diacrítica
Revista do Centro de Estudos Humanísticos

Vol. 40, n.º 1, 2026, pp. 173–194. doi: 10.21814/diacritica.6700

ChatGPT como lingüista: ¿puede la IA identificar la categoría gramatical del léxico disponible?

ChatGPT como linguista: a IA pode identificar a categoria gramatical do léxico disponível?

ChatGPT as a linguist: can AI identify the grammatical category of available lexicon?

Roberto Rubio Sánchez[1]
0000-0001-5043-9781

[1] Universidad de Salamanca, España.

Autor correspondiente: roberto.rubio@usal.es

Resumen. Este estudio, enmarcado en el ámbito de la disponibilidad léxica, explora las posibilidades que ofrece la inteligencia artificial (IA) para automatizar el análisis gramatical del léxico disponible de aprendices de español como lengua extranjera. A partir de una muestra de 876 estudiantes universitarios procedentes de Portugal, Italia y Egipto, se ha trabajado con los datos recogidos en una prueba asociativa compuesta por doce centros de interés. El trabajo se centra en evaluar la eficacia del modelo GPT-4o de ChatGPT para clasificar las unidades léxicas según su categoría gramatical. Para ello, se ha diseñado un modelo personalizado mediante la técnica few-shot, con instrucciones precisas y ejemplos. La clasificación realizada por la IA ha sido posteriormente comparada con una codificación manual de referencia, lo que ha permitido valorar su nivel de fiabilidad, identificar errores recurrentes y ajustar las instrucciones iniciales para mejorar el rendimiento del sistema. Los resultados muestran una alta coincidencia entre ambas clasificaciones, lo que sugiere que este tipo de herramientas puede convertirse en un recurso útil para agilizar tareas analíticas complejas en investigaciones lingüísticas. Además, el estudio analiza la distribución gramatical del léxico disponible en los distintos centros de interés, observando variaciones en función del tipo de estímulo. Las conclusiones destacan el potencial de la IA como asistente metodológico, especialmente en contextos que requieren procesar grandes volúmenes de datos lingüísticos, y abren nuevas líneas para su aplicación futura en estudios de lingüística aplicada.

Palabras-clave: Disponibilidad léxica. Español como lengua extranjera. Análisis automatizado. Gran modelo de lenguaje. Naturaleza gramatical.

Resumo. Este estudo, inserido no âmbito da disponibilidade lexical, explora as possibilidades oferecidas pela inteligência artificial (IA) para automatizar a análise gramatical do léxico disponível de aprendentes de espanhol língua estrangeira. Com base numa amostra de 876 estudantes universitários provenientes de Portugal, Itália e Egito, foram trabalhados os dados recolhidos numa prova associativa composta por doze centros de interesse. O estudo centra-se na avaliação da eficácia do modelo GPT-4o do ChatGPT para classificar as unidades lexicais segundo a sua categoria gramatical. Para tal, foi concebido um modelo personalizado com a

técnica *few-shot*, utilizando instruções precisas e exemplos. A classificação realizada pela IA foi posteriormente comparada com uma codificação manual de referência, o que permitiu avaliar o seu grau de fiabilidade, identificar erros recorrentes e ajustar as instruções iniciais para melhorar o desempenho do sistema. Os resultados revelam uma elevada concordância entre ambas as classificações, o que sugere que este tipo de ferramenta pode tornar-se um recurso útil para agilizar tarefas analíticas complexas em investigações linguísticas. Além disso, o estudo analisa a distribuição gramatical do léxico disponível nos diferentes centros de interesse, observando variações em função do tipo de estímulo. As conclusões destacam o potencial da IA como assistente metodológico, especialmente em contextos que exigem o processamento de grandes volumes de dados linguísticos, e abrem novas perspetivas para a sua aplicação futura em estudos de linguística aplicada.

Palavras-chave: Disponibilidade lexical. Espanhol língua estrangeira. Análise automatizada. Grande modelo de linguagem. Natureza gramatical.

Abstract. This study focuses on lexical availability and explores the potential of artificial intelligence (AI) to automate the grammatical analysis of the available lexicon produced by learners of Spanish as a foreign language. Drawing on a sample of 876 university students from Portugal, Italy, and Egypt, the research is based on data collected through a lexical availability test consisting of 12 prompts, adapted to the aims of the study and the characteristics of the participant population. The main objective is to evaluate the performance of ChatGPT, specifically its GPT-4o model, in classifying lexical items according to their grammatical category. To this end, a custom model was developed using the few-shot technique, incorporating clear instructions and representative examples. The AI-generated classifications were then compared with a manually coded reference set, enabling an assessment of the model's reliability, the identification of recurring errors, and the refinement of prompts to enhance performance. The results reveal a high degree of agreement between the two classifications, suggesting that AI-based tools can be valuable for optimising complex linguistic analysis. In addition, the study examines the grammatical distribution of the available lexicon across the different prompts, noting how it varies depending on the type of stimulus presented. The findings highlight the potential of AI as a methodological assistant, especially in contexts requiring the processing of large volumes of linguistic data, and they open new perspectives for its application in applied linguistics research.

Keywords: Lexical availability. Spanish as a foreign language. Automated analysis. Large language model. Grammatical nature.

1. Introducción

La disponibilidad léxica es un ámbito de la lingüística aplicada que estudia el vocabulario que los hablantes pueden activar de forma inmediata en función de determinados estímulos temáticos denominados 'centros de interés' (CI). Para su análisis, se recurre a pruebas asociativas en las que se solicita a los informantes que enumeren, en un tiempo limitado, las palabras que les vienen a la mente en relación con un tema específico. Las unidades léxicas evocadas en este tipo de ejercicios forman parte del lexicón mental del hablante, el cual puede entenderse de manera resumida como el diccionario que cada persona tiene en su mente (Santos Díaz, 2015, p. 403). Tradicionalmente, los estudios de disponibilidad léxica se han centrado en cuantificar y clasificar los vocablos producidos por los informantes en cada centro, así como establecer comparaciones entre distintos grupos de hablantes, a partir de variables sociolingüísticas

como el sexo, el nivel sociocultural, la procedencia geográfica o la lengua materna (Aabidi, 2019). Además de los recuentos habituales del número total de palabras y de vocablos, el promedio de palabras por informante o el índice de cohesión, algunos trabajos también han abordado el análisis de la naturaleza gramatical de las unidades léxicas que conforman el corpus, con distintos niveles de profundidad (entre otros, (Carcedo González, 2000; Nalesso, 2019; Rubio Sánchez, 2017; Sánchez Gómez, 2005). Sin embargo, esta línea de análisis no resulta tan frecuente, probablemente debido a las dificultades y complejidad de su análisis manual.

En este contexto, nuestro objetivo es ampliar las posibilidades del análisis cualitativo en el estudio del léxico disponible mediante una metodología que automatiza esta tarea, tradicionalmente laboriosa, a través del uso de inteligencia artificial (IA). Concretamente, proponemos emplear un gran modelo de lenguaje (ChatGPT, modelo GPT-4o) para identificar y clasificar gramaticalmente las lexías aportadas por los informantes. Esta herramienta permite obtener de forma rápida y sistemática el número y el porcentaje de vocablos clasificados en cada categoría gramatical (sustantivos, adjetivos, verbos, sintagmas nominales, sintagmas verbales, etc.), reproduciendo con precisión la labor que realizaría un lingüista de forma manual, pero con una apreciable mejora en términos de eficiencia (acelera el proceso de clasificación) y escalabilidad (facilita el tratamiento de grandes volúmenes de datos).

Esta optimización permite efectuar un estudio -*a priori*- más preciso de la incidencia de las categorías gramaticales. Asimismo, el ahorro considerable de tiempo que proporciona esta metodología permite al investigador centrarse en el análisis interpretativo y cualitativo de los resultados, delegando en la herramienta de IA las tareas mecánicas iniciales. De este modo, el modelo actúa como un asistente encargado del trabajo preliminar más laborioso, dejando al lingüista la tarea fundamental de supervisar, corregir posibles imprecisiones y profundizar en la discusión e interpretación de los hallazgos.

2. Marco teórico

Como hemos mencionado, el análisis de la categoría gramatical de las unidades léxicas que integran el corpus ha sido abordado en múltiples investigaciones previas. Si bien la mayoría de estos estudios examinan los sustantivos, adjetivos y verbos en relación con los diferentes CI, algunos investigadores han adoptado una perspectiva más descriptiva, recopilando las lexías simples (como Galloso Camacho, 2003), mientras que otros han incorporado un componente cuantitativo, contabilizándolos y determinando su proporción dentro del conjunto del corpus (Fernández Smith et al., 2008 o Sánchez-Saus Laserna, 2011). Asimismo, existen estudios que amplían el objeto de estudio a los grupos nominales y verbales (Mateo García, 1998 o Samper Hernández, 2001), o que examinan la distribución de las categorías gramaticales según el nivel de competencia en español de los estudiantes (Palapanidi & Mavrou, 2024; Sánchez Gómez, 2005).

Todos los estudios que emplean los CI tradicionales constatan de forma recurrente la predominancia de los sustantivos en las listas de léxico disponible. De hecho, varios autores, como Sánchez-Saus Laserna (2011) o Šifrar Kalan (2014) han observado que en

las listas resultantes de las pruebas clásicas apenas aparecen palabras que no sean sustantivos, en especial sustantivos de referentes concretos, quedando subrepresentados los adjetivos, verbos u otras categorías. Estas autoras consideran que esto se debe a los estímulos ofrecidos, por lo que proponen la integración de otros CI como 'Acciones y actividades habituales', en el que obtienen una mayoría de verbos. A su vez, Sánchez-Saus Laserna (2011) también incorpora en su estudio el centro 'Aspecto físico y carácter', en el que aparecen significativamente más adjetivos (70% del total de unidades recogidas en el centro) que sustantivos. Esta estrategia de rediseño ha demostrado efectividad para diversificar la respuesta léxica.

En esta misma línea, Tomé Cornejo (2015) señala que la lista tradicional de CI es insuficiente para abarcar todas las clases de palabras. En consecuencia, sugiere introducir nuevos estímulos o reformular los existentes con el fin de provocar la aparición de verbos, adjetivos e incluso unidades pluriverbales. De este modo, propone convertir ciertos temas en guiones de acción: por ejemplo, en lugar del centro estático 'Objetos colocados en la mesa para la comida', emplear la consigna 'Poner la mesa', lo que "permitiría captar, además de los sustantivos concretos ..., un buen número de verbos y algunas unidades pluriverbales como puede ser *despejar la mesa*" (Tomé Cornejo, 2015, p. 316).

Un estudio reciente de Palapanidi y Mavrou (2024) confirma esta relación entre la naturaleza del estímulo y la categoría gramatical de las respuestas. Tras analizar siete CI, observaron que la categoría predominante varía según el estímulo: los sustantivos siguen siendo la clase más frecuente; en cambio, en el centro 'Acciones diarias' destacan los verbos y sintagmas verbales. Además, al igual que en el estudio de Sánchez Gómez (2005), en esta investigación se analiza cómo influye el nivel de competencia en español de los informantes en la evocación de diferentes categorías gramaticales. Ambos trabajos muestran que un mayor dominio del idioma favorece la aparición de verbos y sintagmas verbales en los CI.

En suma, los trabajos más relevantes coinciden en retratar un léxico disponible esencialmente nominal, aunque modulado por variables como el diseño de la prueba o el nivel del hablante. Sin embargo, estos análisis requieren una categorización manual, lo que supone una tarea laboriosa y susceptible de errores en la ejecución. Por ello, resulta pertinente explorar herramientas que permitan automatizar la clasificación gramatical del léxico disponible, ámbito en el que se enmarca el presente estudio. Mientras que en disciplinas como las matemáticas, la física o la ingeniería informática ya se han desarrollado investigaciones evaluando estos modelos (véanse, por ejemplo, Dao & Le, 2023; Rodríguez-Echeverría et al., 2024; Wang et al., 2024; entre muchos otros), en el ámbito lingüístico estos análisis son aún escasos. No obstante, existen algunas contribuciones puntuales, como el estudio de Ronan y Schneider (2023), que evaluó el rendimiento de ChatGPT-4 en un examen de lingüística a nivel universitario, obteniendo resultados prometedores en morfología, fonología y sintaxis. Nuestro trabajo se enmarca en esta misma dirección: explorar las potenciales aplicaciones de la IA en la investigación lingüística, al tiempo que examinamos críticamente sus carencias.

3. Marco metodológico

Los estudios de disponibilidad léxica parten de la recolección de datos a partir de una encuesta que se compone de dos secciones. En la primera parte, los informantes proporcionan información personal y académica, la cual se emplea para codificar las variables de análisis. La segunda sección es el núcleo de la encuesta y consiste en presentar a los estudiantes los CI a partir de los cuales deben escribir todas las palabras que les vengan a la mente en dos minutos por campo. Tal como explica Tomé Cornejo (2015), el Proyecto Panhispánico ha estandarizado el uso de 16 centros obligatorios, que se basan en los temas propuestos originalmente en el *Français Fondamental* de Gougenheim. (1964, como citado en Tomé Cornejo, 2015), aunque algunos enunciados han sido adaptados ligeramente en distintas investigaciones. Estos centros abarcan los campos 'Partes del cuerpo', 'La ropa', 'Partes de la casa (sin los muebles)', 'Los muebles de la casa', 'Comidas y bebidas', 'Objetos colocados en la mesa para la comida', 'La cocina y sus utensilios', 'La escuela', 'Iluminación y calefacción', 'La ciudad', 'El campo', 'Medios de transporte', 'Trabajos del campo y del jardín', 'Los animales', 'Juegos y diversiones' y 'Profesiones', y permiten una comparabilidad sistemática entre estudios.

Sin embargo, al analizar investigaciones realizadas con estudiantes de español como lengua extranjera (ELE), se observa una tendencia creciente a modificar el número de centros según los objetivos específicos del estudio o las características del grupo encuestado. Mientras que numerosos trabajos con estudiantes de ELE mantienen los 16 centros del Proyecto Panhispánico como base común, otros estudios incorporan un centro adicional ('Los colores') alcanzando 17, como Mendoza Puertas (2018) o Serfati (2022). Otros trabajos adoptan enfoques más flexibles, aumentando el número de centros, por ejemplo, Sánchez-Saus Laserna (2011, 2016) analiza 18, Gallego Gallego (2014) emplea 20 y Gamazo Carretero (2014) trabaja con 23. A la vez, es frecuente encontrar estudios con estudiantes de ELE que reducen la lista de centros, como Rubio Lastra (2018), que utiliza 15; Rubio Sánchez (2016, 2017), 14; y Šifrar Kalan (2009, 2014), 11.

Esta variabilidad responde, en gran medida, al perfil del informante y a la necesidad de evitar la fatiga cognitiva durante la prueba, especialmente en contextos de aprendizaje formal. Šifrar Kalan (2009, p. 65) subraya que ciertos centros tradicionales resultan poco productivos para estudiantes extranjeros, ya que se refieren a realidades lejanas o poco familiares, generando escasa producción léxica.

Teniendo en cuenta estos antecedentes y la especificidad de nuestra muestra, en esta investigación se ha decidido modificar parcialmente la lista de centros, adaptándola a nuestros objetivos de investigación. Se han descartado cinco centros del inventario clásico: 'Partes de la casa (sin los muebles)', 'Objetos colocados en la mesa para la comida', 'Los muebles de la casa', 'Iluminación y calefacción' y 'Trabajos del campo y del jardín', todos ellos señalados por Bellón Fernández, Samper Padilla y Samper Hernández (2003, p. 59) como áreas con un rendimiento léxico bajo, al tratarse de campos cerrados y poco relevantes para el perfil de los encuestados.

De los once centros restantes, se han conservado siete por su valor comparativo y por representar distintas categorías semánticas, según la clasificación de Hernández Muñoz (2005): 'Partes del cuerpo' (estructura cerrada), 'La ropa', 'Comidas y bebidas',

'Animales' y 'Profesiones' (categorías naturales), 'Juegos y diversiones' (categoría radial) y 'La escuela' (categoría de esquema, como argumenta Tomé Cornejo, 2010).

Además, se han introducido cinco centros nuevos orientados a ampliar la variedad gramatical y temática de las respuestas. En concreto, 'Acciones y actividades habituales' pretende fomentar la producción de verbos, mientras que 'Aspecto físico y carácter' activa mayormente adjetivos. Por otro lado, para explorar cómo afecta un estímulo de corte sociocultural en la categoría gramatical de los vocablos actualizados, se han incluido 'La lengua española', 'España y los españoles' y un centro adaptado según el país de origen de los informantes: 'Portugal y los portugueses', 'Italia y los italianos' o 'Egipto y los egipcios'. En la Tabla 1 se muestran trabajos que han empleado estos CI en sus investigaciones.

Tabla 1. Estudios que emplean los CI añadidos en nuestra investigación

	Acciones y actividades habituales	Aspecto físico y carácter	La lengua española	España	Los españoles	Portugal	Los portugueses	Italia	Los italianos
Rodríguez Menduiña (2010)			x	x	x				
Šifrar Kalan (2014)	x[1]								
Gamazo Carretero (2014)		x[2]		x	x	x	x		
Gallego Gallego (2014)				x					
Sánchez-Saus Laserna (2011, 2016)	x	x							
Rubio Sánchez (2016)	x	x	x	Juntos		Juntos			
Rubio Sánchez (2017)	x	x	x	Juntos				Juntos	
Rubio Lastra (2018)	x	x[3]							
Herreros Marcilla (2020)			x[4]	x	x				
Fernández Gómez (2021)				Juntos					
Palapanidi (2024)				x					
Trigo Ibáñez et al. (2025)				Juntos					

En total, la prueba empleada en esta investigación se compone de 12 CI, cuidadosamente seleccionados para equilibrar la comparabilidad con estudios previos y la adecuación al perfil de los estudiantes no nativos. Esta selección, como demuestra la

[1] Este centro aparece bajo el enunciado 'Acciones que se realizan todos los días' en Šifrar Kalan (2014).

[2] La autora emplea dos CI diferentes: 'Aspecto físico y carácter de los españoles' y 'Aspecto físico y carácter de los portugueses'.

[3] En Rubio Lastra (2018) se muestra como 'Descripción física y carácter'.

[4] Herreros Marcilla (2020) lo denomina 'El idioma español'.

literatura revisada, responde a una estrategia metodológica que prioriza la pertinencia temática, el rendimiento léxico y la naturaleza del grupo encuestado, garantizando así la validez de los datos recogidos en el contexto de ELE. Así, los 12 CI empleados han sido:

a) 'Partes del cuerpo' (CUE)
b) 'La ropa' (ROP)
c) 'La escuela' (ESC)
d) 'Acciones y actividades habituales' (ACC)
e) 'España y los españoles' (ESP)
f) 'Comidas y bebidas' (COM)
g) 'Animales' (ANI)
h) 'Juegos y diversiones' (JUE)
i) 'Aspecto físico y carácter' (ASP)
j) 'Lengua española' (LEN)
k) 'Profesiones' (PRO)
l) 'Portugal y los portugueses' (POR) / 'Italia y los italianos' (ITA) / 'Egipto y los egipcios' (EGI)

3.1. Participantes/Corpus

En este estudio se ha trabajado con un total de 876 estudiantes de español como lengua extranjera. La muestra está conformada por estudiantes universitarios de todos los niveles (inicial, intermedio, avanzado y superior) de diversas instituciones de tres países: Portugal, Italia y Egipto. En el caso de Portugal, han participado 341 informantes de las universidades de Aveiro, Beira Interior, Coímbra, Lisboa y Oporto; en Italia se ha trabajado con 301 estudiantes de las universidades de Bérgamo, Bolonia y Venecia; y en Egipto, la muestra procede íntegramente de la Universidad de El Cairo donde se recopilaron 234 encuestas. Este volumen de participantes adquiere especial relevancia si se considera que, tal como indica Sánchez-Saus Laserna (2016, p. 34), los estudios de disponibilidad léxica suelen trabajar con muestras globales de en torno a los 300 participantes. En nuestro caso, cada grupo nacional ronda esa cifra, lo que equivale prácticamente a haber desarrollado tres estudios paralelos bajo los parámetros convencionales del campo. Esta amplitud muestral no solo aporta solidez estadística a los resultados obtenidos, sino que permite, además, realizar comparaciones significativas entre contextos diversos, favoreciendo un análisis más profundo y contrastivo de los repertorios léxicos de los aprendices de español.

3.2. Procesamiento informático de los datos

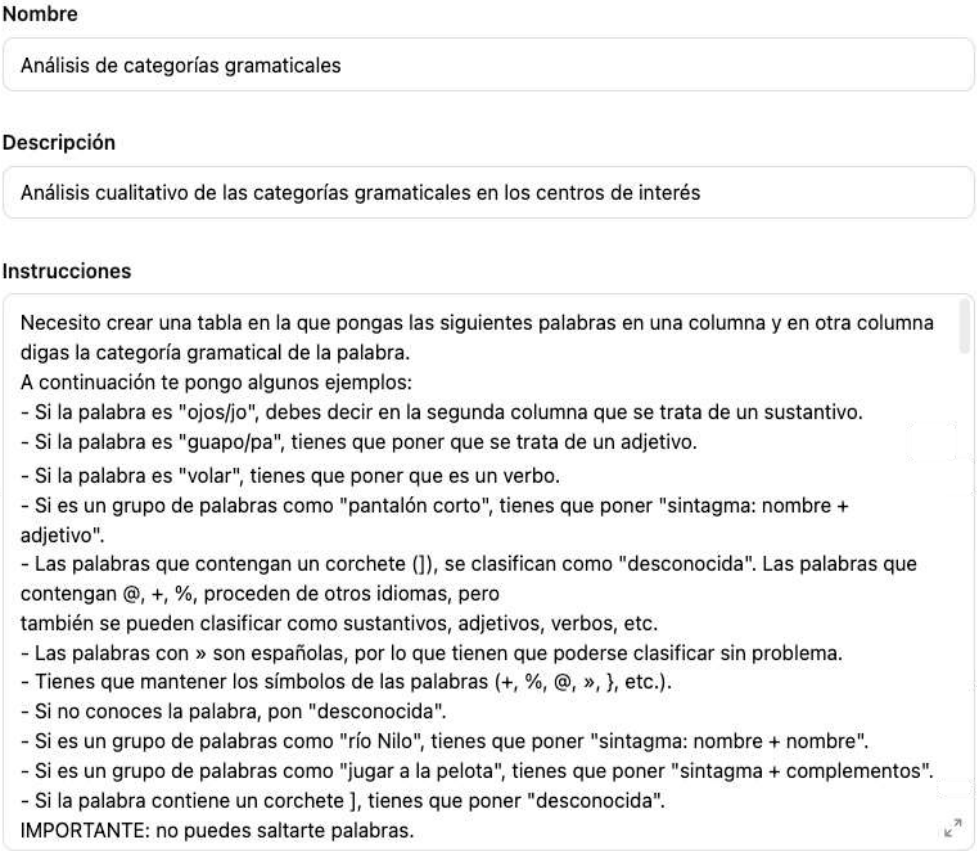
El procesamiento de los datos registrados en las encuestas recopiladas en este estudio se ha llevado a cabo con el programa LexPro, una herramienta desarrollada por la Universidad de Salamanca y la Universidad Miguel Hernández. Este *software* permite obtener automáticamente indicadores como frecuencia, número total de palabras y de vocablos, promedio por informante, índice de cohesión o grado de compatibilidad entre

grupos, así como generar grafos léxicos y detectar vocablos comunes y exclusivos de cada grupo estudiado. Además, trabaja a partir de hojas de cálculo y aplica la fórmula de López-Chavez y Strassburger (2000) para el tratamiento de los datos asociados a los CI, lo que lo convierte en una herramienta eficaz para investigaciones en disponibilidad léxica. Dado que LexPro exige un formato específico para el tratamiento de los datos, se ha utilizado Microsoft Excel para preparar los archivos siguiendo las directrices establecidas en el manual de uso del propio LexPro (Hernández Muñoz et al., 2023).

La edición de los datos se ha realizado siguiendo los criterios propios del Proyecto Panhispánico y los aplicados en estudios de disponibilidad léxica en estudiantes de ELE. Para ello, se han tomado como referencias principales las obras de Fernández Juncal (2013) y Sánchez-Saus Laserna (2016). Sin embargo, una novedad de este estudio es que no se han eliminado los extranjerismos no registrados en fuentes normativas. Esta decisión responde al hecho de que el presente trabajo se enmarca en una investigación más amplia centrada en el análisis de errores, con especial atención a la influencia de la lengua materna (LM) de los informantes. Por esta razón, se ha optado por conservar estos vocablos marcándolos de manera diferenciada: con el signo + cuando se trata de palabras pertenecientes a la LM de los informantes, cuyas formas son lejanas al español en términos de grafía y pronunciación; y con el signo % en el caso de los extranjerismos de otras lenguas extranjeras ausentes en las fuentes normativas.

Una vez obtenidos los listados del léxico disponible registrado por los informantes, se ha procedido a la clasificación manual de las unidades léxicas según su categoría gramatical en cada uno de los CI. Esta tarea constituye la base para desarrollar la propuesta metodológica principal de este trabajo, centrada en integrar la IA en el análisis del léxico disponible y evaluar su fiabilidad y utilidad. Para ello, se ha diseñado un GPT personalizado en la plataforma ChatGPT, en su versión Plus, al que se le han proporcionado las instrucciones específicas orientadas a replicar el proceso de clasificación realizado por el investigador.

Figura 1. Instrucciones del GPT creado para etiquetar palabras según su categoría gramatical



Como se puede observar en la Figura 1, las instrucciones incluían indicaciones sobre el formato en que debe aparecer la respuesta (tabla en dos columnas), el tipo de unidades léxicas que debe considerar (formas simples y complejas), los criterios para la asignación de categorías gramaticales (sustantivo, verbos, adjetivos, etc.) y ejemplos ilustrativos de cada caso. Estas directrices fueron proporcionadas al modelo mediante la técnica *few-shot*, que consiste en mostrarle varios ejemplos representativos para guiar su comportamiento y optimizar la precisión de las respuestas generadas. Según Brown *et al.* (2020), los modelos del lenguaje ofrecen un rendimiento especialmente destacado bajo este tipo de técnica, lo que refleja su habilidad para ajustarse de manera eficiente a contextos diversos y complejos.

Utilizando estas instrucciones, se ha cargado el listado de palabras para cada CI y, de este modo, el modelo ha generado automáticamente la tabla solicitada con la información requerida. A modo de ejemplo, se presenta en la Figura 2 un fragmento de la tabla generada por ChatGPT en el campo ‘Partes del cuerpo’. Dado que los primeros 90 vocablos –aquellos con mayor índice de disponibilidad– corresponden a sustantivos, se ha seleccionado el rango desde el puesto 91 hasta el 115 con el objetivo de visualizar cómo el modelo maneja la clasificación ante una mayor diversidad de categorías gramaticales.

Figura 2. Ejemplo de clasificación de palabra según su categoría gramatical

movimiento	sustantivo
seios+	sustantivo
mujer/es	sustantivo
ouvidos/do+	sustantivo
órganos	sustantivo
(pelo) largo	sintagma: nombre + adjetivo
ropa»	sustantivo
delgado	adjetivo
cílios+	sustantivo
oír	verbo
médico	sustantivo
mãos/mão+	sustantivo
abdomen	sustantivo
piedi+	sustantivo
rótula	sustantivo
pulgar	sustantivo
correr	verbo
umbigo+	sustantivo
caminar	verbo
pantorrilla/s	sustantivo
narinas	sustantivo
columna (vertebral)	sintagma: nombre + adjetivo
espada»	sustantivo

Una vez ejecutada la tarea, se compararon uno a uno los resultados generados por el modelo con la clasificación manual previamente elaborada, con el objetivo de calcular la tasa de acierto y analizar los errores de categorización. Se puede comprobar que el programa clasifica correctamente las lexías en función de su categoría gramatical. No obstante, aunque en este caso todos los vocablos han sido catalogados de forma correcta, esto no siempre ocurre. Por esta razón, resulta fundamental revisar la respuesta proporcionada por el modelo. Además de posibles errores, la estructura de la respuesta puede no ajustarse exactamente a nuestras necesidades, lo que hace esencial el proceso de iteración. Este proceso consiste en la interacción continua con ChatGPT para refinar y dirigir sus respuestas de acuerdo con nuestros objetivos.

Un ejemplo en el que el modelo no devuelve la respuesta esperada es el caso de los extranjerismos acompañados del símbolo %, es decir, aquellos términos provenientes de una lengua extranjera distinta a la LM de los informantes. En estos casos, el modelo los clasifica como palabra 'desconocida'. Sin embargo, si se le recuerda explícitamente que se trata de extranjerismos procedentes del inglés, francés, japonés, etc., debería ser capaz de catalogarlos correctamente según su categoría gramatical. Por este motivo, elaboramos el *prompt* que se observa en la Figura 3.

Figura 3. Ejemplo de iteración con el modelo para mejorar resultados

Tras aplicar esta instrucción, el modelo ahora sí clasifica adecuadamente los extranjerismos de origen distinto a la LM de los estudiantes. En este caso específico, todas las palabras identificadas son sustantivos y, además, el sistema indica que se trata de términos procedentes de otro idioma. Asimismo, después de esta interacción, el modelo no solo incorpora esa información para los extranjerismos con el símbolo %, sino que también extiende dicha apreciación a aquellos señalados con el símbolo +, que corresponden a palabras provenientes de la LM de los informantes. Este ajuste permite obtener un análisis más completo y preciso. Estas correcciones pueden implementarse de manera eficaz cuando el error sigue un patrón identificable, permitiendo su rectificación mediante la actualización del *prompt*. Sin embargo, en casos aislados, la corrección manual sigue siendo la opción más adecuada.

3.3. Procedimiento

A modo de síntesis de lo expuesto hasta este punto y con el propósito de facilitar la replicación de la investigación, se detalla a continuación el procedimiento general seguido, que constó de las siguientes fases presentadas en orden cronológico:

a) Los informantes completaron la encuesta. En primer lugar, completaron la sección correspondiente a los datos personales y, posteriormente, realizaron la prueba de asociación léxica.

b) Las respuestas obtenidas fueron digitalizadas en Microsoft Excel.

c) El archivo resultante se cargó en el programa LexPro, lo que generó un corpus inicial de listas de palabras organizadas por CI. Posteriormente, se procedió a editar los vocablos conforme a los criterios establecidos.

d) Se elaboraron los listados definitivos y se realizó la clasificación manual de las categorías gramaticales de cada lexía en cada CI.

e) Se diseñó el GPT *personalizado* en ChatGPT con instrucciones y ejemplos.

f) Las listas de palabras se introdujeron en el GPT configurado, obteniendo una primera versión de las tablas de salida correspondientes a cada CI (v1).

g) Tras esta primera clasificación automática, se revisaron las tablas para detectar errores recurrentes, se formuló un nuevo *prompt* (Figura 3) y se generó una segunda versión de las tablas (v2).

h) Finalmente, se validaron los resultados y se compararon los aciertos y errores del modelo de ambas versiones (v1 y v2) con la clasificación manual de referencia (realizada por el investigador).

El procedimiento concluyó con la organización de todos los resultados en la Tabla 2, que se presenta en el apartado 4, dedicado al análisis de los resultados.

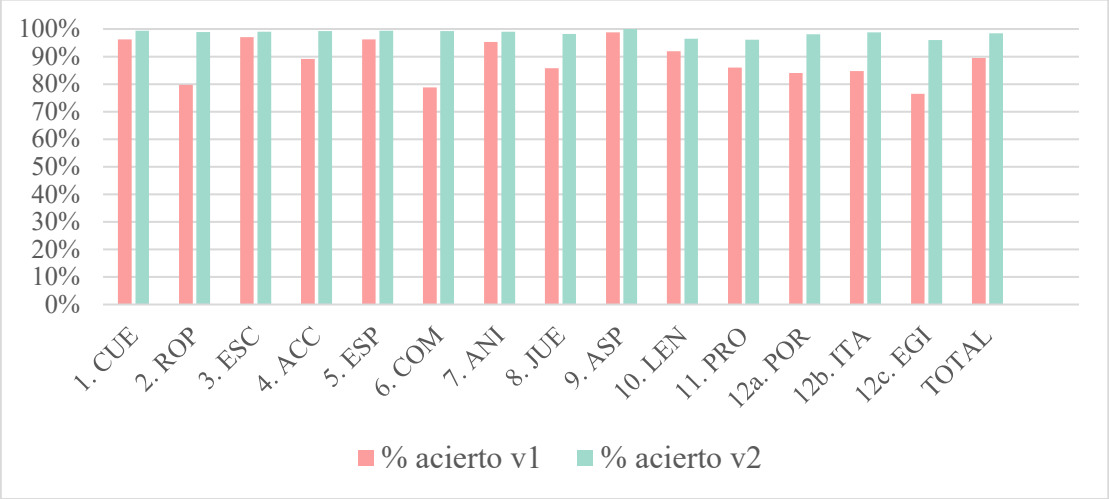
4. Resultados

El objetivo de este estudio es evaluar la fiabilidad de ChatGPT en la categorización gramatical de las palabras, analizando su tasa de acierto y determinando si su uso representa una ventaja en términos de eficiencia o, por el contrario, requiere una inversión de tiempo superior debido a la necesidad de revisión y corrección. A continuación, presentamos los resultados obtenidos en relación con el porcentaje de acierto en cada uno de los CI analizados.

Tabla 2. Porcentajes de acierto y error en la categorización gramatical por parte de ChatGPT

CI	Total palabras	Nº errores (v1)	% error	% acierto	Nº errores (v2)	% error	% acierto
1. CUE	451	17	3,77%	96,23%	3	0,67%	99,33%
2. ROP	474	96	20,25%	79,75%	5	1,05%	98,95%
3. ESC	890	26	2,92%	97,08%	9	1,01%	98,99%
4. ACC	1078	117	10,85%	89,15%	8	0,74%	99,26%
5. ESP	1476	56	3,79%	96,21%	10	0,68%	99,32%
6. COM	944	200	21,19%	78,81%	7	0,74%	99,26%
7. ANI	533	25	4,69%	95,31%	5	0,94%	99,06%
8. JUE	1010	144	14,26%	85,74%	18	1,78%	98,22%
9. ASP	1050	13	1,24%	98,76%	0	0,00%	100,00%
10. LEN	1593	129	8,10%	91,90%	57	3,58%	96,42%
11. PRO	841	118	14,03%	85,97%	33	3,92%	96,08%
12a. POR	963	154	15,99%	84,01%	18	1,87%	98,13%
12b. ITA	825	126	15,27%	84,73%	10	1,21%	98,79%
12c. EGI	455	107	23,52%	76,48%	18	3,96%	96,04%
TOTAL	12 583	1328	10,55%	89,45%	201	1,60%	98,40%
PROM.	898,8	94,86			14,36		

Figura 4. Porcentajes de acierto en la categorización gramatical en la versión 1 y 2



Tras un análisis manual de los resultados[5], tal como muestran la Tabla 2 y la Figura 4, se puede observar que, en un primer momento, el modelo alcanza una tasa de acierto del 89,45%. En otras palabras, clasifica incorrectamente 1328 unidades léxicas de las 12 583 registradas, lo que equivale a un 10,55% de error. Estos valores iniciales ya sugieren un desempeño prometedor y respaldarían el uso de la herramienta para esta tarea. No obstante, los resultados mejoran aún más cuando se ejecuta una interacción más dirigida con el modelo.

Al examinar los datos obtenidos, se han identificado ciertos errores recurrentes en la primera versión de la categorización. Los más habituales están relacionados con los extranjerismos, que el modelo tiende a clasificar como palabras desconocidas; los sintagmas, donde se observan distintos tipos de errores, como la clasificación de un 'sintagma nominal' (por ejemplo, 'diversidad lingüística') como 'sustantivo', la omisión de ciertos elementos (como en 'similar al portugués', que se clasifica como 'sintagma: nombre + preposición') o una categorización demasiado general ('salir de fiesta' etiquetado simplemente como 'sintagma + complementos'). Asimismo, se han identificado imprecisiones en la clasificación de nombres propios, como 'Salamanca', registrado únicamente como 'sustantivo', o 'Marqués de Pombal', etiquetado como 'sintagma: nombre + complementos'.

Después de una instrucción detallada, como la mostrada en la Figura 3, el modelo logra corregir estos errores y optimizar su rendimiento. Como resultado, el número de clasificaciones incorrectas se reduce de 1328 a solo 201, lo que supone una precisión del 98,40% y un margen de error del 1,60%. En consecuencia, los datos evidencian que el uso de esta herramienta resulta altamente eficaz para la tarea en cuestión.

[5] Es preciso señalar que las palabras con una posible doble clasificación, como *catalán*, son etiquetadas correctamente por ChatGPT (en este caso concreto como sustantivo/adjetivo). En nuestro estudio, su clasificación se ha realizado en función del contexto de aparición. Así, en el CI 'España y los españoles', se diferencia *catalán* como sustantivo que hace referencia a la lengua; frente a *catalanes*, adjetivo que se refiere a las personas naturales de esta comunidad autónoma.

Entre los datos más relevantes de la Tabla 2, se observa que la mayor mejora porcentual gracias a la iteración se produce en el centro 'Comidas y bebidas', donde el número de errores disminuye drásticamente de 200 en la primera versión (78,81% de acierto) a solo 7 tras la iteración, alcanzando así un 99,26% de precisión. Esto representa una mejora de en torno al 20% en la tasa de acierto.

Por otro lado, los CI con mayor precisión en la segunda versión son 'Aspecto físico y carácter', que logra una tasa del 100% de aciertos; 'Partes del cuerpo' (99,33%) y 'España y los españoles' (99,32%). En contraste, los campos con menor porcentaje de acierto son 'Lengua española' (96,42%), 'Profesiones' (96,08%) y 'Egipto y los egipcios' (96,04%).

En el caso de 'Lengua española', las imprecisiones se deben a la diversidad de categorías gramaticales presentes, lo que sugiere la necesidad de ajustes adicionales (mayor iteración) para mejorar la clasificación. En 'Profesiones', los errores se explican por la presencia de calcos léxicos, falsos cognados y palabras heterosemánticas, que parecen generar confusión en el sistema. Finalmente, en 'Egipto y los egipcios', la principal dificultad radica en la clasificación de monumentos y festividades nacionales, que no siempre son identificados como sustantivos propios, sino como simples sustantivos. No obstante, estos casos son puntuales y pueden corregirse manualmente sin mayores complicaciones.

Como ya se ha señalado, el presente estudio se ha centrado en comprobar la eficacia del método propuesto sin abordar un análisis detallado de los resultados obtenidos. No obstante, consideramos que este aspecto posee un gran interés y será objeto de futuras investigaciones, en las que también se abordará el estudio de los distintos tipos de sintagmas, como han hecho previamente autores como Carcedo González (2000) o Sánchez Gómez (2005). En el marco de este trabajo, con el fin de facilitar la organización de los datos expuestos en la Tabla 3, se ha optado por agrupar los sintagmas nominales bajo una única categoría (SN), incluyendo estructuras como 'vestido de novia' (sintagma: nombre + preposición + nombre), 'falda larga' (sintagma: nombre + adjetivo), 'falda pantalón' (sintagma: nombre + nombre), etc.; mientras que los sintagmas verbales, como 'escuchar música' (sintagma: verbo + sustantivo); 'levantarse pronto' (sintagma: verbo + adverbio); 'ponerse guapo' (sintagma: verbo + adjetivo, etc.), 'ir de compras' (sintagma: verbo + preposición + sustantivo), etc. se agrupan en la categoría SV.

Figura 5. Porcentajes de las categorías gramaticales

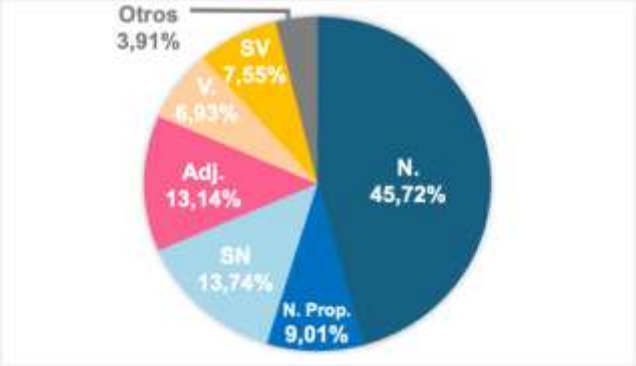


Tabla 3. Distribución y porcentaje de las unidades léxicas según su categoría gramatical en cada centro

Cat. gram.	CUE		ROP		ESC		ACC		ESP		COM		ANI		JUE		ASP		LEN		PRO		POR		ITA		EGI	
	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
N.	309	68,51	326	68,78	594	66,74	175	16,23	512	34,69	614	65,04	429	80,49	472	46,73	164	15,62	597	37,48	615	73,13	346	35,93	388	47,03	212	46,59
N.Prop.	-	-	3	0,63	7	0,79	8	0,74	420	28,46	46	4,87	8	1,50	105	10,40	1	0,10	190	11,93	1	0,12	140	14,54	111	13,45	94	20,66
SN.	37	8,20	50	10,55	103	11,57	17	1,58	261	17,68	186	19,70	25	4,69	138	13,66	102	9,71	287	18,02	121	14,39	190	19,73	146	17,70	66	14,51
Adj.	26	5,76	37	7,81	42	4,72	5	0,46	142	9,62	29	3,07	21	3,94	30	2,97	684	65,14	244	15,32	25	2,97	209	21,70	113	13,70	46	10,11
SAdj.	-	-	2	0,42	-	-	-	-	4	0,27	-	-	-	-	-	-	9	0,86	35	2,20	-	-	14	1,56	2	0,24	4	0,88
Adv.	1	0,22	-	-	-	-	-	-	3	0,20	-	-	-	-	-	-	3	0,29	22	1,38	-	-	2	0,21	-	-	1	0,22
SAdv.	-	-	-	-	-	-	-	-	1	0,07	-	-	-	-	-	-	-	-	1	0,06	-	-	1	0,10	1	0,12	1	0,22
V.	41	9,09	31	6,54	96	10,79	275	25,51	37	2,51	31	3,28	28	5,25	101	10,00	32	3,05	87	5,46	42	4,99	19	1,97	33	4,00	19	4,18
SV.	4	0,89	5	1,05	19	2,13	593	55,01	70	4,74	6	0,64	3	0,56	148	14,65	24	2,29	33	2,07	13	1,55	7	0,73	21	2,55	4	0,88
Prep.	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	4	0,25	-	-	-	-	-	-	-	-
SPrep.	-	-	1	0,21	1	0,11	-	-	2	0,14	1	0,11	-	-	-	-	7	0,67	1	0,06	1	0,12	3	0,31	-	-	1	0,22
Interj.	-	-	-	-	-	-	-	-	9	0,61	-	-	-	-	-	-	-	-	8	0,50	-	-	-	-	-	-	-	-
Pron.	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	10	0,63	-	-	1	0,10	1	0,12	-	-
Orac.	1	0,22	-	-	-	-	-	-	4	0,27	1	0,11	-	-	1	0,10	-	-	32	2,01	-	-	29	3,01	6	0,73	4	0,88
Expr.	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	23	1,44	-	-	-	-	3	0,36	-	-
Conj.	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	6	0,38	-	-	-	-	-	-	-	-
Det.	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	0,13	-	-	-	-	-	-	-	-
Signo	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2	0,13	-	-	-	-	-	-	-	-
Transv.	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	3	0,19	-	-	-	-	-	-	-	-
Prefijo	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0,06	-	-	-	-	-	-	-	-
Descon.	32	7,10	19	4,01	28	3,15	5	0,46	11	0,75	30	3,18	19	3,56	15	1,49	24	2,29	5	0,31	23	2,73	1	0,10	-	-	3	0,66
Total	451	100	474	100	890	100	1078	100	1476	100	944	100	533	100	1010	100	1050	100	1593	100	841	100	963	100	825	100	455	100

Como parte del análisis preliminar, presentamos los datos globales obtenidos (Figura 5 y Tabla 3). Las categorías gramaticales más frecuentes son los sustantivos (45,72%), seguidos de los sintagmas nominales (13,74%), adjetivos (13,14%), nombres propios (9,01%), sintagmas verbales (7,55%) y verbos (6,93%). Estas categorías, que son las más habituales en los estudios de disponibilidad léxica, representan el 96% de nuestra muestra. El 4% restante se distribuye entre diversas categorías: palabras desconocidas, es decir, términos creados *ad hoc*, como 'miejas' o 'tareja' (1,71%); oraciones, como 'mi casa es su casa' o 'no hablan bien inglés' (0,62%); sintagmas adjetivales, como 'seguro de sí mismo' o 'afín al italiano' (0,56%); adverbios y locuciones adverbiales, como 'bien' o 'por supuesto' (0,25%); expresiones, como '¡hasta pronto!' o 'ándale' (0,21%); sintagmas preposicionales, como 'al lado de Portugal' o 'sin mangas' (0,14%); interjecciones, como '¡olé!' o '¡ánimo!' (0,14%); pronombres, como 'vosotros' o 'él/ella' (0,10%); conjunciones, como 'pero' o 'pues' (0,05%); sintagmas adverbiales, como 'cerca de Portugal' o 'siempre de prisa' (0,04%); preposiciones, como 'para' o 'de' (0,03%); palabras transversales, como 'que' o 'mucho' (0,02%); determinantes, 'el, la, los, las' y 'un/a' (0,02%); signos de puntuación, '¿?' y '¡!' (0,02%), y un prefijo, 'uni-' (0,01%).

En términos de distribución, la mayoría de los CI presentan un predominio de sustantivos, con excepción de 'Acciones y actividades habituales' y 'Aspecto físico y carácter'. Precisamente, la decisión de introducir estos CI en la investigación era ofrecer un estímulo que motivara un tipo de respuesta diferente por parte de los informantes y así lo reflejan los resultados. 'Acciones y actividades habituales' destaca en sintagmas verbales (55,01%) y verbos (25,51%), mientras que en 'Aspecto físico y carácter' los adjetivos constituyen la categoría predominante (65,14%). Asimismo, los centros de corte cultural, 'España y los españoles', 'Portugal y los portugueses', 'Italia y los italianos', 'Egipto y los egipcios' y 'Lengua española', también presentan particularidades. Aunque en todos predominan los sustantivos, en ninguno superan el 50% del total. En estos centros también abundan los nombres propios y los adjetivos. Por último, el único CI que registra ejemplos de todas las categorías gramaticales es 'Lengua española', influido por su propia naturaleza como estímulo lingüístico.

5. Discusión

La hipótesis de este estudio planteaba que un modelo de IA como ChatGPT sería capaz de clasificar con alta precisión las unidades léxicas según su categoría gramatical, lo que facilitaría el análisis de grandes volúmenes de datos. Los resultados obtenidos respaldan dicha premisa, dado que la herramienta ha logrado asignar correctamente la categoría gramatical en el 89,45% de las respuestas en una primera versión y, tras una iteración en la que se ha refinado el *prompt*, ha alcanzado una tasa de acierto del 98,4%. Esta mejora evidencia la necesidad de supervisar y ajustar las instrucciones para lograr un rendimiento óptimo y una fiabilidad equiparable a la de un especialista humano. El uso de ChatGPT nos ha permitido procesar un corpus de 12 583 unidades léxicas con gran

velocidad, una tarea que, realizada de modo tradicional, implicaría una inversión considerable de tiempo, como hemos podido comprobar al llevar a cabo la clasificación manual empleada para contrastar las respuestas generadas por el modelo.

En relación con la distribución gramatical de las unidades léxicas registradas, los datos obtenidos en los 7 CI que coinciden con los utilizados en estudios tradicionales (‘Partes del cuerpo’, ‘La ropa’, ‘La escuela’, ‘Comidas y bebidas’, ‘Animales’, ‘Juegos y diversiones’ y ‘Profesiones’) muestran, de forma recurrente, una predominancia clara de sustantivos. Esta tendencia coincide con los resultados reportados en investigaciones anteriores que han cuantificado la naturaleza gramatical de las lexías recopiladas en las encuestas de disponibilidad léxica (por ejemplo, Carcedo González, 2000; o Fernández Smith *et al.*, 2008).

En cuanto a los centros más novedosos, los resultados revelan patrones similares a los observados por Sánchez-Saus Laserna (2011) y Šifrar Kalan (2014). En el centro ‘Actividades y acciones habituales’ se registra una mayor presencia de verbos y sintagmas verbales, mientras que en ‘Aspecto físico y carácter’, la categoría predominante es el adjetivo, que representan el 70% del total de unidades léxicas en el estudio de Sánchez-Saus Laserna (2011) y el 65% en el presente trabajo. Por otro lado, los centros de corte sociocultural (‘España y los españoles’, ‘Portugal y los portugueses’, ‘Italia y los italianos’ y ‘Egipto y los egipcios’), aunque mantienen una predominancia de los sustantivos, presenta también un porcentaje elevado de sustantivos propios y adjetivos. Por último, el centro ‘Lengua española’ destaca por su elevada variación categorial, atribuible a la propia naturaleza lingüística el estímulo. Estos resultados corroboran la observación de Tomé Cornejo (2015) respecto a la influencia que ejerce la renovación de la nómina de estímulos sobre la diversificación de las categorías gramaticales evocadas. La incorporación de nuevos CI parece favorecer una mayor variabilidad categorial en las respuestas obtenidas. Para concluir, es preciso señalar las limitaciones de este estudio. Por un lado, la participación de informantes extranjeros implica la influencia de la LM en sus respuestas, algo que se evidencia en la presencia de extranjerismos no normativos. Estos casos podrían introducir cierto grado de distorsión en la cuantificación categorial de las unidades léxicas. Por otro lado, el modelo GPT-4o ha mostrado algunas imprecisiones en su desempeño. Para entender el origen de estos errores, resulta relevante considerar cómo un modelo de IA interpreta las categorías gramaticales. A diferencia del lingüista humano, cuya categorización se apoya en conocimientos morfosintácticos y en una teoría gramatical, el modelo opera mediante patrones estadísticos derivados de su entrenamiento en grandes volúmenes de texto. Esta diferencia explica ciertos errores observados, especialmente en sintagmas complejos o unidades polisémicas, donde el contexto morfosintáctico resulta determinante. La clasificación automática no constituye un proceso cognitivo, sino probabilístico, lo que evidencia que la IA no “comprende” las categorías gramaticales en sentido estricto, sino que infiere patrones de uso. Esta naturaleza estadística del modelo también explica que, en ocasiones, asignara categorías incorrectas en lugar de marcar las lexías como ‘desconocidas’, tal y como se especificaba en el *prompt.* Estas limitaciones señalan la necesidad de seguir investigando para superar estos obstáculos y manifiestan la obligatoriedad de supervisar los resultados generados por los modelos de IA. Al mismo tiempo, los avances obtenidos en este trabajo abren

nuevas posibilidades para explorar aplicaciones complementarias de ChatGPT en el ámbito del análisis lingüístico. Algunas de estas líneas de investigación se presentan en el apartado final de conclusiones.

7. Conclusiones

El presente trabajo constituye una contribución significativa tanto a la lingüística aplicada como a la incipiente integración de la IA en estudios lingüísticos. Por un lado, se ofrece una aproximación al análisis gramatical del léxico disponible en español, demostrando que la diversificación de los CI puede favorecer una distribución categorial más equilibrada y rica que la observada en estudios tradicionales. Por otro lado, los resultados apuntan a la viabilidad de emplear modelos de IA como ChatGPT para el procesamiento y análisis de datos lingüísticos a gran escala, con niveles de eficacia que son prometedores. Una ventaja añadida del método desarrollado es su aplicabilidad retroactiva, ya que basta con añadir un listado de palabras en el GPT creado para que el modelo realice la tarea de clasificación. Si bien se trata de una propuesta metodológica en desarrollo, consideramos que abre una vía para futuras investigaciones que busquen integrar herramientas de IA en el análisis lingüístico con criterios rigurosos.

En este sentido, las posibilidades de ampliar esta línea de investigación son numerosas. Sería interesante realizar también el análisis sobre la efectividad y fiabilidad del uso de herramientas de IA para clasificar el léxico disponible según otros criterios lingüísticos más allá de la categoría gramatical. Por ejemplo, se podría automatizar la identificación del origen etimológico de los términos registrados (palabras heredadas del latín, anglicismos, galicismos, etc.), determinar el registro de uso de las palabras (coloquial, vulgar, culto, etc.) o el tipo de formación de las lexías (derivada, compuesta, sigla, etc.). Explorar estas dimensiones con el apoyo de la IA permitiría un análisis multidimensional del léxico disponible y enriquecería la comprensión de cómo los hablantes organizan su lexicón mental según distintos parámetros lingüísticos.

Financiamento: Esta pesquisa não recebeu financiamento de nenhuma agência de financiamento público.

Disponibilização de dados: Os dados que sustentam este estudo estão disponíveis mediante solicitação ao autor correspondente.

Referencias

Aabidi, L. (2019). La disponibilidad léxica en español como lengua extranjera: dos décadas de investigación científica. *MarcoELE: Revista de Didáctica Español Lengua Extranjera*, *28*.

Brown, T., Mann, B., Ryder Nick, Subbiah, M., Kaplan, J., Dhariwal, P., Nellakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D., Wu, J., Winter, C., … Amodei, D. (2020). Language models are few-shot learners. En H. Larochelle, M. Ranzato,

R. Hadsell, M. F. Balcan & H. Lin (Eds.), *Advances in neural information processing systems* (Vol. 33). https://doi.org/10.48550/arXiv.2005.14165

Carcedo González, A. (2000). *Disponibilidad léxica en español como lengua extranjera: el caso finlandés (estudio del nivel preuniversitario y cotejo con tres fases de adquisición)*. Universidad de Turku.

Dao, X. Q., & Le, N. B. (2023). *Investigating the effectiveness of ChatGPT in mathematical reasoning and problem solving: evidence from the Vietnamese National High School Graduation Examination*. Cornell University. https://doi.org/10.48550/arXiv.2306.06331

Fernández Gómez, M. I. V. (2021). *Estudio de disponibilidad léxica en estudiantes universitarios portugueses de ELE* [Dissertação de mestrado, Universidade do Minho]. https://repositorium.uminho.pt/entities/publication/cce03ede-5029-4cda-b5d7-f46ecc1449d9

Fernández Juncal, C. (2013). *Léxico disponible de Cantabria. Estudio sociolingüístico*. Ediciones Universidad de Salamanca.

Fernández Smith, G., Rico Martín, A. M., Molina García, M. J., & Jiménez Jiménez, M. Á. (2008). *Léxico disponible de Melilla: estudio sociolingüístico y repertorios léxicos*. Arco Libros.

Gallego Gallego, D. J. (2014). *Léxico disponible de estudiantes de español como lengua extranjera en la comunidad de Madrid* [Tesis doctoral, Universidad de Alcalá]. https://hispanismo.cervantes.es/publicaciones/lexico-disponible-estudiantes-espanol-lengua-extranjera-comunidad-madrid

Galloso Camacho, M. V. (2003). *El léxico disponible de Ávila, Salamanca y Zamora.* Fundación Instituto Castellano y Leonés de la Lengua.

Gamazo Carretero, E. (2014). Estereotipos en el léxico disponible de universitarios portugueses. En A. M. Sainz García (Ed.), *El español como lengua extranjera en Portugal: retos de la enseñanza de lenguas cercanas* (pp. 30–41). Ministerio de Educación, Cultura y Deporte. https://www.educacionyfp.gob.es/dam/jcr:1db67ea5-d7c1-4dcc-8b53-30149a473d74/retos-ensenanza.pdf

Hernández Muñoz, N. (2005). *Hacia una teoría cognitiva integrada de la disponibilidad léxica: el léxico disponible de los estudiantes castellano-manchegos* [Tesis doctoral, Universidad de Salamanca]. https://dialnet.unirioja.es/servlet/tesis?codigo=142575

Hernández Muñoz, N., Tomé Cornejo, C., López García, M., & Bartol Hernández, J. A. (2023). *Manual de LexPro*. https://dispogram.usal.es/

Herreros Marcilla, M. (2020). *Variación sociolingüística en el ámbito neorrural: una perspectiva integradora* [Tesis doctoral, Universidad de Salamanca]. https://gredos.usal.es/handle/10366/144167

López-Chavez, J., & Strassburger, C. (2000). El diseño de una fórmula matemática para obtener un índice de disponibilidad léxica confiable. *Anuario de Letras*, *38*, 227–251.

Mateo García, M. V. (1998). *Disponibilidad léxica en el COU almeriense. Estudio de estratificación social.* Servicio de Publicaciones de la Universidad de Almería.

Mendoza Puertas, J. D. (2018). El léxico disponible de 82 estudiantes coreanos de español como lengua extranjera. *MarcoELE: Revista de Didáctica Español Lengua Extranjera*, *26*. https://marcoele.com/descargas/26/mendoza-lexico_disponible-coreanos.pdf

Nalesso, G. (2019). *Disponibilidad y riqueza léxica en un grupo de estudiantes universitarios italianos de ELE* [Tesis doctoral, Università degli Studi di Padova]. https://dialnet.unirioja.es/servlet/tesis?codigo=319076

Palapanidi, K. (2024). Estereotipos culturales sobre España en el léxico disponible de aprendices griegos de ELE. *El Español Por El Mundo*, *6*(1), 99–108.

Palapanidi, K., & Mavrou, I. (2024). La naturaleza gramatical del léxico disponible en diferentes niveles lingüísticos en español como lengua extranjera. *Revista Signos. Estudios de Lingüística*, *57*(114), 151–167.

Rodríguez Menduiña, P. (2010). *Estereotipos culturales sobre España en estudiantes estadounidenses desde la disponibilidad léxica*. Universidad de Salamanca.

Rodríguez-Echeverría, R., Gutiérrez, J. D., Conejero, J. M., & Prieto, Á. E. (2024). Analysis of ChatGPT performance in computer engineering exams. *IEEE Revista Iberoamericana de Tecnologias Del Aprendizaje*, *19*, 71–80. https://doi.org/10.1109/RITA.2024.3381842

Ronan, P., & Schneider, G. (2023). *Can ChatGPT solve a linguistics exam?* Cornell University. https://doi.org/10.48550/arXiv.2311.02499

Rubio Lastra, M. (2018). Disponibilidad léxica de 52 estudiantes taiwaneses universitarios de ELE A1. *MarcoELE. Revista de Didáctica Español Lengua Extranjera*, *27*. https://marcoele.com/descargas/27/rubio-disponibilidad_lexica_taiwanenes_a1.pdf

Rubio Sánchez, R. (2016). Estudio de disponibilidad léxica en 60 aprendices portugueses de español. En N. Domínguez García, C. Fernández Juncal, & J. L. García Alonso (Eds.), *Innovación y desarrollo en español como lengua extranjera* (pp. 107–128). Ediciones Universidad de Salamanca.

Rubio Sánchez, R. (2017). Acercamiento al léxico disponible de 173 estudiantes italianos preuniversitarios de Español como Lengua Extranjera. In F. Del Barrio de la Rosa (Ed.), *Palabras vocabulario léxico. La lexicología aplicada a la didáctica y a la diacronía* (pp. 143–161). https://doi.org/10.14277/6969-169-0/VP-1-9

Samper Hernández, M. (2001). Disponibilidad léxica en español como lengua extranjera [Memoria de licenciatura, Universidad de Salamanca].

Samper Padilla, J. A., Bellón Fernández, J. J., & Samper Hernández, M. (2003). El proyecto de estudio de la disponibilidad léxica en español. In R. Ávila (Ed.), *Pautas y pistas en el análisis del léxico hispano(americano)* (pp. 27–134). Iberoamericana - Vervuert. http://dialnet.unirioja.es/servlet/articulo?codigo=1703937

Sánchez Gómez, C. (2005). Naturaleza gramatical del léxico disponible en informantes del español como lengua extranjera. *Interlingüística (Alacant, Spain)*, *16*, 977–986.

Sánchez-Saus Laserna, M. (2011). *Bases semánticas para el estudio de los centros de interés del léxico disponible. Disponibilidad léxica de informantes extranjeros en las universidades andaluzas* [Tesis doctoral, Universidad de Cádiz]. https://rodin.uca.es/handle/10498/15862

Sánchez-Saus Laserna, M. (2016). *Léxico disponible de los estudiantes de español como lengua extranjera en las universidades andaluzas*. Editorial Universidad de Sevilla.

Santos Díaz, I. C. (2015). *Evaluación de la competencia léxica bilingüe en estudiantes del Máster Universitario en Profesorado. Análisis de pruebas de disponibilidad léxica y de identificación de tecnicismos en español, inglés y francés* [Tesis doctoral, Universidad de Málaga]. https://riuma.uma.es/entities/publication/148c2fbc-beb3-4981-86b5-d63cd3517523

Serfati, M. (2022). Comparación intralingüística de la disponibilidad léxica de español como lengua extranjera en Marruecos (niveles de enseñanza media y universitaria). *Forma y Función*, *35*(1). https://doi.org/10.15446/fyf.v35n1.88892

Šifrar Kalan, M. (2009). Disponibilidad léxica en español como lengua extranjera: el cotejo de las investigaciones en Eslovenia, Salamanca y Finlandia. *Verba Hispanica*, *17*(1). https://doi.org/10.4312/vh.17.1.165-182

Šifrar Kalan, M. (2014). Disponibilidad léxica en diferentes niveles de español/lengua extranjera. *Studia Romanica Posnaniensia*, *41*(1), 63–85.

Tomé Cornejo, C. (2010). *Aspectos metodológicos de los estudios de disponibilidad léxica. Los centros de interés*. Universidad de Salamanca.

Tomé Cornejo, C. (2015). *Léxico disponible. Procesamiento y aplicación a la enseñanza de ELE* [Tesis doctoral, Universidad de Salamanca]. https://dialnet.unirioja.es/servlet/tesis?codigo=102596

Trigo Ibáñez, E., Dono López, P., & Fernández Gómez, I. V. (2025). Aproximación al léxico disponible en ELE de 42 estudiantes de la Universidade do Minho. En I. C. Santos Díaz & A. M. Ávila Muñoz (Eds.), *Avances y desarrollo de los estudios sobre el léxico disponible: una aproximación desde la variación léxica*. Peter Lang.

Wang, K. D., Burkholder, E., Wieman, C., Salehi, S., & Haber, N. (2024). Examining the potential and pitfalls of ChatGPT in science and engineering problem-solving. *Frontiers in Education*, *8*. https://doi.org/10.48550/arXiv.2310.08773

[recebido em 11 de julho de 2025 e aceite para publicação 10 de novembro de 2025]