

# Hacia una traducción automática inclusiva: la intersección entre inteligencia artificial, terminología LGTBIQ+ y sesgo de género

## Towards Inclusive Machine Translation: the Intersection of Artificial Intelligence, LGBTQIA+ Terminology, and Gender Bias

### Resumen

La traducción automática (TA) y los chatbots de inteligencia artificial (IA) generativa han transformado la comunicación global al facilitar la transmisión de información entre lenguas y, por ende, entre culturas. No obstante, también plantean desafíos éticos debido a los sesgos lingüísticos. En particular, estos sesgos afectan negativamente a la terminología que representa a las mujeres y al colectivo LGTBIQ+ en las traducciones generadas por estas tecnologías. Partimos de la hipótesis de que tanto las tecnologías de traducción automática como los chatbots de inteligencia artificial presentan dificultades para traducir correctamente los marcadores de género y la terminología LGTBIQ+ del inglés al español. En este sentido, con frecuencia optan por el masculino genérico cuando no se proporciona suficiente información contextual o por términos inadecuados, aunque se espera que los chatbots más recientes ofrezcan mejores resultados en este aspecto. Para validar dicha hipótesis, hemos diseñado una metodología de análisis basada en la recogida de datos cuantitativos y cualitativos a partir de traducciones generadas por sistemas de traducción automática convencionales (DeepL y Google Translate) y chatbots de inteligencia artificial como ChatGPT y Gemini. Para evaluar los datos se ha empleado una adaptación de Multidimensional Quality Metrics (MQM), que permitirá obtener un marco estandarizado para medir la calidad de traducciones. Los resultados que se desprenden del análisis muestran la persistencia de un sesgo marcado hacia el género masculino, con una identificación inconsistente del género femenino. De este modo, se puede concluir que el producto resultante de los sistemas basados en inteligencia artificial generativa no presenta mejoras significativas en comparación con los sistemas de traducción automática convencionales. Por ello, es necesario desarrollar tecnologías lingüísticas más inclusivas, equitativas y libres de sesgos, así como fomentar el diseño de sistemas más justos y respetuosos con la diversidad, esenciales para responder a los desafíos de un contexto cada vez más interconectado y globalizado.

### Autoría

**ELENA HERNÁNDEZ FRESNO**  
Universidad de Valladolid, España  
[elena.hernandez.fresno@uva.es](mailto:elena.hernandez.fresno@uva.es)  
<https://orcid.org/0009-0002-4890-3930>

**MARÍA TERESA ORTEGO ANTÓN**  
Universidad de Valladolid, España  
[mariateresa.ortego@uva.es](mailto:mariateresa.ortego@uva.es)  
<https://orcid.org/0000-0003-3379-4700>

### Para citar este artículo:

Hernández Fresno, E.; Ortego Antón, M.T. (2025). Hacia una traducción automática inclusiva: la intersección entre inteligencia artificial, terminología LGTBIQ+ y sesgo de género, *ELUA*, 44, 125-145.  
<https://doi.org/10.14198/ELUA.29144>

Recibido: 28/01/2025  
Aceptado: 10/03/2025

Conflicto de intereses: las autoras declaran que no hay conflicto de intereses.

© 2025 Elena Hernández Fresno  
María Teresa Ortego Antón



Licencia: este trabajo se comparte bajo la licencia de Atribución-NoComercial-CompartirIgual 4.0 Internacional de Creative Commons (CC BY-NC-SA 4.0): <https://creativecommons.org/licenses/by-nc-sa/4.0/>

**Palabras clave:**

traducción automática neuronal; chatbot; sesgo de género; LGTBIQ+; inteligencia artificial; inglés; español; lenguaje inclusivo; DeepL; Google Translate; ChatGPT; Gemini.

**Abstract**

Machine translation (MT) and generative artificial intelligence (AI) chatbots have revolutionized global communication as they facilitate seamless cross-language and rich cross-cultural information transmission. However, they also raise complex ethical challenges due to underlying linguistic and sociocultural biases. These biases negatively impact the accurate and respectful terminology representing women and the LGBTQIA+ community in translations generated by these advanced technological systems. We hypothesize that Machine Translation technologies and Artificial Intelligence chatbots face significant and persistent challenges in accurately translating nuanced linguistic gender markers and culturally specific LGBTQIA+ terminology from English to Spanish. These tools often default to the generic masculine form when contextual information is insufficient or vague, or resort to inadequate, outdated, or inappropriate terms; however, the latest generation of sophisticated chatbots is expected to yield improved and more inclusive results in this sensitive area. To validate the hypothesis, a comprehensive and systematic methodology of analysis has been designed based on rigorous quantitative and in-depth qualitative data collection derived from translations generated by conventional Machine Translation systems (DeepL and Google Translate) and advanced AI chatbots such as ChatGPT and Gemini. This detailed analysis aims to evaluate how these tools handle inclusive and non-discriminatory terminology and their accuracy in faithfully reflecting gender diversity. Multidimensional Quality Metrics (MQM) will be used for the evaluation to provide a standardized and widely recognized framework for objectively measuring translation quality. The findings from the comparative analysis reveal a persistent and problematic bias towards the male gender and inconsistent identification of the female gender. These issues lead to the conclusion that, despite using generative artificial intelligence, no significant or meaningful improvements in translation quality are observed compared to conventional Machine Translation systems. Therefore, it is essential to develop more inclusive, equitable, and unbiased linguistic technologies while actively promoting the design of fairer and more inclusive systems that respect human diversity—both crucial for addressing the urgent challenges of an increasingly interconnected and globalized world.

**Keywords:**

neural machine translation; chatbot; gender bias; LGBTQIA+; artificial intelligence; English; Spanish; inclusive language; DeepL; Google Translate.; ChatGPT; Gemini.

**1. INTRODUCCIÓN**

La traducción automática (TA) se ha consolidado como una herramienta imprescindible para la comunicación global, puesto que permite transmitir de forma rápida y eficiente información entre culturas y lenguas. Prueba de ello son los numerosos estudios realizados en la última década, entre los que destacan aquellos enfocados al producto resultante de sistemas de traducción automática y a su posesición (Koponen y Salmi 2017). Además, el auge de las nuevas tecnologías también ha dado como resultado la aparición de nuevos modelos del lenguaje, integrados en los chatbots de inteligencia artificial (IA) generativa. El trasvase interlingüístico no ha sido ajeno al creciente uso de herramientas como ChatGPT o Gemini, pues también funcionan como traductores automáticos. No obstante, este tipo de tecnología es considerablemente reciente y está en constante cambio, por lo que no hay una gran cantidad de estudios que combinen los chatbots de IA generativa y los sistemas de traducción automática convencionales, aunque este número esté aumentando en la actualidad.

Si bien es un hecho que estas tecnologías están integradas en la sociedad, es fundamental que la traducción automática se adapte a las realidades sociales actuales. En este sentido, resulta particularmente preocupante la presencia de sesgos de género en estas herramientas, puesto que pueden afectar negativamente la representación de grupos históricamente marginados. En los últimos años, han surgido estudios en los que se ha abordado el sesgo de género en la tecnología y, concretamente, en sistemas de traducción automática como Google Translate (Prates *et al.* 2019; Ullman 2022) o chatbots de inteligencia artificial generativa (Ghosh y Caliskan 2023). Así pues, es fundamental tanto abordar los sesgos lingüísticos como concienciar a la sociedad sobre la importancia del lenguaje inclusivo para promover la igualdad y potenciar la representación de las mujeres y del colectivo LGTBIQ+.

En consecuencia, en este trabajo pretendemos evaluar la capacidad de los sistemas de TA y de los chatbots de inteligencia artificial para manejar la terminología especializada y del colectivo LGTBIQ+, así como detectar posibles sesgos de género presentes en el producto resultante. Partimos de la hipótesis de que los modelos o chatbots y los sistemas de traducción automática muestran dificultades a la hora de trasvasar interlingüísticamente los marcadores de género del inglés al español, especialmente si no existe información contextual que pueda ayudar a la identificación del mismo. Por ello, en la mayoría de los casos es probable que se opte por el masculino genérico. Además, por el crecimiento constante y exponencial de los chatbots, esperamos que estos generen una mejor traducción, con menor número de errores en relación con la terminología y con el sesgo de género, específicamente.

Para alcanzar el objetivo establecido y verificar la hipótesis enunciada, emplearemos una metodología que combina el análisis cuantitativo del producto resultante de la traducción automática con una evaluación cualitativa de dicho producto. Para ello, nos serviremos tanto de sistemas de traducción automática como de chatbots de inteligencia artificial, herramientas frecuentemente empleadas por los usuarios. Este estudio busca contribuir al campo de la lingüística computacional y de la ética de la IA, dado que es relevante para desarrollar sistemas que sean no solo precisos, sino que también sean equitativos y respetuosos con la diversidad de género. Así pues, pretendemos promover una mayor conciencia sobre la intersección entre tecnología, lenguaje e inclusión. Asimismo, consideramos que los resultados obtenidos pueden ser de ayuda para desarrolladores y lingüistas, de tal modo que el futuro desarrollo de las tecnologías sea más inclusivo, en este caso, en lo que a lenguaje se refiere. En un mundo cada vez más interconectado, donde la TA desempeña un papel crucial en la mediación de la comunicación global, es imperativo que estas tecnologías evolucionen para reflejar y respetar la diversidad de identidades y expresiones de género.

## **2. INTELIGENCIA ARTIFICIAL Y TRADUCCIÓN**

La traducción automática ha evolucionado a lo largo del tiempo, pero en los últimos años se han producido diversos cambios al respecto. Los sistemas de traducción automática basados en redes neuronales se han posicionado por delante de los sistemas estadísticos y de los basados en reglas (Álvarez-Vidal *et al.* 2021), ya que ofrecen una mayor calidad en comparación con otros sistemas (Sánchez-Gijón *et al.* 2019). Estos se entrenan mediante corpus de gran tamaño

(Forcada 2017) y emplean tanto la IA como redes neuronales (Sánchez Ramos y Rico Pérez 2020), tal y como su propio nombre indica. Diversos estudios se han centrado en el análisis de esta tipología de sistemas (Álvarez-Vidal *et al.* 2021; Baylan y Chatterjee 2018; Bravo-Candel *et al.* 2021; Calixto y Liu 2019; Dwivedi *et al.* 2024, y Wu *et al.* 2016) sirviéndose de diversas métricas para evaluar la calidad de las traducciones, entre las que destacan Multidimensional Quality Metrics (Lommel *et al.* 2014), BLEU (Papineni *et al.* 2002) y METEOR (Banerjee y Lavie 2005).

Asimismo, la inteligencia artificial generativa ha revolucionado el lenguaje con el surgimiento y la rápida evolución de los grandes modelos de lenguaje o grandes modelos lingüísticos (LLM), así como los *large multimodal models* (LMM), los *multimodal large language models* y los *multilingual large language models*. Dado que la mayoría de la literatura está publicada en inglés, estamos ante términos de reciente creación y para los que todavía no existe un equivalente acuñado ampliamente extendido en español, motivo por el que en este trabajo mantendremos dichos términos en lengua inglesa. No obstante, es preciso mencionar las diferencias que existen entre unos y otros. Los LLM son capaces de generar textos a partir de textos, mientras que los LMM pueden procesar no solo textos, sino también vídeos, audios e imágenes (Adrakatti, 2024; Chen *et al.* 2024) de forma simultánea. En este trabajo agrupamos los LLM y los LMM bajo la denominación de chatbots de IA generativa, puesto que estas tecnologías están en constante cambio; por ejemplo, actualmente ChatGPT se considera un LMM, a pesar de que hasta hace poco se incluyera en el grupo de los LLM.

Si bien los estudios en los últimos años se han centrado únicamente en los sistemas de traducción automática, con la llegada de los chatbots de IA se han abierto nuevas líneas de investigación, incluida la comparación de modelos (Wang *et al.* 2023). Por ejemplo, Ferrag y Bentounsi (2024) analizan y comparan las traducciones ofrecidas por Gemini, ChatGPT y Claude. Asimismo, Briva-Iglesias *et al.* (2024) han querido dar un paso más y comparan los LLM con los sistemas de traducción automática neuronal mediante el análisis de la calidad de la traducción de textos jurídicos ofrecida por un sistema de TA neuronal (Google Translate) frente a dos LLM (GPT4 y VICUNA).

### 3. GÉNERO, DIVERSIDAD Y TRADUCCIÓN

Desde hace décadas, los Estudios de Traducción han reconocido que el lenguaje está vinculado con las construcciones sociales de género. Investigadoras como Simon (1996) y Von Flotow (1997) ponen de manifiesto que la práctica traductora puede perpetuar y desafiar las normas de género, y señalan el predominio de un lenguaje masculino en las traducciones, entre otras características. Todo ello se refleja en la traducción automática, pues estas tecnologías se entrenan con una gran cantidad de datos que, a menudo, contienen los mismos sesgos sociales que influyen en el lenguaje. Asimismo, es fundamental considerar la interseccionalidad entre género y otras formas de diversidad con el fin de lograr una traducción inclusiva y representativa; de ahí surge la necesidad de incorporar una terminología LGTBIQ+ adecuada para reflejar con precisión la diversidad de identidades de género en la práctica traductora y, por ende, en la traducción automática.

Diversos estudios sobre el lenguaje LGTBIQ+ en la traducción (Baldo *et al.* 2023) ponen de relieve la complejidad de representar las lenguas vernáculas y las identidades *queer* en distintos idiomas. Una traducción precisa requiere sensibilidad ante la evolución del léxico *queer* y los contextos sociopolíticos

en los que se enmarca. Del mismo modo, Thelwall *et al.* (2022) revisan la evolución de la terminología LGTBIQ+ en la literatura académica. La intersección del sesgo de género en la TA y las complejidades de la terminología *queer* plantean grandes desafíos para la traducción precisa. Los sistemas de TA, optimizados para ser eficientes, frecuentemente utilizan por defecto pronombres binarios y borran los matices de las identidades no binarias. Gromann *et al.* (2023) apuestan por enfoques participativos a la hora de diseñar sistemas de TA que no tengan en cuenta el género, haciendo hincapié en la inclusión de voces no binarias y *queer* en el proceso de desarrollo.

No obstante, en la última década han sido varios los estudios que se han centrado en profundizar en el sesgo de género en la traducción automática. Estos, en su mayoría, analizan el sesgo de género en Google Translate. Prates *et al.* (2019) observan una clara tendencia hacia el masculino, sobre todo en relación con cargos estereotípicamente asociados a la figura masculina como es el caso del ámbito científico. Del mismo modo, Fitria (2021) indica que en el proceso de traducción del inglés al indonesio Google Translate atribuye roles masculinos a profesiones relacionadas con actividades académicas e intelectuales, mientras que asocia roles femeninos a aquellas vinculadas a la docencia y los cuidados, a pesar de que Google Translate emplee una tecnología para reducir el sesgo de género (Johnson 2020). Ciora *et al.* (2021) y Piazzolla *et al.* (2024) incluyen, además, otros sistemas en la evaluación y llegan a conclusiones similares: observan discrepancias en la traducción del género, con un predominio del masculino genérico y de la asignación de roles de género a profesiones de acuerdo con los estereotipos tradicionales.

Por otro lado, los modelos de inteligencia artificial generativa no están exentos de sesgos de género, como se señala en varios estudios; por ejemplo, Kaplan *et al.* (2024) demuestran que ChatGPT no solo replica los sesgos de género presentes en textos producidos por humanos, sino que también los amplifica de manera contextual. En su estudio, generaron cartas de recomendación con ChatGPT utilizando nombres masculinos y femeninos para indicar el género del candidato. Los resultados revelaron una clara diferenciación en las características atribuidas según el género percibido, pues para los nombres femeninos ChatGPT tendía a enfatizar cualidades afiliativas y sociales, mientras que para los nombres masculinos destacaba logros profesionales y habilidades técnicas. Por su parte, Fulgu y Capraro (2024) también detectan sesgos de género en los modelos de GPT, en los que se refuerzan los estereotipos femeninos y, a menudo, se atribuyen erróneamente los masculinos. En este sentido, la traducción llevada a cabo por este tipo de modelos también presenta errores de género. Además, Ghosh y Caliskan (2023) revelan sesgos de género implícitos en ChatGPT, así como errores en la traducción de pronombres neutros tras analizar la traducción entre el inglés y otras seis lenguas (bengalí, farsi, malayo, tagalo, tailandés y turco).

De los estudios previos se desprende la necesidad de mejorar las metodologías de entrenamiento para mitigar los sesgos implícitos y promover sistemas de IA verdaderamente inclusivos, por lo que en este trabajo pretendemos comprobar cómo se trasvasan los marcadores de género y la terminología LGTBIQ+ del inglés al español empleando una serie de herramientas basadas en IA.

#### **4. METODOLOGÍA**

En primer lugar, delimitaremos la muestra de análisis, a continuación, seleccionaremos las herramientas de las que obtendremos el producto resultante, es decir, la traducción, y, en tercer



lugar, estableceremos los parámetros de análisis y el protocolo que empleamos para evaluar la calidad de la traducción.

#### 4.1 Muestra de análisis

Para analizar el producto resultante de la traducción automática, con especial atención en la terminología relacionada con el género, la inclusión y la diversidad sexual y de género, hemos recopilado seis de fragmentos de intervenciones del debate que tuvo lugar el 15 de junio de 2023 en la Cámara de los Comunes del Reino Unido, que versaba sobre el Mes del Orgullo (Cámara de los Comunes 2023). Los motivos que propiciaron la elección de este discurso fueron principalmente dos: a) la temática; y b) el contexto en el que se enmarca el discurso, Reino Unido, país en el que en 2023 gobernaba un primer ministro del Partido Conservador. Así pues, hemos considerado que este factor podría añadir un componente político significativo al análisis y permitir explorar posibles implicaciones ideológicas en la interpretación de este tipo de discursos.

Los seis fragmentos suman un total de 807 palabras y representan diversos estilos discursivos y perspectivas en torno al tema abordado. Entre los fragmentos seleccionados se encuentran las intervenciones orales de cuatro mujeres miembros de la Cámara de los Comunes: Angela Eagle, Nickie Aiken, Charlotte Nichols y Wera Hobhouse. Dado que el objeto de estudio es comprobar la calidad de las traducciones derivadas de la traducción automática, con un énfasis en si las herramientas empleadas son capaces de identificar que los discursos provienen de mujeres, hemos seleccionado fragmentos que incluyeran adjetivos referentes a las propias oradoras como, por ejemplo: *comfortable*, *unbothered* y *ally*, entre otros, como puede observarse en el fragmento que ofrecemos como ejemplo:

I am a feminist, I am a lesbian, and I am a trans ally. I do not believe that allowing trans men and women to live with dignity and respect threatens my rights or my wellbeing in the slightest. We all advance together, or not at all. Even at this late stage, the Government could do the decent thing and abandon their divisive tactics. Instead of endless prevarication, they could publish sensible and inclusive relationships and sex education guidance, which our schools have been waiting for since 2019. They could stop playing dangerous and divisive games with trans people by trying to set their rights against women's rights (Eagle 2023).

Este criterio de selección nos ha permitido analizar de manera específica si las herramientas de traducción emplean la terminología adecuada y si logran mantener las referencias de género y los matices lingüísticos en la lengua meta, aspectos cruciales para evaluar la calidad y precisión de las traducciones.

#### 4.2 Selección de los sistemas de TA y de los chatbots

El principal criterio para seleccionar las cuatro herramientas ha sido su acceso gratuito y su popularidad, dado que primamos emplear herramientas al alcance de cualquier usuario, no necesariamente una empresa de gran envergadura con mayores recursos económicos.



En consecuencia, hemos empleado dos sistemas de traducción automática (DeepL y Google Translate) y dos chatbots de IA (ChatGPT y Gemini). Si bien los sistemas de traducción automática se sirven de la inteligencia artificial, consideramos preciso diferenciarlos de los chatbots, puesto que los sistemas de TA se han desarrollado exclusivamente para traducir. Procedemos a describir cada uno de ellos en detalle.

#### *4.2.1 DeepL*

DeepL, un sistema de traducción automática basado en redes neuronales y en LLM que se sirve de la inteligencia artificial para realizar las traducciones, se lanzó en agosto de 2017 y cuenta con más de treinta lenguas. En este sentido, es preciso señalar que hace distinción entre variedades geográficas, por ejemplo, entre el inglés británico y el inglés americano. Asimismo, cuenta con una versión gratuita y con una versión de pago (DeepL Pro) con varios planes: *starter*, *advanced*, *ultimate* y *enterprise*. En función del plan, ofrece un mayor número de servicios. Todos los planes de pago tienen en común la seguridad de los datos, la traducción ilimitada de textos, la funcionalidad de tratamiento formal o informal para algunas lenguas, así como mayor número de entradas para los glosarios. DeepL permite también traducir archivos de diferentes formatos entre los que se incluyen DOCX, PDF y PPTX. No obstante, es importante señalar que existe un límite en el número de traducciones de documentos, incluso para los suscriptores de DeepL Pro.

A pesar de que DeepL esté disponible como plataforma en línea, desde 2019 también se puede acceder a sus servicios mediante una aplicación de escritorio sin coste, compatible tanto con sistemas Windows como MacOS. Además, ofrece una API que se puede integrar en productos y plataformas externas, disponible en versiones gratuita y de pago, por lo que permite a los desarrolladores incorporar las capacidades de traducción de DeepL en sus propias aplicaciones y servicios.

#### *4.2.2 Google Translate*

Google Translate es un servicio de traducción automática gratuito desarrollado por Google, que permite traducir texto, documentos y sitios web entre más de 100 idiomas. Utiliza técnicas avanzadas de inteligencia artificial, particularmente redes neuronales, para ofrecer traducciones en tiempo real. Desde 2016, se basa en un modelo de traducción automática neuronal que analiza frases completas en lugar de traducir palabra por palabra, lo que resulta en traducciones más contextuales y fluidas (Turovsky 2016, Wu *et al.* 2016).

Google Translate se actualiza constantemente mediante aprendizaje continuo y es accesible de forma gratuita tanto desde su sitio web como desde aplicaciones móviles. Estas características lo convierten en una herramienta valiosa y fácil de usar para usuarios con necesidades diversas, aunque siempre es aconsejable revisar las traducciones en contextos críticos.

#### *4.2.3 ChatGPT*

ChatGPT es un chatbot de IA generativa que presenta diferentes modelos. El que se ha usado en este estudio se corresponde con GPT 4-o, el modelo GPT más avanzado de OpenAI, que se

ha entrenado con datos anteriores a octubre de 2023. GPT 4-o, si bien es un LLM, también entra dentro de la categoría de LMM en tanto que permite texto e imágenes como *input*, aunque genere texto como *output* (OpenAI, s.f.). A pesar de ser de pago, ChatGPT permite a los usuarios disfrutar de un número limitado de interacciones de forma gratuita del modelo GPT 4-o. Por ello, hemos optado por emplear este modelo y no modelos anteriores como GPT 3.5. En este sentido, nos ha permitido traducir los textos de la muestra de análisis con tecnología de última generación.

#### 4.2.4 Gemini 1.5

Gemini es un modelo avanzado de IA desarrollado por Google DeepMind que integra capacidades de IA generativa en diversas aplicaciones, y que está diseñado para procesar y analizar información en texto, imágenes y vídeo, por lo que entra dentro de la categoría de LMM. Asimismo, la plataforma Gemini incluye varios modelos multimodales optimizados para diferentes propósitos, entre los que se incluyen la versión Flash, Nano, Pro y Ultra. No obstante, en este estudio hemos empleado Gemini 1.5 Flash, puesto que el único requisito para poder usarlo es tener una cuenta de Gmail, mientras que el Pro solo ofrece un mes de prueba gratuita.

Si bien Gemini 1.5 Flash presenta una menor capacidad de *tokens* en comparación con los otros modelos, este modelo es más rápido y eficaz, puesto que está diseñado para ofrecer un alto rendimiento en tareas que requieren un procesamiento rápido, como la traducción de idiomas en tiempo real, los resúmenes de datos y el soporte conversacional, entre otros (Subramanya 2024).

Una vez descritas las herramientas, procedemos a describir los parámetros de análisis y el procedimiento para su aplicación.

### 4.3 Parámetros de análisis

Los parámetros de análisis empleados en este estudio se enmarcan en una adaptación de Multidimensional Quality Metrics 2.0 (MQM), un sistema para evaluar la calidad de las traducciones, tanto humanas como automáticas, que categoriza los errores en siete dimensiones que, a su vez, contienen subcategorías (Lommel *et al.* 2024). En el marco de MQM, se distinguen dos variantes principales: MQM Full<sup>1</sup> y MQM Core (Lommel *et al.* 2024). MQM Full constituye el repositorio completo de tipos de errores y ofrece una amplia gama de categorías detalladas para una evaluación exhaustiva de la calidad de la traducción. MQM Core, por su parte, representa un subconjunto preestablecido y ampliamente utilizado de tipos de errores, que se limita a los dos niveles jerárquicos más altos.

En nuestro caso, hemos optado por guiarnos por MQM Core (véase Tabla 1), dado que proporciona una estructura más estandarizada para la evaluación de la calidad de las traducciones, a pesar de que hayamos incluido errores específicos de las subcategorías detalladas en MQM Full en algunos casos, por ejemplo, *overly literal* y *false friend* dentro de la subcategoría de *mistranslation*.

<sup>1</sup> MQM Full: <https://themqm.org/the-mqm-full-typology/> (Fecha de consulta: 29 de noviembre de 2024).

Tabla 1. Dimensiones y subcategoría de errores de MQM Core (MQM Council s.f.)

<i>Terminology</i>	<i>Inconsistent with terminology resource</i>
	<i>Inconsistent use of terminology</i>
	<i>Wrong term</i>
<i>Accuracy</i>	<i>Mistranslation</i>
	<i>Overtranslation</i>
	<i>Undertranslation</i>
	<i>Addition</i>
	<i>Omission</i>
	<i>Do not translate</i>
	<i>Untranslated</i>
<i>Linguistic conventions</i>	<i>Grammar</i>
	<i>Punctuation</i>
	<i>Spelling</i>
	<i>Unintelligible</i>
	<i>Character encoding</i>
	<i>Textual conventions</i>
<i>Style</i>	<i>Organization style</i>
	<i>Third-party style</i>
	<i>Inconsistent with external reference</i>
	<i>Language register</i>
	<i>Awkward style</i>
	<i>Unidiomatic style</i>
	<i>Inconsistent style</i>
<i>Locale conventions</i>	<i>Number format</i>
	<i>Currency format</i>
	<i>Measurement format</i>
	<i>Time format</i>
	<i>Date format</i>
	<i>Address format</i>
	<i>Telephone format</i>
	<i>Shortcut key</i>
<i>Audience appropriateness</i>	<i>Culture-specific reference</i>
	<i>Offensive</i>
<i>Design and markup</i>	<i>Layout</i>
	<i>Markup tag</i>
	<i>Truncation/text expansion</i>
	<i>Missing text</i>
	<i>Link/cross-reference</i>



Además, para este estudio hemos realizado una adaptación de los parámetros teniendo en cuenta las características de la muestra de análisis y el objetivo de nuestro estudio. Por este motivo, hemos omitido dos dimensiones: *locale conventions*, dado que los errores relacionados con carácter ortotipográfico los categorizaremos como errores de convención lingüística, y *design and mark up*, puesto que, al introducir el texto sin formato en las diferentes herramientas, esta dimensión no era relevante, pues no vamos a centrarnos en el diseño, sino en los aspectos puramente lingüísticos.

Por otro lado, hemos omitido algunas subcategorías de las siguientes dimensiones: *terminology*, *linguistic conventions* y *style*. También hemos omitido *inconsistent with terminology resource*, puesto que no hay una base de datos terminológica de referencia.

Por lo que respecta al estilo, únicamente hemos incluido las subcategorías denominadas *language register*, *unidiomatic style* e *inconsistent style*, por su relevancia dentro del análisis discursivo.

En la Tabla 2 se exponen los parámetros empleados tras las modificaciones mencionadas anteriormente.

Tabla 2. Clasificación de errores empleada en este estudio

<i>Terminology</i>	<i>Inconsistent use of terminology</i>	
	<i>Wrong term</i>	
<i>Accuracy</i>	<i>Mistranslation</i>	<i>False friend</i>
		<i>Overly literal</i>
	<i>Overtranslation</i>	
	<i>Undertranslation</i>	
	<i>Addition</i>	
	<i>Omission</i>	
	<i>Do not translate</i>	
	<i>Untranslated</i>	
<i>Linguistic conventions</i>	<i>Grammar</i>	<i>Word-form</i>
		<i>General language collocation</i>
	<i>Punctuation</i>	
	<i>Spelling</i>	
	<i>Unintelligible</i>	
	<i>Textual conventions</i>	
<i>Style</i>	<i>Language register</i>	
	<i>Unidiomatic style</i>	
	<i>Inconsistent style</i>	
<i>Audience appropriateness</i>	<i>Culture-specific reference</i>	
	<i>Offensive</i>	



Para poder aplicar dichos parámetros, en primer lugar, hemos elaborado un documento en Microsoft Excel donde se han segmentado los textos de la muestra con el propósito de facilitar la extracción y el análisis de errores. Se ha creado una hoja de cálculo para cada oradora y en cada hoja hemos incluido una columna específica para cada una de las herramientas de traducción automática evaluadas. Una vez estructurada la hoja, hemos introducido los fragmentos seleccionados en los diversos sistemas de traducción automática y chatbots. Es preciso mencionar que, a pesar de que los diferentes textos se hayan segmentado para facilitar el análisis, al introducirlos para su traducción, hemos empleado las intervenciones completas y no segmentadas, con el fin de que las herramientas empeladas tuvieran el mayor contexto posible. El *prompt* empleado en los chatbots ha sido «*Translate the following text into Spanish*», puesto que queríamos que la traducción se produjera en las mismas condiciones que en los sistemas de traducción automática, a pesar de que los chatbots permitan proporcionar más información. Los resultados obtenidos se han trasladado a las celdas correspondientes a cada segmento. Seguidamente, hemos evaluado el *output*, mediante la aplicación de los parámetros de análisis fundamentados en el modelo MQM y descritos en el punto anterior. En este sentido, la evaluación se ha realizado segmento por segmento, es decir, en primer lugar, hemos evaluado el resultado del primer sistema incluido en la tabla en la hoja de cálculo (DeepL) y se han marcado todos los errores identificados. Seguidamente, se ha evaluado el resultado del mismo segmento ofrecido por el siguiente sistema de traducción incluido en la tabla (Google Translate), hasta completar el análisis de las cuatro herramientas empleadas. Una vez finalizado el análisis de ese segmento, se ha pasado al siguiente y se ha seguido el mismo procedimiento. Este modo de evaluación nos ha permitido analizar el producto resultante con eficacia. Para facilitar la identificación visual, se ha implementado un sistema de código de colores para diferenciar las distintas categorías de errores detectados en los resultados.

TO	A	B	C	D	E
artwork.	DeepL	GT	ChatGPT	Gemini 1.5	
She did not "make me LGBT", but she did let me know that she really would not care if I was, and made me feel, for the first time, comfortable and unbothered about being bisexual.	decirlo, en mis obras de arte.	digamos, en mis obras de arte.	mis obras de arte.	digamos, en mis dibujos.	
When, many years later, I came out to friends and family, I was lucky to be met with the same kind of supportive indifference.	No me «convirtió en LGBT», pero me hizo saber que no le importaba que lo fuera, y me hizo sentir, por primera vez, cómoda y tranquila por ser bisexual.	No me «convirtió en LGBT», pero sí me hizo saber que realmente no le importaría si lo era y me hizo sentir, por primera vez, cómoda y despreocupada por ser bisexual.	Ella no me "hizo LGBT", pero sí me hizo saber que realmente no le importaría si lo era, y me hizo sentir, por primera vez, cómoda y tranquila con ser bisexual.	Ella no "me hizo LGBT", pero sí me dejó saber que realmente no le importaría si lo fuera, y me hizo sentir, por primera vez, cómoda y tranquila con respecto a mi bisexualidad.	
	Cuando, muchos años después, salí del armario ante amigos y familiares, tuve la suerte de encontrarme con el mismo tipo de <b>indiferencia solidaria</b> .	Cuando, muchos años después, se lo dije a mis amigos y familiares, tuve la suerte de encontrarme con el mismo tipo de indiferencia y apoyo.	Cuando, muchos años después, me declaré ante amigos y familiares, tuve la suerte de ser recibida con el mismo tipo de <b>indiferencia comprensiva</b> .	Cuando, muchos años después, salí del armario con mis amigos y familiares, tuve la suerte de encontrarme con el mismo tipo de <b>indiferencia comprensiva</b> .	

Figura 1. Ejemplo de codificación de errores por colores

En cuanto a la contabilización de errores, es importante señalar que, al analizar las traducciones por segmentos individuales, los errores que se repiten en diferentes segmentos se contabilizan como errores individuales en cada ocurrencia. Esta decisión metodológica se basa en el hecho de que cada segmento se considera una unidad de traducción independiente y un error recurrente puede tener un impacto distinto dependiendo del contexto específico de cada segmento. Además, este enfoque permite evaluar la consistencia de las herramientas de traducción automática a lo largo de todo el texto.

Debido a las limitaciones de espacio y para mantener la claridad expositiva, en el análisis se presenta una selección representativa de ejemplos en las tablas y gráficos del artículo. Sin embargo, los resultados y las conclusiones se derivan del análisis completo del corpus. Por lo tanto, en el análisis y la discusión de resultados se hace referencia ocasionalmente a ejemplos o términos que, aunque pueden no aparecer en los segmentos mostrados, son parte integral del análisis global y contribuyen significativamente a las conclusiones de este estudio.

## 5. ANÁLISIS Y RESULTADOS

En la muestra de análisis hemos detectado diversos errores de precisión (*accuracy*), terminología (*terminology*), convenciones lingüísticas (*linguistic conventions*), estilo (*style*) y de adecuación al receptor (*audience appropriateness*), tal y como se puede apreciar en el Gráfico 1.

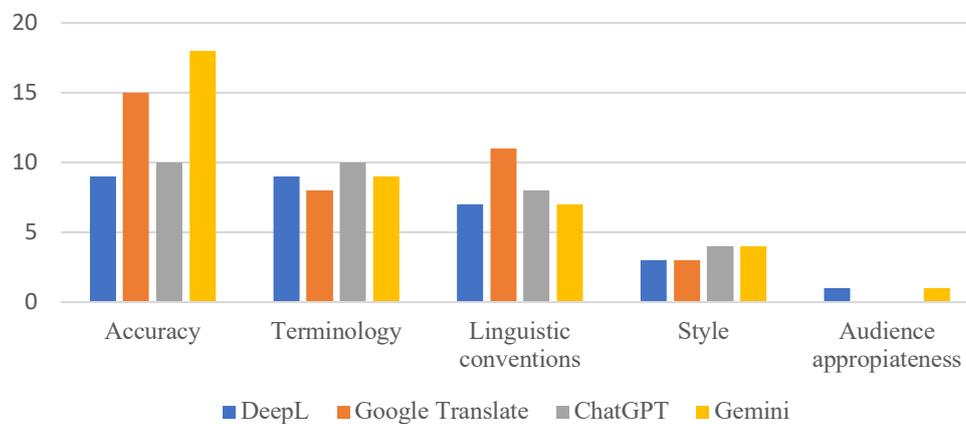


Gráfico 1. Errores por categorías

De la observación del Gráfico 1 se desprende que la mayor parte de los errores se corresponde con la categoría de precisión, seguido de la de terminología y de la de convenciones lingüísticas. Estos resultados siguen una tendencia similar a la obtenida por Peñuelas Gil (2024) en el campo del turismo agroalimentario. Por herramientas, Gemini es la que más errores de precisión presenta, con un total de dieciocho, seguida de Google Translate con quince. Por lo que respecta a la terminología, los resultados no varían demasiado entre herramientas, pues todos cometen entre ocho y diez errores, siendo ChatGPT el que presenta un número más elevado. Sin embargo, en lo que a convenciones lingüísticas se refiere, Google Translate es el que muestra un mayor número de errores (11), seguido de ChatGPT (8). En cuestión de estilo, se puede observar que el número no es muy elevado; los dos chatbots y los dos sistemas de traducción automática neuronal presentan el mismo número de errores, cuatro y tres, respectivamente. Por último, en cuanto a la categoría de adecuación al receptor, solo DeepL y Gemini muestran un único error que, en ambos casos, se corresponde con una mala traducción cultural, concretamente, *year 9*. En este sentido, se considera un error de referencia cultural, puesto que en España el sistema escolar no contiene ningún curso denominado noveno curso. La equivalencia al español de noveno curso sería 2.º de ESO (International School of Navarra, 2023).

Si hacemos un recuento del total de errores, como se refleja en los datos del Gráfico 2, la herramienta que presenta un cómputo total de errores mayor es Gemini (39), seguido de Google Translate (37), de ChatGPT (32) y, por último, de DeepL (29).

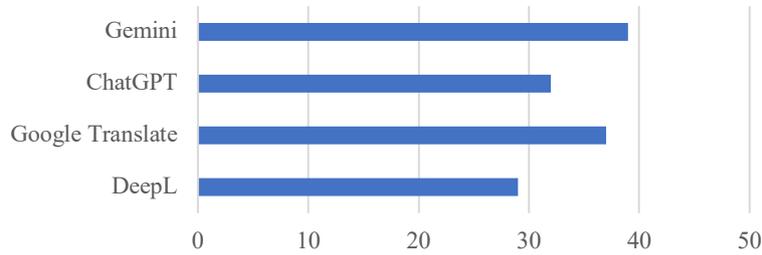


Gráfico 2. Distribución de errores por herramienta

Además, en todos los sistemas y modelos, el error que ha presentado un mayor número de ocurrencias hace referencia a las traducciones erróneas o *mistranslations*, siendo Gemini y Google Translate aquellos con la cifra más alta, con 17 y 15 ocurrencias respectivamente. El segundo tipo de error más frecuente se corresponde con la terminología, especialmente por haber empleado términos incorrectos. Asimismo, tanto los sistemas convencionales como los nuevos modelos presentan diversos errores de puntuación, gramaticales, de estilo y de referencias culturales específicas.

Por lo que respecta a la terminología perteneciente al colectivo LGTBIQ+, observamos que las herramientas de traducción automática seleccionadas han presentado algunas dificultades, aunque en líneas generales el resultado ha sido adecuado. No obstante, ninguna ha modificado el orden de las siglas en LGBT, LGBT+ y LGBTQ+, lo que constituye un error de terminología. Si bien es cierto que en español pueden encontrarse las siglas en el mismo orden que inglés, la forma más adecuada de referirse al colectivo es *LGTBIQ+*, aunque se aceptan las simplificaciones *LGTB* y *LGTBI*.

En este punto nos gustaría detenernos en un ejemplo de la traducción ofrecida para «trans» que se muestra en la Tabla 3.

Tabla 3. Segmento de la intervención de Nickie Aiken

TO	DeepL	Google Translate	ChatGPT	Gemini 1.5
<i>There should no ban, including on education on homosexuality and trans.</i>	No debe haber ninguna prohibición, incluida la educación sobre la homosexualidad y los <b>transexuales</b> .	No debería haber ninguna prohibición, incluida la educación sobre homosexualidad y <b>transexualidad</b> .	No debe haber ninguna prohibición, incluyendo la educación sobre la homosexualidad y las personas <b>trans</b> .	No debe haber ninguna prohibición, ni siquiera sobre la educación sobre la homosexualidad y la <b>transexualidad</b> .

De los datos expuestos en la Tabla 3 se infiere que DeepL no ha empleado el término adecuado a la hora de traducir *trans*. Mientras que Google Translate y Gemini han optado por

transexualidad y ChatGPT, por *personas trans*, DeepL ha empleado «los transexuales», término incorrecto. Cualquiera de las opciones propuesta por el resto de los sistemas y modelos sería apropiada. En cuanto a la precisión en la identificación del género, se observan discrepancias significativas entre las diversas herramientas de traducción analizadas, como se expone en el Gráfico 3. Los datos muestran que Google Translate presenta el mayor número de errores, con un total de siete. Le sigue ChatGPT, con seis errores y Gemini, que acumula cinco. En cambio, DeepL muestra un rendimiento superior en este aspecto, pues tan solo se identifican tres errores.

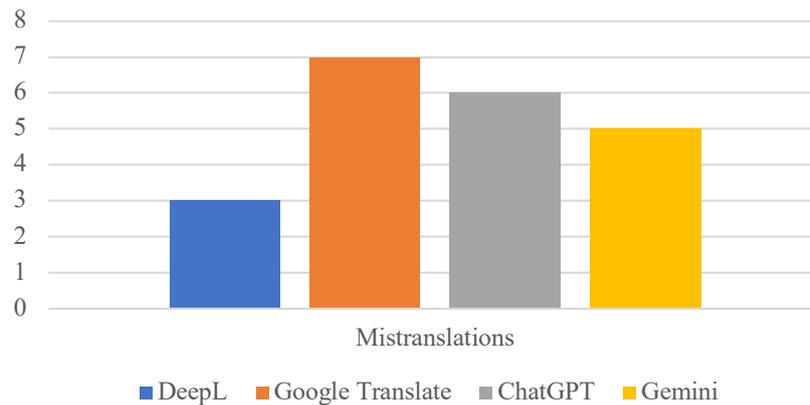


Gráfico 3. Errores de género mal traducido

Al analizar los errores con mayor detalle, detectamos que las herramientas empleadas tienden a optar por el masculino, aunque hay ocasiones en las que identifican correctamente el femenino. Es especialmente llamativo que en términos como *honorable member* o *minister* (ver tablas 4 y 5) estos sistemas se decantan mayoritariamente por la forma masculina, a pesar de que en algunos casos el nombre propio de la persona referida (es decir, un nombre femenino) aparezca en el contexto inmediato.

Tabla 4. Traducción de *minister* en la intervención de Angela Eagle

TO	DeepL	Google Translate	ChatGPT	Gemini 1.5
<i>I worked with her in government as a Minister [...]</i>	Trabajé con ella en el Gobierno como <b>Ministro</b> [...]	Trabajé con ella en el gobierno como <b>Ministro</b> [...]	Trabajé con ella en el gobierno como <b>Ministro</b> [...]	Trabajé con ella en el gobierno como <b>ministra</b> [...]

Tabla 5. Traducción de *honorable Member* en la intervención de Wera Hobhouse

TO	DeepL	Google Translate	ChatGPT	Gemini 1.5
<i>[...] in the words of the hon. Member for Warrington North (Charlotte Nichols) [...]</i>	[...] en palabras de la <b>diputada</b> por Warrington Norte (Charlotte Nichols) [...]	[...] en palabras del <b>Excmo. Miembro</b> de Warrington North (Charlotte Nichols) [...]	[...] como dijo <b>la honorable Miembro</b> por Warrington North (Charlotte Nichols) [...]	[...] en palabras del <b>diputado</b> por Warrington North (Charlotte Nichols) [...]

No obstante, también se presentan buenas traducciones y, a continuación, vamos a incluir tres fragmentos de la muestra de análisis en los que se refleja la traducción del género y de términos del colectivo LGTBIQ+.

Tabla 6. Intervención de Angela Eagle de la muestra de análisis

TO	DeepL	Google Translate	ChatGPT	Gemini 1.5
<i>I am a feminist, I am a lesbian, and I am a trans ally.</i>	Soy feminista, lesbiana y <b>aliada trans.</b>	Soy feminista, soy lesbiana y soy una <b>aliada de las personas trans.</b>	Soy feminista, soy lesbiana y soy <b>aliada de las personas trans.</b>	Soy feminista, soy lesbiana y soy <b>aliada trans.</b>
<i>I do not believe that allowing trans men and women to live with dignity and respect threatens my rights or my wellbeing in the slightest.</i>	No creo que permitir que los hombres y mujeres trans vivan con dignidad y respeto amenace en lo más mínimo mis derechos o mi bienestar.	No creo que permitir que los hombres y mujeres trans vivan con dignidad y respeto amenace mis derechos o mi bienestar en lo más mínimo.	No creo que permitir que los hombres y mujeres trans vivan con dignidad y respeto amenace mis derechos ni mi bienestar en lo más mínimo.	No creo que permitir que los hombres y mujeres trans vivan con dignidad y respeto amenace mis derechos o mi bienestar en lo más mínimo.
<i>We all advance together, or not at all.</i>	O avanzamos todos juntos, o no avanzamos.	Todos avanzamos juntos, o no avanzamos en absoluto.	Avanzamos todos juntos o no avanzamos.	Todos avanzamos juntos, o no avanzamos en absoluto.

En el fragmento de la muestra de análisis que se muestra en la Tabla 6, se puede observar una buena traducción de los términos relacionados con el colectivo e identificación adecuada del género de la oradora. A continuación, ofrecemos algunas apreciaciones en relación con la muestra. En el segmento «*I am a feminist, I am a lesbian, and I am a trans ally*», DeepL es el único que opta por no repetir el verbo en la enumeración y lo traduce como «Soy feminista, lesbiana y aliada trans». A pesar de que en otros contextos la omisión del verbo se podría valorar como la opción más acertada, en este caso, al tratarse de un discurso, consideramos que mantener el verbo es un recurso para dar énfasis.

Por otro lado, Google Translate y Gemini ofrecen una traducción idéntica y errónea de «*We all advance together, or not at all*», puesto que es un calco de la estructura en inglés. En este caso, la opción más adecuada sería la que ofrece DeepL, ya que mantiene la coordinación de las oraciones distributivas (o... o...), a diferencia de ChatGPT que solo mantiene la segunda conjunción distributiva, de tal modo que comete un error gramatical.

No obstante, hemos de reconocer que en algunos fragmentos la identificación del género es correcta, como se recoge en las intervenciones de Charlotte Nichols en la Tabla 7.



Tabla 7. Segmentos de la intervención de Charlotte Nichols

TO	DeepL	Google Translate	ChatGPT	Gemini 1.5
<i>I know the difference that it made to me, as a teenager, when my art teacher, Mrs Tibbatts, was able to broach the subject with me gently after picking up on some unhappy and sapphic themes, shall we say, in my artwork.</i>	<b>Sé la diferencia que supuso para mí</b> , de adolescente, cuando mi <b>profesora</b> de arte, la Sra. Tibbatts, fue capaz de abordar el tema conmigo con delicadeza después de detectar algunos temas infelices y sáficos, por así decirlo, en mis obras de arte.	Sé lo importante que fue para mí, cuando era adolescente, que mi <b>profesora</b> de arte, la señora Tibbatts, pudiera abordar el tema conmigo con delicadeza después de detectar algunos temas infelices y sáficos, digamos, en mis obras de arte.	<b>Sé la diferencia que marcó para mí</b> , como adolescente, cuando mi <b>profesora</b> de arte, la Sra. Tibbatts, pudo abordar el tema conmigo de manera delicada tras captar algunos temas infelices y sáficos, digamos, en mis obras de arte.	<b>Sé la diferencia que supuso para mí</b> , como adolescente, cuando mi <b>profesora</b> de arte, la Sra. Tibbatts, pudo abordar el tema conmigo con delicadeza después de detectar algunos temas infelices y sáficos, digamos, en mis <b>dibujos</b> .
<i>She did not “make me LGBT”, but she did let me know that she really would not care if I was, and made me feel, for the first time, comfortable and unbothered about being bisexual.</i>	No me «convirtió en LGBT», pero me hizo saber que no le importaba que lo fuera, y me hizo sentir, por primera vez, <b>cómoda</b> y <b>tranquila</b> por ser bisexual.	No me “convirtió en LGBT”, pero sí me hizo saber que realmente no le importaría si lo era y me hizo sentir, por primera vez, <b>cómoda</b> y <b>despreocupada</b> por ser bisexual.	Ella no me “hizo LGBT”, pero sí me hizo saber que realmente no le importaría si lo era, y me hizo sentir, por primera vez, <b>cómoda</b> y <b>tranquila</b> con ser bisexual.	Ella no “me hizo LGBT”, pero sí <b>me dejó saber</b> que realmente no le importaría si lo fuera, y me hizo sentir, por primera vez, <b>cómoda</b> y <b>tranquila</b> con respecto a mi bisexualidad.

En los dos segmentos expuestos en la Tabla 7, se puede observar que todas las herramientas empleadas han identificado el género tanto de la oradora como de la profesora a la que hace mención. Si bien dentro del texto hay marcadores que indican que la profesora es mujer (*Mrs.* y *she*), el contexto no aporta indicios del género de la oradora y el único contexto existente hace referencia a encontrarse en clase de arte, por lo que resulta particularmente llamativo. Dada la extensión del segmento, se detectan más errores que no están vinculados a la identificación del género. A modo ilustrativo, procedemos a destacar alguno de ellos. En este sentido, detectamos un error que se corresponde con la subcategoría de estilo no idiomático en la traducción de «sé la diferencia que supuso para mí», en tanto que no suena natural y resulta muy literal la traducción de la expresión *make a difference*. A pesar de que se pueda entender, no sería la expresión más utilizada en español; «marcó un antes y un después» sería una traducción más apropiada. Del mismo modo, detectamos un error de sobretraducción (*overtranslation*) por parte de Gemini, en tanto que traduce *artwork* por «dibujos», un término mucho más específico, y otro error dentro de la subcategoría de traducciones erróneas (*mistranslation*), concretamente por ser demasiado literal, pues traduce «*she did let me know*» por «me dejó saber», una expresión calcada del inglés que carece de sentido en español.

De forma global los resultados reflejan que ChatGPT presenta un menor número de errores que Google Translate, por lo que el producto resultante puede que no dependa tanto del tipo de tecnología empleada (sistemas convencionales de traducción automática basados en redes neuronales o modelos de inteligencia artificial generativa), sino en la estructura y la arquitectura detrás de cada modelo y de los datos de entrenamiento de los mismos. Asimismo, a pesar de que Gemini y Google Translate se basan en arquitecturas teóricamente diferentes, se han identificado casos en los que ambos sistemas ofrecen traducciones similares e, incluso, idénticas en ciertos segmentos. Esta coincidencia sugiere que, aunque Gemini utiliza una arquitectura más avanzada basada en redes neuronales profundas y modelos de lenguaje de gran escala, es posible que comparta algunos componentes o bases de datos con Google Translate.

Por último, una vez analizado el producto resultante del transvase interlingüístico de los diferentes fragmentos de la muestra de análisis con sistemas de traducción automática y con modelos de inteligencia artificial generativa, exponemos las conclusiones.

## 6. CONCLUSIONES

Los resultados obtenidos nos permiten confirmar que el género masculino es la opción más recurrente en todos los sistemas y modelos, es decir, se evidencia una propensión a utilizar el género masculino como opción predeterminada que conlleva a deducir la presencia del sesgo hacia el masculino. Estos hallazgos coinciden con los estudios previos mencionados anteriormente (Ciora *et al.* 2021; Fitria *et al.* 2021; Piazzolla *et al.* 2024, y Prates *et al.* 2019).

Tal y como era previsible, las palabras en las que el género está implícito en la propia unidad léxica como, por ejemplo, *lesbiana*, favorecen una buena interpretación por parte de las herramientas de traducción automática del género del texto original en lenguas sin marcadores de género como el inglés en este caso y evitan traducciones erróneas. Si excluimos las unidades léxicas con género implícito, los resultados no arrojan luz sobre cuáles son las condiciones para que los sistemas identifiquen o no el género en los textos, ya que en algunos segmentos se ha identificado que las oradoras eran mujeres, mientras que, en otros, se ha interpretado que no lo eran. No obstante, sí dan indicios sobre aspectos que pueden influir. Bien es cierto que, cuando se hace alusión a la terminología política, los traductores automáticos optan por emplear el masculino, a pesar de que el contexto sugiera lo contrario, lo que nos lleva a reconocer que en la traducción automática sí hay sesgo de género relacionado con los roles de género tradicionalmente impuestos por la sociedad. Así pues, los resultados ponen de manifiesto la necesidad de mejorar la capacidad de estos sistemas para contextualizar y procesar información de género de manera más precisa y coherente.

En lo relativo a la terminología LGTBIQ+, la traducción ofrecida por los sistemas y modelos es adecuada en la mayor parte de casos, quizá por la gran cantidad de información existente al respecto. Un posible factor que puede incidir en ello está relacionado con la creciente visibilidad y el reconocimiento del colectivo en los últimos años, que ha propiciado un aumento significativo de políticas inclusivas, asociaciones y organizaciones dedicadas a la defensa de sus derechos. Además, la presencia cada vez mayor de estos términos en medios de comunicación, redes sociales y plataformas digitales ha contribuido a su

estandarización y correcta utilización en múltiples lenguas, lo que puede haber facilitado que los traductores automáticos sean más precisos. Sin embargo, cabe señalar que los sistemas y modelos presentan dificultades en la traducción de las propias siglas LGTB y sus variantes, lo que demuestra que todavía hay aspectos que necesitan mejorarse en la traducción de la terminología específica del colectivo.

Por otra parte, desde una perspectiva más general, los resultados ponen de manifiesto la necesidad de la posesición para ofrecer una traducción de calidad. Pues bien, a pesar de que la traducción automática avance y con ella las herramientas al alcance de cualquier usuario, lo cierto es que el producto derivado de esta sigue presentando un gran número de errores que solo pueden solventarse tras una evaluación humana.

A pesar de utilizar inteligencia artificial generativa, no se ven reflejados resultados de mayor calidad en términos de traducción con respecto a traductores automáticos convencionales como DeepL, que, de hecho, ofrece una mejor calidad de traducción según los resultados obtenidos. Por lo tanto, al contrario de lo que se preveía, los LLM todavía tienen que refinarse para generar traducciones más adecuadas. En este sentido, cabe mencionar que quizá un *prompt* más específico en el que se detallen aspectos como el orador y género textual, entre otros, podría favorecer el resultado de las traducciones, pues el chatbot de inteligencia artificial generativa tendría mayor contexto e información de la que partir a la hora de traducir.

En definitiva, se determina que hay sesgo hacia el masculino e inconsistencia en la identificación del femenino, pues, a pesar de que ocasionalmente se reconozca el género femenino, esta identificación no es consistente. Los resultados muestran la necesidad de mejorar los algoritmos de traducción para que sean más sensibles al contexto y precisos en la identificación y traducción del género gramatical y semántico. Asimismo, la metodología expuesta en este estudio podría replicarse empleando otros sistemas, lenguas y muestras.

Para finalizar, nos gustaría dejar constancia de las limitaciones. Dado que la tecnología que subyace a los modelos está en continuo cambio, los resultados podrían haber variado desde su recopilación hasta la fecha de publicación. Consideramos que las líneas de investigación futuras podrían orientarse hacia la realización de un estudio semejante en el que el *input* principal sea audio. Este enfoque permitiría examinar si la modalidad del *input* (texto vs. audio) influye de manera significativa en la identificación del género en contextos tradicionalmente masculinos, lo que podría vislumbrar patrones relativos a la percepción y a los sesgos de género cuando la información se presenta de forma oral en lugar de escrita.

## 7. AGRADECIMIENTOS

Este trabajo se ha desarrollado en el marco de los proyectos nacionales de I+D titulados «VIP II» (PID2020-112818GB-I00), coordinado por la Dra. Gloria Corpas Pastor (Universidad de Málaga), «App para entrenar en posesición de traducción automática neuronal mediante la gamificación en entornos profesionales (GAMETRAPP)» (TED2021-129789B-I00), coordinado por la Dra. Cristina Toledo Báez (Universidad de Málaga) y en el marco del «programa Investigo» del Servicio Público de Empleo Estatal y de la Unión Europea, financiado por la Unión Europea-NextGenerationEU.

## REFERENCIAS BIBLIOGRÁFICAS

- Adrakatti, V. (2024). *Exploring screen summarization with large language and multimodal models* [Tesis de maestría, University of Illinois Urbana-Champaign]. <https://www.ideals.illinois.edu/items/131510>
- Alvarez-Vidal, S., Oliver, A., y Badia, T. (2021). What do post-editors correct? A fine-grained analysis of SMT and NMT errors. *Revista Tradumàtica Technologies de la Traducció*, 19, 131-147. <https://doi.org/10.5565/rev/tradumatica.286>
- Baldo, M., Evans, J., y Guo, T. (13 de diciembre de 2023). Investigating LGBTQ+ issues in translation. *LSE Review of Books*. <https://blogs.lse.ac.uk/lsereviewofbooks/2018/12/13/feature-investigating-lgbtq-issues-in-translation-by-michela-baldo-jonathan-evans-and-ting-guo/>
- Balyan, R., y Chatterjee, N. (2018). Factor-based evaluation for English to Hindi MT outputs. *Language Resources and Evaluation*, 52(4), 969-996. <https://doi.org/10.1007/s10579-018-9426-y>
- Banerjee, S., y Lavie, A. (2005). METEOR: An Automatic Metric for MT Evaluation with Improved Correlation with Human Judgments. En *Proceedings of the ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization* (pp. 65-72). Association for Computational Linguistics. <https://aclanthology.org/W05-0909>
- Bravo-Candel, D., López-Hernández, J., García-Díaz, J. A., Molina-Molina, F., y García-Sánchez, F. (2021). Automatic Correction of Real-Word Errors in Spanish Clinical Texts. *Sensors*, 21(9), 2893. <https://doi.org/10.3390/s21092893>
- Briva-Iglesias, V., Cavalheiro Camargo, J. L., y Dogru, G. (2024). Large language models “Ad Referendum”: How good are they at machine translation in the legal domain? *MonTI*, 16, 75-107. <https://doi.org/10.6035/MonTI.2024.16.02>
- Calixto, I., y Liu, Q. (2019). An error analysis for image-based multi-modal neural machine translation. *Machine Translation*, 33, 155-177. <https://doi.org/10.1007/s10590-019-09226-9>
- Cámara de los Comunes (15 de junio de 2023). Debate on Pride Month. [Transcripción]. UK Parliament. <https://hansard.parliament.uk/Commons/2023-06-15/debates/EA123D7E-5A79-46AC-9EDE-36C945F1FF84/PrideMonth>
- Chen, Z., Xu, L., Zheng, H., Chen, L., Tolba, A., Zhao, L., Yu, K., y Feng, H. (2024). Evolution and Prospects of Foundation Models: From Large Language Models to Large Multimodal Models. *Computers, Materials & Continua/Computers, Materials & Continua (Print)*, 80(2), 1753-1808. <https://doi.org/10.32604/cmc.2024.052618>
- Ciora, C., Iren, N., y Alikhani, M. (2021). Examining covert gender Bias: A case study in Turkish and English machine translation models. En A. Belz, A. Fan, E. Reiter, y Y. Sripada (Eds.), *Proceedings of the 14th International Conference on Natural Language Generation* (Vols. 55-63). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2021.inlg-1.7>
- Dwivedi, R. K., Nand, P., y Pal, O. (2024). Hybrid NMT model and comparison with existing machine translation approaches. *Multidisciplinary Science Journal*, 7(4), 2025146. <https://doi.org/10.31893/multiscience.2025146>
- Eagle, A. (15 de junio de 2023). Intervención en Pride Month. En *Cámara de los Comunes*. [Transcripción]. UK Parliament. <https://hansard.parliament.uk/Commons/2023-06-15/debates/EA123D7E-5A79-46AC-9EDE-36C945F1FF84/PrideMonth>
- Ferrag, F., y Bentounsi, I. A. (2024). The use of Artificial intelligence in academic translation tasks Case study of Chat GPT, Claude and Gemini. *Ziglobitha*, 2(11), 173-192. <https://doi.org/10.60632/ziglobitha.n011.11.vol.2.2024>
- Fitria, T. N. (2021). Gender Bias in translation using Google Translate: Problems and solution. *Language Circle: Journal of Language and Literature*, 15(2). [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3847487](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3847487)

- Forcada, M. L. (2017). Making sense of neural machine translation. *Translation Spaces*, 6(2), 291-309. <https://doi.org/10.1075/ts.6.2.06for>
- Fulgu, R. A., y Capraro, V. (2024). Surprising gender biases in GPT. *Computers in Human Behavior Reports*, 16, 100533. <https://doi.org/10.1016/j.chbr.2024.100533>
- Ghosh, S., y Caliskan, A. (2023). ChatGPT perpetuates gender bias in machine translation and ignores Non-Gendered Pronouns: Findings across Bengali and five other Low-Resource languages. En F. Rossi, S. Das, J. Davis, K. Firth-Butterfield, y A. John (Eds.), *AIES '23: Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society* (pp. 901-912). Association for Computing Machinery. <https://doi.org/10.1145/3600211.3604672>
- Gromann, D., Lardelli, M., Spiel, K., Burtscher, S., Klausner, L. D., Mettinger, A., Miladinovic, I., Schefer-Wenzl, S., Duh, D., y Bühn, K. (2023). Participatory research as a path to community-informed, gender-fair machine translation. En Vanmassenhove, B. Savoldi, L. Bentivogli, J. Daems, y J. Hackenbuchner (Eds.), *Proceedings of the First Workshop on Gender-Inclusive Translation Technologies* (pp. 49-59). European Association for Machine Translation. <https://aclanthology.org/2023.gitt-1.5>
- International School of Navarra. (21 de diciembre de 2023). *Equivalencia entre cursos UK y España - TBSON*. British School of Navarra. <https://isnavarra.es/moving-to-spain-with-children-check-our-table-of-grade-equivalencies-between-british-and-spanish-schools-learn-which-school-year-they-belong-to/>
- Johnson, M. (22 de abril de 2020). A scalable approach to reducing gender bias in Google Translate. *Google Research*. <https://research.google/blog/a-scalable-approach-to-reducing-gender-bias-in-google-translate/>
- Kaplan, D. M., Palitsky, R., Arconada Alvarez, S. J., Pozzo, N. S., Greenleaf, M. N., Atkinson, C. A., y Lam, W. A. (2024). What's in a name? Experimental evidence of gender bias in recommendation letters generated by ChatGPT. *Journal of Medical Internet Research*, 26, e51837. <https://doi.org/10.2196/51837>
- Koponen, M., y Salmi, L. (2017). Post-editing quality: Analysing the correctness and necessity of post-editor corrections. *Linguistica Antverpiensia New Series – Themes in Translation Studies*, 16, 137-148. <https://doi.org/10.52034/lanstts.v16i0.439>
- Lommel, A., Gladkoff, S., Melby, A., Wright, S. E., Strandvik, I., Gasova, K., Vaasa, A., Benzo, A., Marazzato Sparano, R., Foresi, M., Innis, J., Han, L., y Nenadic, G. (2024). The Multi-Range Theory of Translation Quality Measurement: MQM scoring models and Statistical Quality Control. En M. Martindale, J. Campbell, K. Savenkov, y S. Goel (Eds.), *Proceedings of the 16th Conference of the Association for Machine Translation in the Americas (Volume 2: Presentations)* (pp. 75-94). Association for Machine Translation in the Americas. <https://aclanthology.org/2024.amta-presentations.6>
- Lommel, A., Uszkoreit, H., y Burchardt, A. (2014). Multidimensional Quality Metrics (MQM): A framework for declaring and describing translation quality metrics. *Revista Tradumàtica Tecnologies de La Traducció*, 12, 455-463. <https://doi.org/10.5565/rev/tradumatica.77>
- MQM Council. (s.f.). *The MQM CORE Typology*. The MQM. <https://themqm.org/the-mqm-typology>
- OpenAI. (s.f.). *Models*. Open AI Platform. <https://platform.openai.com/docs/models#gpt-4o>
- Papineni, K., Roukos, S., Ward, T, y Zhu W. (2002). BLEU: a method for automatic evaluation of machine translation. En *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*. Association for Computational Linguistics, USA, 311-318. <https://doi.org/10.3115/1073083.1073135>
- Peñuelas Gil, I. (2024). La traducción automática de expresiones multiverbales en el turismo gastronómico: Un estudio de caso. *Hikma*, 23(3), 1-27. <https://doi.org/10.21071/hikma.v23i3.16992>

- Piazzolla, S. A., Savoldi, B., y Bentivogli, L. (2024). Good, but not always Fair: An Evaluation of Gender Bias for three Commercial Machine Translation Systems. *HERMES - Journal of Language and Communication in Business*, 63, 209-225. <https://doi.org/10.7146/hjlc.vi63.137553>
- Prates, M. O. R., Avelar, P. H., y Lamb, L. C. (2019). Assessing gender bias in machine translation: a case study with Google Translate. *Neural Computing and Applications*, 32(10), 6363-6381. <https://doi.org/10.1007/s00521-019-04144-6>
- Sánchez Ramos, M., y Rico Pérez, C. (2020). *Traducción automática: conceptos clave, procesos de evaluación y técnicas de posesición*. Comares.
- Sánchez-Gijón, P., Moorkens, J., y Way, A. (2019). Post-editing neural machine translation versus translation memory segments. *Machine Translation*, 33(1-2), 31-59. <https://doi.org/10.1007/s10590-019-09232-x>
- Simon, S. (1996). *Gender in translation*. Routledge. <https://shorturl.at/VaMog>
- Subramanya, A. (25 de julio de 2024). Gemini's big upgrade: Faster responses with 1.5 Flash, expanded access and more. *The Keyword*. <https://blog.google/products/gemini/google-gemini-new-features-july-2024/>
- Thelwall, M., Devonport, T. J., Makita, M., Russell, K., y Ferguson, L. (2022). Academic LGBTQ+ Terminology 1900-2021: Increasing Variety, Increasing Inclusivity? *Journal of Homosexuality*, 70(11), 2514-2538. <https://doi.org/10.1080/00918369.2022.2070446>
- Turovsky, B. (16 de noviembre de 2016). Found in translation: More accurate, fluent sentences in Google Translate. *Google*. <https://blog.google/products/translate/found-translation-more-accurate-fluent-sentences-google-translate/>
- Ullmann, S. (2022). Gender bias in machine translation systems. En K. Richardson y T. Heffernan (Eds.), *Social and cultural studies of robots and AI* (pp. 123-144). Palgrave Macmillan. [https://doi.org/10.1007/978-3-030-88615-8\\_7](https://doi.org/10.1007/978-3-030-88615-8_7)
- Von Flotow, L. (1997). *Translation and Gender: Translating in the «Era of Feminism»*. University of Ottawa Press. <https://doi.org/10.1353/book6616>
- Wang, L., Lyu, C., Ji, T., Zhang, Z., Yu, D., Shi, S., Tu, Z., y Lab, T. A. (2023). Document-Level Machine Translation with Large Language Models. En H. Bouamor, J. Pino, y K. Bali (Eds.), *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing* (pp. 16646-16661). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2023.emnlp-main.1036>
- Wu, Y., et al. (2016). Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation. *arXiv*. <https://doi.org/10.48550/arxiv.1609.08144>