



Universidad de Valladolid

FACULTAD DE TRADUCCIÓN E INTERPRETACIÓN

GRADO EN TRADUCCIÓN E INTERPRETACIÓN

TRABAJO FIN DE GRADO

Análisis de errores de traducción automática EN > ES en una receta marroquí: comparación entre un sistema de TA neuronal y un LLM

Presentado por Wisal Saidi Saidi

Tutelado por la Dra. M.^a Teresa Ortego Antón

Soria, junio de 2025

A mi familia, por su apoyo y amor incondicional.

A mi tutora, María Teresa Ortego Antón, por su ayuda y orientación académica a lo largo de todo el proceso.

Índice

RESUMEN.....	4
ABSTRACT	4
1. Introducción.....	1
1.1. Competencias	2
1.2. Objetivos	2
1.3. Desarrollo del estudio.....	2
2. Marco teórico	3
2.1. Definición y conceptos de la traducción automática.....	3
2.2. Sistemas de traducción automática.....	4
2.2.1 Sistemas de traducción automática basados en reglas (RBMT).....	4
2.2.2 Sistemas de traducción automática basados en estadística (SMT).....	6
2.2.3 Sistemas de traducción automática basados en ejemplos	6
2.2.4 Sistemas de traducción automática basados en redes neuronales (NMT)	8
2.3. Grandes modelos de lenguaje (LLM).....	9
2.4. Traducción agroalimentaria	10
3. Hipótesis y metodología.....	12
3.1. La selección de los sistemas de traducción automática.....	12
3.2. La muestra de análisis	13
3.3. Metodología del estudio y criterios de evaluación	15
4. Análisis y resultados.....	18
4.1. Análisis de errores de ChatGPT	18
4.2. Análisis de errores de DeepL.....	20
4.3. Comparación entre sistemas de traducción automática.....	22
5. Conclusiones.....	25
6. Bibliografía.....	26

ÍNDICE DE GRÁFICOS

Gráfico 1. Errores en ChatGPT.	18
-------------------------------------	----

Gráfico 2. Errores de DeepL.	20
Gráfico 3. Comparación de subcategorías de errores por sistemas de TA.	22

ÍNDICE DE TABLAS

Tabla 1. Clasificación y definición de errores de traducción según el modelo MQM core.....	17
Tabla 2. Análisis de errores por segmento.....	17
Tabla 3. Errores de ChatGPT.....	18
Tabla 4. Errores de DeepL.....	20
Tabla 5. Recuento de errores por subcategoría de cada sistema de TA.	22
Tabla 6. Errores más recurrentes en ambos sistemas.....	24

RESUMEN

La traducción automática (TA), también conocida como *Machine Translation* (MT, por sus siglas en inglés) tiene sus inicios a mediados del siglo XX, cuando empezaron a desarrollarse los primeros sistemas basados en reglas lingüísticas y ha evolucionado hasta los modelos neuronales actuales y los modelos de lenguaje de gran tamaño (*Large Language Models, LLM*), que son capaces de generar traducciones más fluidas y acertadas. A pesar de estos avances, la TA sigue teniendo limitaciones a la hora de transferir interlingüísticamente textos que contienen una carga cultural, ya que este tipo de contenidos requieren un conocimiento cultural amplio. En este contexto se enmarca el presente Trabajo de Fin de Grado (TFG), que, dentro del área de las tecnologías de la traducción e interpretación inglés↔español, analizará el rendimiento de una selección de sistemas de traducción automática al tratar con textos culturalmente marcados. A partir de una muestra de análisis compuesta por una receta de cocina, se identificarán errores con el uso de una métrica de evaluación adaptada de *MQM core* (*Multidimensional Quality Metrics*) y se reflexionará sobre la capacidad real de estas herramientas para afrontar los desafíos culturales en el proceso de traducción.

Palabras clave: traducción automática, receta, error, MQM, inglés, español.

ABSTRACT

Machine translation (MT) has its origin in the mid-20th century, when the first rule-based systems began to emerge. Since then, it has evolved significantly, giving rise to current neural machine translation systems and, more recently, to Large Language Models (LLMs), which can produce more fluent and accurate translations. Despite these technological advances, machine translation continues to face considerable limitations when dealing with culturally embedded content, as such texts require deep cultural awareness and contextual understanding. This research, situated within the field of English↔Spanish translation and interpreting technologies, explores the performance of several machine translation systems when applied to culturally rich texts. Based on a case study involving a traditional recipe, the analysis draws on an adapted version of the MQM core (Multidimensional Quality Metrics) framework to identify and categorize translation errors.

Keywords: machine translation, recipe, error, MQM, English, Spanish.

1. Introducción

No fue hasta finales de los años cincuenta y principios de los sesenta que la traducción automática comenzó a despertar un interés real, cuando se hicieron visibles las posibilidades de combinar el lenguaje con la informática. Algunos antecedentes teóricos se remontan al siglo XVII, cuando filósofos como Descartes y Leibniz hicieron propuestas en torno a lenguajes universales y lógicos. Sin embargo, la idea de crear un sistema mecánico de traducción no se materializó hasta el siglo XX. Desarrollaron de forma independiente los primeros prototipos en 1933, George Artsrouni en Francia y Petr Troyanskii en Rusia, aunque aún muy alejados de lo que hoy entendemos por traducción automática (Hutchins, 2006, p. 1). Estos sistemas han ido evolucionando a lo largo del tiempo desde enfoques puramente basados en reglas hacia modelos estadísticos, y en la actualidad, hacia modelos neuronales que buscan aproximarse a una traducción más fluida y contextual. Esta evolución, sus fundamentos y sus principales características se abordarán con mayor profundidad en el marco teórico.

En este Trabajo de Fin de Grado, vinculado a las tecnologías de la traducción e interpretación inglés <español, nos propondremos a analizar este tipo de dificultades a partir del estudio de textos que presentan precisamente este tipo de retos. En concreto, centramos nuestra atención en la traducción automática de una receta marroquí, al tratarse de un texto en el que confluyen elementos lingüísticos, sociales y culturales que resultan complejos de trasladar automáticamente sin incurrir en pérdidas o distorsiones del significado.

En una época en la que el uso de herramientas de traducción automática se ha generalizado, incluso en entornos profesionales, consideramos pertinente cuestionar en qué medida estos sistemas son realmente capaces de manejar materiales que requieren algo más que una simple correspondencia gramatical entre lenguas. Como señala Downie (2020, p. 40), si bien estas tecnologías pueden imitar ciertos patrones lingüísticos, aún están lejos de reproducir los procesos interpretativos que caracterizan la práctica traductora humana, especialmente cuando entran en juego aspectos culturales o pragmáticos.

Mediante el análisis de distintos productos generados por sistemas de traducción automática, este trabajo aspira a ofrecer una visión crítica sobre su rendimiento ante un tipo de texto que, aunque aparentemente sencillo, encierra una notable riqueza contextual. Las recetas de origen marroquí, cada vez más difundidas gracias al turismo y la globalización, exigen una traducción que respete no solo el contenido informativo, sino también las particularidades culturales que les confieren su identidad. En este sentido, resulta pertinente indagar si las tecnologías actuales son capaces de preservar dicha dimensión o si, por contrario, tienden a neutralizarla.

1.1. Competencias

A lo largo del Grado en Traducción e Interpretación, se han adquirido y consolidado una serie de competencias básicas, generales, transversales y específicas que han permitido afrontar con solvencia el desarrollo del Trabajo de Fin de Grado.

Entre las competencias básicas y generales, se han desarrollado especialmente CB1, CB2, CB5 y CB6 a lo largo del Grado. Durante la realización del TFG, se han reforzado CB3 y CB4.

En cuanto a las competencias transversales, T1, T2 y T5 han sido fundamentales durante el desarrollo del TFG. T3 y T4 han sido esenciales a lo largo del Grado.

Respecto a las competencias específicas, muchas de ellas se han trabajado de forma continuada a lo largo del Grado. En particular, E1, E2 y E3. Para el desarrollo del TFG han sido esenciales E6 E7, E8, E10, E13, E20 y E22.

1.2. Objetivos

El objetivo principal de este Trabajo de Fin de Grado es comprobar la traducción que ofrece un sistema de TA basado en redes neuronales y un *Large Language Model* (LLM) con el fin de evaluar su rendimiento en la traducción de una receta de cocina. Para ello, se utilizarán dos herramientas automáticas, con el propósito de identificar los errores más recurrentes y valorar cuál de los sistemas utilizados ofrece mejores resultados desde el punto de vista lingüístico y cultural. Con el fin de alcanzar el objetivo general anteriormente formulado, resulta necesario detectar los errores más frecuentes cometidos por los TA en textos con referencias culturales, analizar el comportamiento traductológico de diferentes sistemas ante una misma unidad textual, comparar los resultados de las distintas herramientas y evaluar hasta qué punto los sistemas de TA son fiables.

1.3. Desarrollo del estudio

El presente trabajo se articula en varias etapas que permiten abordar de forma progresiva los distintos aspectos implicados en el análisis. En primer lugar, se presentarán los conceptos fundamentales de la traducción automática, de su evolución histórica y las diferencias entre los modelos neuronales actuales y los LLM. A continuación, se contextualizará el objeto de estudio: las recetas de cocina como género textual, con especial énfasis en su dimensión cultural y en los desafíos que presentan para la traducción. En una tercera etapa, se describirán las herramientas de traducción automática seleccionadas para el análisis. Por último, se expondrán los resultados obtenidos en diferentes sistemas una vez analizados detalladamente los errores y se acompañará de una reflexión crítica sobre el grado de adecuación y eficacia de cada herramienta en la gestión de los contenidos culturales del texto.

2. Marco teórico

En este capítulo se abordan los conceptos clave relacionados con la traducción automática (TA), su evolución y los diferentes sistemas que han surgido a lo largo del tiempo. En primer lugar, encontraremos un apartado en el que se define la traducción automática y se destacan las aportaciones de diversos autores que subrayan los aspectos técnicos y lingüísticos del proceso. Posteriormente, se analiza su evolución desde los primeros sistemas basados en reglas (RBMT), en estadística (SMT), ejemplos (EBMT) y redes neuronales (NMT). Además, en cada apartado se detallarán sus características y métodos de funcionamiento. Por último, dado que el texto elegido para el análisis de errores de los sistemas de traducción corresponde a una receta, consideramos necesario definir el concepto de traducción agroalimentaria y ofrecer una reflexión de las posibles dificultades que nos pueden surgir en este tipo de traducciones.

2.1. Definición y conceptos de la traducción automática

Antes de abordar el análisis específico de traducción automática que se aplican textos con carga cultural, es imprescindible establecer qué se entiende por traducción automática (TA). A lo largo de las últimas décadas, distintos autores han ofrecido definiciones, que aunque coinciden en lo fundamental, destacan distintos aspectos del proceso. En este apartado se recogen tres de las definiciones más representativas, que permiten enmarcar teóricamente el objetivo de estudio del presente trabajo.

Sánchez Ramos y Rico Pérez (2020, p. 1) definen la traducción automática como: «la aplicación de la tecnología informática a la traducción de textos de una lengua a otra sin intervención humana». Esta definición nos muestra que el componente fundamental de la TA es el uso exclusivo de programas informáticos, sin la participación directa de un traductor humano. Esto implica una automatización completa del proceso de traducción. En una segunda definición, la autora amplía esta idea y señala que:

«la TA consiste en la traducción de una lengua a otra mediante un programa informático, que es, al fin y al cabo, quien realiza la tarea sin necesidad de que una persona (el traductor) participe en el proceso. No obstante, este proceso, aparentemente sencillo, implica varias operaciones como las que menciona Berner (2003, p. 6) cuando describe los primeros sistemas de TA» (Sánchez Ramos y Rico Pérez, 2020, p. 2)

Aquí se produce un matiz importante, aunque pueda parecer un procedimiento lineal, la TA implica una serie de operaciones complejas que van más allá de la simple sustitución de palabras. Precisamente, a partir de lo señalado por Sánchez Ramos y Rico Pérez (2020), resulta conveniente incluir la definición de Berner (2003), quien aclara que:

Machine Translation (MT) is the use of computer software to translate text or speech from one natural language into another. Like translation done by humans, MT does not simply involve

substituting words in one language for another, but the application of complex linguistic knowledge: morphology (how words are built from smaller units of meaning), syntax (grammar), semantics (meaning), and understanding of concepts such as ambiguity.

Con esta definición se pone de manifiesto que la traducción automática requiere de tratamiento de distintos niveles lingüísticos y no se limita a un reemplazo mecánico de unidades léxicas. Es decir, para que una traducción automática se acerque a una traducción humana, debe poder interpretar estructuras gramaticales, significados y ambigüedades de forma coherente.

Además de los aspectos técnicos y lingüísticos, otros autores han insistido en el carácter interdisciplinar de esta rama del conocimiento:

[MT] is an interdisciplinary enterprise that combines a number of fields of study such as lexicography, linguistics, computational linguistics, computer science and language engineering [...]. It is based on the hypothesis that natural languages can be fully described, controlled and mathematically coded [...]. (Quah, 2006, p. 57).

Esta definición es especialmente interesante porque nos permite entender la TA como un campo de investigación que requiere colaboración entre múltiples disciplinas.

2.2. Sistemas de traducción automática

Durante las últimas décadas, la traducción automática (TA) ha experimentado una evolución creciente que estaba impulsada por el desarrollo de distintos enfoques computacionales basados en paradigmas teóricos y técnicos diversos. Esta evolución dio lugar a una serie de sistemas que tienen el mismo objetivo: producir traducciones correctas y comprensibles. Sin embargo, difieren en algunos aspectos, como el mecanismo de análisis, la generación de texto y el grado de dependencia del conocimiento de los datos empíricos.

Como señalan Sánchez Ramos y Rico Pérez (2020, p. 11), los esfuerzos por mejorar la calidad de la traducción automática han derivado en la aparición de distintos modelos que incorporan técnicas específicas, cada uno con sus propias fortalezas.

En este apartado desarrollaremos los sistemas de traducción que han marcado la evolución de la traducción automática. Los ordenaremos de forma cronológica y siguiendo la clasificación que se ha adoptado tradicionalmente en los estudios especializados.

2.2.1 Sistemas de traducción automática basados en reglas (RBMT)

Los modelos de traducción automática basados en reglas, conocidos como *Rule-Based Machine Translation* (RBMT), dominaron las primeras décadas de desarrollo tecnológico en este campo. Estos

sistemas «parten de premisas eminentemente lingüísticas para codificar la información de las lenguas origen y meta mediante diccionarios y reglas gramaticales a partir de las cuales se realiza la traducción» (Sánchez Ramos y Rico Pérez, 2020, p. 12). Este enfoque se sustenta en una labor esencialmente humana, en la que, de manera deductiva, los expertos han pensado en cómo automatizar el proceso de traducción automática y han deducido la información necesaria para realizarla (Forcada *et al.*, 2016, p. 151).

Dentro de esta categoría, se distinguen tres modelos principales: los modelos de traducción directa, los modelos de transferencia y los modelos basados en interlengua.

Los primeros sistemas de traducción automática se encuadran en el grupo de los modelos de traducción directa, que realizan la traducción a partir de la consulta de los propios diccionarios bilingües del sistema, así como mediante el uso de reglas sencillas (Shiwen y Xiaojing, 2014; Oliver, 2016;). Estos modelos podrían equipararse a lo que actualmente denominamos traducción palabra por palabra, en la que sobresalen dos procesos fundamentales: la búsqueda del término equivalente de cada palabra en el diccionario bilingüe del sistema de tecnología de la información y la reestructuración de acuerdo con los esquemas de oración del idioma objetivo. Así pues, el análisis semántico y sintáctico no existe (Sánchez Ramos y Rico Pérez, 2020, p. 12). Como consecuencia, los resultados obtenidos por estos modelos no cumplieron las expectativas iniciales, lo que impulsó el desarrollo de alternativas más sofisticadas que evitaran el trasvase literal de términos.

En este contexto surgieron los modelos de transferencia, que incorporaban componentes sintácticos y semánticos de las lenguas implicadas y realizaban un análisis del texto de partida para dar como resultado «una estructura [de la lengua origen] que se transfiere mediante unas reglas a una estructura propia de la lengua de llegada» (Oliver, 2016, p. 164). La traducción en estos sistemas se divide en tres fases: análisis, transferencia y generación, en un proceso que, de alguna manera, se asemeja al propio proceso de traducción humana (Quah, 2006).

Por último, el modelo de interlengua propone un enfoque aún más abstracto. En él, el resultado del análisis se convierte en una representación abstracta, denominada interlengua, que es independiente de las lenguas de partida y llegada, eliminando así el proceso de transferencia (Forcada *et al.*, 2016). Esta representación intermedia captura la información semántica y sintáctica esencial del texto original para luego ser convertida a la lengua meta (Quah, 2006, p. 71). La interlengua, por tanto, funcionaría como un paso intermedio entre las lenguas (Sánchez Ramos y Rico Pérez, 2020, p. 14).

Sánchez Ramos y Rico Pérez (2020, p. 14) concluyen que «los sistemas de TA basados en reglas asumen que la traducción es un proceso que consiste en el análisis y la representación, clara y sencilla, del significado del texto en la lengua origen para así poder permitir que se genere su equivalente en la lengua meta».

2.2.2 Sistemas de traducción automática basados en estadística (SMT)

A partir de la década de 1990, el desarrollo de corpus paralelos amplios y el aumento de la potencia computacional dieron paso al enfoque conocido como *Statistical Machine Translation* (SMT). Su funcionamiento se fundamenta en la probabilidad de que una oración en la lengua origen (LO) tenga una correspondencia determinada en la lengua meta (LM), considerando que esta asociación se deduce a partir del análisis de grandes volúmenes de textos traducidos, conocidos como corpus (Bowker y Buitrago Ciro, 2019, p. 15).

A diferencia de los sistemas basados en reglas, que dependen de diccionarios y estructuras lingüísticas codificadas, los modelos estadísticos no integran conocimiento lingüístico explícito. En su lugar, «se centran en grandes volúmenes de material en lengua de origen y lengua meta, los corpus paralelos o “bitextos”, que son ejemplos reales en lengua origen y lengua meta de los que se adquiere el conocimiento necesario para producir las traducciones» (Sánchez Ramos y Rico Pérez, 2020, p. 16). Así, se sustituye el enfoque deductivo propio de los sistemas RBMT por una metodología de carácter inductivo, en la que «el sistema aprende automáticamente a traducir entre dos lenguas a partir de un corpus paralelo suficientemente grande de oraciones en LO [lengua origen] acompañadas de su traducción a la LM [lengua meta]» (Forcada *et al.*, 2016, p. 177).

Estos sistemas empezaron a reemplazar el desarrollo complejo y costoso de los modelos basados en reglas (Parra Escartín, 2018), partiendo de la idea de que una oración puede tener múltiples traducciones posibles (Sánchez Ramos y Rico Pérez, 2020, p. 16). Cada traducción se valora según su verosimilitud (Forcada *et al.*, 2016).

Aunque Weaver (1949) ya apuntó el potencial de la estadística en la traducción, fue en 1991 cuando el equipo del IBM TJ Watson Research Center aplicó el teorema de Bayes en el sistema *Candide*. Este modelo combinaba datos previos con nuevos textos para predecir futuras traducciones (Quah, 2006, p. 77), utilizando el corpus Hansard, una colección bilingüe de debates parlamentarios canadienses.

Como indica Oliver (2016, p. 177), estos sistemas manejan dos tipos de probabilidad:

- Que una secuencia sea válida en la LM.
- Que una unidad de la LO se traduzca como una unidad concreta en la LM.

2.2.3 Sistemas de traducción automática basados en ejemplos

Los sistemas de TA basados en ejemplos funcionan a partir de corpus paralelos, es decir, colecciones de oraciones en lengua origen y su correspondiente traducción, que sirven como modelos para futuras traducciones. A diferencia de los modelos estadísticos, no operan mediante cálculos de

probabilidad, sino mediante analogías: «aprenden cómo realizar el proceso de trasvase [...] de acuerdo a unos modelos predeterminados de oraciones alineadas» (Oliver, 2016, p. 166).

La idea central es reutilizar oraciones ya traducidas que sean similares a las nuevas para generar la versión traducida. Aunque el mecanismo recuerda al de las memorias de traducción, existe una diferencia esencial: en las memorias es el traductor quien selecciona y edita la sugerencia, mientras que en EBMT es el propio sistema el que produce automáticamente la nueva traducción a partir de ejemplos anteriores (Sánchez Ramos y Rico Pérez, 2020, p. 19).

Estos sistemas surgieron en Japón en la década de 1980. Nagao (1984, p. 175) formuló sus fundamentos teóricos afirmando que el ser humano no traduce una oración sencilla a través de un análisis profundo, sino que «la descompone en frases parciales, traduce cada una por analogía con ejemplos previos y luego recompone la oración completa».

Sánchez Ramos (2020) indica que el funcionamiento de EBMT implica tres fases principales:

- Correspondencia: el sistema busca en el corpus paralelo segmentos similares a los de la oración origen.
- Alineación: establece la relación entre los fragmentos en LO y sus equivalentes en LM.
- Recombinación: une los fragmentos traducidos para formar la nueva oración en la LM.

2.2.3 Sistemas de traducción automática híbridos

Los sistemas anteriores (por reglas, estadísticos o basados en ejemplos) presentan ventajas importantes, como la eliminación del trabajo manual de construir reglas o diccionarios. No obstante, también tienen un inconveniente significativo: no permiten corregir errores de traducción simplemente añadiendo una regla o entrada concreta (Sánchez Ramos y Rico Pérez, 2020, p. 20). Con el objetivo de superar esta desventaja, surgen los sistemas híbridos, que combinan características de los enfoques basados en corpus (como los estadísticos) con elementos de los sistemas basados en reglas. El propósito es aprovechar lo mejor de cada tipo y compensar sus puntos débiles, obteniendo así resultados más precisos (Sánchez Ramos y Rico Pérez, 2020, p. 20).

Existen diferentes formas de llevar a cabo esta combinación (Costa-jussà y Fonollosa, 2015; Oliver, 2016). Por un lado, hay sistemas que mantienen la estructura de los modelos por reglas, pero incorporan corpus para construir o mejorar los resultados, ponderar las salidas o realizar una post-edición estadística (Costa-jussà y Fonollosa, 2015, p. 5). Por otro lado, hay modelos que parten de una arquitectura estadística pero integran reglas lingüísticas o diccionarios durante su entrenamiento (Costa-jussà y Fonollosa, 2015, p. 6). Entre los ejemplos más conocidos de esta aproximación híbrida están *SYSTRAN11* y *PROMT 12* (Sánchez Ramos y Rico Pérez, 2020, p. 20).

2.2.4 Sistemas de traducción automática basados en redes neuronales (NMT)

A partir de 2014, con la aparición de nuevas arquitecturas de redes neuronales profundas, surgió el enfoque conocido como *Neural Machine Translation* (NMT). La traducción automática neuronal, descrita por Cho *et al.* (2014), representa el avance más reciente y significativo en el campo de la traducción automática. Este enfoque ha superado a los sistemas estadísticos, dominando actualmente tanto la investigación como la aplicación comercial en este ámbito (Sánchez Ramos y Rico Pérez, 2020, p. 21).

A diferencia de otros modelos, la TA neuronal se basa en redes neuronales artificiales y en técnicas propias de la inteligencia artificial, como el *deep learning*. Estas redes intentan emular el funcionamiento del cerebro humano mediante conexiones entre nodos que se activan en función de estímulos y aprenden a traducir mediante un proceso de entrenamiento intensivo con grandes volúmenes de datos (Sánchez Ramos y Rico Pérez, 2020, pp. 21-22). En concreto, estas redes se entrenan con corpus paralelos que contienen millones de oraciones traducidas, lo que les permite generar traducciones más fluidas y naturales.

Aunque los primeros intentos de desarrollar este tipo de traducción se remontan a los trabajos de Forcada y Neco (1997) y Castaño y Casacuberta (1997), no fue hasta 2016, con el lanzamiento del traductor neuronal de Google, cuando la TA neuronal se consolidó. La razón principal de este retraso fue la falta de capacidad computacional en las primeras etapas de investigación, una limitación que actualmente se ha superado gracias a la disponibilidad de hardware avanzado como las GPU (Peris Abril y Casacuberta Nolla, 2017; Parra Escartín, 2018).

La arquitectura más común en los sistemas de TA neuronal es la codificador-decodificador con un modelo de atención. Esta estructura está formada por redes neuronales recurrentes que procesan la frase origen y generan una traducción en la lengua meta. El codificador convierte la oración original en una representación vectorial, mientras que el decodificador, con ayuda del modelo de atención, genera la traducción alineando las palabras origen con las palabras destino (Peris Abril y Casacuberta Nolla, 2017, p. 15).

A pesar de sus avances, la TA neuronal no está exenta de desafíos. Requiere una gran cantidad de datos paralelos para su entrenamiento y recursos computacionales potentes. Sin embargo, sus resultados han superado las expectativas iniciales y representan un gran paso respecto a los modelos anteriores (Sánchez Ramos y Rico Pérez, 2020, p. 22). Hoy en día, herramientas ampliamente utilizadas como Google Translate o DeepL se basan en este tipo de tecnología.

2.3. Grandes modelos de lenguaje (LLM)

Los *Large Language Models* (LLM), o modelos de lenguaje de gran tamaño, constituyen una categoría de modelos fundacionales entrenados con volúmenes masivos de datos. Los LLM son capaces de comprender y generar lenguaje natural y otros tipos de contenido, esto les permite abordar una amplia variedad de tareas sin necesidad de un entrenamiento específico para cada una. Según IBM (2023), los LLM representan un cambio de paradigma en el desarrollo de aplicaciones basadas en lenguaje, ya que sustituyen el enfoque tradicional, que se centra en modelos específicos por tarea, por una arquitectura más versátil.

El artículo de IBM destaca que estos modelos han sido desarrollados paralelamente a avances en el aprendizaje automático, las redes neuronales, los algoritmos y la arquitectura *Transformer*, que constituye la base técnica de los LLM. Esta arquitectura, introducida por Vaswani *et al.* (2017), permite a los modelos procesar datos secuenciales (como el texto) de manera eficiente, mediante mecanismos de atención que capturan relaciones contextuales complejas entre tokens.

Language modeling is a long-standing task in natural language processing (Bengio et al., 2000; Mikolov et al., 2010; Khandelwal et al., 2020), which is a task to predict the probability of the next token. Transformer (Vaswani et al., 2017) basically is the backbone of existing LLMs. LLMs show great potential as a universal multitask learner (Zhu et al., 2023, p. 2).

Esta definición resalta tanto la naturaleza predictiva del *language modeling* como el papel central de los LLM como posibles aprendices universales de tareas múltiples.

En la actualidad, los LLM más populares incluyen GPT-3 y GPT-4 de OpenAI, PaLM de Google y LLaMA de Meta. Todos ellos se caracterizan por tener miles de millones de parámetros, lo que les permite capturar patrones lingüísticos complejos. Durante su entrenamiento, los modelos aprenden a predecir el siguiente token basándose en el contexto, con el uso de *corpora* extensos de texto. Este proceso implica la tokenización del texto, la transformación de los tokens en vectores numéricos (*embeddings*) y el ajuste de millones de parámetros internos mediante técnicas de aprendizaje no supervisado. Una vez entrenados, los LLM pueden generar textos coherentes y contextualmente relevantes, traducir idiomas, responder preguntas, resumir documentos y asistir en tareas como la redacción creativa. Además, su rendimiento puede mejorarse mediante técnicas como el *prompt engineering*, el *fine-tuning* o el aprendizaje por refuerzo con retroalimentación humana (*RLHF*).

El potencial de estos modelos se ha puesto a prueba de forma concreta en el ámbito de la traducción automática multilingüe (MMT). En su artículo, Zhu *et al.* (2023) evalúan la capacidad de traducción multilingüe de varios LLM populares, incluidos ChatGPT y GPT-4, en un conjunto de 102 lenguas y 606 combinaciones lingüísticas. Este estudio pone de relieve tanto las fortalezas como las

limitaciones actuales de los LLM aplicados a la MMT. Los autores constatan que las capacidades de traducción de los LLM siguen evolucionando y que GPT-4 alcanza un nuevo techo de rendimiento. No obstante, incluso este modelo presenta dificultades con lenguas poco representadas en los datos de entrenamiento (*low-resource languages*).

Asimismo, el análisis revela que los LLM muestran patrones de funcionamiento distintos a los de los sistemas tradicionales cuando se utilizan para traducción. Por ejemplo, durante el aprendizaje en contexto (*in-context learning*), los modelos pueden ignorar las instrucciones explícitas del enunciado y, en su lugar, beneficiarse de ejemplos multilingües como guía más eficaz en tareas de traducción con escasos recursos. Lo más relevante del estudio es que los LLM son capaces de adquirir habilidades de traducción de forma eficiente en términos de recursos, lo que refuerza su valor potencial en este campo: «*LLM can acquire translation ability in a resource-efficient way, which indicates the promising future of LLM in multilingual machine translation*» (Zhu et al., 2023, p. 9).

Estos resultados confirman que los LLM no solo representan una evolución tecnológica en el tratamiento del lenguaje, sino que también abren nuevas posibilidades para la traducción automática multilingüe, especialmente en contextos donde los recursos lingüísticos son limitados.

2.4. Traducción agroalimentaria

La traducción agroalimentaria puede definirse como una rama de la traducción especializada que se ocupa de los textos producidos en el sector agroalimentario, el cual comprende actividades del sector primario (agricultura, ganadería, pesca, silvicultura), secundario (transformación de materias primas) y terciario (comercio, turismo, hostelería, etc.) (Rivas Carmona & Veroz González, 2018, p. 8). Esta amplia cobertura implica una notable diversidad textual: textos técnicos, científicos, publicitarios o divulgativos, entre otros, lo que convierte a este ámbito en un terreno fértil para la práctica traductora. Además, su creciente relevancia en los mercados internacionales refuerza la necesidad de una traducción precisa y adecuada tanto desde un punto de vista lingüístico como cultural.

Tal y como destaca Ortego Antón (2024a, p. 70), «*one of the pillars of the Spanish economy is the agri-food industry and, more specifically, the meat industry, which has become the most relevant one in terms of turnover and direct employment*». Asimismo, «*the main economic engines of rural regions located in Southern Sparsely Populated Areas (SSPA) are meat companies, which have a family structure, a small size, and a farming tradition*» (Ortego Antón, 2024a, p. 70). Debido a su voluntad de internacionalizar sus productos, estas empresas requieren servicios de redacción y traducción del español al inglés.

Ahora bien, aunque podría pensarse que los sistemas de traducción automática neuronal son una solución adecuada para las pequeñas y medianas empresas, lo cierto es que:

Genres are characterized by different patterns depending on the target culture. Hence, the promotion of a given product needs to be sensitive to cross-cultural differences to guarantee that target texts satisfy the standards and expectations of the target community, not only regarding the meaning but also register, style, geographical variant, etc. (Durán Muñoz & Corpas Pastor, 2020, p. 164).

Para afrontar estos retos, el grupo de investigación ACTRES ha emprendido estudios contrastivos hispano-ingleses basados en corpus, como el corpus CLANES o los estudios centrados en productos como el vino, el aceite de oliva, las infusiones o los embutidos, entre otros (Rabadán *et al.*, 2021a, 2021b; Moreno Pérez & López Arroyo, 2021; Ortego Antón, 2024b).

Otro aspecto que genera dificultad durante el trasvase interlingüístico de contenido agroalimentario, como destacan Zhang y Torres-Hostench (2024), es el de los retos culturales (*cultural challenges*), especialmente cuando el texto debe ayudar al consumidor a decidir si desea probar un plato: «*The translation should help the customer to decide whether to order the dish or not. If possible, it would be better to enumerate the main ingredients and the cooking method. Thus, the translation may entail including new items (e.g., cooking method) that were not in the original name of the dish*» (Zhang y Torres-Hostench, 2024, p. 54). Esta recomendación resulta directamente aplicable a nuestro caso, ya que la traducción de recetas marroquíes puede requerir la incorporación de información adicional que no siempre está presente en el original, con el fin de hacer el plato comprensible y atractivo para el público hispanohablante.

Además, estas autoras amplían el marco de dificultades más allá de los ingredientes poco comunes, que tradicionalmente se han considerado el principal obstáculo de la traducción gastronómica. Subrayan que existen otras variables, además de las mencionadas anteriormente (el exotismo y la adecuación cultural), como la precisión lingüística y la palatabilidad (capacidad de hacer que el plato resulte apetecible) (Zhang y Torres-Hostench, 2024, p. 54). En nuestro trabajo, centrado en recetas de la tradición culinaria marroquí, consideramos que los desafíos más relevantes serán precisamente *exoticism* y *cultural accuracy*, ya que ambos influyen directamente en cómo se percibe el plato traducido. Aunque dicho trabajo se refiera en concreto a la traducción de la gastronomía china, su análisis proporciona herramientas conceptuales y estrategias que pueden extrapolarse a otras tradiciones culinarias.

En consecuencia, esta información puede ayudarnos a anticipar y abordar las dificultades específicas que surgirán en la traducción de recetas marroquíes, un tipo de texto profundamente marcado por elementos culturales y denominaciones propias que no siempre tienen equivalencia directa en la lengua de llegada.

3. Hipótesis y metodología

Tras haber expuesto el marco teórico que sustenta este trabajo, en el presente capítulo enunciamos la hipótesis y la metodología que guiarán nuestro análisis. En primer lugar, se indicará qué sistemas de traducción automática serán empleados para traducir el texto seleccionado, que en este caso corresponde a una receta marroquí. A continuación, se describirán las características lingüísticas y culturales de dicha receta, se destacarán los elementos que pueden presentar mayor complejidad para los traductores automáticos. Por último, se explicará el procedimiento metodológico adoptado, basado en la aplicación de una métrica de evaluación por errores, el modelo MQM *Core* (*Multidimensional Quality Metrics*). Esta herramienta ofrece un enfoque sistemático y preciso para valorar la calidad de las traducciones generadas, lo cual resultará esencial para el análisis comparativo que desarrollaremos en los siguientes apartados. Con ello, se cierra la parte teórica y se da paso a la fase metodológica del estudio.

3.1. La selección de los sistemas de traducción automática

A la hora de seleccionar los sistemas de traducción automática que analizaremos en este trabajo, podríamos habernos decantado por alguno de los tres enfoques principales que han marcado la evolución de la traducción automática: los sistemas basados en reglas, los estadísticos o los híbridos. Sin embargo, hemos optado por centrarnos exclusivamente en los sistemas basados en redes neuronales, por ser actualmente los que ofrecen mejores resultados tanto desde un punto de vista lingüístico como funcional. Tal y como señalan Sánchez Ramos y Rico Pérez (2020, p. 26), la traducción automática neuronal supone «un paso más allá de lo hasta ahora investigado en TA», y presenta varias ventajas fundamentales frente a los sistemas anteriores. En primer lugar, destaca por proporcionar una mayor calidad en los resultados en términos de fluidez. Diversos estudios (Lommel, 2017) coinciden en que la TA neuronal genera traducciones que suenan más naturales que las producidas por sistemas estadísticos. No obstante, como se advierte en la literatura, una traducción fluida no es suficiente si no transmite fielmente el contenido original, por lo que la precisión sigue siendo un criterio indispensable.

En segundo lugar, los sistemas neuronales pueden ofrecer buenos resultados sin necesidad de contar con grandes volúmenes de datos paralelos, esto los convierte en herramientas útiles en contextos con recursos lingüísticos limitados o lenguas menos representadas.

Por último, cabe destacar la capacidad de estos sistemas para realizar traducciones entre pares de lenguas que no han sido explícitamente entrenadas, gracias a un fenómeno conocido como *zero-shot translation*. Como explica Lommel (2017, p. 3): «*For example, if you do not have Greek–Finnish training data, but you do have Greek–English and English–Finnish, the system can combine them to translate from Greek to Finnish without going via a pivot language*».

Estas características justifican sobradamente la elección de los sistemas neuronales como objeto de análisis en este trabajo. En concreto, hemos seleccionado dos herramientas representativas de este enfoque: DeepL y ChatGPT. Ambas se basan en redes neuronales profundas y están ampliamente reconocidas por su capacidad de adaptación a distintos contextos, así como por su accesibilidad, ya que ofrecen versiones gratuitas que permiten a cualquier usuario emplearlas sin coste alguno.

A continuación, describiremos de forma breve cada uno de los sistemas seleccionados:

- ChatGPT es un modelo de lenguaje de inteligencia artificial desarrollado por OpenAI, basado en la arquitectura de *Transformer*. Surgió como parte de los avances en los modelos *Generative Pre-trained Transformer* (GPT). Utiliza grandes volúmenes de texto para su entrenamiento para mejorar su capacidad de comprensión y generación de lenguaje natural sin supervisión humana directa. Su impacto ha sido notable, especialmente en campos como la educación. No obstante, el modelo enfrenta retos en cuanto a la precisión de sus respuestas, la ética del uso de IA y la necesidad de marcos regulatorios adecuados (Diego Olite *et al.*, 2023).
- DeepL es una herramienta de traducción automática desarrollada por la empresa alemana DeepL SE, ubicada en Colonia. Lanzada en agosto de 2017, ha destacado rápidamente por su alta calidad de traducción y ha establecido nuevos estándares en el ámbito de la traducción automática. El traductor de DeepL se basa en redes neuronales artificiales, que son entrenadas con millones de textos traducidos para ofrecer traducciones precisas (Kutyłowski, 2017). Sin embargo, su ventaja competitiva frente a otros traductores automáticos de grandes empresas tecnológicas radica en las innovaciones metodológicas introducidas por su equipo de investigación, particularmente en cuatro áreas clave del desarrollo de redes neuronales, lo que mejora significativamente la calidad de las traducciones generadas (DeepL).

Una vez descritos los sistemas que emplearemos en nuestro análisis, procedemos a delimitar la muestra de análisis.

3.2. La muestra de análisis

Para llevar a cabo la evaluación de los dos sistemas de traducción automática seleccionados, se ha optado por seleccionar un texto perteneciente al ámbito de la traducción especializada en el sector alimentario, concretamente una receta de cocina. La elección de este tipo de texto responde al creciente interés por analizar cómo se comportan los motores de traducción automática ante contenidos marcadamente culturales y de carácter instructivo. En este caso, hemos seleccionado una receta tradicional ma-

rroquí conocida como *Moroccan Diced Vegetable Soup, Chorba Fassia*, escrita por Nada Kiffa y publicada en el portal gastronómico *Taste of Maroc*)¹. Esta página web cuenta con una sección dedicada exclusivamente a platos típicos marroquíes, donde se recopilan algunas de las recetas más representativas del país, explicadas paso a paso. Cabe señalar que el sitio no ofrece opciones de traducción a otros idiomas, por lo que se deduce que los textos han sido redactados originalmente en inglés y no existen versiones oficiales en español. Esto nos permite trabajar con un texto auténtico y sin intervención previa de traductores humanos, pues favorece una evaluación más objetiva del rendimiento de los sistemas automáticos.

La receta seleccionada destaca por su contenido cultural, es un texto idóneo para analizar cómo los traductores automáticos manejan elementos culturales específicos, ingredientes poco comunes y técnicas de cocina tradicionales. El texto cumple con las convenciones lingüísticas propias del género instructivo: ofrece instrucciones detalladas sobre cómo preparar un plato específico. Incluye una lista de ingredientes, utensilios, pasos secuenciados, tiempos y observaciones complementarias.

A continuación, pasaremos a describir las características comunicativas de la receta seleccionada. Como ya hemos mencionado anteriormente, el emisor es una autora marroquí experta en gastronomía, que combina el conocimiento culinario con una intención divulgativa y cultural. El receptor es un público general interesado en la cocina marroquí, posiblemente angloparlante, con un nivel medio de competencia en técnicas culinarias. No se trata de un público especializado, sino más bien de lectores que buscan experimentar con recetas tradicionales desde un enfoque accesible. El canal es escrito y digital, pues la receta se difunde a través de una página web que actúa como plataforma de divulgación gastronómica y cultural. Se favorece el uso de hipervínculos, imágenes, listas desplegables y otros recursos que complementan la información verbal. El registro del texto es semiformal: adopta una estructura normativa típica del género (introducción, ingredientes, pasos, notas) y utiliza terminología específica del ámbito culinario, aunque mantiene un tono accesible y cercano. Se alterna un lenguaje más técnico, por ejemplo, en la descripción de ingredientes o utensilios, con expresiones coloquiales o personales (*I'm more on the lazy side of the family*). En cuanto a la función comunicativa principal, predomina la función conativa o apelativa, ya que el objetivo del texto es instruir al lector en la preparación de un plato concreto. Esta función se manifiesta a través de verbos imperativos y estructuras secuenciales que indican qué acciones realizar. No obstante, también están presentes la función referencial al contextualizar la receta dentro de la tradición culinaria de Fez y explicar el origen y propiedades de ciertos ingredientes y la función expresiva, visible en las intervenciones de la autora.

Por su parte, el texto responde plenamente a las convenciones del género receta: incluye una lista precisa de ingredientes con medidas, instrucciones en orden cronológico y tiempos de cocción. Asimismo, incorpora explicaciones sobre ingredientes culturalmente marcados como el *smen*, esto

¹ <https://tasteofmaroc.com/moroccan-diced-vegetable-soup-chorba-fassia/> (fecha de consulta: 7 de abril de 2025).

acentúa la densidad cultural del contenido y lo convierte en un reto especialmente adecuado para la traducción automática.

En conjunto, la receta constituye un texto instructivo, procedimental y cultural, en el que confluyen elementos funcionales y etnográficos.

3.3. Metodología del estudio y criterios de evaluación

En cuanto al protocolo de recogida de datos, se traducirá el texto original mediante los dos sistemas de traducción automática seleccionados: ChatGPT y DeepL. A continuación, se llevará a cabo una evaluación comparativa de las traducciones generada con base en el modelo MQM (*Multidimensional Quality Metrics Core*)². Este modelo se define de la siguiente manera:

Multidimensional Quality Metrics (MQM) is a framework for analytic Translation Quality Evaluation (TQE). It can be applied to both human translation, machine translation and AI-generated translation. As part of a quality management system, MQM can be used to identify quality issues in translation products, classify them against a shared, open and standardized error typology, and generate quality measures that can be used to gauge how well the translation product meets quality requirements.

El modelo de MQM core consta de ocho categorías: *Terminology, Accuracy, Linguistic conventions, Style, Locale conventions, Audience appropriateness, Design and markup* y *Custom*. A continuación, pasaremos a definir brevemente cada categoría y sus respectivas subcategorías:

- *Terminology: «Errors arising when a term does not conform to normative subject field or organizational terminology standards or when a term in the target content is not the correct, normative equivalent of the corresponding term in the source content».* En esta categoría podemos encontrar errores de los siguientes tipos: *Inconsistent with terminology resource, inconsistent use of terminology* y *wrong term*.
- *Accuracy: «Errors occurring when the target content does not accurately correspond to the propositional content of the source text because of distortion, omission, or addition to the message».* Dentro de esta categoría, podemos encontrar las siguientes subcategorías: *Mistranslation, Overtranslation, Undertranslation, Addition, Omission, Do not translate* y *Untranslated*.
- *Linguistic conventions: «Errors related to the linguistic well-formedness of the text, including problems with grammaticality, idiomaticity, and mechanical correctness».* En esta categoría encontramos los siguientes tipos de errores: *Grammar, punctuation, spelling, Unintelligible, Character encoding* y *textual conventions*.

² <https://themqm.org/the-mqm-typology/> (fecha de consulta: 20 de abril de 2025)

- *Style*: «Errors occurring in a text that are grammatically acceptable but are inappropriate because they deviate from organizational style guides or exhibit inappropriate language style». En este apartado existen los siguientes tipos de errores: *Organization style*, *Third-party style*, *Inconsistent with external, reference*, *Language register*, *Awkward style*, *Unidiomatic style*, *Inconsistent style*.
- *Locale conventions*: «Errors occurring when the translation product violates locale-specific content or formatting requirements for data elements». En este apartado existen las siguientes subcategorías de error: *Number format*, *Currency format*, *Measurement format*, *Time format*, *Date format*, *Address format*, *Telephone format* y *Shortcut key*.
- *Audicende appropriateness*: «Error where content Inappropriately uses a culture-specific reference that will not be understandable to the intended audience». En esta subcategoría existen los siguientes tipos de errores: *Culture-specific reference* y *Offensive*.
- *Design and markup*: «Errors related to the physical design or presentation of a translation product, including character, paragraph, and UI element formatting and markup, integration of text with graphical elements, and overall page or window layout». En esta categoría podemos encontrar las siguientes subcategorías: *Layout*, *Markup tag*, *Truncation/text expansión*, *Missing text* y *Link/cross-reference*.
- *Custom*: «Any other issue».

En nuestro análisis de errores no emplearemos las ocho categorías que propone el modelo de MQM core, ya que al tratarse de una receta, muchos de los elementos de este modelo no aparecerán en nuestro texto. A continuación, indicaremos la adaptación de este modelo que utilizaremos para nuestro análisis de errores y dentro de estos modelos qué subcategorías se emplearán. Las categorías que emplearemos serán: *Terminology*, *Accuracy*, *Style* y *Locale conventions*. Para comenzar, en el apartado de *Terminology* utilizaremos la subcategoría de *wrong term* en los casos en los que exista incoherencia en el uso de la terminología o se empleen términos incorrectos. En cuanto a *Accuracy*, se emplearán las subcategorías de *Mistranslation*, *Addition*, *Omission* y *Untranslated*. Aquí se analizará la precisión cuando el contenido del texto meta no se traslade la manera correcta y conlleve a estos errores. En *Style* se evaluará la subcategoría de *Inconsistent style*, es decir, si el registro o el estilo a lo largo del texto varía. En *Linguistic conventions*, utilizaremos la subcategoría *Grammar*. Por último, en *Locale conventions*, analizaremos errores de *Measurement format* y *Time format*, puesto que se trata de una receta, podemos encontrar errores en elementos de medida y tiempos. A continuación, hemos creado una tabla en la que definiremos las categorías que se utilizarán según el modelo MQM core.

Terminology	Wrong term	Use of term that it is not the term a domain expert would use or because it gives rise to a conceptual mismatch.
-------------	------------	--

Accuracy	Mistranslation	Error occurring when the target content that does not accurately represent the source content.
	Addition	Error occurring in the target content that includes content not present in the source.
	Omission	Error where content present in the source is missing in the target.
	Untranslated	Error occurring when a text segment that was intended for translation is omitted in the target content.
Style	Inconsistent style	Characteristic of text that uses a level of formality higher or lower than required by the specifications or general language conventions.
Locale conventions	Measurement format	Inappropriate measurement format for its locale.
	Time format	Error involving incorrect time format for its locale.
Linguistic conventions	Grammar	Error that occurs when a text string (sentence, phrase, other) in the translation violates the grammatical rules of the target language.

Tabla 1. Clasificación y definición de errores de traducción según el modelo MQM *core*.

Todas estas categorías de errores se cuantificarán en una tabla comparativa para facilitar el análisis de los resultados y poder determinar el rendimiento de cada sistema.

Una vez descrita la tipología de errores, para llevar a cabo el análisis dividiremos la receta original en inglés en 31 segmentos. Cada uno de estos segmentos aparece en la columna nombrada «SEGMENTO EN». A continuación, se presentan las traducciones generadas por los sistemas ChatGPT y DeepL, seguidas de una clasificación de los errores indicados. Estos errores han sido categorizados según la tabla que hemos realizado basada en el modelo MQM *core*. En la Ttabla 2 se ofrece cómo clasificamos los errores de traducción más significativos, lo que permitirá establecer comparaciones entre ambos sistemas y valorar su grado de fiabilidad.

SEGMENTO EN	ES CON CHATGPT	ERROR 1 SEGÚN MQM	ERROR 2 SEGÚN MQM	ERROR 3 SEGÚN MQM	ES CON DEEPL	ERROR 1 SEGÚN MQM	ERROR 2 SEGÚN MQM	ERROR 1 SEGÚN MQM
-------------	----------------	-------------------	-------------------	-------------------	--------------	-------------------	-------------------	-------------------

Tabla 2. Análisis de errores por segmento.

Explicada la metodología de análisis procedemos a ofrecer los datos de su aplicación en el siguiente capítulo.

4. Análisis y resultados

Tras haber presentado la metodología de nuestro trabajo, procedemos a exponer el análisis de errores detectados en las traducciones generadas por los sistemas de traducción automática seleccionados.

4.1. Análisis de errores de ChatGPT

Se ha llevado a cabo un recuento del total de errores siguiendo la clasificación de categorías y subcategorías de la Tabla 1 para ChatGPT. Procedemos a describir los errores por categorías, de mayor a menor ocurrencias.

	ChatGPT
Terminology	16
Untranslated	7
Addition	2
Omission	3
Style	3
Measurement format	0
Time format	3
Mistranslation	16
Grammar	6

Tabla 3. Errores de ChatGPT.

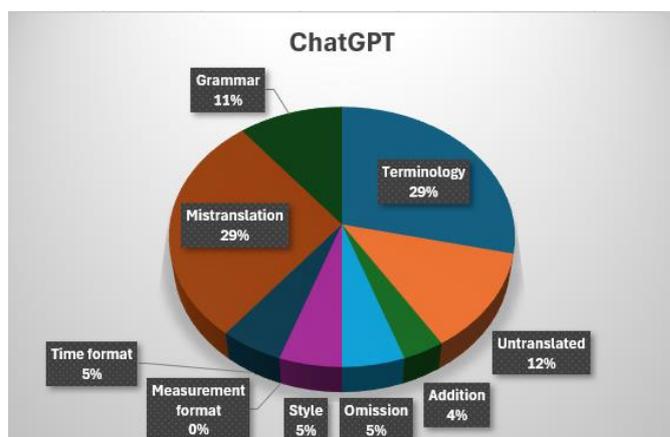


Gráfico 1. Errores en ChatGPT.

Podemos observar que la categoría de *Terminology* presenta el mayor porcentaje de errores, junto a *mistranslation*, ambas con el 29 %. Ofrecemos un ejemplo de error terminológico: «*This hearty Moroccan vegetable soup makes a delicious and nutritious one-pot meal* se ha traducido como «Esta contundente sopa marroquí de verduras constituye una comida completa, deliciosa y nutritiva en una sola olla». Lo que se quiere dar a entender en la versión original es un plato único y completo, no se refiere a la metodología ni a los utensilios, por lo tanto la expresión «en una sola olla» sería incorrecta. Una manera de evitar este error sería parando la frase después de «nutritiva» o sustituyendo «comida completa» por «plato único». Este error se repite en otros fragmentos a lo largo del texto.

Respecto a la categoría de *Mistranslation*, por ejemplo, en el segmento original «*All the vegetables in this soup should be finely chopped, no kidding!*», se proporciona la traducción «Todas las verduras de esta sopa deben estar finamente picadas, ¡no es broma!». Sin embargo, es incorrecta porque se ha traducido de forma literal, cuando la intención de la autora era reforzar la idea de que las verduras deben estar bien picadas. Una opción adecuada podría ser «es imprescindible que las verduras estén muy bien picadas...». Lo importante, en este caso, es reflejar la insistencia de la autora.

Otro ejemplo de *Mistranslation* se pone de manifiesto en el segmento «*broken vermicelli*». La traducción que se ha proporcionado para este segmento es «fideos rotos» y, en otra ocasión, lo traduce como «fideos troceados». Sin embargo, es un calco, ya que en español, este término no se utiliza. En su lugar se puede traducir como fideos o fideos vermicelli como nos muestra el diccionario de anglicismos y otros extranjerismos de Suárez (2019, p. 212).

La siguiente categoría que vamos a comentar es *Untranslated*, con el 12 % del total. Esta categoría implica la no traducción de elementos que deberían haberse traducido, ya que tienen un equivalente directo en español. Las palabras que no se han traducido a lo largo de la receta son «smen y ghee». Por ejemplo, en el fragmento «*1/2 tsp smen - (or ghee or blue cheese, see notes)*», ChatGPT ofrece esta opción de traducción: «1/2 cucharadita de smen (o ghee o queso azul; ver notas)», en tanto que una traducción adecuada sería: «1/2 cucharadita de mantequilla clarificada (o queso azul; ver notas)».

La siguiente categoría es *Grammar*, con el 11 % de errores sobre el total. Por ejemplo, en el fragmento «*So to keep the chorba adequately soupy, adjust water accordingly and correct the seasoning for later serving*», ChatGPT ofrece la traducción «Por eso, para que la chorba mantenga la consistencia adecuada de sopa, ajusta la cantidad de agua y corrige el sazón antes de servir». El error reside en «el sazón», ya que, esta palabra es femenina. La forma correcta sería «corrige la sazón». Además, en la versión original del texto, tenemos este fragmento que indica lo siguiente: «*Add the oil and smen then the meat and/or the bones, onions...*». La traducción que nos ofrece ChatGPT es la siguiente: «Añade el aceite y el smen, luego la carne y/o los huesos, la cebolla...». En este caso, podemos ver como en el original indica claramente «*onions*» en plural, sin embargo, en la traducción aparece «cebolla» en singular. Este es un error significativo, ya que puede afectar al curso de la receta.

Las siguientes categorías hemos decidido agruparlas, puesto que las tres, tienen el 5 % cada una sobre el total y se corresponden con *Time format*, *Style* y *Omission*.

Comenzaremos con *Time format*, en esta categoría se clasifican los formatos incorrectos del tiempo. En nuestro caso, lo utilizaremos para los tiempos de cocción. En el texto original, la información sobre el tiempo de duración de la receta se encuentra dispuesta de la siguiente manera: «*1 hr 10 mins*». ChatGPT, ofrece la traducción: «1 hr 10 mins». Sin embargo, esto es incorrecto según las normas de ortografía española. La manera correcta sería «1 h 10 min». Para indicar la hora con una «h» es suficiente, lo mismo pasa con los minutos, se indica «min», sin importar que sea uno o más, no se pone nunca en plural.

Seguimos con *Style*, uno de los problemas que se han encontrado a lo largo de la receta es que, en cada fragmento se utilizaba un registro diferente, es decir, en ocasiones se utiliza un tono formal, con formas verbales propias del tratamiento de cortesía (usted), mientras que en otros se recurre a un tono más informal y cercano (tú). Por ejemplo: «planea prepararla cuando quiera vaciar el frigorífico y

aprovechar las últimas verduras que quedan de la compra semanal». Otro ejemplo se pone de manifiesto en el segmento «ve aligerando la sopa a medida que lo necesites y según tu preferencia». Como se puede observar, hay una inconsistencia. En textos dirigidos al público general, como es el caso de las recetas, lo habitual es optar por un tratamiento uniforme, que en español peninsular suele corresponder al uso de tú, especialmente cuando se busca cercanía y claridad.

En la categoría de *Omission*, tenemos en el original «*Moroccan Diced Vegetable Soup – Chorba Fassia*». La traducción ofrecida es la siguiente: «Sopa marroquí de verduras en dados – Chorba fassia». En este caso, se omite el verbo «cortadas». Esta expresión resulta ambigua y poco natural. Es una omisión relevante, ya que en este caso, es necesario el verbo para mantener la coherencia semántica y la corrección sintáctica.

Por último, la categoría de *Addition*, corresponde a un porcentaje del 4 % sobre el total de errores. En el texto original tenemos «*It's brought to a boiling point on a low heat...*», se proporciona la traducción «Se lleva a ebullición a fuego muy lento...». El «muy» en este caso es una adición innecesaria.

No se ha identificado ningún error en la categoría *Measurement format*. Con esto damos por finalizado el análisis de errores de este sistema de traducción y nos disponemos a comenzar el análisis del siguiente.

4.2. Análisis de errores de DeepL

Se ha llevado a cabo un recuento del total de errores siguiendo la clasificación de categorías y subcategorías de la Tabla 1 para DeepL. Procedemos a describir los errores por categorías, de mayor a menor ocurrencias.

	DeepL
Terminology	21
Untranslated	4
Addition	0
Omission	8
Style	4
Measurement format	3
Time format	0
Mistranslation	20
Grammar	6

Tabla 4. Errores de DeepL.

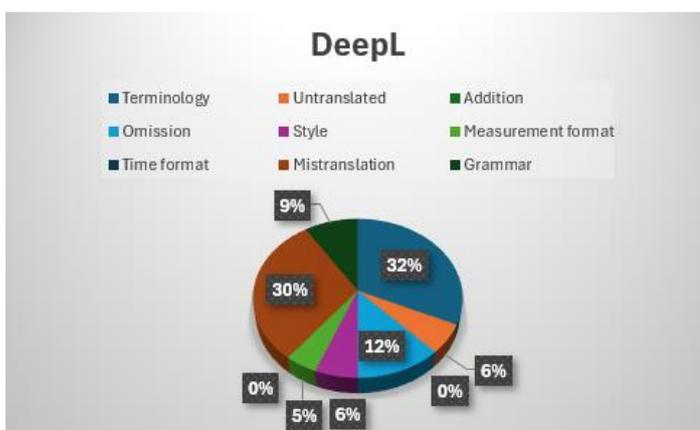


Gráfico 2. Errores de DeepL.

Terminology representa el mayor porcentaje de errores (32 %). Por ejemplo, en el caso de «*Add the oil and smen then the meat and/or the bones...*», DeepL lo traduce como «Añade el aceite y el aceite de oliva, la carne y/o los huesos...». Como podemos observar, traduce «*smen*» como «aceite de oliva», sin embargo, la traducción correcta es «mantequilla clarificada».

Mistranslation constituye el 30 % de los errores detectados. Por ejemplo, la versión original aparece como «*The turnip plays a vital role in this soup and in Morocco, the turnip family is quite rich*», DeepL nos ofrece «El nabo desempeña un papel fundamental en esta sopa y en Marruecos, la familia del nabo es bastante rica». En este caso, el problema reside en traducir «*rich*» como «rica». Se entiende que en este contexto se refiere más bien a que la familia del nabo es variada. Este error sería un calco, por lo que se tendría que clasificar en esta categoría.

Omission constituye el 12 % de los errores detectados. Por ejemplo, en el segmento «*Be sure to dice those veggies quite small...*», se proporciona la traducción «Asegúrese de cortar las verduras en dados pequeños...». En el fragmento original, se enfatiza en la importancia de cortar los dados en un tamaño muy pequeño cuando se utiliza «*quite*». Sin embargo, ese énfasis no queda reflejado en la traducción, ya que debería haberse utilizado el adverbio «muy», es decir, «Asegúrese de cortar las verduras en dados muy pequeños...». En este caso, se estaría omitiendo contenido del texto original.

Grammar constituye el 9 % de los errores detectados. Por ejemplo, en el segmento «*What Makes a Good Chorba Fassia. A good Chorba Fassia depends on the quality of the vegetables*», se proporciona la traducción «Qué hace que un Chorba Fassia sea bueno. Un buen Chorba Fassia depende de la calidad de las verduras». El error en este fragmento se manifiesta en el uso incorrecto del género gramatical. Estas dos palabras son femeninas, por lo que lo correcto sería decir «una buena Chorba Fassia».

Untranslated constituye el 6 % de los errores detectados. En esta categoría, se han registrado los mismos errores que en el otro sistema de traducción automática. Las palabras que no se han traducido a lo largo de la receta son «*smen* y *Ghee*». Por ejemplo, en el segmento «*You may replace smen by a small piece of blue cheese, mostly the white creamy part*», se proporciona la traducción «Puede sustituir el smen por un trocito de queso azul, sobre todo la parte blanca cremosa».

Style también constituye el 6 % de los errores detectados. Por ejemplo, en el segmento «*If you are making a batch with potential leftover for the coming days, add the vermicelli to the quantity of soup you are planning to have the same day and leave the rest without pasta. You can still do the same the next day with your remaining soup*», se proporciona la traducción «Si vas a hacer una tanda con posibles sobras para los días siguientes, añade los fideos a la cantidad de sopa que tengas pensado tomar ese mismo día y deja el resto sin pasta. Puede hacer lo mismo al día siguiente con la sopa restante». En este caso, el error de estilo se manifiesta en una clara alternancia pronominal: mientras que el primer enunciado utiliza el tratamiento informal de segunda persona del singular (vas, tengas), el segundo

recurre al tratamiento formal de tercera persona del singular (puede). Esta oscilación entre registros resulta inadecuada desde el punto de vista estilístico y rompe la coherencia textual.

Measurement format constituye el 5 % de los errores detectados. Por ejemplo, en el segmento «*1 cup cubed lamb, beef or chicken - (cut less than 1" thick, see notes)*», se proporciona la traducción «1 taza de dados de cordero, ternera o pollo - (cortados de menos de 1" de grosor, ver notas)». El error en este caso radica en la conservación de la unidad de medida de una pulgada (1") en lugar de adaptarla al sistema métrico decimal. Una traducción correcta sería «menos de 2,5 cm de grosor».

No se ha identificado ningún error en las categorías *Time format* y *Addition*.

4.3. Comparación entre sistemas de traducción automática

Se presenta una tabla comparativa (Tabla 5. Recuento de errores por subcategoría de cada sistema de TA) que recoge los errores detectados en las traducciones generadas por ChatGPT y DeepL, clasificados según las categorías del modelo MQM previamente descritas. Esta comparación permite visualizar de forma clara las diferencias cualitativas y cuantitativas entre ambos sistemas de traducción automática en el contexto del texto analizado.

	ChatGPT	DeepL
Terminology	16	21
Untranslated	7	4
Addition	2	0
Omission	3	8
Style	3	4
Measurement format	0	3
Time format	3	0
Mistranslation	16	20
Grammar	6	6

Tabla 5. Recuento de errores por subcategoría de cada sistema de TA.

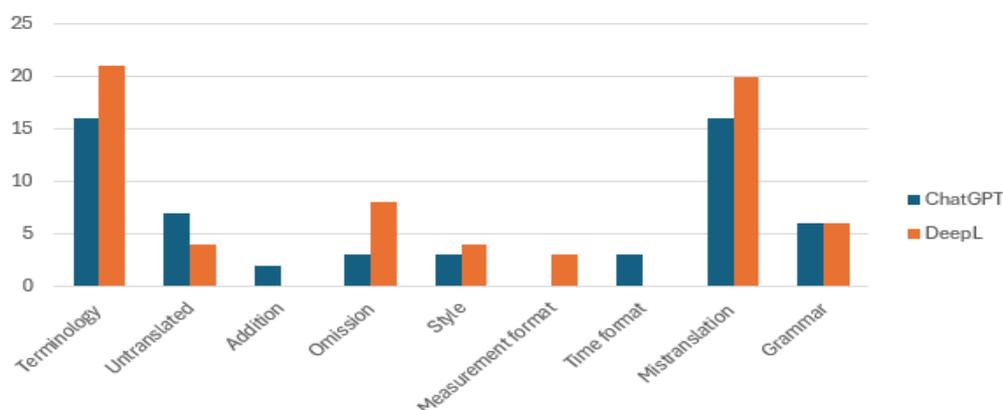


Gráfico 5. Comparación de subcategorías de errores por sistemas de TA.

El Gráfico 3 nos muestra una comparación entre los dos sistemas de traducción automática. En general, las cifras de errores son bastante similares en ambos sistemas y ambos cometan errores en casi

todas las categorías. Para empezar, se puede apreciar que ChatGPT y DeepL coinciden en las categorías de errores (*Terminology* y *Mistranslation*). En la categoría de *Terminology*, ChatGPT presenta un menor número de errores (16 frente a 21 de DeepL), resultados que se replican en el caso de *Mistranslation* (16 en ChaGPT y 20 en DeepL). Respecto a los errores de *Omission*, DeepL presenta también un número más elevado (8 frente a 3 de ChatGPT). No obstante DeepL no comete errores de *Addition*, mientras que ChatGPT sí (2). En cuanto a la categoría de *Untranslated*, DeepL muestra mejores resultados (4 en DeepL frente a 7 de Chat). En *Measurement format*, se presentan únicamente errores de DeepL (3), lo contrario pasa con *Time format*, solo se presentar errores de ChatGPT (3). En Grammar ambos traductores contienen la misma cantidad de errores (6), aunque el tipo de error no es siempre el mismo en ambos sistemas. Por último, en *Style*, se presenta un número similar de errores (3 en ChatGPT y 4 en DeepL).

En conjunto, los datos permiten observar que, si bien ambos sistemas comparten ciertos aspectos mejorables, especialmente en la *Terminology* y en *Mistranslation*, ChatGPT ofrece un mejor rendimiento en la mayoría de las categorías analizadas. Cabe destacar que varios de los errores detectados en este sistema de traducción se repiten a lo largo del texto, lo que refuerza su impacto en la calidad global de la traducción. Además, es importante señalar que ChatGPT podría mejorar sus resultados si se emplea un *prompt* más específico o contextualizado, es decir, tiene una mayor flexibilidad del sistema frente a DeepL, cuya capacidad de adaptación es más limitada. A pesar de ello, ambos traductores automáticos continúan mostrando limitaciones en aspectos clave como el estilo o la precisión léxica. Esto confirma la necesidad de una revisión humana especializada, especialmente en textos culturalmente marcados como es el caso de esta receta.

Por último, se ha querido aportar una tabla de las categorías de errores con una mayor ocurrencia que se han detectado en ambos sistemas de traducción.

	ERRORES	ChatGPT	DeepL
<i>Terminology</i>	<i>one-pot meal</i>	comida en una sola olla	comida de una sola olla
	<i>braising step</i>	el paso del braseado	la fase de estofado
	<i>too chunky</i>	demasiado enteras	tienen demasiados trozos
	<i>Stir to combine</i>	remueve para mezclar	remover para mezclar
	<i>strained</i>	filtra	cuela(correcto)
	<i>Cheesy taste</i>	quesoso	sabor a queso (correcto)
	<i>field</i>	cosecha	campo (correcto)
<i>Mistranslation</i>	<i>no kidding</i>	¡no es broma!	¡no es broma!
	<i>finely chopped</i>	finamente picadas	finalmente cortadas
	<i>broken vermicelli</i>	fideos rotos	fideos rotos
	<i>time to age</i>	tiempo para envejecer	tiempo de maduración
	<i>The older the smen</i>	Cuanto más antiguo	cuanto más viejo

<i>to render their liquid</i>	suelten su jugo	suelten su líquido
<i>plate</i>	plato (correcto)	curso
<i>Bowl of pure goodness</i>	un cuenco de pura bondad	un plato de pura bondad
<i>richer in flavors</i>	más rica en sabores	más rica en sabores
<i>bring to a simmer</i>	llévala a fuego lento	seguir cocinando a fuego lento

Tabla 6. Errores más recurrentes en ambos sistemas.

Los datos extraídos del análisis de errores permiten establecer una visión global sobre el rendimiento de los sistemas de traducción automática que se han seleccionado. La identificación y clasificación de los errores, y la distribución por categorías y sistemas de TA constituye una base sólida para el desarrollo de las conclusiones. En el siguiente capítulo, se abordarán dichas conclusiones a partir de los resultados presentados, con el objetivo de evaluar la eficacia de cada sistema en este caso concreto.

5. Conclusiones

La traducción automática ha experimentado un avance notable en los últimos años, lo que la ha convertido en una herramienta habitual en múltiples ámbitos, incluida la traducción gastronómica. Sin embargo, el presente trabajo ha demostrado que, cuando se trata de textos con contenido cultural, como las recetas marroquíes, tanto los sistemas de traducción automática neuronales como los *Large Language Models* presentan limitaciones significativas. Por tanto, podemos afirmar que hemos alcanzado el objetivo general propuesto, así como los objetivos específicos.

De forma más concreta, hemos podido constatar que la traducción de la terminología especializada y de las expresiones idiomáticas no siempre se resuelve de forma adecuada, pues muchos de estos elementos no constan de equivalentes directos en la lengua meta. Las traducciones que nos ofrece DeepL tienden a ser más uniformes y estables, pero también más literales y se adaptan menos al contenido cultural. ChatGPT, en cambio, presenta mayor capacidad de adaptación. Es un sistema en el que es posible la interacción con el usuario y admite retroalimentación, esto permite conseguir resultados más concretos mediante especificaciones y encargos. La utilización *prompts* específicos, permite afinar la respuesta y obtener una traducción satisfactoria. Por otro lado, la variabilidad en los resultados compromete su fiabilidad.

Por lo tanto, aunque ambos sistemas pueden ser de utilidad como apoyo al trabajo traductor, no son capaces de sustituir el juicio humano en la interpretación de elementos culturales complejos. A raíz de ello, se proponen varias vías de mejora: la incorporación de glosarios con un repertorio lingüístico más amplio, integrar conexiones con bases terminológicas como IATE y el desarrollo de modelos que se ajusten más al registro y al contexto pragmático. Asimismo, se considera que el papel de la posesión humana debe seguir ocupando un lugar central en el proceso traductor asistido por inteligencia artificial.

Futuros estudios podrían ir encaminados a ampliar la muestra de análisis, a incorporar más lenguas, especialmente en el caso de Deep (ya que dispone de una menor variedad de idiomas que ChatGPT), y a aplicar métricas de calidad a los sistemas de traducción como la que hemos utilizado para nuestro trabajo (MQM *core*) que permitan valorar con mayor precisión el rendimiento de cada sistema según distintos tipos textuales y según los errores repetidos. En nuestro caso particular, hemos tenido la oportunidad de evaluar con especial atención los matices culturales y lingüísticos implicados, lo que refuerza la necesidad de una traducción automatizada que no solo sea precisa, sino también culturalmente competente.

6. Bibliografía

- Berner, S. (2003). Lost in Translation: cross-lingual communication, and virtual academic communities. En *5th Annual Conference on World Wide Web Application, 10-12 septiembre, Durban, Sudáfrica*.
- Bowker, L., y Buitrago Ciro, J. (2019). *Machine Translation and Global Research: Towards Improved Machine Translation Literacy in the Scholarly Community*. Leeds: Emerald Publishing Limited. DOI: <https://doi.org/10.1108/9781787567214>
- Castaño, M. A., Casacuberta, F. y Vidal, E. (1997). Machine Translation using Neural Networks and Finite-State Models. En *Proceedings of the 7th Conference on Theoretical and Methodological Issues in Machine Translation and Natural Languages*. <https://aclanthology.org/1997.tmi-1.19.pdf> (Fecha de consulta: 06/06/2025).
- Cho, K., Van Merriënboer, B., Gulcehre, C., Bougares, F., Schwenk, H. Y Bengio, Y. (2014). Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. En *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Doha: Association for Computational Linguistics, 1724-1734. <https://arxiv.org/pdf/1406.1078> (Fecha de consulta: 06/06/2025).
- Costa-jussà, M. R. y Fonollosa, J. A. R. (2015). Latest trends in hybrid machine translation and its applications. *Computer Speech & Language*, 32(1), 3–10. <https://doi.org/10.1016/j.csl.2014.11.001>
- Diego Olite, F., Morales Suárez, I. y Vidal Ledo, M. J. (2023). Chat GPT: origen, evolución, retos e impactos en la educación. *Educación Médica Superior*, 37(2), 1-23. <https://ems.sld.cu/index.php/ems/article/view/3876/1508> (Fecha de consulta: 06/06/2025).
- Downie, J. (2020). *Interpreters vs Machines*. Oxon/New York: Routledge.
- Durán Muñoz, I. y Corpas Pastor, G. (2020). Corpus-Based Multilingual Lexicographic Resources for Translators: an Overview. En Domínguez Vázquez, M. J., Mirazo Balsa, M. Valcárcel Riveiro, C. (Eds.), *Studies on Multilingual Lexicography* (pp. 159-178). Berlín: De Gruyter.
- Forcada, M. L., y Ñeco, R. P. (2005). Recursive hetero-associative memories for translation. En Bertino, E., Gao, W., Steffen, B. y Yung, M. (Eds), *Lecture Notes in Computer Science* (pp. 453–462). Berlín: Springer. DOI: <https://doi.org/10.1007/bfb0032504>
- Forcada, M., Sánchez Martínez, F., y Pérez Ortiz, J. A. (2016). *Manual de informática y de tecnologías para la traducción*. Alicante: Departamento de Lenguas y Sistemas Informáticos de la Universidad de Alicante. <https://rua.ua.es/dspace/handle/10045/53085> (Fecha de consulta: 06/06/2025).
- Hutchins, J. (2006). Example-based machine translation: a review and commentary. *Machine Translation*, 19(3-4), 197–211. DOI: <https://doi.org/10.1007/s10590-006-9003-9>

- IBM. (2023, 2 de noviembre). *What are large language models (LLMs)?*, <https://www.ibm.com/think/topics/large-language-models>
- Kiffa, N. (2017, 8 de octubre). Moroccan Diced Vegetable Soup - Chorba Fassia. *Taste of Maroc*. <https://tasteofmaroc.com/moroccan-vegetable-soup-chorba-fassia/> (Fecha de consulta: 06/06/2025).
- Kutyłowski, J. (2017) DeepL Publisher. *DeepL.com*. <https://www.deepl.com/es/publisher> (Fecha de consulta: 06/06/2025).
- Lommel, A. (2017). *Neural MT: sorting fact from fiction*. Common Sense Advisory.
- Moreno-Pérez, L. y López-Arroyo, B. (2021). Atypical corpus-based tools to the rescue: How a writing generator can help translators adapt to the demands of the market. *MonTI*, 13, 251–279.
- Nagao, M. (1984). A Framework of a Mechanical Translation Between Japanese and English by Analogy Principle. En Elithorn, A. y Banerji, R. (Eds.), *Artificial and Human Intelligence*. Amsterdam: Elsevier. <https://mt-archive.net/70/Nagao-1984.pdf> (Fecha de consulta: 06/06/2025).
- Oliver, A. (2016). *Herramientas tecnológicas para traductores*. Barcelona: Editorial UOC.
- Ortego Antón, M. T. (2024a). The design of TorrezoTRAD: The Semiautomatic Spanish-English writing and translation aid tool. En Peñuelas Gil, I. y Ortego Antón, M. T. (Eds.) (2024). *Interpreting and Translation for Agri-Food Professionals in the Global Marketplace* (pp. 69-84). Berlín: De Gruyter.
- Ortego Antón, M. T. (2024b). Metodología para el diseño de un asistente semiautomático de redacción y de traducción de fichas descriptivas de embutidos del español al inglés. *Cadernos de tradução*, 44(1), 1-20.
- Parra Escartín, C. (2018). Evolución de la traducción automática. *La Linterna Del Traductor*, 16. <https://lalinternadeltraductor.org/n16/traduccion-automatica.html> (Fecha de consulta: 06/06/2025).
- Peris Abril, Á. y Casacuberta Nolla, F. (2017). Traducción automática neuronal. *Tradumàtica*, 15. https://revistes.uab.cat/tradumatica/article/view/n15-casacuberta-peris/pdf_48
- Quah, C. (2006). *Translation and Technology*. Londres: Palgrave Macmillan. DOI : <https://doi.org/10.1057/9780230287105>
- Rabadán, R., Pizarro, I. y Sanjurjo-González, H. (2021a). Authoring support for Spanish language writers: A genre-restricted case study. *RESLA*, 34(2): 677–717.
- Rabadán, R., Ramón, N. y Sanjurjo-González, H. (2021b). Pragmatic Annotation of a Domain Restricted English-Spanish Comparable Corpus. *Bergen Language and Linguistics Studies*, 11(1), 209–23. DOI: <https://doig.org/10.15845/bells.v11i1.3445>.
- Rivas Carmona, M. del M. y Veroz González, M. A. (2018). *Agroalimentación: lenguajes de especialidad y traducción*. Granada: Editorial Comares

- Sánchez Ramos, M. y Rico Pérez, C. (2020). *Traducción automática: conceptos clave, procesos de evaluación y técnicas de posesición*. Granada: Editorial Comares.
- Shiwen, Y. y Xiaojing, B. (2014). Rule-Based Machine Translation. En C. Sin-wai (Ed.), *The Routledge encyclopedia of translation technology* (pp. 186-200). Londres: Routledge.
- Suárez, D. (2019). *Diccionario de anglicismos y otros extranjerismos*. <https://defensadelidioma.com/wp-content/uploads/2018/09/diccionario-de-anglicismos-y-otros-extranjerismos.pdf>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A., Kaiser, Ł. y Polosukhin, I. (2017). Attention Is All You Need. En *31st Conference on Neural Information Processing Systems (NIPS 2017)*. <https://arxiv.org/pdf/1706.03762> (Fecha de consulta: 06/06/2025).
- Weaver, W. (1949). Memorandum on translation from one language to another, and on the possibility of contributing to this process by the use of modern computing devices. *The Rockefeller Foundation Archives*.
- Zhang, X. y Torres Hostench, O. (2023). La traducción al español de las técnicas culinarias chinas. *Hermēneus. Revista de Traducción e Interpretación*, 24, 483–514. DOI: <https://doi.org/10.24197/her.24.2022.483-514>
- Zhu, W., Liu, H., Dong, Q., Xu, J., Huang, S., Kong, L., Chen, J. y Li, L. (2023). Multilingual Machine Translation with Large Language Models: Empirical Results and Analysis. En Duh, K., Gómez, H. y Bethard, S. (Eds.), *Findings of the Association for Computational Linguistics : NAACL 2024* (pp. 2765-2781). Ciudad de México: Association for Computational Linguistics. <https://arxiv.org/pdf/2304.04675> (Fecha de consulta: 06/06/2025).