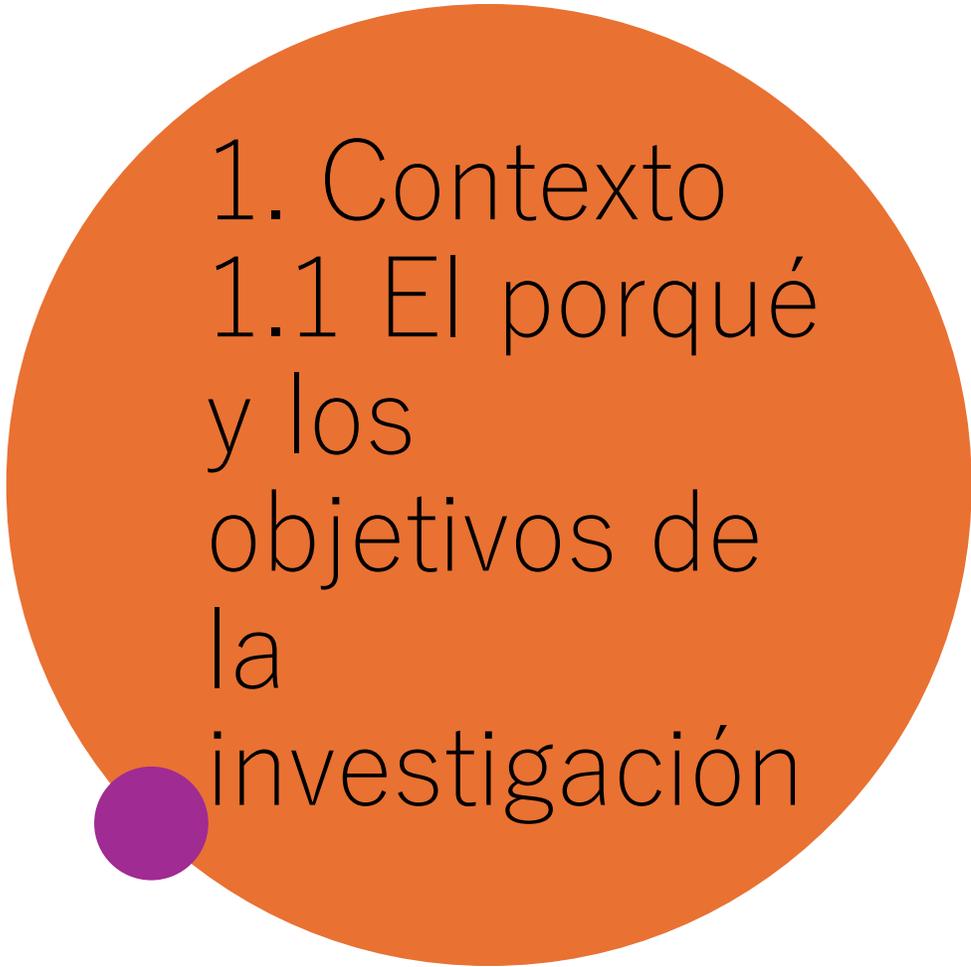


TADESKA
9 de noviembre
Motoko Hirai y Ainhoa
Bujanda

[Taller 2]
**El potencial del uso de ChatGPT
en la corrección de redacciones en
español: un estudio de caso con
estudiantes japoneses de nivel B1**

Agenda

1. Contexto
 - 1.1 El porqué y los objetivos de la investigación
 - 1.2 Los datos
2. Metodología
 - 2.1 Los prompts
3. Los resultados
 - 3.1 Cómo evalúa ChatGPT una de las tareas de la prueba 3 del examen DELE B1
 - 3.2 Cómo corrige ChatGPT los errores gramaticales
 - 3.3 Los errores gramaticales más comunes que suelen cometer los aprendientes en este contexto.
4. Limitaciones de la investigación y perspectivas de futuro
5. Discusión



1. Contexto

1.1 El porqué y los objetivos de la investigación

El porqué de la investigación:

Buscamos una respuesta a la pregunta: ¿Puede la Inteligencia Artificial ser una herramienta útil tanto para la enseñanza de ELE por parte del docente como para el aprendizaje de español por parte del aprendiente?

Si vas a utilizarla, ¿para qué la quieres?

En nuestro caso, se centra en la corrección de los textos escritos por los aprendientes, ya que se trata de una tarea compleja para los docentes por dos razones fundamentales: por el trabajo que supone en sí mismo y por la exigencia de hacerlo de manera rigurosa.

1. Contexto

1.1 El porqué y los objetivos de la investigación



El objetivo a largo plazo de la investigación es explorar en qué medida la IA puede contribuir a la enseñanza de la competencia discursiva, en concreto, en la corrección de las producciones escritas de los aprendientes de ELE en las universidades japonesas.



El objetivo a corto plazo sería aprender a manejarla y seguir desarrollando la competencia digital.

1. Contexto

1.2 Los datos

Pedimos a los profesores colaboradores que realizaran una tarea de la Prueba 3 del modelo oficial del examen DELE B1, disponible en la página web del Instituto Cervantes, con el fin de analizar la viabilidad del uso de ChatGPT en la evaluación y corrección de redacciones de estudiantes japoneses de ELE de nivel B1.

Profesoras colaboradoras del nivel B1 (junto con estudiantes cuya especialidad es la lengua española o la filología hispánica):

- Profesora Paula Letelier
- Profesora Alba Medialdea
- Profesora Montserrat Sanz

OPCIÓN 1

Lea el siguiente mensaje publicado en un blog dedicado a los viajes:

En nuestro blog vamos a abrir una sección especial sobre viajes de invierno. Cuéntenos un viaje que usted haya hecho en esa estación del año y sus impresiones sobre él.

Escriba un comentario para enviar al blog en el que cuente:

- el lugar al que fue;
- por qué eligió ese destino y con quién fue;
- cuándo fue y cuánto duró el viaje;
- qué fue lo que más le gustó y por qué;
- alguna anécdota que recuerde del viaje.

Nuestro más sincero agradecimiento

A las profesoras
Paula Letelier
Alba Medialdea
Montserrat Sanz

A los profesores de la Universidad Ritsumeikan, gracias a quienes contamos con los datos de los aprendientes de español como segunda lengua extranjera, que aún quedan por analizar.

2. Metodología

- 1) Analizar la viabilidad del uso de ChatGPT en la corrección y evaluación de redacciones de estudiantes japoneses de ELE de nivel B1. Para ello, elaborar ejemplos de prompts eficaces que permitan optimizar dicha corrección y evaluación, hasta lograr que ChatGPT realice su tarea de forma adecuada.
- 2) Revisar y analizar los resultados proporcionados por ChatGPT-4 de pago
- 3) Reflexionar sobre posibles estrategias para mejorar la enseñanza de la redacción, teniendo en cuenta el desarrollo de la competencia discursiva en las clases de ELE en universidades japonesas.

2. Metodología

2.1. Los prompts

- Los prompts.
- En julio, cuando trabajábamos en esta investigación, no llegamos a conseguir que ChatGPT generara el informe en Word ni en PDF.
Sin embargo, la evaluación y la corrección se realizaron normalmente sin problema, excepto en estos dos casos:
 - 1) A partir de algunas redacciones, ChatGPT creó por su cuenta el texto, lo evaluó y lo corrigió. → Para evitarlo, pedimos la corrección de cada redacción por separado.
 - 2) Una vez ha cambiado el modo de evaluación: de “1 a 4” en lugar de “0 a 3”.

3.1. Evaluación del modelo ChatGPT-4

- Para este estudio, se empleó la rúbrica DELE B1 de expresión e interacción escritas (Criterios → Adecuación, Coherencia, Corrección, Alcance Léxico y Calificación Holística), con puntuaciones de **0 a 3**.
- A efectos de comparación entre ChatGPT-4 y la corrección humana, se estableció la siguiente **binarización de resultados**:
 - **A** (Aprobado) → puntuaciones **2 o 3**
 - **S** (Suspendido) → puntuaciones **0 o 1**

Categoría	3	2	1	0
Adecuación	Escribe textos claros y precisos, ampliándolos con detalles tanto en temas concretos como abstractos. Redacta cartas, mensajes y notas, empleando el registro adecuado al contexto. Desarrolla con eficacia todos los puntos de orientación dados.	Escribe textos claros y sencillos. En el caso de las cartas, mensajes, notas..., respeta las convenciones de género básicas (inicio y cierre del texto...) y utiliza las fórmulas básicas de cortesía (saludo, despedida...). Desarrolla con claridad la gran mayoría de los puntos de orientación dados, aunque alguno falte o no esté tratado adecuadamente.	Escribe textos muy breves y básicos sobre su entorno más inmediato o aspectos de la vida cotidiana. En algunos casos, la información aparece desordenada o incompleta, lo que obliga a una relectura para su comprensión. Redacta cartas, mensajes y notas sencillas y breves vinculadas a necesidades básicas o para transmitir información personal. Utiliza aún con vacilaciones los exponentes funcionales más frecuentes, normas de cortesía elementales o fórmulas de saludo y tratamiento («muchas gracias»; «hola, ¿cómo estás?») o falta algún detalle importante (por ejemplo, saludo o despedida en las cartas). Menciona solo alguno de los puntos de orientación o no los desarrolla suficientemente.	El texto producido se limita a una serie de frases sencillas aisladas sobre sí mismo y sobre otras personas o temas de su entorno más próximo. En algunos casos el texto resulta incomprensible. Comete errores de formulación en fórmulas sencillas y cotidianas relativas a saludos, despedidas, presentaciones y expresiones del tipo: «por favor», «gracias», «lo siento...», que adapta al discurso. Hay errores de registro y faltan detalles importantes. El texto no sigue los puntos de orientación dados o se aleja de ellos, ni se ajusta a la extensión indicada: Tarea 1: menos de 70 palabras. Tarea 2: menos de 90 palabras.
Coherencia	Escribe textos claros, coherentes y estructurados con un uso limitado pero adecuado de los mecanismos de cohesión para enlazar frases en un discurso planificado, teniendo en cuenta el efecto que puede producir en el receptor. Sintetiza, evalúa y varía la información de otras fuentes ensayando nuevas combinaciones y expresiones, marcando la relación existente entre las ideas. Utiliza de forma adecuada los signos de puntuación aunque puede cometer algún error.	Escribe textos breves y cohesionados, ordenados mediante una secuencia lineal de elementos sencillos, utilizando organizadores de la información («primero», «luego», «después») y conectores básicos frecuentes («y», «también», «por eso», «entonces», «pero», «porque...»), aunque el texto presente alguna deficiencia o limitación en su estructura.	Escribe una serie de oraciones breves enlazadas con conectores muy sencillos y básicos («y», «pero», «porque»). Su discurso, en algunos casos memorizado, desordenado o incompleto, obliga a una relectura para su comprensión.	El texto se limita a una serie de palabras o grupos de palabras enlazadas con conectores muy básicos y lineales («y», «pero»). El discurso no mantiene una estructura organizada que permita seguir los razonamientos del candidato.
Corrección	Mantiene un buen control gramatical que utiliza en situaciones cotidianas, aunque todavía puede cometer algún error no sistemático o pequeños fallos en la estructura de la frase que no producen malentendidos. La ortografía es razonablemente correcta aunque puede cometer algún error por influencia de la lengua materna en el léxico menos frecuente.	Muestra un control razonable de elementos lingüísticos básicos y estructuras habituales que utiliza para satisfacer necesidades inmediatas y en temas predecibles o de interés personal. Puede cometer algún error en la ortografía de las palabras, pero que no interfiere en la transmisión de la idea principal del texto.	Utiliza estructuras gramaticales sencillas. Comete errores básicos, aunque se entienda lo que quiere transmitir siempre que el mensaje esté relacionado con una situación comunicativa frecuente. Comete errores ortográficos sistemáticos que en algunos casos dificultan la comprensión del mensaje.	Muestra un control limitado incluso de algunas estructuras gramaticales muy básicas y sencillas, o se sirve de frases cortas (probablemente memorizadas) relacionadas con necesidades básicas e inmediatas. Comete abundantes errores gramaticales y ortográficos (concordancias, error en la elección de la persona del verbo), que dificultan la comprensión del mensaje y obligan a la relectura continua.
Alcance	Domina un vocabulario amplio, que incluye algunas expresiones idiomáticas y coloquiales, que le permite describir situaciones impredecibles, explicar los puntos principales de una idea o un problema con razonable precisión y expresar pensamientos sobre temas abstractos y culturales. Puede cometer alguna imprecisión léxica de escasa importancia.	Tiene el vocabulario suficiente para desenvolverse en intercambios cotidianos próximos a su entorno más inmediato. Esto le permite solicitar información, hacer valoraciones, expresar deseos y dar instrucciones. Puede cometer errores si utiliza estructuras o vocabulario más complejo.	Tiene un vocabulario limitado que utiliza para transmitir información básica en situaciones que respondan a necesidades muy concretas y cotidianas que resulta insuficiente para transmitir el mensaje. Presenta imprecisiones que no afectan a la comunicación.	Utiliza un repertorio de palabras muy básicas y frases aisladas que no son suficientes para transmitir la información requerida ni para que se produzca la comunicación. Comete constantes imprecisiones léxicas e interferencias de otras lenguas.
Escala holística	Aporta a la información requerida detalles que hacen que la organización y formulación del mensaje cumplan sobradamente con los objetivos comunicativos planteados. Utiliza un repertorio lingüístico suficiente para hacer descripciones claras y precisas, expresar opiniones, puntos de vista y desarrollar argumentos sin aparentes limitaciones. El resultado es un texto claro y detallado.	Aporta la información requerida de forma comprensible y logra transmitir los mensajes. Se expresa de manera clara en intercambios de información sobre temas cotidianos aunque, si los temas son abstractos, puede producirse alguna vacilación o faltar algún detalle. Utiliza un repertorio lingüístico suficiente para expresarse de manera adecuada sobre situaciones cotidianas y temas de su interés. A pesar de algunos errores, vacilaciones o repeticiones construye oraciones sencillas y lineales, con palabras clave en un discurso comprensible y claro.	Aporta parte de la información requerida de manera sencilla. Su brevedad y poca claridad resultan insuficientes para transmitir el mensaje. Utiliza un repertorio lingüístico limitado compuesto por estructuras sintácticas y expresiones memorizadas en un discurso con errores elementales que dificultan la comprensión y que, en algunos casos no dejan clara la idea general.	Aporta solo algunos datos que resultan insuficientes para transmitir los mensajes. El escrito se limita a una serie de oraciones muy breves y simples en un discurso desorganizado y con errores abundantes que dificultan la comprensión del mensaje.

Rúbrica de creación propia, extraída de: *Instituto Cervantes (2014). Guía del examen DELE B1. Madrid: Instituto Cervantes.*

Evaluación del modelo ChatGPT-4

Interpretación:

- **TP (Verdadero Positivo)** → ambos (ChatGPT-4 y humano) califican la redacción como *aprobada* (coincidencia positiva).
- **FN (Falso Negativo)** → ChatGPT-4 considera *suspendida* una redacción que el humano calificó como *aprobada*.
- **FP (Falso Positivo)** → ChatGPT-4 aprueba una redacción que el humano consideró *suspendida*.
- **TN (Verdadero Negativo)** → ambos coinciden en que la redacción está *suspendida*.

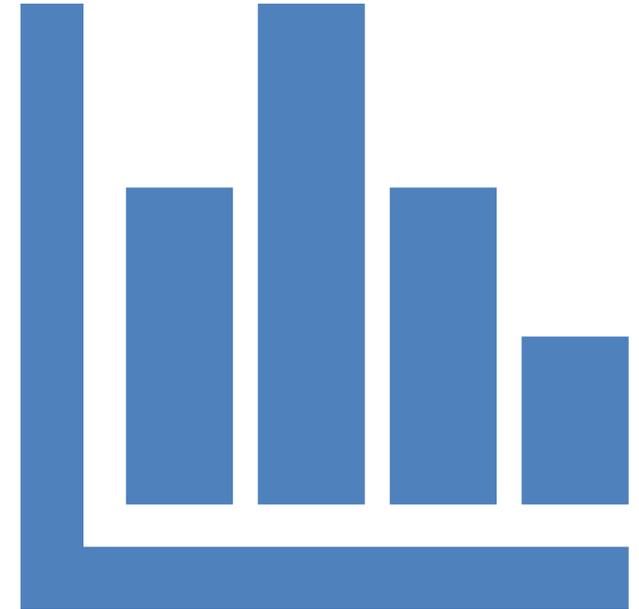
3.1. Evaluación del modelo ChatGPT-4

De esta manera, las coincidencias y discrepancias entre ChatGPT-4 y el corrector humano pueden representarse mediante la siguiente matriz de confusión:

	Humano A (Aprobado)	Humano S (Suspendido)
ChatGPT-4 A (Aprobado)	TP (Verdadero Positivo) ChatGPT-4 = A y Humano = A	FP (Falso Positivo) ChatGPT-4 = A y Humano = S
ChatGPT-4 S (Suspendido)	FN (Falso Negativo) ChatGPT-4 = S y Humano = A	TN (Verdadero Negativo) ChatGPT-4 = S y Humano = S

Matriz elaborada en base a la explicación del libro Ctrl + Blablablá: El ABC de la lingüística computacional (Martín-Pérez González, 2023).

Interpretación de Métricas de Evaluación



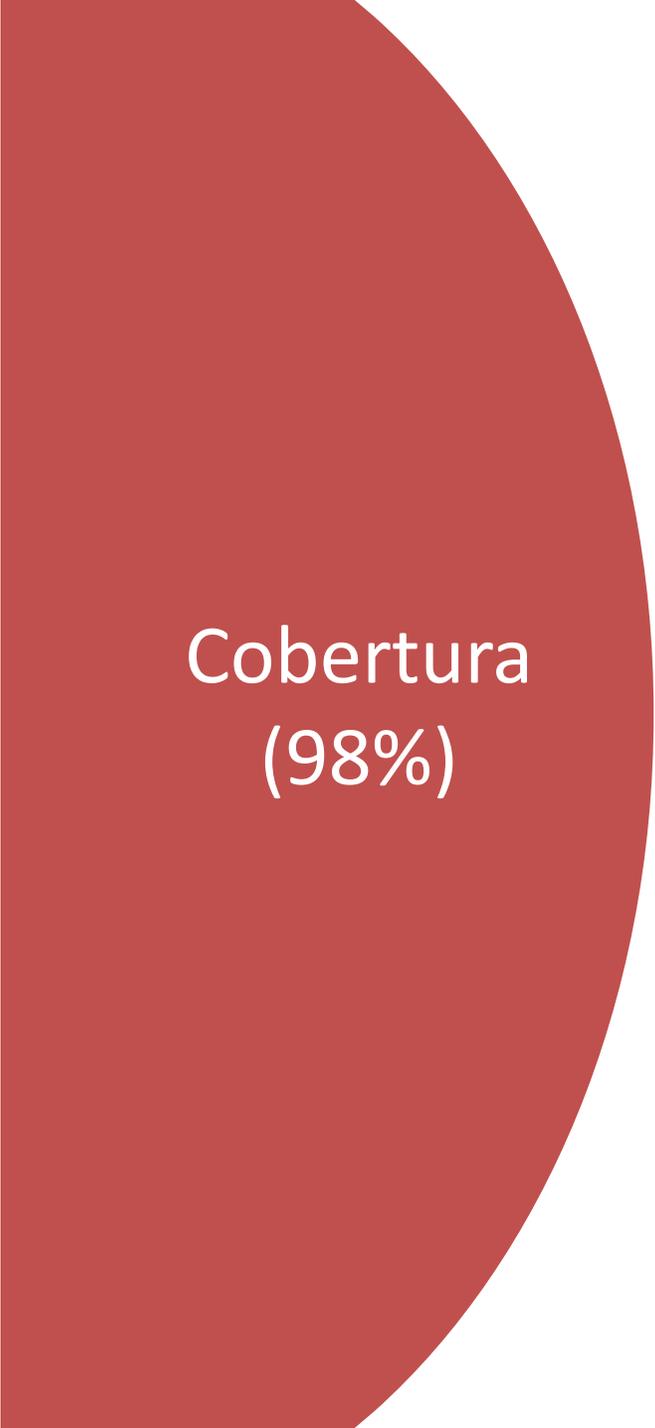
Precisión, Cobertura y F1 Score

Martín-Pérez González, L. (2023). *Ctrl + Blablablá: El ABC de la lingüística computacional*.

Pie de Página

Precisión (55%)

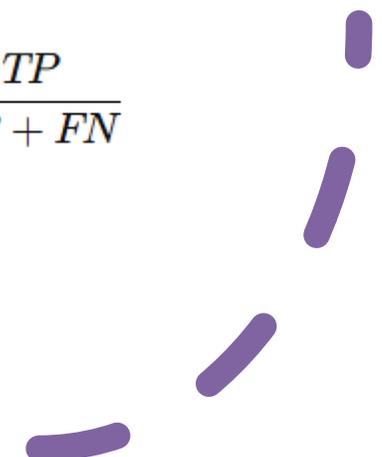
- Mide la proporción de redacciones **correctamente clasificadas como aprobadas** respecto al total de redacciones que ChatGPT clasificó como aprobadas
- Fórmula:
$$\text{Precisión} = \frac{TP}{TP + FP}$$
- En este estudio: 46 aciertos de 84 correcciones \rightarrow ~55%.
- Interpretación: de cada 10 redacciones que el modelo consideró aprobadas, solo 5 o 6 coincidieron con la evaluación humana.
- El resto fueron **falsos positivos** (redacciones que ChatGPT aprobó pero el corrector humano suspendió).

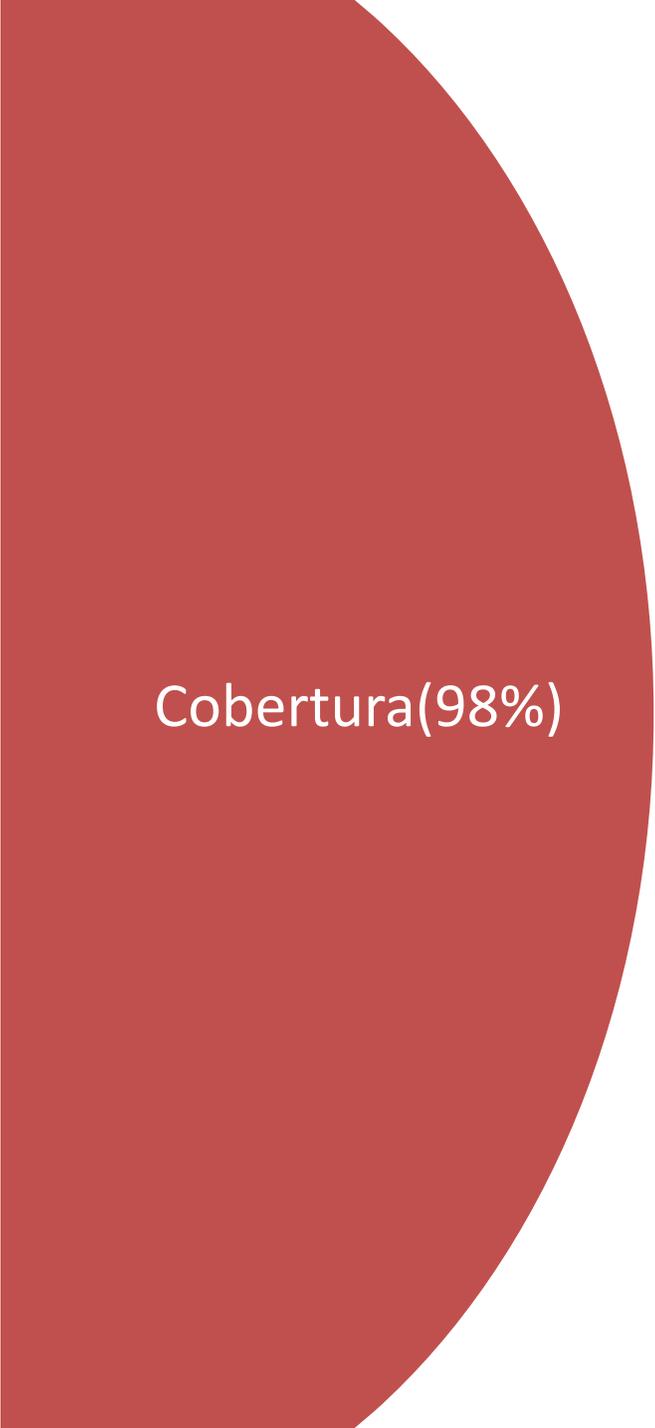


Cobertura
(98%)

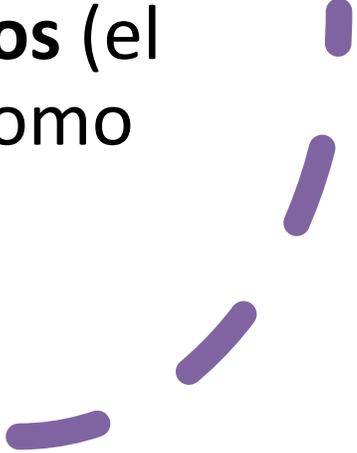
- Mide la proporción de **redacciones correctamente identificadas como aprobadas** respecto al total de redacciones **que realmente estaban aprobadas** según la evaluación humana.

- Fórmula:
$$\text{Cobertura} = \frac{TP}{TP + FN}$$





Cobertura(98%)

- En este estudio: ChatGPT identificó correctamente 9,8 de cada 10 redacciones aprobadas ($\approx 98\%$).
 - Interpretación: el modelo logró reconocer casi todos los casos aprobados. Solo el 2% de los textos que correctores humanos consideraron aptos fueron **falsos negativos** (el modelo los clasificó como suspendidos)
- 

F1 Score
(69%)

- Balance entre Precisión y Cobertura.
- Fórmula: $F1 = 2 \times \frac{(\text{Precisión} \times \text{Cobertura})}{(\text{Precisión} + \text{Cobertura})}$
- En el estudio: Precisión media (55%) y Cobertura casi perfecta (98%).
- El F1 queda en 69%.
- ChatGPT **detecta la mayoría de los casos aprobados**, pero su **criterio es permisivo**, lo que reduce la precisión general.
- En otras palabras, reconoce casi todos los aciertos, aunque **tiende a aprobar más de lo debido**.

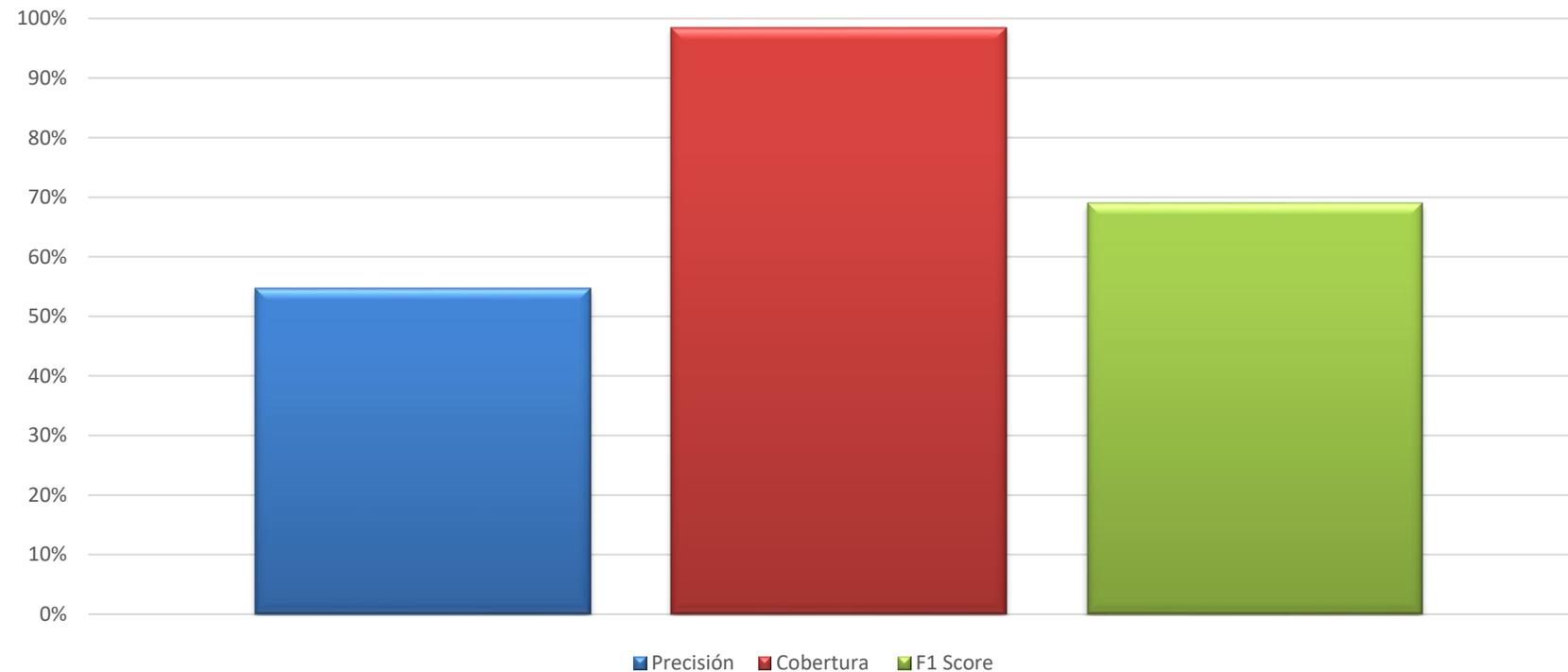
Resumen del modelo analizado

- Precisión: Corrige bien, pero a menudo **aprueba textos que no debería**. Tiende a aceptar como correctas redacciones con errores leves o de coherencia.
- Cobertura: Excelente, reconoce prácticamente **todas las redacciones aprobadas por los correctores humanos**.
- F1 Score: Balance bueno (69%), útil pero **tiene tendencia a sobre-aprobar**, mostrando más casos positivos de los que existen en la realidad.

Interpretación de los datos

El modelo es muy sensible (encuentra todo), lo que lo hace útil para detectar casos aprobados, pero **demasiado permisivo al corregir**. De esta forma, aparecen más correcciones “aceptadas” de las que realmente deberían aprobarse.

Evaluación de ChatGPT-4 en la corrección de redacciones DELE B1



3. Los resultados

3.2. Cómo corrige ChatGPT los errores gramaticales

Conclusiones:

ChatGPT-4 de pago detecta en principio todos los elementos gramaticales básicos y, además, ajusta el grado de detalle y el foco de la explicación según la “influencia en la superación del nivel B1” o la “coherencia discursiva”.

Sin embargo, en ciertos aspectos como la distinción entre pretérito indefinido e imperfecto, o el uso de ser y estar, algunos errores pueden considerarse “comprensibles” según el contexto y por tanto no se señalan como fallos estrictos.

Y ocasionalmente, también puede equivocarse.

3. Los resultados

3.2. Cómo corrige ChatGPT los errores gramaticales

3.2.1 Los errores básicos se detectan e indican de forma sistemática.

- **Errores ortográficos y de acentuación** (Ejemplo: *veldura* → *verdura*).
✘ En algunos casos no corrige el uso de mayúscula en *Febrero* cuando debería ser *febrero*.
- **Errores de género y número en sustantivos y de concordancia con artículo y adjetivo** (Ejemplo: *el torre* → *la torre*).
- **Errores de conjugación verbal** (Ejemplo: *ví* → *vi*).

3. Los resultados

3.2. Cómo corrige ChatGPT los errores gramaticales

3.2.2 A veces no se detectan, aunque en su mayoría sí se detectan

- **Uso de las preposiciones**

Ejemplos:

Visité Saga *a ver mis abuelos* → no se menciona la necesidad de la preposición *a*, que se coloca antes del objeto directo de persona (*ver a mis abuelos*).

por avión → no se señala como error. El uso correcto es *en avión*.

visitar a lugar → el verbo *visitar* es transitivo y no requiere la preposición *a* delante de un lugar, ya que se trata de un objeto directo.

- **El verbo gustar**

Ejemplo:

Me gusta naturaleza. → Normalmente se necesita el artículo determinado antes del sustantivo (*Me gusta la naturaleza*), pero aunque no se coloque, ChatGPT no lo marca como error.

3. Los resultados

3.2. Cómo corrige ChatGPT los errores gramaticales

3.2.3. Muchas veces no se detectan como errores aunque en el modelo de la redacción están corregidos.

- **Ser/estar**

ChatGPT rara vez recomienda usar el verbo estar para referirse al sabor de una comida recién probada cuando el estudiante emplea ser en este contexto. Por lo general lo deja tal cual: Era muy delicioso en lugar de Estaba muy delicioso / Estaba muy rico.

3. Los resultados

3.2. Cómo corrige ChatGPT los errores gramaticales

3.2.3. Muchas veces no se detectan como errores, aunque en el modelo de la redacción estén corregidos.

- **Pretérito indefinido / pretérito imperfecto**

No se detecta como error este tipo de uso en pasado: Fue maravilloso, aunque hizo mucho frío, pero en el modelo lo corrigen como: Aunque hacía mucho frío, fue una experiencia maravillosa.

✂ Cuando los estudiantes hablan del viaje realizado (como en esta tarea), normalmente narran en pasado. Precisamente en este contexto, si los aprendientes emplean el tiempo presente, ChatGPT lo detecta y lo señala como error.

Ejemplo: El viaje fue muy divertido porque ~~podemos~~ descansamos mucho.

3. Los resultados

3.2. Cómo corrige ChatGPT los errores gramaticales

3.2.4. Rara vez, pero ChatGPT también comete errores

Un estudiante escribió esta frase en su redacción:

La razón por que elegí Hong Kong es que quería ir Disney Land de otro país.

ChatGPT detectó un error en por que y lo corrigió como porque, explicando que hay que tener cuidado con la confusión entre la conjunción causal porque y por que, que no significan lo mismo.

Sin embargo, en realidad la frase correcta es:

La razón por la que elegí Hong Kong es que quería ir a Disney Land de otro país.

3. Los resultados

3.3. Los errores gramaticales más comunes que suelen cometer los aprendientes en este contexto (redacción sobre el viaje realizado)

3.3.1 verbo transtivo/intranstivo

Se equivocan frecuentemente el uso del verbo visitar como un verbo intranstivo

3.3.2 ser/estar (al expresar el sabor de la comida recién probada)

3.3.3 pretérito indefinido e imperfect

Cuando evalúan del viaje realizado, en español suelen emplear el indefinido, pero los estudiantes, el imperfecto. Lo que apoya Yamamura (2019).

3.3.4

Algo similar con 3.3.3, pero los estudiantes suelen emplear los dos primeros, que son incorrectos y los nativos el último

Lo que más me gustó ~~era~~

Lo que más me gustó ~~es~~

Lo que más me gustó fue

Errores pragmáticos

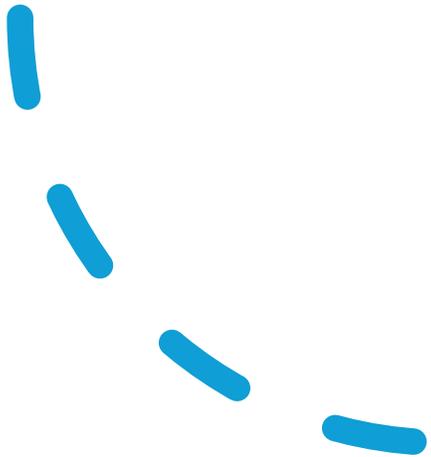
En cuanto a la corrección de los errores pragmáticos y los relativos a la construcción del texto por parte de ChatGPT, los estamos analizando y trabajando en ello.

Compartiremos los resultados en algún congreso próximo.

4. Limitaciones de la investigación y perspectivas de futuro

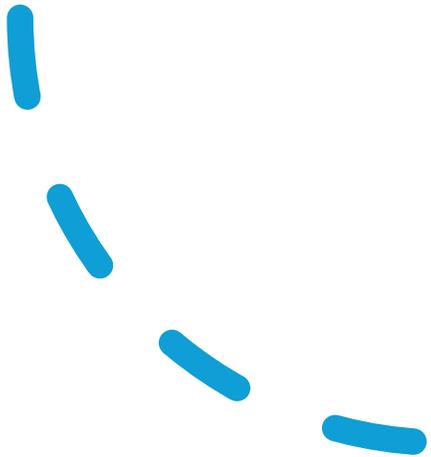
El manejo de Chat GPT por parte del docente

Evolución rápida de la IA (aparición de ChatGPT-5)



4. Limitaciones de la investigación y perspectivas de futuro

Va a ser algo natural trabajar con la IA, por lo que habrá muchos temas de investigación.



Discusión sobre el uso de la IA como herramienta en una clase de lengua extranjera

- Conclusiones
- Abrir un espacio de diálogo sobre el papel de la IA, especialmente ChatGPT, en el ámbito de ELE.



Nuestras conclusiones

- Es útil para la corrección de textos de los estudiantes.
- El tiempo invertido es igual o incluso mayor.
- La productividad no aumenta, pero la calidad de la retroalimentación mejora notablemente.
- El docente aprende mucho al revisar lo que ha hecho ChatGPT, ya que revisa y evalúa críticamente las sugerencias del modelo.

“Alucinaciones”

“Sin embargo, estas herramientas presentan un tipo de error particular conocido como **“alucinaciones”**, que consiste en la generación de respuestas cuyo contenido sigue patrones inexistentes, creando resultados inesperados o incorrectos.”

(Spinak, 2023)

Por eso, la supervisión del docente es esencial.

Discusión sobre el uso de la IA como herramienta en una clase de lengua extranjera

- ¿Qué opinan sobre el uso de la IA para corregir textos de aprendientes de español?
- ¿La usarían en su práctica docente?

Mentimeter:

<https://www.menti.com/alzawtnspv4v>



Docente + IA = Mayor calidad educativa



EL OBJETIVO NO ES SUSTITUIR
AL PROFESOR, SINO
POTENCIAR SU TRABAJO.



NO SOLO SIRVE PARA CORREGIR TEXTOS: SU
POTENCIAL EN ELE ES ENORME. PUEDE
AYUDAR A DISEÑAR MATERIALES, CREAR
EJERCICIOS PERSONALIZADOS Y GENERAR
INTERACCIONES NATURALES PARA
PRACTICAR EL IDIOMA.



PODEROSA HERRAMIENTA SI
SE UTILIZA CON SENTIDO
CRÍTICO Y PEDAGÓGICO.

¿Qué ventajas y riesgos observan en el uso de la IA en clases de ELE?

- ¿Qué ventajas ven en el uso de la IA en clases de lengua extranjera?
- ¿Qué riesgos o dificultades perciben?
- ¿Qué papel debe tener el docente frente a la IA?

Referencias bibliográficas

Martín-Pérez González, L. (2023). *Ctrl + Blablablá: El ABC de la lingüística computacional*. Pie de Página.

Spinak, E. (2023, 20 de diciembre). ¿Es que la Inteligencia Artificial tiene alucinaciones? *SciELO en Perspectiva*.

<https://blog.scielo.org/es/2023/12/20/es-que-la-inteligencia-artificial-tiene-alucinaciones/>

Yamamura, Hiromi. (2019). 日本語母語話者にとっての「過去の出来事」・「過去の状況」とスペイン語の「点過去」・「線過去」の使い分け. *言語科学*, 54, 37–54.



Muchas gracias por su atención

Si tienen alguna pregunta u observación...

Estaremos encantadas de escucharlos.

motocohirai@gmail.com

bujanda8@gmail.com

