/02/2024 Aceptado: 31/07/2024 Publicado: 30/12/2024

1. Introducción

La inteligencia artificial (IA) ha revolucionado numerosos campos y su impacto en la lingüística no es la excepción. Los Grandes Modelos de Lenguaje (LLM, por sus siglas en inglés), como GPT-3 y GPT-4, han transformado la manera en que interactuamos con el texto, de modo que proporcionan respuestas que imitan sorprendentemente el lenguaje humano. Estos modelos no solo comprenden y generan texto, sino que también reflejan la forma en que interpretan el mundo. Este artículo examina cómo los LLM articulan sus respuestas basados en sus modelos de mundo, utilizando una perspectiva léxico-estadística.

La noción de "modelo de mundo" ha sido fundamental en diversas disciplinas, desde la robótica hasta la ecología (Carvajal, 2013). Un modelo de mundo es una representación simplificada y abstracta de un entorno que permite comprender y predecir comportamientos o fenómenos que ocurren en él. En el ámbito de la inteligencia artificial, estos modelos permiten a los sistemas de IA simular y planificar acciones basadas en su percepción del entorno (Pérez Parejo, 2004).

En el contexto del procesamiento del lenguaje natural, los LLM han sido entrenados con vastas cantidades de datos textuales para reconocer y generar texto, basándose en patrones lingüísticos aprendidos. Sin embargo, su capacidad para comprender el mundo más allá del texto es limitada, lo

que ha llevado a la evolución hacia modelos más integrales, conocidos como Grandes Modelos del Mundo (LWM, por sus siglas en inglés), que buscan integrar datos textuales, visuales y auditivos para una comprensión más completa (Cheng *et al.,* 2023).

El objetivo de este artículo es comparar la articulación de respuestas de diferentes Grandes Modelos de Lenguaje (LLM) basados en sus modelos de mundo, utilizando medidas léxico-estadísticas como la diversidad léxica y la densidad léxica. Este análisis es crucial para entender las capacidades y limitaciones de los LLM en la representación del mundo, lo que, a su vez, tiene implicaciones para su aplicación en diversas áreas. Justificar esta investigación radica en la necesidad de mejorar la comprensión de cómo los modelos de lenguaje procesan y generan conocimiento del mundo, y cómo estas representaciones pueden ser optimizadas para aplicaciones más complejas y precisas.

El artículo presenta el marco conceptual, que incluye una revisión de los conceptos de "modelos de mundo" y "modelos de lenguaje", así como su evolución en el campo de la IA. Luego, se plantea la metodología utilizada para comparar la diversidad y la densidad léxica de varios LLM. En los resultados se discuten las implicaciones de los hallazgos para, finalmente, ofrecer conclusiones y reflexiones que contribuyan a futuras investigaciones y aplicaciones prácticas.

2. Marco conceptual

2.1. Modelos de mundo

Un modelo de mundo es una representación simplificada y abstracta de un entorno o conjunto de condiciones específicas, que se utiliza en diversas disciplinas para entender, analizar y predecir comportamientos o fenómenos dentro de ese entorno (Joly, 1988). Los modelos pueden ser tanto físicos (i.e. maquetas o simulaciones espaciales) como matemáticos (utilizando ecuaciones y fórmulas para representar relaciones y dinámicas) o computacionales, con modelos creados y analizados mediante *softwares* que pueden incluir simulaciones complejas y visualizaciones. Además, se utilizan en campos como la robótica, la inteligencia artificial, la literatura, la lingüística y muchos otros, lo que permite la planificación, la toma de decisiones y la resolución de problemas en contextos específicos. Por ejemplo, en inteligencia artificial, un modelo de mundo puede ayudar a un robot a entender su entorno y planificar acciones.

El concepto de "modelo" es complejo de definir, debido a que puede abordarse desde diferentes perspectivas. Sobre esta polisemia diserta Carvajal (2013), quien presenta esta variedad de sentidos.

Desde lo cotidiano, se puede decir que un modelo es un objeto que se imita, por ejemplo, un modelo para hacer un mueble. Otra acepción hace referencia a la exposición de un producto para su venta. Desde el punto de vista ético, representa la búsqueda de la perfección o el ideal, pero que no se alcanza, como cuando se hablaba del alumno ideal. Una perspectiva artística adopta la concepción de modelo como referencia a objetos, animales, personas o paisajes que intentan reproducirse (Carvajal, 2013).

Carvajal (2013) aclara que, desde el punto de vista epistemológico, el modelo puede considerarse como una especie de descripción de la realidad, que, por lo general, está al alero de una teoría. En otras palabras, se trata de una construcción mental y como tal, puede variar, tal como puede cambiar la realidad o la percepción de la realidad. El modelo, por lo tanto, nunca es el mundo real.

Asimismo, los modelos de mundo son esquemas conceptuales por los cuales los seres humanos intentan organizar el conocimiento que se alcanza a través de la experiencia (Carvajal, 2013).

Desde el ámbito de la crítica literaria, Asensi (2016) delimita las diferencias entre la teoría de los modelos de mundo y la teoría de los modelos de mundo posibles. Y explica que, en la teoría de los modelos de mundo, la referencia del texto se sitúa en el mundo, mientras que en la teoría de los mundos posibles, la referencia textual se sitúa fuera del mundo, tomando así una postura crítica. En este mismo sentido, Torres-Bravo (2020) respalda los postulados de este autor poniendo en relevancia la relación entre la teoría de los modelos de mundo y el pensamiento crítico y destaca las posibilidades didácticas que derivan de los postulados. Por su parte, Pérez Parejo (2004), desde la semiótica de la cultura, se refiere al concepto de modelos de mundo propuesto por la Escuela de Tartu-Moscú, a finales de los años sesenta, que establece que cada cultura, situada sobre un lenguaje, crea un modelo de mundo y explica que:

> Se entiende por modelos de mundo la percepción cultural que el sujeto tiene del mundo al que pertenece, un nuevo mundo resultante que, al haber pasado ya por el filtro de la cultura, difiere del mundo real, ya que la cultura proporciona inconscientemente unas estructuras de percepción que deforman el objeto. (p. 50)

Según esta Escuela, los modelos de mundo son cambiantes como todo aspecto derivado de la cultura. Esta idea es una reelaboración del concepto de "visión del mundo" (*Weltansicht*) propuesto por Wilhelm von Humboldt, quien hace depender el pensamiento del lenguaje (von Humboldt, 1991). Este autor, considerado el gran referente de la lingüística moderna, sostenía que la lengua era la manifestación externa del espíritu de los pueblos (Pape, 2010).

2.2. Modelos de lenguaje

Las ideas propuestas por von Humboldt (1991) sugieren plantear nuevas hipótesis en torno al lenguaje, que conversan con los modelos de mundo y su relación epistémica de la concepción de este. Deutscher (2011) destaca que la lengua se clasifica en dos ámbitos distintos: el de las etiquetas y el de los conceptos. Las etiquetas, por una parte, reflejan convenciones culturales, mientras que los conceptos reflejan la naturaleza. Si bien esto resulta interesante, la idea de una clasificación binaria es algo simplista, debido a que surge de manera natural la pregunta sobre cómo se clasificarían los términos más abstractos o aquellos en los cuales los límites se tornan más difusos (Deutscher, 2011).

Desde las discusiones surgidas a partir del relativismo lingüístico (hipótesis de Sapir-Whorf), que sostenía que la lengua puede influir en el pensamiento y en las percepciones, la diferencia entre las lenguas no solo está en los sonidos y en los signos, sino también en la visión del mundo (Matthews, 2014).

Si lo anterior se demostrara en un ejemplo, se podría imaginar que la imposibilidad de dialogar con un león no se debe a que el ser humano desconoce el idioma "leonés" (si es que existiera tal lenguaje), sino a que ambos no comparten la forma de percibir el mundo. Al respecto, Wittgenstein (1922) señaló que el límite del lenguaje humano se corresponde con el límite de su realidad. Con este aforismo, se manifiesta que con nuestras palabras no solo nos referimos a objetos o entidades, sino que nos relacionamos con el mundo y con los demás seres humanos.

Las lenguas clasifican el mundo de manera que agrupan cosas similares o, al menos, percibidas como similares. En el verbo *percibir,* se oculta el modelo de mundo (Deutscher, 2011).

2.3. El modelo de mundo y modelo de lenguaje en la IA

Se denomina modelo de mundo a un sistema de inteligencia artificial (IA) que construye una representación interna de un entorno y la utiliza para simular acontecimientos futuros dentro de ese entorno. Hasta ahora estos modelos de mundo apuntan a entornos muy limitados y controlados, como videojuegos, simuladores o conducción automática. El objetivo final es la creación de modelos de mundo generales que representen y simulen una amplia gama de situaciones, como las que se dan en el mundo real.

Por otro lado, la IA ha desarrollado grandes modelos de lenguaje (LLM, por sus siglas en inglés), los cuales son capaces de reconocer y generar texto, entre otras tareas. Los LLM son alimentados con grandes cantidades de datos textuales y se basan en el aprendizaje automático, concretamente con redes neuronales y procesamiento de lenguaje natural (NLP, por sus siglas en inglés) (Ozdemir, 2023).

2.3.1. De los grandes modelos de lenguaje a los grandes modelos de mundo

Desde el lanzamiento de ChatGPT en noviembre de 2022, los grandes modelos de lenguaje (LLM) han llamado la atención debido a su gran rendimiento en una amplia gama de tareas de lenguaje. La capacidad de los LLM para comprender y generar lenguaje se adquiere al procesar grandes cantidades de datos textuales.

Los modelos lingüísticos de gran tamaño (LLM) se refieren principalmente a modelos lingüísticos neuronales basados en transformadores, que contienen millones de parámetros y que se entrenan previamente con datos de texto masivos, como PaLM, LLaMA y GPT-4 (Minaee *et al.,* 2024). Estos autores realizan una revisión de los primeros modelos neuronales de lenguaje pre-entrenados, ya que son la base de los grandes LLM y los comparan en torno a sus características generativas, potencialidades y limitaciones.

Los LLM como GPT-3 y GPT-4 han revolucionado la forma de interactuar con la información. Al procesar grandes cantidades de datos de texto, estos modelos se han convertido en expertos para comprender y generar texto similar al que produciría un ser humano, lo que permite avanzar en áreas que van desde la creación de contenidos hasta el servicio de atención al cliente.

La dependencia del texto como única entrada limitaba su comprensión del mundo a una perspectiva textual. Esto ha sido superado con la siguiente etapa en el desarrollo de la IA que integró entradas multimodales; es decir, datos de sonido y visuales. Esto permitió a la IA no solo procesar datos de texto, sino también imágenes y sonidos, lo que le proporcionó una comprensión más precisa y enriquecida del entorno y del ser humano.

Herramientas como DALL-E y CLIP son capaces de generar imágenes a partir de texto, de manera que logran así romper las barreras entre lo textual y lo visual. La herramienta Sora es un

modelo de IA capaz de crear escenas de vídeo realistas a partir de instrucciones de texto, por lo que a veces es casi imposible discernir entre un escena de video real de una creada por la IA.

El ámbito de la inteligencia artificial (IA) podría estar al borde de un nuevo salto evolutivo, pasando de los Grandes Modelos de Lenguaje (LLM) a un concepto innovador y expansivo, denominado los Grandes Modelos de Mundo (LWM, por sus siglas en inglés). Nos adentramos así en el viaje desde los LLM centrados en el texto hasta la integración multimodal de los LLM, que llevará a los LWM, que integrarán la totalidad de nuestras experiencias físicas y digitales.

Los Grandes Modelos de Mundo (LWM) pueden representar el futuro de la IA, yendo más allá del texto, el sonido y las imágenes para incluir la totalidad de nuestras realidades físicas y digitales. Los LWM procesarán datos del mundo real procedentes de diversas fuentes, como sensores, cámaras, etc., para interpretar el mundo e interactuar con él de un modo que refleje la percepción y la cognición humanas en toda su complejidad. Para ello, se integran datos visuales, auditivos y hasta físicos, pues se incluirán sensores no humanos como infrarrojos, radares, escáneres térmicos y otros datos. Esto permite la toma de decisiones en tiempo real.

Un ejemplo de lo anterior es el uso de los modelos de aprendizaje automático para procesar datos a fin de mejorar la eficiencia y la precisión de los diagnósticos médicos en corto tiempo. Así, se ayudaría a mejorar los resultados de salud y las experiencias de los pacientes. En ese sentido, se podría hipotetizar que las máquinas inteligentes utilizarán el 'lenguaje' para interpretar e interactuar con nuestro mundo de manera más integrada.

Los teléfonos móviles pueden combinar a la perfección los mundos digital y físico. Al aprovechar los datos de dispositivos de realidad virtual y aumentada, estos modelos ofrecerán experiencias que trascenderán las interfaces tradicionales, como *smartphones,* televisores y computadores. Esta integración no solo cambiará la experiencia del usuario, sino que también proporcionará a la IA una visión más completa del ser humano, su comportamiento y su entorno; es decir, enriquecerá su modelo de mundo.

2.4. Riqueza léxica desde la diversidad y la densidad

Dado que los LLM basan su funcionamiento en grandes modelos lingüísticos, no es de extrañar que, para realizar un análisis de sus resultados, se utilicen técnicas léxico-estadísticas. Es más, esta área de la lingüística ha resultado de gran ayuda para el desarrollo de estudios asociados a la competencia léxica a través de los cuales se ha abordado operaciones, a veces de gran complejidad, las cuales toman como unidades de trabajo las palabras y los vocablos; la palabra como unidad del texto y el vocablo como unidad del léxico (López Morales, 2002). Tanto la diversidad léxica como la densidad léxica son medidas que tienen la ventaja de ser fácilmente operacionalizables y medibles, gracias a los avances en técnicas computacionales de análisis y a los estudios de corpus (Riffo *et al.,* 2019; Johansson, 2008).

Müller (1973, como se citó en Riffo *et al*., 2019) planteó que la estructura de un vocabulario incluye elementos cuantitativos simples: el número de palabras en un texto y la frecuencia de cada una de ellas. Además, destacó aspectos cualitativos, como la naturaleza gramatical de las palabras y las relaciones de asociación tanto gramaticales o semánticas, como paradigmáticas y sintagmáticas.

Cuantificar el vocabulario de un texto implica dos operaciones distintas que pueden ser sucesivas o simultáneas: a) el recuento de las palabras que componen el texto, cuyo número, representado por "N", proporciona una medida de la extensión del texto, y b) el recuento de los vocablos empleados en el texto, cuyo número, representado por "V", mide la extensión del vocabulario. Acorde con esta aproximación se encuentran trabajos como los de Kubát y Milička (2013) y Justice *et al.* (2014), quienes abordan la riqueza léxica con medidas de diversidad y densidad en el estudio de las producciones textuales de aprendientes de segundas lenguas y estudios de género.

Por su parte, Read (2010) sugiere que la riqueza léxica es una medida estadística que asume que la buena escritura se caracteriza por una variedad de palabras diferentes en lugar de un número limitado de palabras repetidas. La diversidad léxica se refiere al número de palabras diferentes utilizadas en un texto; un rango mayor indica una mayor diversidad (Johansson, 2008; López Morales, 2002; McCarthy y Jarvis, 2010). La medida aplicada en este caso es el índice de relación tipo-token (TTR), donde *type* corresponde a las palabras distintas del texto y *token* al número total de palabras.

Otra característica relevante es el porcentaje de palabras léxicas, o de contenido (verbos, sustantivos, adjetivos y algunos adverbios), en comparación con las llamadas palabras gramaticales o funcionales (artículos, preposiciones, conjunciones, entre otros). Esta medida es conocida como densidad léxica (Read, 2010; Johansson, 2008). La densidad léxica es un indicador de la calidad informativa del texto: un alto índice de densidad léxica indica que el texto contiene más palabras de contenido, lo que proporciona mayor información (Altmann *et al.,* 2009; Riffo *et al*. 2019; Vine *et al*., 2021). La fórmula para el cálculo de la densidad léxica es el número de palabras léxicas/ número total de palabras.

3. Metodología

El propósito de esta investigación es comparar el léxico de textos generados por seis Grandes Modelos de Lenguaje (LLM). Para ello, se empleó un diseño cuasi-experimental comparativo, que permite evaluar cómo estos modelos representan el mundo a través de la riqueza léxica, medida en términos de diversidad y densidad léxica.

El mundo de los LLMs hoy es enorme. Son cientos los grandes modelos de lenguaje que se han venido creando desde su invención, desde aquellos que requieren grandes máquinas para ser entrenados o, en algunos casos, para funcionar, hasta modelos pequeños que se diferencian principalmente por la cantidad de tokens que utilizan. Entre los más conocidos encontramos a Chat GPT-4o, Copilot, Llama, Mistral, Phi-3, Gemma, etc.

3.1. Selección de modelos

Se seleccionaron de forma dirigida seis LLM: ChatGPT-4o, Copilot, Gemma 2B, Llama 3 7B, Phi 3 3B, y Gemini 1.0 Pro. El criterio es, principalmente, la facilidad de uso con la herramienta LM Studio, software que permite utilizar varios a la vez comparándolos entre ellos. ChatGPT y Copilot se escogieron por contar con los investigadores con una cuenta pagada. Las características de estos modelos se describen en la Tabla 1.

En el caso de Copilot, aunque se basa en LLM previamente entrenados, como GPT-4, su inclusión en esta investigación está justificada porque ha sido optimizado específicamente para tareas de codificación e integración en entornos de desarrollo integrado (IDE), lo que lo distingue de otros LLM en términos de su aplicación práctica y su adaptación a contextos específicos de uso. Esta especialización permite evaluar cómo un LLM adaptado a una tarea concreta maneja la generación de texto y la diversidad léxica en comparación con otros modelos más generalistas.

Tabla 1
Modelos LLM utilizados en la investigación de alta precisión

Modelo	Parámetros (Billones)	Capacidad de Memoria	Optimización para Tareas	Velocidad de Respuesta	Disponibilidad	Precisión
Chat GPT-4	1,5	Alta	Conversación	Rápida	Amplia	Alta
Copilot	0,1	Media	Codificación	Rápida	Integrada en IDEs	Alta
Gemini 1.0 Pro	2	Alta	Análisis de Datos	Rápida	Corporativa	Muy Alta
Llama 3 7B	7	Alta	Generación de Texto	Media	Investigación	Alta
Gemma 2B	2	Media	Análisis de Sentimiento	Media	Corporativa	Media
Phi 3 3B	3	Media	NLP General	Rápida	Desarrolladores	Media

3.2. Generación de texto

Para la generación del texto, se le solicita a cada LLM que genere un documento con una extensión máxima de 10 páginas, en las cuales deben describir, a partir de su conocimiento, cómo conciben al ser humano, abordando sus complejidades. Para esto se ha utilizado el mismo Prompt en cada uno.

"Necesito que describas al ser humano, con todas sus complejidades y características como ser vivo, social y emocional. Describe cómo es su relación con el resto del entorno en este planeta y universo, la importancia que tiene este en el ecosistema, sus debilidades y fortalezas como especie. No debe sobrepasar tu descripción las 10 páginas. Debes considerar que lo que busco es representar el modelo de mundo que tienes sobre el ser humano".

3.3. Análisis

Las diferencias lingüísticas de cada texto se analizan desde la perspectiva de la riqueza léxica, específicamente, utilizando las medidas de diversidad léxica y densidad léxica. La diversidad léxica, como se mencionó anteriormente, se refiere a la variedad de palabras diferentes usadas en un texto en relación con el número total de palabras. Es una medida que registra la cantidad de palabras únicas que se usan en un texto en comparación con el total de palabras. La densidad léxica mide la cantidad de palabras con significado (léxicas) en un texto en relación con el número total de palabras. Las palabras léxicas incluyen sustantivos, verbos, adjetivos y adverbios, mientras que se excluyen las palabras funcionales como preposiciones, artículos, conjunciones y pronombres.

4. Resultados y discusión

La diversidad y densidad léxica son medidas utilizadas para evaluar la riqueza y calidad del vocabulario (Johansson, 2008; Altmann *et al.,* 2009; Riffo *et al*., 2019; Vine *et al*., 2021). Para realizar el cálculo de ambas medidas, este trabajo ha utilizado Chat GPT- 4 o, accediendo a la librería *NLTK (Natural Language Toolkit)* como herramienta para determinar las palabras únicas y con significado. Para este análisis, entonces, se especificó en el *prompt* que debía utilizar NLTK en los cálculos de la diversidad y densidad léxica. La Tabla 2 proporciona los datos sobre la diversidad y densidad léxica obtenidos a partir de los seis LLM seleccionados, utilizando las siguientes fórmulas:

- Diversidad léxica

$$DivL = \frac{V}{N}$$

Donde

V: vocablos o palabras únicas del texto

N: número total de palabras del texto.

- Densidad léxica

$$DenL = \frac{Vs}{N}$$

Donde

Vs: número de palabras únicas (vocablos) con significado semántico

N: número total de palabras el texto

Tabla 2
Resultados de diversidad y densidad léxica

Modelo	Número de palabras	Número de palabras únicas	Número total de palabras con significado	Diversidad léxica	Densidad léxica
Chat GPT 4o	542	285	429	0.5251	0.7915
Copilot	140	85	80	0.6075	0.5714
Gemma 2B	246	128	159	0.5189	0.6463
Llama 3 7B	257	149	155	0.5806	0.6031
Phi 3 3B	370	200	257	0.5397	0.6946
Gemini 1.0 Pro	391	218	258	0.5581	0.6609

La diversidad léxica se refiere a la variedad de palabras únicas utilizadas en un texto. Según la Tabla 2, Copilot presenta la mayor diversidad léxica (0.6075), seguido de Llama 3 7B (0.5806) y Gemini 1.0 Pro (0.5581). La alta diversidad léxica de Copilot sugiere que este modelo utiliza

un vocabulario amplio y variado, lo cual es beneficioso para tareas que requieren creatividad y variabilidad en el lenguaje (Altmann *et al.,* 2009).

En el ámbito de la generación de textos, es preciso indicar que la diversidad léxica es esencial para la generación de textos con características más naturales y menos repetitivas. Los modelos con alta diversidad léxica pueden generar contenido que se asemeja más a la producción humana, lo que es crucial para aplicaciones en procesamiento de lenguaje natural (PLN), como la redacción automática y la traducción de idiomas (Nippold, 2017). Además, una alta diversidad léxica puede mejorar la comprensión y el involucramiento del usuario al interactuar con sistemas de IA.

Por su parte, la densidad léxica mide la proporción de palabras con significado (léxicas) en un texto, excluyendo palabras funcionales como lo son artículos y preposiciones. Según los resultados, ChatGPT-4 presenta la mayor densidad léxica (0.7915), seguido de Phi 3 3B (0.6946) y Gemma 2B (0.6463). Una alta densidad léxica indica que el texto contiene más palabras de contenido, proporcionando así mayor información y relevancia en cada oración (Riffo *et al*., 2019; Vine *et al*., 2021).

La alta densidad léxica de ChatGPT-4 sugiere que este modelo es eficaz en la generación de textos informativos y concisos. Esto es especialmente valioso en aplicaciones donde la precisión y la relevancia del contenido son cruciales, como en la generación de informes técnicos y resúmenes de texto (McGregor *et al.,* 2013). Además, la alta densidad léxica puede ser un indicador de la capacidad del modelo para manejar tareas complejas que requieren un uso preciso y adecuado del lenguaje.

La comparación entre diversidad y densidad léxica revela diferencias que podrían ser significativas entre los modelos analizados. Copilot, a pesar de su alta diversidad léxica, presenta una densidad léxica menor (0.5714), lo que sugiere un equilibrio entre palabras funcionales y léxicas. Esto podría deberse a un enfoque en la diversidad del vocabulario para mejorar la capacidad de generación de texto variado, aunque con menos profundidad informativa (Larsen y Nippold, 2007).

Por otro lado, ChatGPT-4, con su alta densidad léxica y diversidad moderada (0.5251), parece estar optimizado para generar texto que es tanto variado como altamente informativo. Este equilibrio es ideal para tareas que requieren precisión y relevancia en el contenido, haciendo de ChatGPT-4 un modelo versátil para diversas aplicaciones en PLN (Goodwin y Ahn, 2010).

Gemma 2B y Phi 3 3B muestran un buen balance entre diversidad y densidad léxica, lo que indica que son capaces de generar textos con un vocabulario variado y un alto contenido informativo. Estos modelos pueden ser adecuados para aplicaciones que requieren tanto diversidad en el lenguaje como profundidad en la información, como en la redacción de artículos académicos y la creación de contenido educativo (Nippold, 2017).

Los resultados de este análisis tienen implicaciones para el desarrollo futuro de LLM. La capacidad de un modelo para equilibrar diversidad y densidad léxica es crucial para su eficacia en diferentes aplicaciones. Modelos como ChatGPT-4, que muestran un alto rendimiento en ambas métricas, son particularmente valiosos para tareas que requieren un lenguaje natural, preciso e informativo (Nippold, 2014).

Además, estos resultados destacan la importancia de la personalización y la optimización de LLM para tareas específicas. Por ejemplo, un modelo con alta diversidad léxica puede ser más adecuado para aplicaciones creativas y de entretenimiento, mientras que un modelo con alta densidad léxica es más apropiado para aplicaciones técnicas y científicas. La elección del modelo adecuado puede mejorar significativamente la calidad y efectividad del texto generado (Hirschman, 2000).

Sin embargo, no se debe olvidar que esta comparación se basa únicamente en medidas cuantitativas de diversidad y densidad léxica, sin considerar aspectos cualitativos del contenido generado. La evaluación cualitativa por parte de expertos podría proporcionar una visión más completa de la calidad del texto generado por cada modelo (Kamhi y Koenig, 1985). Por otro lado, el estudio no considera la variabilidad en la calidad del entrenamiento de los modelos. Los LLM pueden ser entrenados con diferentes conjuntos de datos y configuraciones de parámetros, lo que puede influir en su rendimiento.

Los modelos más avanzados hacen un uso cada vez más eficiente de la información dentro del contexto, por ejemplo eliminando datos no útiles, mejorando la predicción de tokens (Minaee *et al.*, 2024). Futuras investigaciones deberían considerar otras variables para proporcionar una evaluación más robusta y precisa de los modelos.

5. Conclusiones

Una vez analizados los datos obtenidos de los modelos estudiados, se ha llegado a las siguientes conclusiones:

El modelo Copilot es el que, hasta la actualidad, presenta en este estudio mayor diversidad léxica, lo que sugiere que podría estar optimizado para generar textos con un vocabulario muy diverso, de lo que se puede concluir que puede ser de utilidad para aplicaciones que requieren de léxico diverso y creatividad en el lenguaje. Este podría ser el caso de la creación de textos originales donde la imaginación juega un papel importante.

Por otro lado, Chat GPT-4o exhibe, al día de hoy, mayor densidad léxica, además de estar muy bien equilibrado, dando por resultado un texto que es tanto variado como altamente informativo, lo que lo posiciona en un modelo ideal para la realización de tareas que requieren de alta precisión y relevancia en el contenido, como en el caso de generación y tratamiento de textos de alta densidad terminológica, como textos técnicos o traducciones especializadas.

Estos resultados indican que cada modelo tiene distintas fortalezas. En primer lugar, la elección del modelo a usar puede depender del balance entre la diversidad del vocabulario y la relevancia de las palabras utilizadas según las necesidades específicas de la aplicación.

Comprendiendo que la densidad y la diversidad léxica medidas bajo *type token/ ratio* presentan limitaciones, especialmente las referidas a la dependencia de la medida en la extensión total del texto, resulta interesante que, pese a que el *prompt* permitía una gran extensión (10 páginas, no se solicita un número de palabras totales, para examinar el comportamiento de los LLM), Chat GPT-4o es el LLM que genera un resultado con el texto más extenso (542 palabras) y, a la vez, con mayor número de palabras con alto contenido léxico-semántico.

Lo anterior implica que el texto generado tiene una densidad léxica mayor a las de los otros LLM utilizados, y supera incluso los resultados de otras investigaciones en las cuales se ha estudiado la riqueza léxica de personas en torno a temáticas especializadas (Riffo *et al.,* 2019; Vine *et al.,* 2021).

Por otra parte, al considerar que el modelo del mundo que se forma la IA se expresa a través del lenguaje, una forma de aproximarse a su análisis puede venir de la diversidad y de la densidad léxica, por tanto, es una forma de acceder a la competencia léxica del LLM. Es así como se puede observar a partir de los resultados, que un modelo como Chat GPT-4o parece ser el más equilibrado, pudiendo estar describiendo de una forma más significativa un concepto determinado como el de "ser humano".

La transición de los LLM a los LWM representa un cambio de paradigma en la IA, que pasa de entender el mundo a través del texto a "experimentarlo" como lo hacen los humanos, en toda su complejidad. Esta evolución promete desbloquear nuevas capacidades y aplicaciones, de manera que cambia fundamentalmente la forma en que interactuamos con la tecnología y percibimos el mundo que nos rodea.

El camino hacia las LWM no es solo un avance tecnológico, sino un paso más hacia la creación de máquinas que comprendan el mundo e interactúen con él de una forma realmente humana. Sin embargo, aún no se puede vislumbrar en su totalidad la magnitud del impacto que puedan tener estos avances en áreas como la educación, por ejemplo, en el desarrollo del pensamiento crítico del estudiantado (Torres-Bravo, 2020).

Otro punto a tener en cuenta es la definición de modelo de mundo desde las teorías de modelo de mundo revisadas en este artículo y desde la concepción de modelo desde el punto de vista de la IA. Ahora que nos encontramos al borde de esta nueva era, es crucial recorrer este camino centrándonos en consideraciones éticas y en el impacto social, a fin de garantizar que los beneficios de los LWM sean accesibles y positivos para todos.

5.1. Limitaciones y proyecciones del experimento

Los modelos se seleccionaron en forma dirigida, sin considerar la diversidad actual, entrenados con bases más completas y con más memoria, lo que podría generar diferencias significativas para modelar el mundo.

Además, es importante considerar la limitación del experimento, porque se registran solamente los parámetros de la densidad léxica y la diversidad léxica para medir el modelo del mundo. Falta considerar un análisis cualitativo, donde expertos puedan medir la calidad y profundidad del contenido generado.

Referencias

Altmann, E. G., Pierrehumbert, J. B. y Motter, A. E. (2009). Beyond Word Frequency: Bursts, Lulls, and Scaling in the Temporal Distributions of Words. *PLoS ONE, 4*(11), 1-7. https://doi.org/10.1371/journal.pone.0007678

Asensi, M. (2016). Teoría de los modelos de mundo y teoría de los mundos posibles. *Actio Nova: Revista de teoría de la literatura y literatura comparada*, (0), 38-55. https://doi.org/10.15366/actionova2016.0.003

Carvajal, Á. (2013). Teorías y modelos: formas de representación de la realidad. *Revista Comunicación*, *12*(1), 33-46.

Cheng, M., Durmus, E. y Jurafsky, D. (2023). Marked Personas: Using Natural Language Prompts to Measure Stereotypes in Language Models. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 1504–1532). Toronto, Canada. Association for Computational Linguistics.

Deutscher, G. (2011). *El prisma del lenguaje. Cómo las palabras colorean el mundo*. Editorial Ariel.

Goodwin, A. P. y Ahn, S. (2010). A meta-analysis of morphological interventions: effects on literacy achievement of children with literacy difficulties. *Annals of dyslexia*, *60*(2), 183–208. https://doi.org/10.1007/s11881-010-0041-x

Hirschman, M. (2000). Language repair via metalinguistic means. *International Journal of Language & Communication Disorders*, *35*(2), 252–268. https://doi.org/10.1080/136828200247179

Johansson, V. (2008). Lexical diversity and lexical density in speech and writing: a develop-mental perspective. *Working Papers in Linguistics*, *53*, 61-79. https://journals.lub.lu.se/index.php/LWPL/article/view/2273/1848

Joly, F. (1988). *La cartografía*. Oikos-Tau.

Justice, L. M., Schmitt, M. B., Murphy, K. A., Pratt, A. y Biancone, T. (2014). The 'robustness' of vocabulary intervention in the public schools: targets and techniques employed in speech-language therapy. *International journal of language & communication disorders*, *49*(3), 288–303. https://doi.org/10.1111/1460-6984.12072

Kamhi, A. G. y Koenig, L. A. (1985). Metalinguistic Awareness in Normal and Language-Disordered Children. *Language Speech and Hearing Services in Schools, 16*, 199-210. https://doi.org/10.1044/0161-1461.1603.199

Kubát, M. y Milička, J. (2013). Vocabulary Richness Measure in Genres. *Journal of Quantitative Linguistics*, *20*(4), 339-349. https://doi.org/10.1080/09296174.2013.830552

Larsen, J. A. y Nippold, M. A. (2007). Morphological analysis in school-age children: dynamic assessment of a word learning strategy. *Language, speech, and hearing services in schools*, *38*(3), 201–212. https://doi.org/10.1044/0161-1461(2007/021)

López Morales, H. (2002). Los índices de 'Riqueza léxica' y la enseñanza de lenguas. En *Del texto a la lengua: la aplicación de los textos a la enseñanza-aprendizaje del español* (pp. 15-28). https://dialnet.unirioja.es/descarga/articulo/5419218.pdf

Matthews, P. (2014). *Sapir-Whorf hypothesis*. En P. H. Matthews (Hrsg.), *The Concise Oxford Dictionary of Linguistics* (3.° ed.). Oxford University Press.

Mccarthy, P. y Jarvis, S. (2010). MTLD, vocd-D, and HD-D: a validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods*, *42*, 381-392.

McGregor, K. K., Oleson, J., Bahnsen, A. y Duff, D. (2013). Children with developmental language impairment have vocabulary deficits characterized by limited breadth and depth. *International journal of language & communication disorders*, *48*(3), 307–319. https://doi.org/10.1111/1460-6984.12008

Minaee, S., Mikolov, T., Nikzad, N., Chenaghlu, M., Socher, R., Amatriain, X. y Jianfeng, G. (2024). Large Language Models: A Survey. *Arxiv*. https://doi.org/10.48550/arXiv.2402.06196

Nippold, M. A. (2014). *Language sampling with adolescents: Implications for intervention*. (2nd ed.). Plural Publishing.

Nippold, M. A. (2017). Building a literate lexicon in adolescents: Selecting words for language intervention. 14th International Congress for the Study of Child Language (IASCL), University of Lyon, France.

Ozdemir, S. (2023). *Quick Start Guide to Large Language Models: Strategies and Best Practices for Using ChatGPT and Other LLMs*. Addison-Wesley Professional.

Pape, C. (2010). *Sprachliche Relativität und ihre Implikationen für den Englischunterricht* [Tesis de licenciatura, Humboldt-University of Berlin]. GRIN Verlag, https://www.grin.com/document/169274

Pérez Parejo, R. (2004). Modelos de mundo y tópicos literarios: la construcción ficcional al servicio de la ideología del poder. *Revista De Literatura*, *66*(131), 49–76. https://doi.org/10.3989/revliteratura.2004.v66.i131.140

Read, J. (2010) *Assessing vocabulary*. (9. ed). Cambridge University Press.

Riffo, K. F., Osuna, S. H. y Lagos, P. S. (2019). Descripción de la diversidad y densidad léxicas en noticias escritas por estudiantes de periodismo. *Revista Brasileira de Linguística Aplicada*, *19*(3), 499-528. https://doi.org/10.1590/1984-6398201914113

Torres-Bravo, L. (2020). Teoría de los modelos de mundo: lectura literaria y posibilidades didácticas. *Enunciación*, *25*(2), 292-305. https://doi.org/https://doi.org/10.14483/22486798.16634

Vine, A., Fuentes Riffo, K., Neira Martínez, A. y Poza Molina, C. (2021). Lexical richness on written texts by foreign spanish language learners: A contribution to the specific notions of the curricular plan of the Cervantes Institute. *Nueva revista del Pacífico*, (75), 201-226.

von Humboldt, W. (1991). Escritos sobre el lenguaje. (A. Sánchez Pascual, Trad.). Península.

Wittgenstein, L. (1922). *Tractatus Logius-Philosophicus, con una introducción de Bertrand Russell.* Trench Trübner.

Contribución del autor

Gabriela Kotz, Pedro Salcedo y Karina Fuentes han participado en la concepción y diseño del artículo; Pedro Salcedo, en la recolección de datos; Gabriela Kotz, Pedro Salcedo y Karina Fuentes, en el análisis e interpretación de datos Gabriela Kotz; Pedro Salcedo y Karina Fuentes, en la redacción del artículo; Gabriela Kotz en la revisión crítica del artículo. Gabriela Kotz, Pedro Salcedo y Karina Fuentes aprueban la versión que se publica.

Agradecimientos

Agradecemos a la Universidad de Concepción, en especial a las Facultades de Educación y de Humanidades y Arte por su apoyo institucional, y a los Departamentos de Metodología de la Investigación e Informática Educacional y al Departamento de Idiomas Extranjeros por las facilidades entregadas para el desarrollo de este trabajo.

También merecen un especial agradecimiento los editores de la Revista Lengua y Sociedad por animarnos a publicar y, sobre todo, la sección de corrección de estilo por su ayuda y paciencia en la revisión de este artículo.

Financiamiento

La investigación se realizó sin financiamiento.

Conflicto de intereses

Los autores no presentan conflicto de interés.

Correspondencia: gkotz@udec.cl

Trayectoria académica de los autores

Gabriela Kotz es profesora de alemán egresada de la Universidad Nacional de Córdoba, Argentina, es doctora en lingüística por la Universidad de Concepción, Chile y docente del Departamento de Idiomas Extranjeros de la Facultad de Humanidades y Arte de la misma Universidad, donde imparte cursos de alemán para la carrera de Traducción / Interpretación en Idiomas Extranjeros y dicta cursos en el programa de postgrado de lingüística. Trabaja en lingüística aplicada y lexicología, donde se encuentran sus principales líneas de investigación. Actualmente se desempeña en el cargo de Profesor Asociado y sus intereses investigativos actuales giran en torno al léxico, las emociones y al proceso de enseñanza/aprendizaje en idiomas extranjeros.

Pedro Salcedo es profesor de Matemática y Física y Magister en Ciencias de la Computación por la Universidad de Concepción, además es Doctor en Inteligencia Artificial por la Universidad de Nacional de Educación a Distancia (UNED), España. Es miembro de claustro de los Doctorados en Educación, Psicología y Lingüística, además de integrante del comité académico del Doctorado en Inteligencia Artificial del consorcio Cruch Biobío, Chile. Trabaja en Ingeniería del Conocimiento, IA en educación, didáctica de las matemáticas, lingüística aplicada y lexicología, donde se encuentran sus principales líneas de investigación. Actualmente se desempeña como Profesor Titular en la Universidad de Concepción, Chile y sus intereses investigativos actuales giran en torno a la Computación Afectiva (la IA de las Emociones), integración de las TIC en el aula, competencias TIC y Lingüística Aplicada.

Karina Fuentes es doctora en Lingüística por la Universidad de Concepción. Máster en Periodismo Digital por la Universidad Autónoma de Madrid. Licenciada en Comunicación Social y periodista por la Universidad de Concepción. Se desempeña como académica en el Departamento de Metodología de la Investigación e Informática Educacional de la Facultad de Educación de la Universidad de Concepción. Investigadora en el proyecto Anid, Fondecyt Entrenando la comprensión lectora online mediante una plataforma virtual. Evidencias empíricas desde la neurociencia, 2024-2027.