

# Lexicografía innovadora: Generando una gramática de abajo arriba para acercarse a los aprendices de lengua

SVEN TARP

Universitat Rovira i Virgili (España) y Universidad de Aarhus (Dinamarca)

st@cc.au.dk

<https://orcid.org/0000-0003-1941-9082>

## Resumen

Este trabajo presenta una nueva visión para la enseñanza de la gramática española a estudiantes extranjeros, desarrollada a través de la producción de un asistente de escritura digital con el apoyo de la inteligencia artificial generativa. Basado en la redefinición de la lexicografía como diccionografía y glosografía, el proyecto explora nuevas glosas digitales generadas mediante la IA para crear una gramática para aprendices. Esta gramática se distingue de los enfoques tradicionales al adoptar una orientación ascendente, es decir, de abajo arriba, que parte de los errores detectados en los textos de los aprendices para construir explicaciones gramaticales contextualizadas. El asistente de escritura utiliza la IA para diversas tareas, incluyendo la identificación de errores y la generación de explicaciones gramaticales. Se generan tanto explicaciones largas, que abordan en profundidad las reglas gramaticales, como glosas cortas, que ofrecen soluciones rápidas y concisas. La combinación de estas glosas sigue una metodología hegeliana invertida, donde se parte de lo individual (el error) para llegar a lo universal (la regla gramatical), facilitando un aprendizaje más intuitivo y práctico. El desarrollo de este asistente destaca la importancia de la colaboración interdisciplinaria y el papel crucial de los lexicógrafos en la era digital. Se espera que esta herramienta mejore significativamente la competencia escrita de los estudiantes de español, aunque el desafío futuro será fomentar habilidades lingüísticas más profundas y creatividad.

## **Palabras clave**

Asistente de escritura; glosografía digital; gramática; IA generativa; modelo lingüístico.

## **Innovative lexicography: Generating a bottom-up grammar to get closer to language learners**

### **Abstract**

This paper presents a new approach to teaching Spanish grammar to foreign language learners, developed through the production of a digital writing assistant supported by generative artificial intelligence. Based on the redefinition of lexicography as dictionography and glossography, the project explores new digital glosses generated by AI to create a grammar for learners. This grammar distinguishes itself from traditional approaches by adopting a bottom-up orientation, starting from errors detected in learners' texts to construct contextualized grammatical explanations. The writing assistant uses AI for various tasks, including error identification and the generation of grammatical explanations. Both long explanations, which address grammatical rules in depth, and short glosses, which offer quick and concise solutions, are generated. The combination of these glosses follows an inverted Hegelian methodology, where it starts from the individual (the error) to reach the universal (the grammatical rule), facilitating a more intuitive and practical learning process. The development of this assistant highlights the importance of interdisciplinary collaboration and the crucial role of lexicographers in the digital age. It is expected that this tool will significantly improve the writing competence of Spanish learners, although the future challenge will be to foster deeper linguistic skills and creativity.

### **Keywords**

Digital glossography; generative AI; grammar; language model; writing assistant.

Recibido el 30/04/2025

Aceptado el 25/06/2025

### **Cómo citar/how to cite**

Tarp, S. (2025). Lexicografía innovadora: Generando una gramática de abajo arriba para acercarse a los aprendices de lengua. *Revista Internacional De Lenguas Extranjeras / International Journal of Foreign Languages*, (23), 76-94. <https://doi.org/10.17345/rile23.4197>

## 1. Introducción

Inspirado en las reflexiones de Al-Kasimi (1977), este trabajo se basa en dos supuestos. El primero es que cada cultura y época desarrollan sus propios productos lexicográficos, adaptados a sus posibilidades tecnológicas y destinados a satisfacer las necesidades de su sociedad. El segundo supuesto es que el actual empuje tecnológico, con la inteligencia artificial generativa a la cabeza, no solo implica nuevos métodos de compilación para los productos tradicionales, sino también formas totalmente nuevas de abordar las necesidades de los usuarios. Por tanto, las nuevas tecnologías conllevan nuevos tipos de productos lexicográficos diferentes a los anteriores, por lo que los lexicógrafos deben estar atentos al constante flujo de innovaciones tecnológicas y a las posibilidades que ofrecen para rejuvenecer la disciplina.

Partiendo de esta premisa, se ha desarrollado una nueva visión de la enseñanza de la gramática española a estudiantes extranjeros. Esta visión ha ido tomando forma en el proceso de producción de una herramienta de escritura digital con la ayuda multifacética de la inteligencia artificial. La nueva visión se basa en la redefinición que Tarp & Gouws (2023) hicieron de la lexicografía como una disciplina con dos ramas principales: la diccionografía y la glosografía. Esta última incluye no solo las glosas que históricamente dieron origen a los diccionarios en muchas culturas, como señalan McArthur (1986), Hanks (2013) y Benati & Händl (2019), y de las que Ruiz Asencio et al. (2020) ofrecen unos bellos ejemplos encontrados en el Monasterio de Santo Domingo de Silos. También incluye los nuevos tipos de glosas que surgen en la era digital y que Gouws & Tarp (2024) vinculan al futuro mismo de la lexicografía como práctica cultural.

La nueva visión del aprendizaje gramatical se ha construido sobre tres de estas nuevas glosas digitales que, a su vez, se generan mediante tres métodos muy distintos de interacción con la inteligencia artificial que se explicarán más adelante. Cuando estas glosas lexicográficas de nuevo tipo se combinan de manera adecuada, pueden adoptar la forma de una gramática para aprendices que difiere en diseño, contenido y enfoque didáctico no solo de los manuales de gramática tradicionales, sino también de las mini-gramáticas integradas en los libros de texto y diccionarios, así como de las notas o explicaciones gramaticales dirigidas a palabras individuales en los diccionarios para aprendices.

Las glosas en cuestión se están elaborando en el marco de un proyecto de asistente de escritura apoyado por un modelo lingüístico basado en la IA, explicado en su versión inicial por Tarp (2023) y, después de la integración de la IA generativa en el proyecto, por Huete García & Tarp (2024), Tarp & Nomdedeu Rull (2024) y Li et al. (2024). Este proyecto es tanto de *investigación* e interés académico como de *producción*, por lo que se lleva a cabo en colaboración con la empresa danesa Ordbogen A/S, que aporta informáticos y servidores para el proyecto. Su objetivo es ayudar a los estudiantes de español como segunda lengua a escribir textos en español, no solo para mejorar su redacción concreta e inmediata, sino también para adquirir destrezas de escritura más avanzadas en esta lengua mediante la asimilación gradual de su gramática.

El apartado 2 contiene algunas reflexiones generales sobre lengua, gramática y enseñanza. En el siguiente apartado se presentan las características básicas del nuevo

tipo de gramática que se pretende alcanzar, así como las exigencias de las explicaciones sobre las que se construye. El apartado 4 explica, en líneas generales, el papel de los lexicógrafos en el proyecto y cómo se utiliza la IA, mientras que los apartados 5 y 6 describen cómo se aprovecha esta tecnología para generar las explicaciones largas y cortas, respectivamente. A continuación, el apartado 7 introduce tres conceptos dialécticos hegelianos para comprender mejor cómo esta herramienta puede facilitar el aprendizaje gramatical de abajo arriba. Finalmente, el apartado 8 contiene las conclusiones y algunas perspectivas.

## 2. Reflexiones generales sobre la gramática y su enseñanza

La gramática, como sistema de reglas establecidas por lingüistas y otros estudiosos, es el resultado de la observación, abstracción y generalización de las regularidades estructurales que caracterizan la lengua y cohesionan sus palabras. Sin embargo, la lengua es un fenómeno natural y vivo. La crean personas en contextos sociales que han ido cambiando y evolucionando constantemente a lo largo de miles de años. Por tanto, la lengua no es estática. Como todos los demás fenómenos de este mundo, está sujeta a las leyes de la evolución, y también lo está su gramática. Cualquier gramática oficial es, desde una perspectiva histórica, un sistema rígido que tiende a quedarse obsoleto desde el momento en que las reglas gramaticales se formulan y se elevan a la categoría de norma. Una gramática dictada desde arriba puede retrasar el proceso hasta cierto punto, pero no puede detenerlo. Más bien, una imposición de este tipo provoca que la gramática oficial se distancie cada vez más de la gramática viva utilizada por la gente “de abajo”, quienes son los verdaderos creadores lingüísticos. Esto es especialmente cierto si la gramática oficial se inspira en ideas aristocráticas, puristas y excesivamente rígidas. La historia no carece de ejemplos de ello.

Los seres humanos hablaban miles de lenguas durante decenas de miles de años antes de que se formularan las primeras reglas gramaticales. Los niños de todas las culturas aprendieron y siguen aprendiendo su lengua materna básicamente a través de la socialización y la interacción con sus padres, familiares y otros niños. Muchos adultos dominan su lengua materna a la perfección, pero son incapaces de explicar las reglas gramaticales que utilizan intuitivamente para expresarse con claridad. Los estudios gramaticales pueden ser útiles y contribuir al conocimiento de la lengua y sus posibilidades expresivas, pero no son la fuente principal del desarrollo lingüístico materno de los niños y adolescentes.

Sin embargo, en el caso del aprendizaje de una segunda lengua, la enseñanza de la gramática se ha convertido hoy en día en algo habitual. Según el sistema didáctico empleado, los aprendices de L2 estudian y memorizan las distintas reglas gramaticales y luego intentan utilizarlas en diferentes tipos de ejercicios o tareas de escritura. Esta práctica no deja de resultar un tanto sorprendente, ya que el proceso de adquisición de la lengua materna demuestra que la competencia lingüística no surge necesariamente de la instrucción normativa.

La diferencia entre la gramática normativa formulada desde arriba y la gramática intuitiva utilizada por los hablantes “de abajo” plantea algunas cuestiones importantes en el ámbito de la enseñanza de segundas lenguas. Aunque la enseñanza gramatical ocupa un lugar central en la mayoría de los programas didácticos, no siempre se reflexiona con la debida profundidad sobre el tipo de gramática que se transmite, ni sobre las condiciones en que esa transmisión resulta más significativa para el aprendiz. En la práctica, el conocimiento gramatical suele transmitirse a través de tres géneros textuales diferentes: los manuales de gramática, los libros de texto y los diccionarios. Cada uno de ellos desempeña un papel específico dentro del proceso de aprendizaje, pero también presenta limitaciones notables, sobre todo a la hora de resolver dudas concretas surgidas durante el uso activo de la lengua, que es “el problema principal al que el aprendiz se enfrenta”, según Nomdedeu Rull (2023: 193). A continuación, se comentan brevemente las características de estos tres géneros, lo que permitirá introducir posteriormente una propuesta alternativa que parte de un enfoque diferente, más cercano a la realidad lingüística del aprendiz.

1. Los *manuales de gramática* tienen como objetivo principal presentar el sistema gramatical de forma ordenada y exhaustiva. Suelen estar estructurados de manera sistemática y ofrecen una visión panorámica de la lengua, lo que puede resultar útil para adquirir una comprensión general de sus reglas. Sin embargo, presentan ciertas limitaciones: a menudo son monolingües, no son fáciles de consultar debido a índices poco funcionales y rara vez explican los fenómenos gramaticales a nivel de las palabras individuales, que es precisamente donde suelen surgir muchas dudas.
2. Los *libros de texto* abordan algunos aspectos de la gramática a través de pequeñas notas explicativas integradas en las respectivas lecciones o a través de apéndices finales con mini-gramáticas parciales. Su ventaja es que están adaptados al nivel del aprendiz y acompañan su progresión a lo largo del proceso de enseñanza-aprendizaje. No obstante, tampoco están concebidos como herramientas de consulta eficientes y, al igual que los manuales, no suelen ofrecer explicaciones detalladas sobre el funcionamiento gramatical de palabras concretas, más allá de los ejemplos ilustrativos que proporcionan.
3. Los *diccionarios para aprendices*, en algunos casos, incluyen notas gramaticales sobre ciertas palabras, así como mini-gramáticas resumidas en los apéndices. Por su formato, están diseñados como obras de consulta, y por tanto representan teóricamente la herramienta idónea para tratar la gramática a nivel de palabra individual, como ya propuso y defendió Sweet (1899) en su famosa “guía para profesores y aprendices”. Sin embargo, en el caso del español, su tratamiento gramatical deja mucho que desear, motivo por el cual han sido objeto de críticas por parte de diversos lexicógrafos como Alvar Ezquerro (1982), Gutiérrez Cuadrado (1994), Bargalló Escrivà

(1999, 2001, 2010), Maldonado González (2003, 2012), Santamaría Pérez (2004), Nomdedeu Rull & Tarp (2018, 2024), Dam-Jensen & Tarp (2019), Nomdedeu Rull (2023) y Nomdedeu Rull & Barcroft (2023), entre otros.

Los tres géneros mencionados anteriormente y utilizados habitualmente para enseñar gramática en contextos de segunda lengua comparten un enfoque común: parten de una *concepción descendente* del aprendizaje y dominio de la gramática. Es decir, presentan esta última como un sistema previamente estructurado que debe transmitirse con la expectativa de que el aprendiz interiorice reglas abstractas y luego las aplique en situaciones comunicativas concretas.

La novedad de la era digital, que aún no se ha aprovechado plenamente, es que, gracias a las nuevas tecnologías, ahora es posible explicar la gramática en cualquier contexto en el que se usen las palabras, no solo en los textos preconcebidos de los libros de texto. En otras palabras, ahora es posible crear *un cuarto tipo de gramática*, de *orientación ascendente* —es decir, *de abajo arriba*— que parte de un problema concreto detectado en un texto escrito por el propio aprendiz y culmina en la formulación de una regla gramatical que lo explica de forma clara, concisa y contextualizada, tal y como reclaman Tarp & Gouws (2019).

### 3. Características básicas del nuevo tipo de gramática

Mediante una negación dialéctica, se pueden destacar, a partir de la discusión anterior, algunas características fundamentales que debe tener el nuevo tipo de gramática orientada al aprendiz de español como segunda lengua. Este modelo es distinto a los tres géneros tradicionales, tanto en su diseño como en su finalidad didáctica, y representa una elaboración ulterior de un modelo inicial propuesto por Li et al. (2024: 146):

- Se trata de una gramática *bilingüe*, redactada en la lengua materna del aprendiz, lo que permite una mayor claridad en la explicación de los fenómenos gramaticales y reduce la carga cognitiva asociada a la comprensión metalingüística en L2.
- Tiene una *única función lexicográfica*: ayudar al aprendiz durante la redacción de textos en español mediante un asistente de escritura digital. Su objetivo principal no es ofrecer una descripción general del sistema gramatical, sino facilitar una comprensión inmediata y contextualizada de los posibles errores que puedan surgir en el proceso de producción escrita.
- Es de *acceso fácil e inmediato*: el usuario puede consultar las explicaciones haciendo clic sobre las palabras o secuencias de texto marcadas como problemáticas por el modelo lingüístico basado en IA. Esta interfaz se integra de manera natural en el flujo de trabajo del aprendiz.
- No es *ni sistemática*, como los manuales de gramática o las mini-gramáticas de los libros de texto, *ni centrada en palabras individuales*, como las notas

explicativas de los libros de texto o los datos gramaticales de los diccionarios para aprendices. Su estructura se adapta a las necesidades concretas del usuario en tiempo real.

- Se compone de un *conjunto de explicaciones breves*, muchas de las cuales abordan problemas gramaticales comunes que afectan a grupos amplios de palabras o secuencias. Este conjunto funcional permite al aprendiz identificar patrones de uso y de error que van más allá del caso puntual.

Como se desprende de estos cinco enunciados, el nuevo tipo de gramática que integra la herramienta de escritura se construye sobre un conjunto de glosas que pueden visualizarse haciendo clic sobre una palabra o secuencia marcada por dicha herramienta para explicar problemas concretos en un texto escrito por un aprendiz del español. Esto plantea grandes exigencias en cuanto a su diseño e implica, según Li & Tarp (2024: 400-401), que las explicaciones deben redactarse como:

textos breves y didácticos, con lenguaje claro y sencillo, sin demasiada terminología especializada, que proporcionen la información más relevante sobre el problema lingüístico concreto y que estén estructurados de forma que al lector le resulte fácil hacerse una idea general y captar la esencia del problema.

Es obvio que las explicaciones deben tener un enfoque didáctico, pero los demás criterios merecen un comentario aparte, ya que son aspectos que también hay que tener en cuenta. El punto de partida es que el aprendiz está escribiendo un ensayo o un texto similar en una segunda lengua para comunicar alguna idea en clase o en su tiempo libre. Por su parte, el asistente de escritura señala los posibles problemas en el texto y propone correcciones, además de ofrecer algunas aclaraciones que pueden despertar el interés del aprendiz. Sin embargo, como el objetivo inmediato del aprendiz no es aprender gramática sino intentar comunicarse de manera comprensible y correcta, es evidente que las aclaraciones deben concebirse de forma que no distraigan su atención del contenido de la comunicación, por lo que deben ser cortas, transparentes y redactadas en un lenguaje directo y fácil de entender, sin extenderse en una explicación gramatical que vaya más allá del problema en cuestión. En otras palabras, la filosofía subyacente de las explicaciones es inducir la comprensión gramatical mediante el uso concreto del lenguaje, y no que el aprendiz luego pueda transmitirla en un lenguaje académico salpicado de términos especializados.

#### 4. Tareas lexicográficas y usos de la IA en el proyecto

La lexicografía, con su larga tradición de centrarse en las necesidades de información de los usuarios y crear productos que respondan a dichas necesidades, puede jugar sin duda un papel muy relevante en la creación de asistentes de escritura y productos similares que se beneficien de la IA y otras tecnologías digitales. De hecho, en el proyecto en

curso esta disciplina desempeña un papel más o menos importante en la ejecución de las siguientes ocho tareas: 1) diseño general del asistente, 2) composición de un corpus para su entrenamiento, 3) aportación de datos sintéticos para el mismo fin, 4) generación de datos de validación para medir su rendimiento, 5) determinación de los tipos de problemas que deben explicarse, 6) generación de las glosas largas, 7) diseño de la plantilla para las glosas cortas y 8) ensayo del asistente de escritura con usuarios reales después de su lanzamiento.

Hasta el momento, la IA “normal” se ha integrado en el modelo lingüístico de soporte (tarea que corresponde a los informáticos), mientras que la IA generativa se utiliza para resolver las tareas 2, 4, 5 y 6 por parte de los lexicógrafos. Además, se prevé que también se use en el futuro para aportar datos a la plantilla incluida en la tarea 7, como se discutirá más adelante. Huete García & Tarp (2024) y Tarp & Nomdedeu Rull (2024) ya han tratado exhaustivamente el uso de la IA generativa para crear un corpus de entrenamiento (tarea 2) y los datos de validación (tarea 4), respectivamente, por lo que nos centramos a continuación en las tareas 5 y 6, es decir, la determinación de los problemas gramaticales que merecen explicaciones y la generación de estas últimas.

A este respecto, Li et al. (2024) documentan minuciosamente la ausencia de corpus y manuales de gramática en español que ofrezcan o permitan obtener una lista de los problemas gramaticales más frecuentes que deben explicarse en un asistente de escritura como el que se discute en este artículo. Existen efectivamente algunos corpus con errores cometidos por estudiantes de español como segunda lengua, como el COWS-L2H, pero este corpus, que es el más útil dentro de lo que cabe, no categoriza los errores más allá de las siete categorías principales y tampoco permite calcular su frecuencia, cf. Hernández Muñoz et al. (2027). Es decir, que los corpus y manuales existentes, a pesar de sus múltiples usos y aplicaciones, no han previsto un tipo de gramática para estudiantes de español como el que se está desarrollando en el marco del proyecto en curso.

Por eso, los tres autores tuvieron que diseñar su propia metodología para obtener la información requerida, ya que tampoco confiaban en la introspección a pesar de su dilatada experiencia en la enseñanza del español como lengua segunda. Para dar este salto hacia lo desconocido, recurrieron a la ayuda de la IA generativa, no para que esta hiciera el trabajo por sí sola, sino para que les sirviera de inspiración y lograran elaborar conjuntamente listas de problemas relevantes en una fructífera interacción entre dos tipos diferentes de inteligencia que los autores describen como una simbiosis. Basándose en esta experiencia y en otra similar en su propio proyecto, Li & Tarp (2024: 403) sintetizaron la siguiente metodología para detectar categorías y subcategorías de errores relevantes que requieren explicaciones en asistentes de escritura para aprendices de L2:

- En primer lugar, le pedimos al chatbot que nos diera una lista de tipos de error relevantes, lo que hizo sorprendentemente bien.
- A continuación, pulsamos el botón de regeneración para ver si nos daba sugerencias más útiles, lo que ocurrió en la mayoría de los casos. Seguimos haciéndolo hasta que empezó a repetirse y no apareció nada nuevo.

- Si el resultado no era satisfactorio, también probamos a modificar o reformular el prompt para mejorar la respuesta.
- El siguiente paso fue utilizar nuestros propios conocimientos gramaticales y nuestra experiencia docente para añadir algunos tipos de error o dividir en dos algunos de los que nos proporcionaba el chatbot.
- Para comprobarlo, consultamos el corpus para ver si realmente contenía errores pertenecientes a todas las subcategorías propuestas. Si no era así, ignorábamos la subcategoría por el momento.
- Por último, refinamos los términos utilizados para describir las distintas subcategorías, ya que el chatbot no era coherente en este sentido.

Como ejemplo, y aplicando esta metodología, Li et al. (2024: 151) consiguieron elaborar la siguiente lista de 25 subcategorías de errores dentro de la categoría general de desacuerdo de género:

### **1. Sustantivo y adjetivo**

- 1.1. Desacuerdo entre sustantivo y adjetivo pospuesto
- 1.2. Desacuerdo entre sustantivo singular y adjetivo pospuesto no variable
- 1.3. Desacuerdo entre sustantivo singular y adjetivo antepuesto
- 1.4. Desacuerdo entre sustantivo coordinado del mismo género y adjetivo
- 1.5. Desacuerdo entre sustantivos coordinados de distinto género y adjetivo
- 1.6. Desacuerdo entre adjetivo coordinado singular y sustantivo plural

### **2. Sustantivo y determinante**

- 2.1. Desacuerdo entre artículo y sustantivo
- 2.2. Desacuerdo entre artículo masculino singular y sustantivo femenino
- 2.3. Desacuerdo entre sustantivo coordinado y artículo
- 2.4. Desacuerdo entre sustantivo y determinante demostrativo
- 2.5. Desacuerdo entre sustantivo y determinante definido
- 2.6. Desacuerdo entre sustantivo y determinante indefinido
- 2.7. Desacuerdo entre sustantivo y determinante posesivo

### **3. Sustantivo y pronombre**

- 3.1. Desacuerdo entre sustantivo y pronombre personal átono
- 3.2. Desacuerdo entre antecedente y pronombre posesivo
- 3.3. Desacuerdo entre antecedente y pronombre relativo

### **4. Sustantivo y participio**

- 4.1. Desacuerdo entre sustantivo y participio

### **5. Pronombre y adjetivo**

- 5.1. Desacuerdo entre pronombre personal y adjetivo
- 5.2. Desacuerdo con adjetivo después de pronombres *alguien* y *nadie*

### **6. Adverbio y adjetivo/adverbio**

- 6.1. Desacuerdo con adjetivo tras cuantificador adverbio medio

6.2. Desacuerdo con adjetivo/adverbio tras adverbios indefinidos *algo* y *nada*

## 7. Número y sustantivo

7.1. Desacuerdo con sustantivo después de cardinal o decimal

7.2. Desacuerdo con cardinal terminado en *un* o *una* antes de sustantivo

7.3. Desacuerdo con construcciones partitivas como '*uno de los* + sustantivo'

## 8. Adjetivo y sujeto en cláusula con *que*

8.1. Desacuerdo en construcciones con '*lo* + adjetivo + cláusula con *que*'

Como se puede ver, no se trata de una lista típica desde el punto de vista lingüístico. Sus características específicas se deben a que se crea a partir de los tipos de errores que pueden observarse en los textos escritos; es decir, es una lista que sirve de base para una gramática ascendente. Además, es de uso exclusivamente interno, de modo que los informáticos que no estén familiarizados con la terminología lingüística pueden utilizarla para codificar los distintos tipos de errores y vincular las correspondientes explicaciones al modelo lingüístico que soporta la herramienta de escritura. En el siguiente apartado discutimos cómo se generan estas explicaciones.

## 5. Generando explicaciones (glosas) largas

No cabe duda de que un profesor de español como lengua extranjera con años de experiencia conoce en profundidad los errores que cometen sus estudiantes y puede, por lo general, explicarlos fácilmente en clase. Sin embargo, es mucho más difícil escribir una explicación concisa que aborde el meollo de la cuestión y que sea fácil de comprender para los estudiantes. Además de seleccionar los aspectos clave que deben tratarse, hay que determinar la estructura y el estilo más adecuados para los estudiantes, lo que requiere una profunda reflexión previa que, incluso para profesores de idiomas y lexicógrafos experimentados, suele llevar tiempo. Por ello, merecía la pena explorar si la IA generativa podría ayudar a acortar el proceso y evitar tener que escribir las explicaciones desde cero.

Ya al generar el material de validación, el chatbot con el que se trabajaba había ofrecido a veces explicaciones del problema en cuestión que no se le habían solicitado. Algunas de estas explicaciones tenían una estructura sencilla e inspiradora sobre la que se podía trabajar para crear el modelo de explicaciones deseado en términos de diseño, estilo y estructura de argumentación. Tras algunos experimentos, se consiguió diseñar un prompt que produjo el resultado deseado: una explicación de cada problema gramatical particular que, en la mayoría de los casos, solo requería pequeños ajustes para cumplir su propósito. La metodología aplicada para realizar estos ajustes consta de seis pasos, resumidos por Li & Tarp (2024: 410-411):

1. Leer rápidamente las explicaciones generadas por la IA.
2. Eliminar algunos ruidos de fondo del texto para no desviar la atención del estudiante del punto principal.

3. Cambiar algunas palabras y frases para mejorar la legibilidad.
4. Seleccionar datos relevantes de los ejemplos y añadirlos a la explicación inicial si es inevitable utilizar cierta terminología técnica.
5. Perfeccionar la estructura argumentativa de la explicación de los ejemplos para que esta sección sea más lógica, directa y concisa.
6. Mejorar el diseño general para que el lector capte rápidamente la esencia del problema tratado.

La Figura 1 muestra un ejemplo de esta clase de explicación “larga” tras de su edición por parte de los lexicógrafos (para mayor conveniencia de los lectores, la explicación de este y demás ejemplos está en español).

Como se desprende de la Figura 1, la estructura general de la explicación consiste en dar primero una breve regla gramatical, mostrar después ejemplos incorrectos y correctos del problema gramatical en cuestión y, por último, explicar estos ejemplos. A veces, como en este caso, si se considera pertinente, también se añade una nota gramatical de lectura opcional, lo que hace que la explicación sea un poco más extensa. Esta

Figura 1. Explicación de desacuerdo entre sustantivo singular y adjetivo antepuesto

Algunos pocos adjetivos, como ‘bueno’ o ‘malo’, cambian de forma cuando se anteponen a un sustantivo singular, como ‘año’ o ‘día’.

De este modo, los adjetivos como ‘bueno’ y ‘malo’ se escriben ‘buen’ y ‘mal’ cuando se ponen delante del sustantivo al que modifican, p. ej. ‘año’ y ‘día’.

**Ejemplos incorrectos:**

- Te deseo un bueno año.
- Juan ha tenido un malo día.

**Ejemplos correctos:**

- Te deseo un buen año.
- Juan ha tenido un mal día.

**Explicación:**

En los ejemplos incorrectos, los adjetivos ‘bueno’ y ‘malo’ presentan la terminación típica del género masculino en español con la vocal ‘-o’.

Sin embargo, en este caso, al colocarse delante de un sustantivo masculino singular como ‘año’ y ‘día’, deben cambiarse a ‘buen’ y ‘mal’, como se ve en los ejemplos correctos.

**Nota:**

Las formas adjetivas ‘buen’ y ‘mal’ sólo se utilizan delante de sustantivos masculinos singulares. Nunca se usan delante de sustantivos femeninos, como ‘amiga’, o plurales, como ‘amigos’ o ‘amigas’.

Otros adjetivos que también cambian de forma delante de sustantivos masculinos singulares son ‘primero’ y ‘tercero’, que deben cambiarse por ‘primer’ y ‘tercer’, respectivamente.

**MENOS**

estructura de argumentación parece lógica, sencilla y fácil de seguir, por lo que se ha elegido como la más adecuada para los estudiantes.

También se nota el uso de algunos términos técnicos como *sustantivo*, *adjetivo*, *masculino*, *femenino*, *singular* y *plural*, sin los cuales sería muy difícil explicar la regla gramatical. Aunque se supone que la mayoría de los estudiantes de español conocen estos términos (al ser bastante básicos y frecuentes), se ha hecho un esfuerzo para relacionarlos con las palabras clave que aparecen en los ejemplos –en este caso, *buen* y *bueno*, *mal* y *malo*, además de *año* y *día*–, con el fin didáctico de ejemplificar su significado y facilitar la comprensión. La inclusión de estos detalles, que también se han generalizado en las demás explicaciones, forma parte del trabajo de redacción de los lexicógrafos, al igual que el uso de negrita, colores y subrayados para optimizar el diseño y su dimensión didáctica. La idea es ir escribiendo explicaciones como esta para todas las subcategorías de errores identificadas con la metodología descrita en el apartado 4.

## 6. Generando explicaciones (glosas) cortas

Además de la explicación larga, también se debe hacer una glosa corta para visualizarse junto con la sugerencia alternativa cuando el usuario hace clic en las palabras problemáticas marcadas por el asistente de escritura. Su propósito difiere del de la explicación larga y debe entenderse en el marco de los dos tipos tradicionales de aprendizaje, el incidental y el intencional, ya que el primero no solo tiene que ver con la lectura como se discute en la mayor parte de la literatura académica, entre otros Krashen (1989), Hulstijn (2013) y Leow & Zamora (2017), sino también con la escritura de textos como argumenta Graham (2020), entre otros.

A este respecto, Tarp (2022) presenta una nueva definición de los dos conceptos desde una perspectiva lexicográfica *centrada en el flujo del proceso de escritura* (o lectura cuando se revisa el texto producido). Si la ayuda que ofrece la lexicografía no interrumpe este flujo y desvía la atención del aprendiz del contenido del texto, se permite e incluso refuerza la posibilidad de aprender de forma incidental. Esta ayuda “discreta” es el objetivo principal de la explicación corta, mientras que la consulta de la explicación larga necesariamente interrumpe el flujo y requiere que el aprendiz se concentre en su contenido, en este caso la regla gramatical, por lo que pasa al aprendizaje intencional, es decir, el deseo de aprender algo nuevo.

De lo anterior se desprende que la redacción de la explicación corta requiere una cuidadosa reflexión en cuanto a contenido, estilo y extensión. Originalmente, como discutía Tarp (2023: 120-121), estaba previsto escribir explicaciones cortas similares a las de Grammarly, ProWritingAid y LanguageTool, que hacen uso genérico de expresiones aproximadoras como *puede que*, *posible* y *parece* [*may be*, *possible* y *seems*], al estilo de *Puede que la palabra X se utilice incorrectamente* [*The word X may be used incorrectly*], citada de Grammarly. Sin embargo, el proyecto en curso no solo consiste en producir un asistente de escritura digital que se está desarrollando de forma bastante satisfactoria,



en pruebas, confianza que el chatbot viene creando al demostrar un altísimo grado de precisión en la mayoría de los ensayos. Indudablemente, esto también supondría entrenar ulteriormente el modelo lingüístico para que consiga identificar los errores en los textos escritos por aprendices con una tasa de aciertos cercana al cien por cien.

El segundo reto consiste en que, con esta técnica, se le permitiría al chatbot acceder al texto completo del usuario o a parte de él. Sin embargo, un acceso incontrolado podría suponer un grave problema de seguridad y una amenaza para la privacidad, ya que nunca se sabe en qué manos pueden acabar estos datos, sobre todo si se usan chatbots como ChatGPT, Claude, Gemini o Grok. Ante esta situación, la mejor alternativa sería desarrollar una IA segura propia en la que los datos creados por los usuarios no se usarían para entrenar otro modelo lingüístico, ni saldrían nunca de los servidores de la empresa o institución que ofrece el asistente de escritura y se autodestruirían al final. Aunque esta solución excede con creces el ámbito de la lexicografía e incluye importantes aspectos de ética, es perfectamente posible si hay voluntad, según los informáticos, lo que sugiere que la idea desarrollada en este apartado es viable y ofrece una nueva perspectiva para la enseñanza de lenguas, que se discutirá a continuación.

## 7. Combinando glosas en una perspectiva hegeliana

El asistente de escritura en construcción funcionará de la siguiente manera si se aplica la idea discutida en el apartado anterior: cuando un aprendiz de español comete un error gramatical, el editor de texto lo marca y ofrece al instante una pequeña glosa (una o más palabras) que constituye una sugerencia alternativa. Si el error no es trivial, el asistente también proporciona otra glosa digital que explica resumidamente en qué consiste el problema, así como una opción para acceder a una glosa más larga que explica la regla gramatical general implicada. La Figura 3 muestra los tres tipos de glosas activados juntos (en la versión final, la glosa o explicación larga reemplazará a la corta cuando se visualice).

Al llegar a este punto, quedó claro para los lexicógrafos que, sin pretenderlo, habían creado algo que podía considerarse una metodología hegeliana invertida. En varios de sus libros sobre dialéctica e historia, el gran filósofo alemán Hegel (1770-1831) discutía sobre cómo el “espíritu” (Geist), que para él era el punto de partida, se iba materializando en la sociedad humana hasta expresarse en el Estado prusiano de la primera mitad del siglo XIX. Para ello, introducía y manejaba con gran talento los tres conceptos dialécticos de *lo universal*, *lo particular* y *lo individual*, siendo lo particular el mediador entre lo universal (el espíritu) y lo individual (la sociedad) para que lo universal se materializara gradualmente en lo individual, configurándose así el devenir histórico (cf. Hegel, 1807, 1821).

Ahora bien, al reflexionar profundamente sobre el resultado mostrado en la Figura 3, surge fácilmente la idea de relacionar los conceptos de lo individual, lo particular y lo universal con los tres tipos de glosa: 1) la sugerencia alternativa, 2) la explicación corta y 3) la explicación larga, respectivamente. Sin embargo, a diferencia del método hegeliano, en el que lo particular media entre lo universal y lo individual para que lo universal se materialice gradualmente en lo individual, en este caso se parte de lo individual



sustantivo *momento* combinado con *malo*, media o hace de puente entre el error y la regla gramatical. Además, en otros casos, ninguna de las palabras incorrectas escritas por el aprendiz sería idéntica a las palabras subrayadas en los ejemplos para ilustrar el problema, por lo que la distancia entre error y explicación sería aún mayor y resultaría más necesario contar con un puente que las conectara.

En cualquier caso, al aplicar la metodología dialéctica de Hegel al revés, se consigue el objetivo de inducir la gramática a través del uso práctico de la lengua, lo que resulta una forma más natural de aprender gramática que leer y memorizar primero las reglas para luego aplicarlas a los textos escritos. En este sentido, la nueva gramática también se ha puesto al revés en comparación con las gramáticas tradicionales basadas en reglas, con los dos pies firmemente plantados en el lenguaje escrito, que a menudo es problemático y propenso a errores, que los aprendices cometen en mayor o menor medida en su proceso de aprendizaje de la L2.

## 8. Conclusiones y perspectivas

La discusión desarrollada en los apartados anteriores corrobora inequívocamente la idea de que las nuevas tecnologías digitales, y en particular la inteligencia artificial generativa, conllevan no solo nuevos métodos de trabajo, sino también productos lexicográficos totalmente nuevos. Estos avances se basan en una distinción clave entre diccionografía y glosografía como las dos ramas principales de la lexicografía, tanto en su dimensión disciplinar como en su práctica cultural.

Se ha mostrado cómo las nuevas glosas digitales pueden integrarse en un asistente de escritura respaldado por un modelo lingüístico de IA, lo que da lugar a un nuevo tipo de gramática de L2 de abajo arriba. Asimismo, se ha subrayado la importancia de que los lexicógrafos estén atentos al constante flujo de innovaciones tecnológicas y a las posibilidades que estas ofrecen para situar a la disciplina a la altura de las nuevas necesidades que van surgiendo entre sus usuarios. Finalmente, se pone de manifiesto que la nueva era digital requiere una estrecha colaboración interdisciplinaria entre lexicógrafos, informáticos y otros expertos para desarrollar los nuevos productos.

Las nuevas tecnologías no están aquí para sustituir a los lexicógrafos, sino para ayudarlos a hacer su trabajo aún mejor e inspirarlos para concebir productos que satisfagan a sus usuarios de forma más adecuada y adaptada a la época actual. La experiencia, competencia y sensibilidad de los lexicógrafos y otros expertos siguen siendo imprescindibles para garantizar un buen resultado.

Una vez completado, el asistente de escritura mejorará sin duda la calidad de los textos escritos por los aprendices de español como segunda lengua. Sin embargo, aunque esta redacción es suficiente para resolver determinadas tareas relacionadas con la enseñanza del español, los aprendices deben aspirar a más. El gran reto de la nueva era de la IA es cómo motivar a los aprendices para que desarrollen habilidades, sentido crítico y creatividad lingüística más allá de la mera redacción. Este es el propósito con el que se ha concebido y desarrollado el nuevo tipo de gramática de L2. Solo el tiempo, las

futuras pruebas con usuarios reales y la investigación complementaria dirán si tendrá éxito. El reto está lanzado.

## Referencias

- AL-KASIMI, Ali (1977). *Linguistic and Bilingual Dictionaries*. Leiden: EJ Brill Publishers.
- ALVAR EZQUERRA, M. (1982). Diccionario y gramática. *Lingüística Española Actual*, IV, 1982, 151-207.
- BARGALLÓ ESCRIVÀ, Maria (1999). La información gramatical en los diccionarios didácticos monolingües del español. En M. Neus VILA RUBIO, M. Á. CALERO FERNÁNDEZ, Rosa M. MATEU, Montse CASANOVAS CATALÁ & José Luis ORDUÑA LÓPEZ (eds.) (1999). *Así son los diccionarios*. Lleida: Edicions de la Universitat de Lleida, 15-42.
- BARGALLÓ ESCRIVÀ, Maria (2001). La información gramatical en los diccionarios especializados. En Maria BARGALLÓ ESCRIVÀ, Esther FORGAS BERDET, Cecilio GARRIGA ESCRIBANO, Ana RUBIO & Johannes Schnitzer (eds.) (2001). *Las lenguas de especialidad y su didáctica*. Tarragona: Universitat Rovira i Virgili, 81-90.
- BARGALLÓ ESCRIVÀ, Maria (2010). La información gramatical explícita en los diccionarios monolingües del español. En Elisenda BERNAL, Sergi TORNER & Janet DECESARIS (eds.) (2010). *Estudis de Lexicografia 2003-2005*. Barcelona: IULA-UPF, 17-37.
- BENATI, Chiara & HÄNDL, Claudia (eds.) (2019). *From Glosses to Dictionaries: The Beginnings of Lexicography*. Newcastle upon Tyne: Cambridge Scholars Publishing.
- COWS-L2H. *Corpus of Written Spanish of L2 and Heritage Speakers*. <https://github.com/ucdaviscl/cowsl2h>
- DAM-JENSEN, Helle & TARP, Sven (2019). El tratamiento de adjetivos en gramáticas y diccionarios de español. *Revista Internacional de Lenguas Extranjeras*, 10, 13-47. <https://doi.org/10.17345/rile10.13-47>
- GOUWS, Rufus H. & TARP, Sven (2024). Despite Current Challenges: Lexicography has a bright future. *Lexicographica*, 40, 187-212. <https://doi.org/10.1515/lex-2024-0009>
- GRAHAM, Steve (2020). The Sciences of Reading and Writing Must Become More Fully Integrated. *Reading Research Quarterly*, 55(S1): 535-544.
- GUTIÉRREZ CUADRADO, Juan (1994). Gramática y diccionario. *Actas del Congreso de la lengua española*. Madrid: Instituto Cervantes, 637-656.
- HANKS, Patrick (2013). Lexicography from Earliest Times to the Present. En Keith ALLAN (ed.) (2013). *The Oxford Handbook of the History of Linguistics*. Oxford: Oxford University Press, 503-536. <http://dx.doi.org/10.1093/oxfordhb/9780199585847.013.0023>
- HEGEL, G. W. F. (1807/2010). *Fenomenología del espíritu*. Madrid: Fondo de Cultura Económica.
- HEGEL, G. W. F. (1821/2001). *Principios de la filosofía del derecho*. Madrid: Alianza Editorial.
- HERNÁNDEZ MUÑOZ, Natividad, TOMÉ CORNEJO, Carmela, BELÍO-APAOLAZA, Helena Sofía, RECIO DIEGO, Álvaro & DAVIDSON, S. (2027). Anotación de errores

- gramaticales en el corpus de aprendices de español COWS-L2H. *Revista Española de Lingüística Aplicada/Spanish Journal of Applied Linguistics*, 40(1), (en prensa).
- HUETE-GARCÍA, Ángel & TARP Sven (2024). Training an AI-based Writing Assistant for Spanish Learners: the Usefulness of Chatbots and the Indispensability of Human-assisted Intelligence. *Lexikos*, 34, 21-40. <https://doi.org/10.5788/34-1-1862>
- HULSTIJS, J. H. (2013). Incidental Learning in Second Language Acquisition. En A.C. Chapelle (ed.) (2013). *The Encyclopedia of Applied Linguistics*. Nueva York: Wiley-Blackwell, 2632-2637.
- KRASHEN, Stephen (1989). We Acquire Vocabulary and Spelling by Reading: Additional Evidence for the Input Hypothesis. *Modern Language Journal*, 73. 440-464.
- LEOW, Ronald & ZAMORA, Celia (2017). Intentional and Incidental L2 Learning. En S. Loewen & M. Sato (eds.) (2017). *The Routledge Handbook of Instructed Second Language Acquisition*. Nueva York: Routledge, 33-49.
- LI, Qian & TARP, Sven (2024). Using Generative AI to Provide High-quality Lexicographic Assistance to Chinese Learners of English. *Lexikos*, 34, 397-418. <https://doi.org/10.5788/34-1-1944>
- LI, Qian, TARP, Sven & NOMDEDEU RULL, Antoni (2024). The Necessary Symbiosis: How ChatGPT Co-authored a New Type of Learner's Grammar. *Círculo de lingüística aplicada a la comunicación*, 100: 145-160. <https://doi.org/10.5209/clac.97409>
- MALDONADO GONZÁLEZ, Concepción (2003). La lexicografía didáctica monolingüe en español. En M. A. MARTÍN-ZORRAQUINO & J. L. ALIAGA-JIMÉNEZ (eds.) (2003). *La lexicografía hispánica ante el siglo XXI. Balance y perspectivas*. Zaragoza: Sansueña Industrias Gráficas, 129-150.
- MALDONADO GONZÁLEZ, Concepción (2012). Los diccionarios en el mundo ELE: ayer, hoy y mañana (una reflexión desde la propia experiencia). *Revista Internacional de Lenguas Extranjeras*, 1, 151-179. <https://doi.org/10.17345/rile1.11>
- MCCARTHUR, Tom (1986). *Worlds of Reference. Lexicography, Learning and Language from the Clay Tablet to the Computer*. Cambridge: Cambridge University Press.
- NOMDEDEU RULL, Antoni (2023). Sin noticias aún de datos gramaticales útiles en diccionarios para aprendices. En Cecilio GARRIGA ESCRIBANO, Sandra IGLESIA MARTÍN, José Antonio MORENO VILLANUEVA & Antoni NOMDEDEU RULL (eds.) (2023). *Lligams: Texts dedicats a María Bargalló Escrivà*. Tarragona: Publicacions URV, 191-201. <https://llibres.urv.cat/index.php/purv/catalog/download/544/565/1245-1?inline=1>
- NOMDEDEU RULL, Antoni & BARCROFT, Joe (2023). Los diccionarios de ELE. En Sergi TORNER, Paz BATTANER & Irene RENAU (eds.) (2023). *Lexicografía hispánica / The Routledge Handbook of Spanish Lexicography*. Londres y Nueva York: Routledge, 448-466.
- NOMDEDEU RULL, Antoni & TARP, Sven (2018). Hacia un modelo de diccionario en línea para aprendices de español como LE/L2. *Journal of Spanish Language Teaching*, 5(1), 50-65. <https://doi.org/10.1080/23247797.2018.1459279>
- NOMDEDEU RULL, Antoni & TARP, Sven (2024). *Introducción a la Lexicografía en Español: Funciones y Aplicaciones*. Londres: Routledge.

- RUIZ ASENCIO, José Manuel, RUIZ ALBI, Irene & HERRERO JIMÉNEZ, Mauricio (2020). *Las Glosas Silenses. Estudio crítico y edición facsímil. Versión castellana del Penitencial*. Burgos: Instituto Castellano y Leonés de la Lengua.
- SANTAMARÍA PÉREZ, María Isabel (2004). La información gramatical en los diccionarios monolingües de aprendizaje: reflexiones a partir del verbo. *ELUA: Estudios De Lingüística. Universidad De Alicante*, (Anexo 2), 559-580. <https://doi.org/10.14198/ELUA2004.Anexo2.27>
- SWEET, Henry (1899). *The Practical Study of Languages: A Guide for Teachers and Learners*. Londres: Oxford University Press. <https://archive.org/details/practicalstudyof00sweeuoft/page/n5/mode/2up>
- TARP, Sven (2022). A Lexicographical Perspective to Intentional and Incidental Learning: Approaching an Old Question from a New Angle. *Lexikos*, 32(2), 203-222. <https://doi.org/10.5788/32-2-1703>
- TARP, Sven (2023). Eppur si muove: Lexicography Is Becoming Intelligent! *Lexikos*, 33(2), 107-131. <https://doi.org/10.5788/33-2-1841>
- TARP, Sven & H. GOUWS, R. (2019). Lexicographical Contextualization and Personalization: A New Perspective. *Lexikos*, 29, 250-268. <https://doi.org/10.5788/29-1-1520>
- TARP, S. & H. GOUWS, Rufus (2023). A Necessary Redefinition of Lexicography in the Digital Age: Glossography, Dictionography and the Implications for the Future. *Lexikos*, 33, 425-447. <https://doi.org/10.5788/33-1-1826>
- TARP, Sven & NOMDEDEU RULL, Antoni (2024). Who Has the Last Word? Lessons from Using ChatGPT to Develop an AI-based Spanish Writing Assistant. *Círculo de lingüística aplicada a la comunicación*, 97, 309-321. <https://doi.org/10.5209/clac.91985>