

Autocorrección en el aula de ELE a través de ChatGPT 3.5

ÁNGEL DE LA TORRE SÁNCHEZ

Università degli Studi di Macerata (Italia)

A1.delatorresanchez@unimc.it

<https://orcid.org/0000-0001-5052-525X>

Resumen

El artículo aborda una experiencia educativa que explora las posibilidades autocorrectivas de ChatGPT 3.5 para promover la autonomía en contextos de aprendizaje de español como lengua extranjera. Se proporciona un análisis de las principales teorías y aplicaciones didácticas, que sirven como base para analizar los resultados obtenidos. Se examinan muestras de autoevaluación realizadas por estudiantes del nivel C1 a partir de la escritura de textos argumentativos, analizando detalladamente los pasos seguidos por el alumnado para la autocorrección, así como la retroalimentación proporcionada por ChatGPT 3.5. La investigación persigue identificar los criterios y las deficiencias de las sugerencias de corrección de textos por parte de ChatGPT 3.5. El corpus del estudio comprende los intercambios entre estudiantes y la herramienta, con el objetivo de entender cómo dichos intercambios contribuyen a mejorar la calidad de los textos y a fomentar la autonomía en el proceso de escritura y revisión. Tras analizar el uso de ChatGPT 3.5, se infiere que, si bien puede ser útil como guía en la redacción, su enfoque en correcciones gramaticales limita su eficacia. Además, requiere instrucciones precisas, halla dificultades con una gran variedad de elementos lingüísticos y su enfoque superficial puede obstaculizar la mejora del texto de manera autónoma por parte del alumnado.

Palabras clave

ELE; ChatGPT 3.5; escritura; autocorrección; retroalimentación.

Self-correction in the Spanish as a foreign language classroom through ChatGPT 3.5

Abstract

The article examines an educational experience exploring the self-correction capabilities of ChatGPT 3.5 to foster autonomy in Spanish as a foreign language learning contexts. It includes an analysis of key theories and applications that serve as a foundation for evaluating the results. The study analyzes self-evaluation samples conducted by C1-level students based on their argumentative writing, detailing the steps they followed for self-correction and the feedback provided by ChatGPT 3.5. The research seeks to identify the criteria and shortcomings in ChatGPT's text correction suggestions. The study corpus consists of interactions between students and the tool, aiming to understand how these exchanges contribute to improving text quality and to fostering autonomy in writing and revision. The analysis concludes that, while ChatGPT 3.5 can serve as a useful writing guide, its emphasis on grammatical corrections limits its overall effectiveness. Furthermore, it requires precise instructions, struggles with various linguistic elements, and its superficial approach may impede students' autonomy in improving their texts.

Keywords

Spanish as a Foreign Language; ChatGPT 3.5; writing; self-correction; feedback.

Recibido el 26/11/2024

Aceptado el 23/06/2025

Cómo citar/how to cite

de la Torre Sánchez, Á. (2025). Autocorrección en el aula de ELE a través de ChatGPT 3.5. *Revista Internacional De Lenguas Extranjeras / International Journal of Foreign Languages*, (23), 27-49. <https://doi.org/10.17345/rile23.4090>

1. Introducción

La integración de la inteligencia artificial generativa (IAG) en el ámbito educativo ha abierto un vasto campo de posibilidades, transformando la manera en que se abordan algunos procesos de enseñanza y aprendizaje. Entre las diversas aplicaciones en la educación, destaca el uso de ChatGPT 3.5 como una herramienta gratuita y prometedora para mejorar la escritura y la retroalimentación (o *feedback*) en el aula (Hong, 2023). La retroalimentación escrita constituye un componente esencial en la enseñanza de lenguas, aunque su desarrollo suele requerir tiempo, lo que puede generar un distanciamiento cognitivo entre la producción y la revisión del texto (Cassany, 2018). En este contexto, ChatGPT 3.5 ofrece retroalimentación instantánea sobre problemas lingüísticos y de organización, con ejemplos y sugerencias concretas. Sin embargo, la eficacia y las implicaciones pedagógicas de utilizar ChatGPT 3.5 en el proceso de escritura y corrección aún no han sido completamente exploradas, especialmente en el contexto de estudiantes de español como lengua extranjera (ELE). Esta brecha en la investigación motiva este estudio, que busca analizar la retroalimentación de ChatGPT 3.5 en estudiantes de ELE, su asimilación y su impacto en la autonomía para corregir textos.

Partiendo de esta premisa, el estudio se plantea si una herramienta accesible y gratuita como ChatGPT 3.5 puede proporcionar retroalimentación eficaz y si los estudiantes son capaces de asimilarla para fomentar su autonomía en la revisión de textos escritos.

Para abordar estas interrogantes, se llevó a cabo una actividad de producción escrita seguida de una sesión de autocorrección utilizando ChatGPT 3.5. La muestra consistió en diez estudiantes de nivel intermedio-alto (C1 según el MCER) de español, a quienes se propuso una tarea de producción de textos argumentativos sobre temas previamente tratados en clase. A través de la observación y análisis de las interacciones entre el alumnado y la herramienta, se busca comprender cómo ChatGPT 3.5 puede ser utilizado como una herramienta efectiva para mejorar la calidad de la escritura y fomentar la autonomía del estudiante en el proceso de corrección.

Este estudio contribuye a llenar un vacío en la literatura sobre el uso de IAG en la enseñanza de ELE, proporcionando ideas valiosas sobre el potencial de ChatGPT 3.5 como herramienta de autocorrección en el aprendizaje de la escritura en ELE. Los resultados obtenidos pueden tener implicaciones significativas para el diseño de intervenciones pedagógicas centradas en el uso de la tecnología para mejorar las habilidades de escritura de los estudiantes de ELE.

2. ChatGPT 3.5 como herramienta de autocorrección

La integración de herramientas como ChatGPT 3.5 en el aprendizaje de idiomas ha generado un gran interés en la comunidad educativa debido a su capacidad para mejorar diversos aspectos de la adquisición de lenguas. En particular, se ha explorado el potencial de las tecnologías basadas en IAG para proporcionar retroalimentación personalizada

y apoyar el desarrollo de competencias lingüísticas, especialmente en el ámbito de la interacción escrita (Rudolph et al., 2023; Woo et al., 2023; Yan, 2024) y oral (Zheng et al., 2024). Dichos trabajos muestran que algunos sistemas automatizados como ChatGPT 3.5 ofrecen una retroalimentación más dinámica, precisa y centrada en el estudiante, en comparación con las herramientas tradicionales de evaluación automatizada de la escritura (conocidas con las siglas *AWE*, del inglés *automated writing evaluation*). De hecho, la retroalimentación generada por IAG fomenta un proceso más interactivo, lo que permite a los estudiantes involucrarse en revisiones iterativas que mejoran su escritura a lo largo del tiempo, lo que puede ser particularmente útil para resolver errores recurrentes (Yan, 2024). A diferencia de los métodos convencionales de corrección, que a menudo suponen que la retroalimentación por sí sola llevará a una mejora en el aprendizaje, ChatGPT 3.5 ofrece un enfoque más flexible y contextual, que promueve una mayor participación del estudiante y un aprendizaje más autónomo.

En este sentido, en el proceso de aprendizaje de idiomas, la retroalimentación juega un papel crucial a la hora de mejorar los procesos relacionados con la interacción, tanto oral como escrita, puesto que brinda la posibilidad de reducir la brecha entre las expectativas del profesor y las habilidades del estudiante (Sheen y Ellis, 2011). A esto se suma la evidencia que muestra que los estudiantes no reciben suficiente instrucción en escritura ni retroalimentación por parte de los profesores (Butterfuss et al., 2022).

En la definición de Bailini (2024: 220), el proceso de retroalimentación implica ofrecer respuestas a la producción lingüística del estudiante. Esta retroalimentación puede presentarse de diversas formas: oral o escrita, enfocada en todos los errores o solo en ciertas categorías lingüísticas específicas. Además, puede ser directa, proporcionando la corrección explícita, o bien indirecta, señalando la presencia de un error sin ofrecer la solución de inmediato. Este proceso puede ocurrir entre el estudiante y el profesor, o entre compañeros de clase y, a su vez, actualmente es común ofrecer retroalimentación mediante herramientas tecnológicas. Sin embargo, según Bailini (2024: 221), es importante diferenciar entre la retroalimentación electrónica mediado por la tecnología, donde las herramientas -tanto síncronas como asíncronas- se perciben como medios que facilitan la interacción entre individuos, y la retroalimentación electrónica automatizada, donde estas herramientas se perciben como interfaces que interactúan directamente con el usuario.

Diversas investigaciones dedicadas a la eficacia de las diferentes formas de entrega de la retroalimentación, como las de Chen (2016), Zhang y Hyland (2018) y Yilmaz et al., (2024), han indicado que los estudiantes experimentan una mayor sensación de distancia, autonomía y flexibilidad al recibir retroalimentación automatizada. No obstante, se han identificado ciertas desventajas. Por ejemplo, no toda la retroalimentación proporcionada es precisa, como señalan Chapelle et al. (2015). Además, estudios como el de Dikli y Bleyle (2014) sugieren que los profesores experimentados tienden a ofrecer retroalimentación más precisa y útil. Por otro lado, los aprendientes no siempre asimilan la retroalimentación automatizada ni saben cómo utilizarla sin la orientación del docente. A su vez, experimentan confusión si la retroalimentación proporcionada por el sistema automático difiere de la que reciben de sus profesores. En este sentido, si bien

las herramientas automatizadas pueden ayudar a reducir errores lingüísticos, esto podría implicar una simplificación de la complejidad del texto (Li et al., 2015).

Por lo tanto, como proponen Huhta et al. (2023: 251) y Bailini (2024: 224), aunque la evaluación automatizada de escritura puede desempeñar un papel importante en respaldar y complementar la retroalimentación y las discusiones con otros individuos, como los profesores y compañeros, no puede sustituir por completo el papel de ambos grupos como mediadores al explicar las razones detrás de las debilidades identificadas, los conceptos subyacentes involucrados y los pasos específicos a seguir.

Los expertos en el ámbito de la retroalimentación han manifestado que uno de los principales retos de cara al futuro es el desarrollo de los sistemas de tutoría inteligentes que sean cada vez más capaces de ofrecer una retroalimentación detallada, gradual y circunstancial, compaginándolo con la potencialidad de ofrecer retroalimentación personalizada que poseen asistentes virtuales basados en modelos de lenguaje grande como ChatGPT 3.5 (Bailini, 2024: 239). En ese sentido, la introducción del *chatbot* ChatGPT 3.5 en noviembre de 2022 marcó un hito significativo en la accesibilidad a la IAG. Este modelo, que ya no necesita presentaciones dada la profusión de investigaciones en el ámbito, emplea un amplio corpus de texto, compuesto por una gran cantidad de términos provenientes de libros, artículos de prensa, Wikipedia, blogs y otros sitios web. Una vez que han accedido a la herramienta, los usuarios pueden realizar preguntas y dar instrucciones para recabar grandes cantidades de texto de manera fácil a través de una interfaz sencilla. Su valor radica en la capacidad para procesar rápidamente el lenguaje natural y facilitar una comunicación instantánea. Esta herramienta puede analizar y comprender los aspectos lingüísticos de manera profunda para generar texto, asegurando la coherencia y la cohesión en la información. Sus usos en el aula pueden resultar muy variados: motor de búsqueda, oponente socrático, entrenador de colaboración, guía en el aprendizaje, tutor personal, codiseñador, explorador, compañero de estudio, motivador y evaluador dinámico, entre otras (UNESCO, 2023). Aun así, es necesario mencionar que dichas potencialidades se ven limitadas a un puñado de lenguas, diez en el caso de ChatGPT 3.5 en 2024.

Como pilar para este trabajo, se han tenido en cuenta algunos de los temas fundamentales que son recurrentes en la literatura especializada. En primer lugar, la IAG permite fortalecer las habilidades interactivas tanto orales como escritas, así como el desarrollo de elementos como el vocabulario y la gramática (Huang et al., 2023). Además, se destacan los beneficios del aprendizaje personalizado, la adaptabilidad a las necesidades individuales y la oportunidad de práctica ilimitada. En este sentido, hay dos conceptos fundamentales que vertebran el presente trabajo: la autonomía del aprendiente y la (auto)evaluación. En este sentido, uno de los impactos más inmediatos de ChatGPT 3.5 es el fomento de la autonomía (UNESCO, 2021: 23), al resolver dudas y obtener retroalimentación inmediata y personalizada adaptada al ritmo del estudiante (Muñoz-Basols y Fuertes Gutiérrez, 2024: 356), sin experimentar ansiedad o vergüenza, como podría suceder en un entorno de aula (Bárkányi, 2024). Así pues, ChatGPT 3.5 ofrece un entorno de aprendizaje libre de presiones y juicios, lo cual indudablemente beneficia el proceso de aprendizaje. Para aprovechar al máximo estas oportunidades, es

esencial que el alumnado comprenda cómo funcionan las herramientas que utiliza, incluyendo sus limitaciones, y reciba orientación sobre qué actividades pueden enriquecer su uso. De lo contrario, el aprendizaje podría verse afectado si se utiliza la IAG sin un propósito claro.

Por otra parte, la IAG proporciona numerosos beneficios en el ámbito de la (auto) evaluación: los estudiantes pueden utilizar estas herramientas para realizar un seguimiento de su progreso, repasar y prepararse para las evaluaciones. Además, ayuda a los docentes al ahorrarles tiempo en el diseño, la personalización y la corrección de las pruebas, así como en la preparación de la retroalimentación (Bailini, 2024; Kukulska-Hulme et al., 2021).

En este sentido, algunos estudios han intentado demostrar las posibilidades de ChatGPT 3.5 a la hora de llevar a cabo autocorrecciones en comparación con otras herramientas. Entre los resultados, se puede constatar que ChatGPT 3.5 tiene mayores dificultades a la hora de realizar correcciones en oraciones de cierta extensión, tiende a cambiar la expresión superficial de ciertas frases o la estructura de las oraciones, manteniendo la corrección gramatical, es decir, produce menos problemas de corrección insuficiente que otras herramientas, pero más sobrecorrecciones (Wu et al., 2023). A su vez, posee grandes ventajas en entornos multilingües y de bajos recursos lingüísticos, como es el caso de principiantes, lo que lo hace especialmente interesante para el aprendizaje de lenguas (Fang et al., 2023).

3. Metodología

La literatura resalta la capacidad de ChatGPT 3.5 para identificar y corregir problemas de lenguaje, ofreciendo retroalimentación instantánea (Hong, 2023), lo que puede reducir uno de los aspectos más comunes a la hora de proporcionarla: la frustración del alumnado (Cassany, 2018). Sin embargo, la efectividad de ChatGPT 3.5 varía según el nivel de competencia lingüística de los estudiantes (Fang et al., 2023) y plantea desafíos éticos relacionados con derechos de autor y sesgos. Ante la falta de estudios, se realizó una investigación en acción en la que se llevó a cabo un análisis de la retroalimentación ofrecida por ChatGPT 3.5 a las producciones escritas de estudiantes de ELE y, al mismo tiempo, tratará de dar cuenta de la capacidad de asimilación de dicha retroalimentación por parte del alumnado con el fin de comprender la manera en que dicho proceso promueve la autonomía en la corrección de textos escritos.

Para ello, se llevó a cabo un análisis cuantitativo y cualitativo a partir de una actividad de producción escrita y posterior autocorrección con el uso de ChatGPT 3.5. La actividad consistió en la redacción de un texto argumentativo de 350-400 palabras, basado en una temática previamente tratada en clase. Los estudiantes disponían de dos textos de referencia, en español y en italiano. El primer tema fue el grafiti como forma de arte y, el segundo, la presencia del machismo en nuestra sociedad. En este sentido, el análisis cuantitativo considera los criterios de corrección utilizados en el aula para focalizarse en los datos obtenidos a partir de las correcciones ofrecidas por la herramienta,

mientras que el análisis cualitativo se centra en estudiar los intercambios de ChatGPT 3.5 con el alumnado para tratar de obtener una visión más amplia sobre el fenómeno de la autocorrección a través de ChatGPT 3.5 y su implicación en los procesos de aprendizaje del alumnado.

La muestra comprende 10 estudiantes de género femenino que escribieron 10 producciones escritas a partir de una de las dos temáticas mencionadas anteriormente. Todas ellas afirmaron no haber usado nunca la herramienta ChatGPT 3.5 ni otras similares. El objetivo final del curso es la consecución de un nivel C1 de español según el MCER, por lo que el nivel de las estudiantes participantes en la tarea se sitúa en el umbral entre el B2 y el C1. Así pues, inicialmente se realizó la tarea de producción escrita, en la cual el profesor decidió no intervenir mientras el alumnado redactaba su texto, como se recomienda en la literatura (Cassany, 2018: 180), ya que el objetivo era entender cuál era el impacto que ChatGPT 3.5 tenía en la autocorrección de las estudiantes simulando un entorno de escritura sin la tutorización del profesor.

Tras la escritura, se ofreció una explicación sobre el potencial didáctico y las posibilidades de uso tomando como base la guía creada por Pujolá et al. (2023). Tras dichas explicaciones, se les instó a que probaran la herramienta con diferentes *prompts* para que vieran el modo en que razona y qué tipo de retroalimentación ofrece. Seguidamente, se les explicó el objetivo de la tarea y se invitó a las estudiantes a llevar a cabo la autocorrección. Dada la importancia del *prompt*, cuyas implicaciones y resultados concretos obtenidos en la tarea se explicarán más adelante, se decidió llevar a cabo una reflexión colectiva sobre cuáles serían las mejores prácticas y qué detalles tendría que incluir, aunque se dejó libertad a las alumnas para redactar sus propios *prompts*.

El objetivo final era el asimilar las indicaciones, sugerencias, comentarios y correcciones que habían obtenido en la interacción con ChatGPT 3.5 para llevar a cabo una corrección autónoma de sus propios textos y, finalmente, presentar una versión definitiva del texto revisada por ellos mismos. El docente conservó las conversaciones de las estudiantes con la herramienta, así como el texto de partida y el texto definitivo, puesto que dichos intercambios constituyen el corpus objeto de estudio principal del presente trabajo. Las conversaciones tuvieron lugar en sus cuentas personales que, posteriormente, se pidió que enviaran al docente para su análisis.

En este sentido, el corpus está constituido por los *prompts* proporcionados por las estudiantes y las respuestas ofrecidas por parte de la herramienta a cada intervención. Los textos han sido escritos por estudiantes de ELE, sin embargo, el objeto del análisis no es el de incidir sobre un elemento lingüístico específico, sino el entender de qué manera ChatGPT 3.5 puede ser utilizado como herramienta de revisión y autocorrección por parte de estudiantes de ELE. El análisis se centra en el modo en que la herramienta ofrece sugerencias y correcciones en tiempo real.

La metodología, por lo tanto, se basa en el análisis textual de los intercambios con el fin de entender de qué manera los intercambios entre las estudiantes y la herramienta contribuyen a los fenómenos mencionados anteriormente. Para verificar si las correcciones de ChatGPT 3.5 eran adecuadas para el contexto dado, se tomaron en cuenta cuatro criterios de corrección: alcance, corrección, coherencia y adecuación. La elección de

dichos criterios está en línea con la idea de que el tratamiento del error no debe limitarse a aspectos ortográficos, morfosintácticos o léxicos, sino que también tiene en cuenta el error pragmático y textual, con el fin de que el alumnado sea competente tanto en la corrección formal del idioma como en la adecuación sociolingüística y cultural (Muñoz-Basols y Bailini, 2018: 100). A su vez, dichos criterios gozan de amplio consenso en el ámbito de la evaluación de la expresión escrita (Cassany, 2005: 22-23) y han sido adoptados por entidades certificadoras como el Instituto Cervantes para la corrección de sus pruebas.

4. Resultados

4.1 La importancia del *prompt*

A la hora de interactuar con ChatGPT 3.5, es fundamental entender qué es un *prompt*. Los *prompts* son las instrucciones que se introducen en el chat y se perfilan como las herramientas fundamentales para dirigir la interacción con modelos de lenguaje grande como ChatGPT 3.5 (Morales-Chan, 2023). El modo en el que se formulan supone un factor crucial para garantizar conversaciones fructíferas. Así pues, a la hora de interactuar con Chat GPT 3.5, es importante proporcionar un contexto claro y preciso, puesto que de esta manera se garantizará en mayor medida que la herramienta comprenda el tema o la tarea y pueda generar una respuesta adecuada. Es fundamental formular preguntas concretas y específicas, evitando la vaguedad o la confusión, ya que podrían resultar en respuestas imprecisas. Así pues, la calidad de la respuesta de la IAG depende directamente de la información proporcionada, por lo que es fundamental ofrecer un marco claro para obtener respuestas precisas y relevantes.

Entre los elementos que debe contener dicho *prompt*, no hay una propuesta única. Entre los principales, es necesario incluir el tema, el estilo, el tono, el contexto y la información relevante (Morales-Chan, 2023: 2). En el ámbito de ELE, Pujolá et al. (2023: 9) proponen incluir la acción, el formato o extensión, el contexto, la audiencia y extras. A pesar de las divergencias, lo importante es saber adaptarse a la situación que se presenta, puesto que cada tarea presenta un contexto y unos requerimientos diferentes de los que habrá que informar a la herramienta para obtener una respuesta adecuada.

De esta manera, se informó a las estudiantes sobre la importancia del *prompt* en ChatGPT 3.5 a través de algunos ejemplos básicos que podían probar ellas mismas para notar las diferencias. *Prompts* como “cómo se cocina una paella” no reciben la misma retroalimentación que otros como “escribe una receta de paella” o “mañana tengo una fiesta y voy a cocinar una paella para 4 personas, haz una lista de la compra y explícame el procedimiento para cocinar una paella”, por citar algunos ejemplos. A través del material publicado por Pujolá et al. (2023), se intentó explicar a las estudiantes las posibilidades educativas que tenía ChatGPT 3.5, especialmente en su vertiente autocorrectiva, dando cuenta, al mismo tiempo, de los riesgos que podía presentar un uso incorrecto (Álvarez Bernat, 2023). Este fue uno de los escollos de mayor calado a la hora de

trabajar con las estudiantes en clase, puesto que presenta algunas problemáticas. El docente es clave para desarrollar el pensamiento crítico sobre la IAG, asegurando que los estudiantes aprovechen su potencial más allá de *prompts* básicos como “corrige este texto”.

En las muestras del presente estudio, aparecen *prompts* de apertura muy variados. El grado de contextualización es variable entre ellos, a pesar de haber asistido a la misma formación inicial sobre el argumento y no poseer experiencia previa en el uso de la herramienta. Enumeramos a continuación todos los *prompts* tal como fueron redactados por las participantes:

Tabla 1. *Prompts* iniciales utilizados por las estudiantes

Texto 1	<i>Corrige el siguiente texto argumentativo sobre el micromachismo de nivel C1 mencionando y comentando los errores gramaticales, la coherencia y la claridad en la exposición.</i>
Texto 2	<i>corrige el siguiente texto</i>
Texto 3	<i>Corrige este texto de nivel c1</i>
Texto 4	<i>Corrige el texto</i>
Texto 5	<i>Corrige el siguiente texto argumentativo de nivel c1</i>
Texto 6	<i>corrige este texto argumentativo de nivel c1 con los siguientes criterios: corrección, coherencia, léxico</i>
Texto 7	<i>explícame cómo puedo mejorar el siguiente texto argumentativo, es un texto de nivel C1 y los criterios de corrección son: coherencia, adecuación, léxico y corrección</i>
Texto 8	<i>necesito corregir el siguiente texto</i>
Texto 9	<i>Soy una estudiante de la universidad, en Italia y estoy estudiando lenguas en el último año. Por eso tengo un nivel C1 de español. He escrito un texto argumentativo sobre el “micromachismo” y querría que tú me corrijas el texto, mirando el léxico, la coherencia, la adecuación y la corrección, y despues, que por cada categoría (LÉXICO, COHERENCIA, ADECUACIÓN Y PUNTUACIÓN) me explikas la motivación de la corrección.</i>
Texto 10	<i>Soy una estudiante de nivel C1 de español. Como podría mejorar ese texto?</i>

Como se observa, la acción es el elemento que predomina en los *prompts* de las alumnas. Desde el punto de vista expresivo, predomina el imperativo, presente en los primeros 7 casos, mientras que en el número 8 se emplea la primera persona y en el número 9 y 10 una formulación más propia de la interacción, con peticiones a través de subordinadas. A su vez, algunas estudiantes detallaron y contextualizaron de manera profusa, aunque de manera inconsistente, dando detalles sobre el género textual (1, 2, 3, 5, 6, 7, 9) y el nivel (1, 3, 5, 6, 7, 9, 10). Los criterios de corrección, que las estudiantes han podido leer en clase junto al profesor y que están disponibles para su consulta en la plataforma a la que tienen acceso, aparecen en algunos de los *prompts* (1, 6, 7, 9), aunque de manera algo confusa, como en el ejemplo número 9, en el que el usuario ha cambiado los criterios iniciales. El único texto que se apoya en elementos extra en el *prompt* inicial es el número nueve, en el que aparece el contexto académico y geográfico de la estudiante.

Como es natural, la retroalimentación que las estudiantes obtendrán a partir de ese primer *prompt* será muy diversa. Dado que el objetivo no es obtener una corrección del texto definitiva, sino interactuar con ChatGPT 3.5 para lograr una retroalimentación que acompañe su evaluación formativa, entendida como herramienta de aprendizaje que forma parte de las actividades que se proponen en el aula con el objetivo de estimular en el aprendiente la concienciación y la reflexión sobre la lengua (Bailini, 2024: 221), sería ineficaz a la hora de fomentar dicha formación no incidir en las respuestas para obtener mayores detalles sobre la corrección. En este sentido, “es de vital importancia que no nos conformemos con la primera respuesta generada por ChatGPT 3.5 y que usemos nuestro pensamiento crítico, experiencia y formación para conseguir el mejor de los resultados” (Pujolá et al., 2023: 14). Considerando dichas reflexiones, el profesor instó a las estudiantes cuyos *prompts* resultaban demasiado carente de contexto a volver a intentarlo con otros adicionales que respondieran al objetivo de la actividad y a sus necesidades como estudiantes.

Así pues, la retroalimentación que recibieron las estudiantes fue variada. Las participantes que habían usado un *prompt* con cierta contextualización obtuvieron una serie de correcciones, instrucciones y recomendaciones para mejorar su texto. En cambio, las que empezaron el diálogo con un *prompt* básico, lograron tan solo una corrección desprovista de comentarios. Los comentarios y correcciones que ofreció ChatGPT 3.5 a las estudiantes seguían un patrón bastante similar. En los casos de mayor contextualización, como se muestra en la tabla 2 la herramienta proporcionaba un texto introductorio inicial que reformulaba la instrucción proporcionada, si bien este no era un patrón uniforme, ya que no todos los textos con alta contextualización incluían este pasaje parafrástico.

Tabla 2. Ejemplos de textos introductorios ofrecidos por ChatGPT 3.5

T. 1	<i>Tu texto trata un tema importante, el de la desigualdad de género y los micromachismos, pero tiene algunos errores gramaticales y de claridad que podrían corregirse para mejorar la exposición. Aquí están las correcciones y comentarios:</i>
T. 7	<i>El texto que proporcionaste es un argumento bien estructurado que aborda el debate sobre si el graffiti es considerado arte o vandalismo. Aquí tienes algunas sugerencias para mejorarlo:</i>

A continuación, la herramienta ofreció una serie de correcciones y sugerencias personalizadas sobre el texto que seguían un patrón evidente. Por un lado, los *prompts* que incluían algún tipo de criterio de corrección recibían retroalimentación relacionada con dichos criterios. Por otra parte, los *prompts* que no incluían criterios de corrección, sino que solicitaban mejoras o consejos, recibían una respuesta en la que aparecían con frecuencia tres elementos: claridad y coherencia, corrección gramatical y, por último, estilo

En la tabla 3, se pueden ver algunos de los comentarios relativos a los criterios de corrección introducidos por las estudiantes:

Tabla 3. Comentarios ofrecidos por ChatGPT 3.5 relativos a los criterios de corrección introducidos por las estudiantes

T. 6	<p>1. Corrección gramatical y ortográfica:</p> <ul style="list-style-type: none"> - <i>Se corrigieron errores de puntuación y se mejoró la estructura de las oraciones para mayor claridad y fluidez en la lectura.</i> - <i>Se ajustaron algunos tiempos verbales para mantener la coherencia en el texto.</i> <p>2. Coherencia y claridad:</p> <ul style="list-style-type: none"> - <i>Se reorganizó el contenido para que las ideas estén presentadas de manera más lógica y secuencial. Esto ayuda al lector a seguir el argumento con mayor facilidad.</i> - <i>Se eliminaron repeticiones innecesarias para evitar redundancias y hacer el texto más conciso.</i> <p>3. Léxico y estilo:</p> <ul style="list-style-type: none"> - <i>Se utilizaron términos más precisos y variados para enriquecer el léxico del texto y evitar la repetición de palabras.</i> - <i>Se sustituyeron algunas expresiones por sinónimos más adecuados, lo que contribuye a una mejor expresión y estilo en el texto.</i> <p>4. Aclaración de ideas:</p> <ul style="list-style-type: none"> - <i>Se agregaron explicaciones adicionales para ayudar a clarificar conceptos o ejemplos, como la referencia al puñal de gran tamaño pintado en una pared y su impacto en los niños.</i>
-------------	--

Por otra parte, cuando no se ofrecían criterios de corrección, la retroalimentación ofrecida era la siguiente:

Tabla 4. Retroalimentación sin comentarios de ChatGPT 3.5 relativos a los criterios

T. 8	<p>1. Claridad y coherencia: <i>Simplifiqué algunas frases para hacerlas más claras y fáciles de entender. Esto ayuda a que el mensaje sea más accesible para cualquier lector. Por ejemplo, en lugar de "para comunicar o simplemente para informarse y quedarse actualizado", cambié a "comunicarse o mantenerse informadas y actualizadas".</i></p> <p>2. Redacción más precisa: <i>Ajusté algunas frases para mejorar la precisión y eliminar redundancias. Por ejemplo, cambié "Es una plataforma utilizada sobre todo por los jóvenes, pero también por las personas adultas y por los chicos" a "Es una plataforma principalmente utilizada por jóvenes, aunque también por adultos y niños". Esto hace que la idea sea más clara y concisa.</i></p> <p>3. Eliminación de errores gramaticales: <i>Corregí algunos errores gramaticales menores para mejorar la fluidez del texto. Por ejemplo, cambié "Esa es solo una de los muchos ataques y molestias que recibe cada utente todos los días" a "Esa es solo una de las muchas molestias y ataques que recibe cada usuario todos los días".</i></p> <p>4. Reestructuración de frases: <i>Reorganicé algunas frases para mejorar la fluidez y la coherencia del texto. Por ejemplo, cambié "Entre los 15 y los 18, Instagram era para mí un infierno" a "Personalmente, entre los 15 y los 18 años, Instagram fue un infierno para mí".</i></p>
-------------	--

Durante la experiencia de diálogo con ChatGPT 3.5, uno de los aspectos sobre el que las estudiantes hicieron mayor hincapié fue en la necesidad de obtener ejemplos concretos de las correcciones. Como se pudo ver en la tabla 3, fue habitual que la herramienta ofreciera solamente correcciones, sugerencias o mejoras de carácter general, sin mostrar el ejemplo concreto relacionado con dicho comentario. De esta manera, el intercambio con ChatGPT 3.5 fue muy útil para comparar las expresiones o frases que la

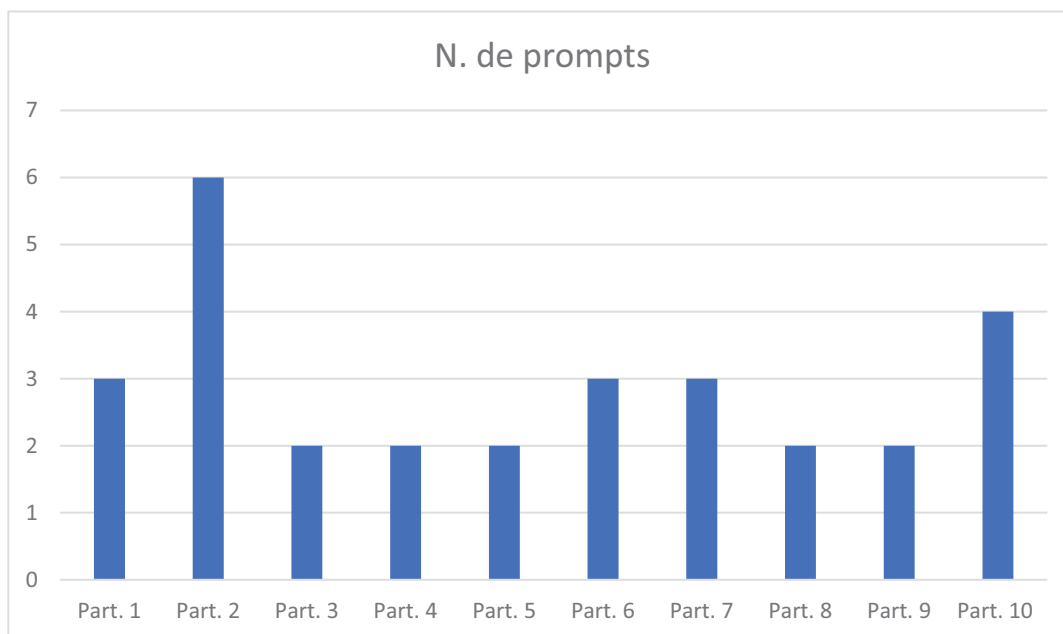
herramienta consideraba necesarias de modificar, como se muestra en algunos de los ejemplos de la tabla 5.

Tabla 5. Comentarios ofrecidos por ChatGPT 3.5 tras la solicitud de ejemplos concretos

T. 1	<ol style="list-style-type: none"> 1. "platicar" - El verbo "platicar" es común en algunos dialectos del español, pero es más apropiado usar "hablar" en este contexto. 2. "entre más" - La forma correcta es "cuanto más" o "cuantos más". 3. "más hay la oportunidad" - La estructura correcta es "hay más oportunidad".
T. 2	<p><i>Original: "Una modalidad de pintura libre o arte visual callejero, por lo general ilegal, generalmente realizado en las superficies de los espacios urbanos: paredes, portones, muros, etc".</i></p> <p><i>Mejora: "El grafiti, una forma de expresión visual callejera, se caracteriza por ser una pintura libre que a menudo se realiza en superficies urbanas como paredes, portones y muros, aunque su práctica suele ser ilegal."</i></p>
T. 3	<p><i>"Algunas personas piensan que el grafiti es un símbolo de arte contemporáneo mientras que otros los consideran como un acto vandálico."</i></p> <p><i>Cambié "piensan que" por "consideran" para hacer la expresión más precisa y formal. Además, eliminé "los" antes de "consideran" para mejorar la claridad gramatical.</i></p>

Estos ejemplos ponen de manifiesto la importancia de interactuar y no conformarse con la primera respuesta. Además, es conveniente recordar que "ChatGPT puede producir textos con errores, sesgados o con información poco precisa, por ello es necesario leer sus respuestas con una actitud crítica, además de poseer el conocimiento suficiente para poder advertirlo" (Pujolá et al., 2023: 34). Por ejemplo, la estudiante que escribió el texto 1 se sorprendió al ver que ChatGPT 3.5 censuraba el uso del verbo "platicar" como sinónimo de "hablar", entre otros casos. Esta retroalimentación puede suponer también una oportunidad para reflexionar sobre los motivos por los que la ChatGPT 3.5 selecciona uno frente a otro y para tratar aspectos de variación diatópica en el aula.

En cuanto al número de *prompts* con el que las estudiantes se sintieron satisfechas para dar paso a la autocorrección del texto que habían escrito, el resultado fue variado, pues oscila entre 2 y 6, con una media de 2,9 por participante, como se muestra en la figura 1. Aunque el análisis detallado del número de *prompts* y las razones tras ese dato excede los objetivos de este trabajo, creemos que es una cuestión interesante para futuras investigaciones. Este aspecto es relevante ya que podría revelar estrategias individuales en el uso de herramientas de autocorrección como ChatGPT 3.5, mostrando diferentes niveles de autonomía y enfoques hacia el aprendizaje. Por ejemplo, un mayor número de *prompts* podría estar relacionado con estudiantes que buscan afinar sus textos a través de iteraciones, mientras que un menor número podría indicar una mayor satisfacción inicial con las sugerencias de la herramienta o una estrategia más focalizada.

Figura 1. Número de *prompts* introducidos por participante

4.2 Las correcciones de ChatGPT 3.5

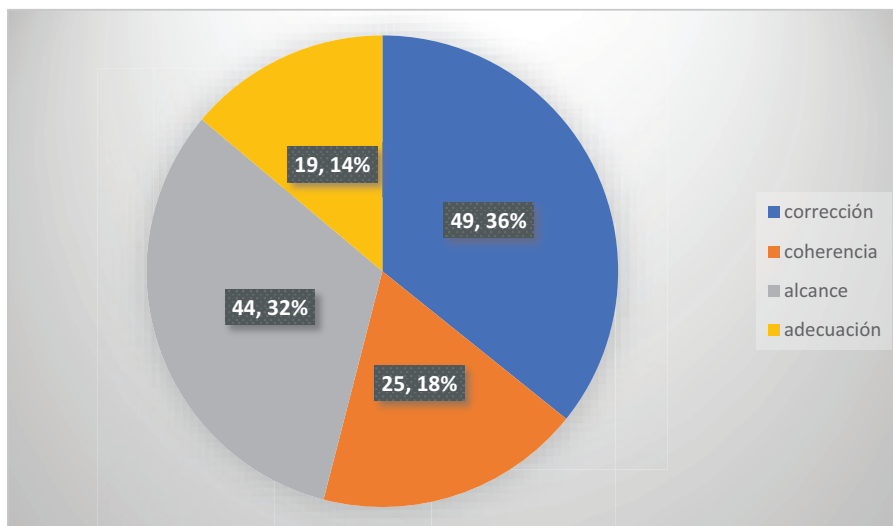
Es esencial analizar cómo ChatGPT 3.5 corrige y sugiere mejoras a las estudiantes, especialmente en contextos universitarios donde la revisión y reescritura de textos implica clarificar ideas, mejorar redacción y transformar el conocimiento (Uribe-Álvarez y Álvarez-Angulo, 2022). La revisión debe ir más allá de los aspectos gramaticales, abordando también deficiencias complejas a través de experiencias didácticas que faciliten la comprensión y ejecución de esta fase (Salas Acuña y Amador Solano, 2023).

Los resultados experimentales presentes en la literatura demuestran que ChatGPT 3.5 es capaz de detectar errores. A su vez, muestra mayores ventajas en entornos multilingües y de recursos limitados, donde su rendimiento puede potenciarse mediante el uso de *prompts* de cadena de pensamiento (*chain-of-thought prompts*, un método de *prompt* que alienta a la IAG a explicar su razonamiento). En el ámbito del aprendizaje del inglés, diferentes análisis de la capacidad de ChatGPT 3.5 para corregir varios tipos de errores gramaticales a nivel de documento indican que tiene un desempeño deficiente en la mayoría de correcciones, como la concordancia, la correferencia, los errores de tiempo entre oraciones y los errores de límites entre oraciones (Fang et al., 2023: 10). Sin embargo, en ocasiones, ChatGPT 3.5 introduce cambios adicionales en la expresión superficial de ciertas frases o estructuras, manteniendo la corrección gramatical (Wu et al., 2023).

No conocemos estudios realizados sobre autocorrección de la interacción escrita a través de ChatGPT 3.5 en el aula de ELE. Sin embargo, en estudios realizados en el ámbito de la escritura académica universitaria (Salas Acuña y Amador Solano, 2023), las correcciones suelen estar relacionadas con la estructura sintáctica (orden y concordancia gramatical, por ejemplo), con el léxico (repeticiones, imprecisiones), la cohesión (uso de

conectores) o la ortotipografía. En el estudio presentado, desde un punto de vista numérico, se observa una sobrerrepresentación de la retroalimentación relacionada con los criterios de corrección y alcance, mientras que la coherencia y la adecuación recibieron menos sugerencias y modificaciones de ChatGPT 3.5, como muestra la figura 2.

Figura 2. Frecuencia de correcciones según criterios



En los siguientes apartados se analizarán las correcciones realizadas por ChatGPT 3.5 a partir de las interacciones con las estudiantes. Para ello, se seguirán los criterios de corrección explicados en la metodología (corrección, coherencia, alcance y adecuación) y se mostrarán las ambigüedades, sesgos y lagunas en el proceso de autocorrección, con el objetivo de ofrecer una visión completa del proceso mediado por esta herramienta.

4.2.1 Corrección

Las correcciones más habituales incluyen aspectos sintácticos, como cambios en el orden de la frase (**más hay la oportunidad de cuestionarnos y reflexionar* → *más oportunidades hay de cuestionarnos y reflexionar*), régimen preposicional regido por la estructura argumental (**patrones muy sutiles los cuales no nos damos cuenta* → *patrones muy sutiles de los cuales no nos damos cuenta*; **requieren de un considerable lapso de tiempo* → *requieren un considerable lapso de tiempo*), contraste entre modo indicativo y subjuntivo con el consiguiente matiz significativo (**los que creen que los grafites sean un tipo de arte* → *aquellos que consideran que el grafiti es una forma de arte*) así como las vinculadas a cuestiones ortográficas (como los acentos gráficos, con un elemento en particular que genera numerosas dificultades: el interrogativo indirecto).

Es relevante señalar que, en ocasiones, ChatGPT 3.5 realiza correcciones en algunas interacciones, mientras que en otras el mismo elemento permanece inalterado. Un ejemplo de ello es el uso de la mayúscula tras los dos puntos en oraciones que continúan en el mismo párrafo, incluso cuando ha corregido otros elementos de la frase, como la puntuación:

N.º Texto	Texto original	Corrección de ChatGPT 3.5
T. 8	<i>Primero: Parar de compararse a otras personas y pensar en sí mismo, focalizar toda la atención en nosotros y amar lo que somos.</i>	<i>Primero: dejar de compararse con otras personas y centrarse en uno mismo, enfocando toda la atención en quiénes somos y amándonos a nosotros mismos</i>
T. 2	<i>La pregunta que nos ponemos hoy es ¿los grafitis son una forma de arte o de vandalismo?</i>	<i>La pregunta que nos planteamos hoy es: ¿Los grafitis son una forma de arte o simplemente un acto de vandalismo?</i>

Además, las correcciones que plantea contienen errores gramaticales que no estaban presentes en el texto original:

N.º Texto	Texto original	Corrección de ChatGPT 3.5
T. 6	<i>¿Cómo se lo explicas a un niño una acción tal?</i>	<i>¿Cómo se les explica una acción así a un niño?</i>

Es significativo mencionar, sin embargo, que “ChatGPT 3.5 no explica la gramática bien”, tal y como apunta Román-Mendoza (2023: 15), a pesar de que es uno de los elementos que corrige con mayor frecuencia. En este sentido, en algunos casos, las estudiantes albergan dudas sobre las correcciones de ChatGPT 3.5 y preguntan de manera explícita el motivo de alguna corrección. A modo de ejemplo, en la tabla 6 mostramos el diálogo que surge entre ChatGPT 3.5 y la estudiante a partir de la siguiente corrección con el elemento que genera dificultad subrayado.

N.º Texto	Texto original	Corrección de ChatGPT 3.5
T. 10	<i>Aunque <u>estamos</u> en el año 2024, el machismo sigue siendo una forma de violencia de la que no se habla lo suficiente.</i>	<i>Aunque <u>estemos</u> en el 2024 el Machismo es una forma de violencia de la que no se habla bastante.</i>

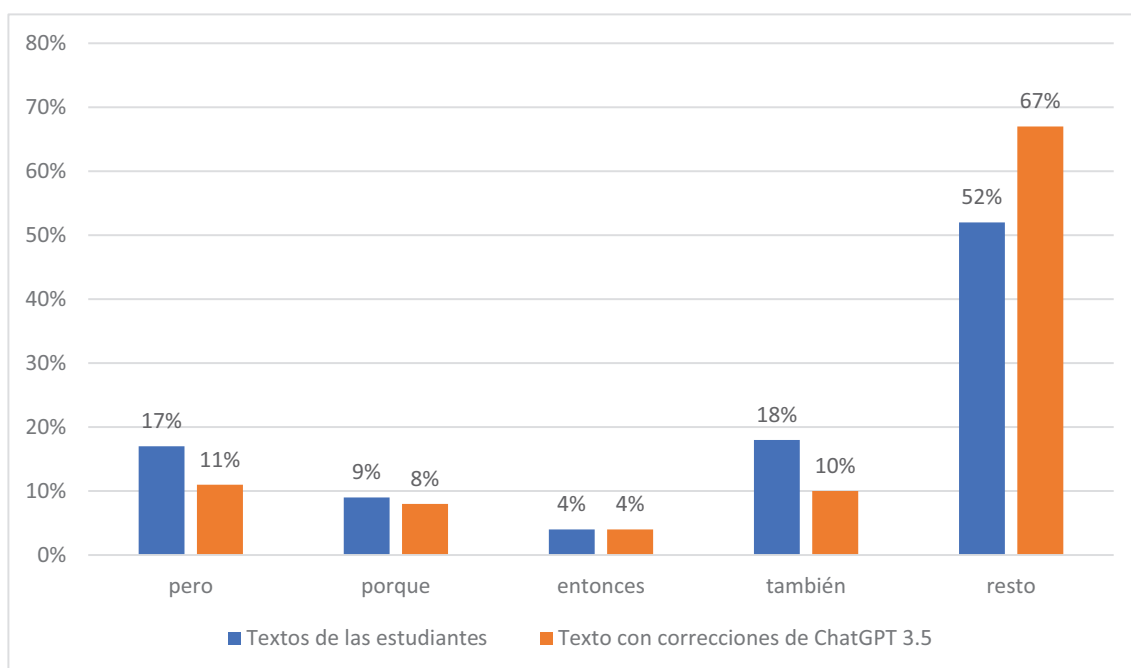
Tabla 6. Diálogo entre ChatGPT 3.5 y una estudiante en torno a una corrección sintáctica (T.10)

Estudiante (T.10):	<i>¿Porqué correjiste “estamos” con “estemos”?</i>
ChatGPT 3.5	<i>La corrección de “estamos” a “estemos” se debe al uso del subjuntivo, que es más apropiado en este contexto. Cuando se expresa una situación hipotética, como en este caso donde se habla de una situación general (el hecho de estar en el año 2024) que no implica certeza absoluta, se utiliza el subjuntivo en español. Entonces, en lugar de “estamos”, se usa “estemos” para indicar esta idea de hipótesis o posibilidad.</i>
Estudiante	<i>Pero es verdad che estamos viviendo en el 2024, no es una hipótesis o una situación no real</i>
ChatGPT 3.5	<i>Tienes razón, me disculpo por la confusión. En este caso, al referirnos a una situación real y presente, “estamos” es la forma correcta. La elección de “estemos” fue un error. Gracias por señalarlo.</i>

4.2.2 Coherencia

En este ámbito, algunas de las correcciones más habituales son la sustitución de conectores por otros de mayor complejidad (*además* → *asimismo*; *por lo tanto* → *por ende*; *para mí* → *desde mi punto de vista*, etc.), aunque a veces sin un criterio claro (*generalmente realizado* → *que a menudo se realiza*). En las correcciones de ChatGPT 3.5, las sustituciones relacionadas con el nivel en el MCER de los conectores son poco comunes, ya que conectores como “pero”, “también”, “porque” y “entonces”, que corresponden a los niveles A1 y A2, aparecen con frecuencia en los textos de las estudiantes sin recibir ningún comentario de la herramienta, como se puede observar en la figura 3.

Figura 3. Porcentaje de aparición de marcadores discursivos



A su vez, son frecuentes las correcciones relativas a la puntuación, como la adición de comas tras los conectores u otros elementos que lo exigen (**sí son pequeños, pero [...]* → *sí, son pequeños, pero [...]*; eliminación cuando no es necesario (**que han sido aceptadas tanto por el público, como por las autoridades* → *que han sido aceptadas tanto por el público como por las autoridades*).

Por otra parte, algunos textos con redacción confusa son modificados de manera tan excesiva que se terminan convirtiendo en un texto completamente diferente, posiblemente fuera del alcance de los conocimientos de dicha estudiante, como se puede ver en el párrafo que enlazamos a continuación. Es interesante notar, con respecto a este párrafo, que se reproduce el patrón que había notado Román-Mendoza (2023: 14), puesto que su análisis confirmó que ChatGPT 3.5 “interpreta” siempre los *prompts* que se introducen, incluso cuando contienen, desde una perspectiva normativista, lo que podría considerarse errores. Un fenómeno similar sucede en el siguiente caso.

N.º Texto	Texto original	Corrección de ChatGPT 3.5
T. 5	<i>Los que creen que los grafitos son arte opinan que es una forma de expresión de los artistas de las calles que, por ejemplo, no se pueden permitir los gastos de ser artistas en otros ámbitos</i>	<i>Quienes ven el grafiti como arte argumentan que es una expresión legítima de los artistas callejeros que, por ejemplo, no pueden costear los gastos asociados con otras formas de arte, como la compra de materiales necesarios para otras formas de pintura.</i>

4.2.3 Alcance

Desde el punto de vista del alcance, las modificaciones y sugerencias se presentan de manera menos abundante. Entre ellas, encontramos incorrecciones (**La pregunta que nos ponemos hoy es* → *La pregunta que nos planteamos hoy es*; **removerlos* → *eliminarlos*) e imprecisiones (*foco* → *enfoque*). Es significativo, en relación con los italo hablantes que aprenden español (y viceversa), mencionar el caso de los falsos amigos, uno de los elementos de mayor dificultad para los aprendices de este par de lenguas (Bailini, 2016: 70 y 110). ChatGPT 3.5 es capaz de reconocer falsos amigos en casos muy evidentes (**disturbios alimentarios*, sustituido por *trastornos alimentarios*), mientras que en frases de mayor complejidad erra en la corrección, creando una frase con sentido diferente al original. Sucede, de manera significativa para el tema que nos ocupa, en el caso de una frase que presenta tres de estos falsos amigos. Aparece, en primer lugar, el caso de *dependiente*, en español, y *dipendente*, en italiano. En segundo lugar, *subir*, en español, y *subire*, en italiano y, por último, *molestia*, en español, y *molestia* en italiano. El primer escollo lo salva adecuadamente, mientras que en la segunda parte de la frase se mantiene el término *molestia*, que en italiano se refiere al *acoso o abuso sexual*. De esta manera, el término *subire* es interpretado por la herramienta dentro del marco de las molestias físicas, sin llegar a interpretar que la estudiante se refería al acoso sexual.

Nº. Texto	Texto original	Corrección de ChatGPT 3.5
T. 8	<i>Según el texto, un ex <u>dependiente</u> de Meta denuncia el hecho de que su hija <u>subió molestias</u> en la plataforma</i>	<i>Según el texto, un ex <u>empleado</u> de Meta denunció que su hija <u>experimentó molestias</u> en la plataforma.</i>

Otros elementos que generan grandes dificultades son los modismos y la fraseología, como sucede en el siguiente caso, que sería fácilmente reconocible para un docente que trabaja en contextos de enseñanza de español a italo hablantes.

Nº. Texto	Texto original	Corrección de ChatGPT 3.5
T. 8	<i>[...] siempre recordarse que no es todo <u>plata</u> lo que brilla.</i>	<i>[...] recordar siempre que no todo lo que brilla es <u>oro</u>.</i>

ChatGPT 3.5 reconoce, de manera parcial, que esta paremia no contiene la palabra *plata* en español, sino *oro*, pero no le ofrece a la estudiante la paremia correcta ni

tampoco una explicación sobre el uso de la misma. La intención de usar la paremia por parte de la estudiante es clara, por lo cualquier docente habría ofrecido una explicación para ayudarle a la hora de usarla.

En ocasiones, la corrección lleva a cabo adiciones que no estaban presentes en el texto original, poniendo en boca de las estudiantes palabras que no han pronunciado:

N.º Texto	Texto original	Corrección de ChatGPT 3.5
T. 7	<i>Estas obras definidas Grafiti pueden desafiar las normas establecidas.</i>	<i>Estas obras, definidas por su espontaneidad y audacia, desafían las normas establecidas.</i>

Sin embargo, se ocasionan mayores lagunas y dificultades a la hora de proporcionar retroalimentación sobre el léxico para el nivel requerido, carencia sobre la cual actuaría cualquier docente. Verbos como *mover* o *querer*, de niveles inferiores, permanecen inalterados (*al escribir sobre este tema me gustaría mover el enfoque; quiero lanzar una provocación*). A su vez, repeticiones evidentes (*desafortunadamente*, presente dos veces en un mismo párrafo), son pasadas por alto, si bien es posible encontrar excepciones cuando la repetición es notoria (“en la frase *como personas que se divierten en pintar los muros solo para divertirse sin motivo* eliminé la repetición de *divertirse* para evitar redundancia”, afirma ChatGPT 3.5 en un intercambio con la estudiante que escribió el texto n.º 3). Por otra parte, en ocasiones los textos presentan términos o expresiones inexistentes en la lengua española, que ChatGPT 3.5 no corrige ni menciona como problemáticos, como el término **incorrectitud* o el modismo **poner luz*.

N.º Texto	Texto original	Corrección de ChatGPT 3.5
T. 9	No tenemos que callarnos sino subrayar la disparidad y la <i>incorrectitud</i> de lo que se está diciendo.	No debemos quedarnos callados, sino más bien resaltar la disparidad y la <i>incorrectitud</i> de lo que se está diciendo.
T. 10	Creo que sea fundamental <i>poner luz</i> sobre la limitación no explícita sufrida por las mujeres.	Creo que es fundamental <i>poner luz</i> sobre la limitación no explícita sufrida por las mujeres.

4.2.4 Adecuación

En cuanto a la adecuación, el registro suele ser uno de los factores con mayor presencia, como ocurre con la sustitución o adición de términos para conferir mayor formalidad y naturalidad al texto (por ejemplo, *¿Los grafitis son una forma de arte o de vandalismo?*, frente a *¿Los grafitis constituyen una forma de arte o de vandalismo?*; lo *hacen* empleando dinero público, sustituido por lo que *conlleva* un gasto de dinero público). En algunos casos ChatGPT 3.5 incluso menciona explícitamente que se debe a matices estilísticos: “cambié *decorado de las ciudades* por *paisaje urbano* para una expresión más natural y reemplacé *gente* por *personas* para mayor formalidad”. Es interesante notar, en este caso, una corrección frecuente relacionada con los verbos *existir* y *haber*, cuyo grado de naturalidad varía en los textos según ChatGPT 3.5, aunque no ofrece una explicación clara al respecto (por ejemplo, en el texto 1, *que existe entre mujeres y varones* es corregido con

que hay entre mujeres y varones, mientras que en el texto 4, hay ejemplos concretos en la vida diaria que demuestran que nuestra sociedad es machista es corregido con existen ejemplos concretos en la vida diaria que demuestran que nuestra sociedad es machista).

También se encuentran correcciones relacionadas con el registro, tanto diatópico (por ejemplo, de *platicar* a *hablar*), como diafásicas (*lo de cerrar esta brecha* a *cerrar esta brecha*) o combinaciones de ambas (*entre más se discute sobre algo* a *cuanto más se discute sobre algo*).

Para concluir el análisis, es relevante señalar que los textos que presentan un nivel alto de corrección reciben, en nuestra opinión, una retroalimentación insatisfactoria, ya que las correcciones y sugerencias, a menudo, carecen de un criterio claro y no permiten que la estudiante utilice la herramienta para autocorregirse y mejorar su destreza en la expresión escrita. En este sentido, se observa cómo se enfoca más en valorar y discutir las polémicas relacionadas con el vandalismo, lo que desvía la atención de los objetivos lingüísticos deseados.

N.º Texto	Texto original	Corrección de ChatGPT 3.5
T. 7	<i>Quizás el problema no sea tanto si el grafiti es arte o vandalismo, sino cómo podemos reconciliar estas dos perspectivas opuestas. La educación, el diálogo y la colaboración entre artistas, autoridades y población pueden ser la base para encontrar soluciones que satisfagan los deseos tanto de los artistas como de la sociedad en general.</i>	<i>El verdadero desafío puede no ser determinar si el grafiti es arte o vandalismo, sino cómo reconciliar estas dos perspectivas aparentemente opuestas. La educación, el diálogo y la colaboración entre artistas, autoridades y la sociedad en general pueden servir como la base para encontrar soluciones que satisfagan tanto las necesidades de los artistas como las de la comunidad en su conjunto.</i>

5. Discusión

El análisis realizado nos permite ofrecer algunas interpretaciones de los resultados obtenidos. En primer lugar, la calidad de la respuesta de ChatGPT 3.5 está directamente relacionada con la calidad de la información proporcionada. Es fundamental ofrecer un contexto claro y preciso para obtener respuestas adecuadas. Los *prompts* deben incluir el tema, el estilo, el tono, el contexto y la información relevante. Además, es fundamental saber adaptarse a la situación que se presenta, puesto que cada tarea presenta un contexto y unos requerimientos diferentes. A su vez, los *prompts* iniciales utilizados por las estudiantes fueron muy variados, aunque predominaba el elemento de acción. Por este motivo, la retroalimentación que recibieron las estudiantes también fue variada. Las participantes que habían usado un *prompt* con cierta contextualización obtuvieron una serie de correcciones, instrucciones y recomendaciones para mejorar su texto. En cambio, aquellas que comenzaron el diálogo con un *prompt* básico lograron tan solo una corrección desprovista de comentarios. La retroalimentación ofrecida por ChatGPT 3.5 a las estudiantes seguía un patrón bastante similar. En los casos de mayor contextualización, la herramienta proporcionaba un texto introductorio que reformulaba la instrucción original. A continuación, ofrecía una serie de correcciones y sugerencias personalizadas sobre el texto, siguiendo un patrón claro.

Con respecto a las correcciones, el análisis muestra que ChatGPT 3.5 tiene capacidad para detectar errores, aunque esta es muy heterogénea. Siguiendo los criterios de corrección utilizados en la investigación, podemos afirmar que ChatGPT 3.5 realiza ajustes relacionados con la sintaxis, la concordancia gramatical y la ortografía, pero, en ocasiones, mantiene errores o introduce otros que no estaban presentes en el texto original. Esto puede generar confusión entre las estudiantes, quienes, a veces, cuestionan las correcciones.

ChatGPT 3.5 también tiende a modificar el texto para darle mayor fluidez, si bien no ofrece una explicación clara del motivo de dicha corrección lo que dificulta que las estudiantes comprendan el motivo de los cambios.

En cuanto a la coherencia, las correcciones de ChatGPT 3.5 suelen incluir la adición o eliminación de conectores y ajustes de puntuación. Sin embargo, en algunos casos, las correcciones transforman significativamente el texto original, lo que puede exceder el alcance de los conocimientos del aprendiente. Las ausencias de ciertos conectores o la falta de criterio en su sustitución son notables.

En términos de alcance, ChatGPT 3.5 realiza menos modificaciones y sugerencias. Cuando lo hace, a veces introduce incorrecciones o no corrige términos inadecuados, como en el caso de falsos amigos entre el español y el italiano. La herramienta también tiene dificultades con modismos y frases idiomáticas, ya que no siempre proporciona la corrección adecuada o la explicación correspondiente.

En lo concerniente a la adecuación, ChatGPT 3.5 suele modificar el registro para hacerlo más formal y natural, pero sin un criterio evidente para el usuario. Las correcciones relacionadas con el registro diatópico y diafásico son comunes, pero la herramienta a veces agrega términos o expresiones que no estaban presentes en el texto original.

Así, considerando las preguntas que motivan la presente investigación, es posible afirmar que ChatGPT ofrece retroalimentación para corregir las producciones del alumnado y demuestra una capacidad notable para detectar y corregir ciertos errores. Sin embargo, su efectividad es limitada por la falta de coherencia y claridad en algunas de sus correcciones y sugerencias. Por lo que respecta al fomento de la autonomía, la herramienta puede ser útil para mejorar aspectos básicos de la escritura, pero no sustituye la guía pedagógica necesaria para desarrollar completamente las habilidades de redacción del alumnado. Supone, por tanto, una herramienta complementaria a la retroalimentación del profesorado.

En este sentido, consideramos que los resultados arrojan datos esperanzadores para implementar nuevas herramientas para fomentar la autocorrección y la autonomía de aprendizaje a la hora de revisar y reescribir textos. No obstante, es necesario incluir la instrucción explícita sobre la herramienta en el aula para aprovechar al máximo sus potencialidades.

6. Conclusiones

En el presente artículo se ha presentado una investigación cuyo objetivo es conocer el alcance y las potencialidades de las IAG, particularmente de la herramienta gratuita ChatGPT 3.5, con el fin de fomentar la autocorrección y la autonomía del aprendizaje en

el aula de ELE. Para ello, mediante una metodología de investigación-acción, se llevó a cabo una actividad parcialmente guiada en el aula en la cual el alumnado realizó un ejercicio de expresión escrita y, posteriormente, una tarea de autocorrección usando ChatGPT 3.5.

Tras el análisis, se observa que, aunque esta herramienta tecnológica ofrece ciertas ventajas, la presencia y orientación del profesorado sigue siendo preferible como guía en el proceso de escritura. ChatGPT 3.5 puede desempeñar roles útiles como guía, tutor o compañero durante la redacción de textos. Sin embargo, su enfoque se centra en las correcciones gramaticales, lo que puede limitar su eficacia en términos de enseñanza y mejora del texto. Si bien ChatGPT 3.5 es útil para la autocorrección, debe complementarse con la orientación docente, indispensable para un aprendizaje eficaz. A su vez, su eficacia varía según el nivel de habilidad del aprendiente y la claridad de los *prompts* proporcionados. En definitiva, la incorporación de herramientas como ChatGPT 3.5 en el aula de ELE requiere una integración pedagógica crítica que complemente, pero no sustituya, el papel del docente en el desarrollo de la competencia escrita.

Como posible línea de investigación futura, se podría ampliar esta investigación a otros géneros textuales más específicos y variados. Así pues, explorar cómo ChatGPT 3.5 y otras IAG se adaptan y mejoran en estos contextos textuales diversificados permitiría entender mejor sus capacidades y limitaciones. Además, sería relevante investigar su aplicación en diferentes niveles lingüísticos y diferentes contextos geográficos para evaluar cómo pueden fomentar la autonomía y el desarrollo de habilidades lingüísticas a lo largo del proceso educativo. Comparar su desempeño con otras IAG de acceso gratuito podría proporcionar una visión valiosa sobre la eficacia relativa de estas en la enseñanza de la escritura y la comunicación efectiva en diversos contextos.

Referencias bibliográficas

- ÁLVAREZ BERNAT, Naiara (2023). *Los riesgos de la inteligencia artificial en la educación: el caso del ChatGPT*. [Trabajo de Fin de Grado, Universitat Jaume I]
- BAILINI, Sonia (2016). *La interlengua de lenguas afines: el español de los italianos, el italiano de los españoles* (Vol. 7). LED Edizioni Universitarie.
- BAILINI, Sonia (2024). Evaluación y feedback en entornos virtuales de aprendizaje. En J. MUÑOZ-BASOLS, M. FUERTES GUTIÉRREZ y L. CERESO (Eds.), *La enseñanza del español mediada por tecnología. De la justicia social a la Inteligencia Artificial (IA)* (pp. 218-244). Routledge.
- BÁRKÁNYI, Zsuzsanna. (2024). Ansiedad y aprendizaje virtual. En J. MUÑOZ-BASOLS, M. FUERTES GUTIÉRREZ y L. CERESO (Eds.), *La enseñanza del español mediada por tecnología. De la justicia social a la Inteligencia Artificial (IA)* (pp. 160-185). Routledge.
- BUTTERFUSS, Reese, ROSCOE, Rod. D., ALLEN, Laura K., MCCARTHY, Kathryn S. y MCNAMARA, Danielle S. (2022). Strategy Uptake in Writing Pal: Adaptive Feedback and Instruction. *Journal of Educational Computing Research*, 60(3), 696–721. <https://doi.org/10.1177/073563312111045304>

- CASSANY, Daniel (2018). Expresión escrita. En J. MUÑOZ-BASOLS, E. GIRONZETTI, M. LACORTE (Eds.), *The Routledge Handbook of Spanish Language Teaching* (pp. 168-182). Routledge.
- CASSANY, Daniel (2005). *Expresión escrita en L2/ ELE*. Arco/Libros.
- CHAPELLE, Carol, COTOS, Elena y LEE, Jooyoung (2015). Validity arguments for diagnostic assessment using automated writing evaluation. *Language Testing*, 32(3), 385–405. <https://doi.org/10.1177/0265532214565386>
- CHEN, Tsui ping (2016). Technology-supported peer feedback in ESL/ EFL writing classes: A research synthesis. *Computer Assisted Language Learning*, 29(2), 365–397. <https://doi.org/10.1080/09588221.2014.960942>
- DIKLI, Semire. y BLEYLE, Susan (2014). Automated essay scoring feedback for second language writers: How does it compare to instructor feedback? *Assessing Writing*, 22, 1–17. <https://doi.org/10.1016/J.ASW.2014.03.006>
- FANG, Tao, YANG, Shu, LAN, Kaixin, WONG, Derek F, HU, Jinpeng, CHAO, Lidia S. y ZHANG, Yue (2023). Is ChatGPT a highly fluent grammatical error correction system? a comprehensive evaluation. *arXiv preprint arXiv:2304.01746*.
- HONG, Wilson Cheong Hin (2023). The impact of ChatGPT on foreign language teaching and learning: Opportunities in education and research. *Journal of Educational Technology and Innovation*, 5(1), 37-45.
- HUANG, Xinyi, ZOU, Di, CHENG, Gary, CHEN, Xieling y XIE, Haoran. (2023). Trends, research issues and applications of artificial intelligence in language education. *Educational Technology & Society*, 26(1), 112-131.
- HUHTA, Ari, HARSCH, Claudia, LEONTJEV, Dmitri y NIEMINEN, Lea (2023). *The diagnosis of writing in a second or foreign language*. Taylor & Francis.
- KUKULSKA-HULME, Agnes, BOSSU, Carina, COUGHLAN, Tim, FERGUSON, Rebecca, FITZGERALD, Elizabeth, GAVED, Mark, HERODOTOU, Christothea et al. (2021). *Innovating Pedagogy 2021: Open University Innovation Report 9*. The Open University.
- LI, Jinrong, LINK, Stephanie y HEGELHEIMER, Volker (2015). Rethinking the role of automated writing evaluation (AWE) feedback in ESL writing instruction. *Journal of Second Language Writing*, 27, 1–18. <https://doi.org/10.1016/j.jslw.2014.10.004>
- MORALES-CHAN, Miguel Ángel (2023). Explorando el potencial de Chat GPT: Una clasificación de Prompts efectivos para la enseñanza. *GES Department, Universidad Galileo, Guatemala*, 1–8, <https://biblioteca.galileo.edu/tesario/handle/123456789/1348>
- MUÑOZ-BASOLS, Javier y BAILINI, Sonia (2018). Análisis y corrección de errores. En J. MUÑOZ-BASOLS, E. GIRONZETTI, E., M. LACORTE (Eds.), *The Routledge Handbook of Spanish Language Teaching* (pp. 94-108). Routledge.
- MUÑOZ-BASOLS, Javier y FUERTES GUTIÉRREZ, Mara (2024). Oportunidades de la Inteligencia Artificial (IA) en la enseñanza y el aprendizaje de lenguas. En J. MUÑOZ-BASOLS y M. FUERTES GUTIÉRREZ (Eds.), *La enseñanza del español mediada por tecnología* (pp. 343-365). Routledge.

- PUJOLÀ FONT, Joan Tomas, GONZÁLEZ ARGÜELLO, María Vicenta y MENA OCTAVIO, Manuela (2023). *ChatGPT en la clase de ELE*. Universitat de Barcelona, <https://diposit.ub.edu/dspace/handle/2445/203129>
- ROMÁN-MENDOZA, Elena (2023). Formular preguntas para comprender las respuestas: ChatGPT como agente conversacional en el aprendizaje de español como segunda lengua. *MarcoELE: Revista de Didáctica Español Lengua Extranjera*, (36), 6.
- RUDOLPH, Jürgen, TAN, Samson y TAN, Shannon (2023). ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? *Journal of applied learning and teaching*, 6(1), 342-363.
- SALAS ACUÑA, Erick F. y AMADOR SOLANO, María Gabriela (2023). Usos de ChatGPT® para la revisión de textos académicos: algunas consideraciones. *Innovaciones Educativas*, 25(Especial), 59-77.
- SHEEN, Younghee y ELLIS, Rod (2011). Corrective feedback in language teaching. En E. Hinkel (Ed.), *Handbook of Research In Second Language Teaching And Learning*, Vol 2, (pp. 593-610). Routledge, 593-610.
- UNESCO. (2023). *ChatGPT: An Artificial Intelligence in Higher Education*. UNESCO. Document code: ED/HE/IESALC/IP/2023/12.
- UNESCO. (2021). *Inteligencia artificial y educación: guía para las personas a cargo de formular políticas*. UNESCO. <https://unesdoc.unesco.org/ark:/48223/pf0000379376>
- URIBE-ÁLVAREZ, Graciela y ÁLVAREZ-ANGULO, Teodoro (2022). Presentación. En María Teresa MATEO-GIRONA, Graciela. URIBE-ÁLVAREZ y Silvia Eva AGOSTO-RIERA (Eds.). *Revisión y reescritura para la mejora de los textos académicos* (pp. 9-12). Octaedro.
- WOO, David James, SUSANTO, Hengky, YEUNG, Chi Ho, GUO, Kai y FUNG, April Ka Yeng (2023). Exploring AI-Generated text in student writing: How does AI help?. *arXiv preprint arXiv:2304.02478*.
- WU, Haoran, WANG, Wenxuan, WAN, Yuxuan, JIAO, Wenxiang y LYU, Michael (2023). ChatGPT or Grammarly? Evaluating ChatGPT on grammatical error correction benchmark. *arXiv preprint arXiv:2303.13648*.
- YAN, Da (2024). Comparing individual vs. collaborative processing of ChatGPT-generated feedback: Effects on L2 writing task improvement and learning. *Language Learning & Technology*, 28(1), 1-19. <https://hdl.handle.net/10125/73597>
- YILMAZ, Yucel, GRANENA, Gisela, CANALS, Laia y MALICKA, Aleksandra (2024). The effects of feedback type and explicit associative memory on the effectiveness of delayed corrective feedback in computer-mediated communication. *Language Learning & Technology*, 28(1), 1-26. <https://hdl.handle.net/10125/73588>
- ZHANG, Zhe y HYLAND, Ken (2018). Student engagement with teacher and automated feedback on L2 writing. *Assessing Writing*, 36, 90-102. <https://doi.org/10.1016/J.ASW.2018.02.004>
- ZHENG, Chunping, CHEN, Xu, ZHANG, Huayang y CHAI, Ching Sing (2024). Automated versus peer assessment: Effects on learners' English public speaking. *Language Learning & Technology*, 28(2), 210-228. <https://hdl.handle.net/10125/73577>