María Luisa Carrió Pastor
Clara Ureña Tormo (eds.)
Análisis del discurso y enseñanza de lenguas en la era de la inteligencia artificial
Horizontes Universidad
Octaedro

Análisis del discurso y enseñanza de lenguas en la era de la inteligencia artificial

María Luisa Carrió Pastor
y Clara Ureña Tormo (eds.)

Análisis del discurso y enseñanza de lenguas en la era de la inteligencia artificial

Octaedro

Colección Horizontes-Universidad
Título: *Análisis del discurso y enseñanza de lenguas en la era de la inteligencia artificial*

Primera edición: julio de 2025

© María Luisa Carrió Pastor y Clara Ureña Tormo (eds.)

© De esta edición:
Ediciones OCTAEDRO, S.L.
C/ Bailén, 5 – 08010 Barcelona
Tel.: 93 246 40 02
octaedro@octaedro.com
www.octaedro.com

Esta publicación está sujeta a la Licencia Internacional Pública de Atribución/Reconocimiento-NoComercial 4.0 de Creative Commons. Puede consultar las condiciones de esta licencia si accede a: https://creativecommons.org/licenses/by-nc/4.0/

ISBN: 978-84-1079-114-5

Maquetación: Fotocomposición gama, sl
Diseño y producción: Octaedro Editorial

Publicación en acceso abierto - *Open access*

Sumario

Introducción 9
María Luisa Carrió Pastor; Clara Ureña Tormo

SECCIÓN 1.
ANÁLISIS DEL DISCURSO EN LA ERA
DE LA INTELIGENCIA ARTIFICIAL

1. Metodología y herramientas posdigitales para el análisis de un repertorio lingüístico minoritario. El estudio de caso del arbëresh en Italia a través de los mediagramas 17
Alain Cicconetti

2. Anotando sintácticamente lengua oral 49
Eva María Domínguez Noya; María Paula Santalla del Río

3. Las formas de tratamiento en la comunicación gubernamental francesa y española. Análisis contrastivo de las cuentas de X/Twitter de los presidentes Emmanuel Macron y Pedro Sánchez 75
Sonia Madrid Cánovas; Stéphane Patin

4. ChatGPT como herramienta de documentación para intérpretes en el ámbito futbolístico: ¿realidad o quimera? 101
Alfonso Peñaranda Latorre; María Teresa Ortego Antón

5. ¿A punto de caramelo? La traducción automática de expresiones multiverbales del turismo gastronómico 129
Isabel Peñuelas Gil; María Teresa Ortego Antón

6. Discurso político emitido en redes sociales: estudio de caso sobre las publicaciones de Rocío Monasterio (Vox) e Irene Montero (Unidas Podemos) 157
Hanae Trola Skalli; Sara Gómez Díaz; Gérard Fernández Smith

SECCIÓN 2.
ESTUDIOS SOBRE LA ENSEÑANZA DE LENGUAS
EN LA ERA DE LA INTELIGENCIA DIGITAL

7. Percepciones sobre un posible uso de entornos virtuales en el aula de ELE 185
Marta Baró-Miró; Françoise Olmo-Cazevieille; Beatriz Rey

8. ChatGPT como herramienta de corrección y reflexión para la producción textual en lengua alemana en el ámbito universitario 207
Almudena Mallo Dorado

9. Docentes y ChatGPT. Una comparativa en la evaluación de la expresión escrita de los maestros en formación 239
Maria Navío-Inglés; Jesús Guzmán Mora

10. Hacia una comunicación académica más efectiva: aplicación del modelo de lenguaje de gran tamaño ChatGPT a la generación de títulos de trabajo fin de grado 257
Beatriz Rodríguez Cuadrado; Ruth Rodríguez Cuadrado

5

¿A punto de caramelo? La traducción automática de expresiones multiverbales del turismo gastronómico

Isabel Peñuelas Gil
María Teresa Ortego Antón
CITTAC, Universidad de Valladolid

Resumen
El avance de los sistemas de traducción automática (STA) ha revolucionado la relación que la sociedad tiene con la traducción, algo especialmente perceptible en ámbitos como el turismo gastronómico donde es frecuente encontrar textos traducidos automáticamente que, aunque buscan facilitar la comunicación, en ocasiones tienen el efecto contrario. Una de las principales causas son las expresiones multiverbales (Corpas, 2013) o EMV, unidades lingüísticas extremadamente frecuentes en las lenguas naturales (Jackendoff, 1977) que están arraigadas en culturas concretas y cuyo significado no siempre se infiere del de sus componentes. La abundancia de EMV y el uso frecuente de STA en el turismo plantean interrogantes sobre su impacto en la comunicación en este sector. El presente estudio pretende identificar patrones de comportamiento y posibles soluciones a partir del análisis de los errores más comunes observados al traducir automáticamente una muestra de EMV del gastroturismo. Para ello, compilamos un corpus en español siguiendo el protocolo propuesto por Seghiri (2017), el cual se ha aplicado con éxito en diversos estudios posteriores (Ortego, 2019, 2024a, 2024b; Peñuelas, 2024a, 2024b), del que se extrajeron ejemplos reales de empleo de las EMV y sus contextos. Estos se tradujeron haciendo uso de una selección de STA populares en el mercado. Los resultados, analizados mediante una modificación del modelo de evaluación de Ortiz Boix (2016), revelaron problemas recurrentes de falso sentido, equivalencia terminológica y traducción literal, lo que pone de manifiesto deficiencias en el rendimiento de los STA con respecto al tratamiento de las EMV y las características intrínsecas de estas.

Palabras clave: traducción automática, turismo gastronómico, expresiones multiverbales, español-inglés.

1. El turismo gastronómico y la traducción

El turismo constituye uno de los pilares fundamentales de la economía y, según datos de 2022, España se sitúa como el segundo país que más turistas internacionales recibe al año (71,6 millones de turistas internacionales a lo largo del año según el Ministerio de Industria, Comercio y Turismo [2023]), solo por detrás de Francia (Organización Mundial del Turismo [OMT], 2023, p. 6). Este crecimiento del público internacional ha llevado a un aumento de la disponibilidad de traducciones de textos de corte turístico (como pueden ser sitios web, folletos, guías, menús…) destinadas a expandir los servicios y productos proporcionados a un público internacional, de habla no hispana, más amplio.

Más allá de los monumentos históricos y las playas soleadas, la comida y la bebida se han convertido en un atractivo fundamental para aquellos que desean sumergirse en la cultura local. Además, es un activo económico que beneficia la economía sostenible, el consumo de proximidad y el comercio local. En el caso de España, un país conocido por su rica tradición culinaria, el turismo gastronómico desempeña un papel crucial en la promoción de su patrimonio cultural y «representa una tipología turística que, a través de la valorización de los recursos culinarios locales, expresa autenticidad, espectacularidad, sensorialidad, erudición del consumidor, experiencias y coproducción del servicio» (Hernández, Di-Clemente y López, 2015, p. 410).

A medida que los turistas llegan de diferentes partes del mundo, surge la necesidad de traducir la riqueza gastronómica de España a otros idiomas. La traducción no solo implica convertir palabras, sino también transmitir la esencia cultural y las emociones asociadas con cada plato. La elección de las palabras y las descripciones puede afectar la percepción del turista y su experiencia culinaria (Organización Mundial de Turismo y Basque Culinary Center, 2019).

En este sector, si bien una parte de las traducciones son realizadas por profesionales especializados, también es común ver

que, en el caso de pequeñas y medianas empresas con unos recursos limitados, estas se elaboran haciendo uso de los sistemas de traducción automática (STA). La aparición y popularización de los STA en línea gratuitos en las últimas décadas ha democratizado su empleo y, en consecuencia, facilitado el proceso de comunicación para muchos (Moorkens, 2022, p. 137). Los STA son herramientas de fácil acceso a nivel mundial, gracias a las cuales una gran cantidad de información que antes podía resultar ilegible para una parte de la población –por la escasez de profesionales con un par de lenguas determinado, por cuestiones monetarias o, simplemente, por la cantidad de material a traducir–, ahora está a disposición de la gran mayoría.

No obstante, esta práctica puede desembocar problemas de comunicación que no siempre se consideran al decantarse por el uso de la traducción automática (TA) y que contrasta, como indica Álvarez Jurado (2020, p. 4), con el incremento del nivel de exigencia por parte del turista con respecto a los textos que consume.

Ante este panorama, uno de los principales obstáculos que impiden que el producto resultante se asemeje a la calidad de una traducción humana se corresponde con las expresiones multiverbales (Corpas, 2013) o EMV, unidades lingüísticas extremadamente frecuentes en las lenguas naturales (Jackendoff, 1977), que están arraigadas en culturas concretas y cuyo significado no siempre se infiere de la suma de sus componentes. La abundancia de EMV y el uso frecuente de STA en el turismo gastronómico plantean interrogantes sobre su impacto en la comunicación en este sector.

En consecuencia, en el presente estudio, que tiene por objetivo identificar patrones de comportamiento y posibles soluciones a partir del análisis de los errores más comunes observados al traducir automáticamente una muestra de EMV del gastroturismo, acotamos el concepto de EMV (apartado 2) y su traducción en los STA para, a continuación, establecer la metodología de análisis: la delimitación de la muestra de análisis extraída a partir de un corpus monolingüe, la selección de los STA y la clasificación de los errores durante el trasvase interlingüístico (apartado 3), el análisis y los resultados (apartado 4) y las conclusiones (apartado 5).

2. La traducción automática de las EMV

La traducción no ha sido ajena a los avances tecnológicos y, fruto del progreso en áreas como la lingüística computacional, la estadística o la inteligencia artificial (IA), surge la traducción automática (TA), que se define como la aplicación de la tecnología informática a la traducción de textos de una lengua a otra sin intervención humana (ISO, 2017). Aunque pueda parecer reduccionista, la TA se caracteriza por la interdisciplinariedad, dado que, como indica Quah (2006, p. 57), intervienen áreas como la lexicografía, la lingüística, la lingüística computacional, la informática y la ingeniería lingüística. Asimismo, la traducción automática se basa en el postulado de que las lenguas naturales pueden ser descritas, controladas y codificadas.

Con la proliferación de los STA en el siglo XXI, también encontramos dos corrientes de pensamiento especialmente prominentes. En las etapas iniciales de la popularización y expansión de estos sistemas, muchos profesionales veían la traducción automática como una amenaza. Como explica Oliver (2016, p. 155), lo hacían bajo la premisa de que la calidad de estos sistemas llegaría, en un futuro próximo, a un nivel suficiente como para hacer innecesaria la intervención humana. Por el contrario, el público no experto percibió estos sistemas de una manera positiva, pues, la aparente falta de intervención humana –una concepción que ha sido repetida por la prensa popular de manera regular (Bowker, 2023, p. 97), lo que ha llevado al público general a generar ideas erróneas sobre el funcionamiento y uso de estas herramientas– facilitaba el acceso a la información, eliminando así las barreras comunicativas (Austermühl y Kortenbruck, 2012, p. 153), de una manera «fácil, rápida y sin costes aparentes» (Vieira, 2020, p. 98).

Para comprender la realidad a la que se enfrentan los sectores que conviven con un uso regular de la TA, particularmente en el caso del sector turístico, es necesario comprender las carencias de estos planteamientos. Ante todo, está la noción, ampliamente aceptada por todos los profesionales de la mediación lingüística, de que esta va más allá de la sustitución de palabras en una lengua (lengua origen, LO) por sus equivalentes en otra (lengua meta, LM). Para que el acto comunicativo que representa una traducción se considere apropiado a la situación, un traductor

debe poseer un conocimiento que va más allá del idioma y de las características del lenguaje de especialidad, debe prestar atención a los factores culturales, tal y como expresan Pérez Blanco e Izquierdo (2021, p. 148) al indicar que la propiedad del lenguaje, así como el conocimiento de las convenciones culturales –en todos los niveles– y de tipología textual son fundamentales.

Con la implementación de los avances de la IA aplicados al procesamiento del lenguaje natural (PLN), los STA presentan mejoras respecto a los sistemas anteriores, especialmente gracias a la posibilidad de utilizar corpus de entrenamiento de mayor tamaño según mejoraba la capacidad de procesamiento de los ordenadores. Sin embargo, estas tecnologías siguen necesitando de la intervención humana para asegurar la calidad de sus traducciones, de ahí que hayamos visto surgir nuevas especializaciones como son la labor de preedición y posedición. La mera aparición de estas nuevas labores es indicador de que la TA por sí misma no supone una amenaza como veíamos que algunos profesionales creían inicialmente. No obstante, si un usuario no profesional decide optar por la TA, lo más probable es que recaiga en el uso de la traducción resultante sin modificaciones (*raw translation*) y no cuente con una ayuda profesional que la revise y el producto resultante contenga errores.

Los errores más superficiales, como pueden ser la falta de concordancia en la persona o número gramaticales, son fácilmente detectables si uno traduce a su lengua materna o una lengua intermedia con la que está muy familiarizado. Sin embargo, cuestiones como fraseología, humor, dobles sentidos y tonos (como son, por ejemplo, el sarcasmo o la ironía), las cuales tienen la capacidad de modificar por completo el significado de un texto, pueden pasar desapercibidas.

Ante este panorama, en este trabajo nos centraremos en el análisis de determinadas formas de unidad fraseológica a las que nos referiremos de aquí en adelante como expresiones multiverbales (EMV)[1] y que entenderemos como combinaciones de al menos dos palabras en las cuales el hablante no tiene libertad

1. Existe una falta de unidad al referirse al objeto de estudio de la fraseología por parte de los expertos (Penadés, 2015, p. 15), que denominan a estas unidades empleando términos como «colocaciones, expresiones fijas, fraseologismos, unidades multiverbales, frasemas, etc.», como si fueran intercambiables entre sí (Mitkov et al., 2018, p. 3), si bien no es una visión compartida por todos los lingüistas.

total a la hora de elegir sus componentes, puesto que ya han sido convencionalizadas por su uso dentro de un contexto lingüístico determinado, cuando se detectan y se analizan desde la perspectiva de la lingüística computacional (Peñuelas, 2024b, p. 73).

Dado su carácter multipalabra y cultural, que además presenta con frecuencia un sentido idiomático, se trata de unos elementos lingüísticos complejos de traducir. Esto es especialmente así en el caso de los traductores automáticos que, como se adelantaba, se limitan únicamente al trasvase entre lenguas, por lo que los significados no aparentes de las expresiones –es decir, cuando la suma del significado de los elementos individuales no equivale al significado de la EMV como conjunto–, los posibles dobles sentidos o ambigüedades que generan, así como las implicaciones en una cultura que pueden tener, no se verán representadas en el producto de una TA. Además, se trata de expresiones extremadamente recurrentes que representan un porcentaje significativo del vocabulario habitual de los hablantes nativos de una lengua natural (Jackendoff, 1977; Biber et al., 1999; Sag et al., 2002). Dicha recurrencia implica que el uso apropiado de las EMV resulta indispensable para alcanzar una expresión natural en un lenguaje cualquiera.

Estos factores han hecho que las EMV, más allá de simbolizar uno de los elementos básicos de los textos turísticos en cualquiera de sus representaciones, también se hayan convertido en uno de los puntos de interés en el campo del PLN, pues suponen una de las mayores trabas en el progreso de las tecnologías de la traducción:

> The successful computational treatment of MWUs is essential for Natural Language Processing, including Machine Translation and Translation Technology; the inability to detect MWUs automatically may result in the incorrect (and even unfortunate) automatic translation and may jeopardise the performance of applications. (Mitkov et al., 2018, p. 3)

Dado el crecimiento de la traducción automática de textos turísticos y las reflexiones arriba representadas, no es descabellado decir que el presente y futuro de ambos ámbitos están estrechamente ligados al avance en cuestiones del procesamiento de las EMV.

2.1. El procesamiento de expresiones multiverbales en los STA

A medida que los sistemas de TA han avanzado, también lo ha hecho la investigación sobre cómo los distintos paradigmas se enfrentan a las EMV y a los resultados obtenidos. Cada paradigma procesa de manera diferente la información lingüística con la que se crea, así como los textos de entrada que se van a someter al proceso de traducción, lo que modifica inevitablemente el resultado obtenido. En la actualidad, entre los STA más comunes en la traducción para fines específicos encontramos los sistemas estadísticos (SMT, por sus siglas en inglés), neuronales (NMT) e híbridos (HMT).

Tanto los SMT como los NMT requieren del uso de corpus para su entrenamiento. Por su parte, y por norma general, los motores híbridos actuales también basan su arquitectura en el uso de corpus, ya que son el producto de la combinación de resultados de otros paradigmas (Costa-jussà y Fonollosa, 2015, p. 4) y, tradicionalmente, estos motores se han basado de forma parcial en los sistemas SMT y, más recientemente, han aprovechado también la tecnología neuronal. Por lo tanto, los tres paradigmas parten en la mayoría de los casos de la tecnología de corpus. No obstante, a pesar de este punto en común, todos ellos procesan estos datos de manera diferente.

Los sistemas estadísticos significaron un gran progreso en cuestiones de TA debido a la relación positiva entre la calidad de los resultados y el esfuerzo necesario para crear los sistemas. Los SMT dominaron el mercado hasta la aparición de la tecnología neuronal, no obstante, hoy en día siguen teniendo su utilidad en ámbitos especializados.

Como su propio nombre indica, extraen información estadística de los grandes corpus utilizados en la fase de entrenamiento del motor (Brown et al., 1988, p. 71) y se «sustentan en la probabilidad (alta o baja) de que a una oración en la lengua de origen le corresponda una traducción en la lengua meta» (Sánchez y Rico, 2020, p. 15). Por supuesto, este es el mismo principio que utilizan para la traducción de EMV. Sin embargo, las bases teóricas de este enfoque –primero basado en la alineación de pares de palabras (Brown et al., 1988, 1990) y, posteriormente, basado en la «alineación de frases» (Zens, Och y Ney, 2002; Koehn, Och y Marcu, 2003)– no atienden a las necesidades específicas de la traducción de EMV.

El primer caso presenta dos obstáculos: por un lado, la discordancia entre el carácter idiomático subyacente de las EMV y la división palabra por palabra de la alineación que entiende todos los elementos constituyentes de una EMV como elementos individuales; por el otro, su incapacidad para procesar correspondencias «many-to-many» (Mitkov et al., 2018, p. 16). Este segundo obstáculo queda solventado en el modelo basado en la alineación de frases. Sin embargo, estas frases, más conocidas como *n-gramas*, no siempre equivalen a EMV y pueden resultar en combinaciones con una importancia limitada desde el punto de vista lingüístico. Se trata de un problema que motores SMT con otros modelos de alineación, específicamente aquellos que incluyen información sintáctica (Yamada y Knight, 2001), mejoran y es la tendencia que se ha seguido en los últimos años. No obstante, como señalan Mitkov *et al*. (2018, p. 17):

> In the state-of-the-art PB-SMT systems, the correct translation of MWUs occurs therefore only on a statistical basis if the constituents of MWUs are marked and aligned as parts of consecutive phrases (n-grams) in the training set and it is not generally treated as a special case where correspondences between source and target may not be so straightforward, i.e. it does not consist of consecutive many-to-many source-target correspondences. (Mitkov *et al*., 2018, p. 17)

Los sistemas neuronales, por su parte, despuntaron a mediados de la década de 2010 y se presentaron no solo como uno de los paradigmas más innovadores, sino como un paradigma capaz de progresar rápidamente (Sánchez y Rico, 2020, p. 12). Se basan en redes neuronales artificiales, un modelo computacional que «comprises a large number of highly interconnected processing elements that work in unison to solve specific problems» (Bowker y Buitrago, 2019, p. 45).

Al tratarse de sistemas relativamente nuevos, la investigación de cómo estos gestionan y traducen las EMV es aún limitada. Sin embargo, cuentan con una ventaja potencial en este aspecto: los sistemas neuronales aprenden al traducir, es decir, utilizan una tecnología de aprendizaje automático que hace que la red neuro-

nal sea capaz de identificar patrones complejos en los textos de entrenamiento[2] y los utiliza para aprender a traducir nuevos textos de forma automática (Bowker, 2023, p. 96).

Desde el punto de vista de las EMV esto implica que, con el tiempo, los motores NMT podrían aprender a traducir cualquier expresión. Sin embargo, para llegar a ese punto es necesario hacer un uso adecuado de estas herramientas para que no vuelvan a aprender los mismos errores, perpetuándolos así. Se trata de una noción de gran importancia y novedosa en el campo que lleva a reflexiones como la siguiente:

> A NMT system might indeed produce an idiomatic translation, but this is generally because the data it has learned from contain hundreds or maybe thousands of examples of that very translation. An NMT system ... does not know it is being idiomatic, or using a cultural equivalent, when it correctly translates [an idiom]. (Kenny, 2022, p. 39)

Aunque estos sistemas han supuesto un paso adelante en cuanto al procesamiento de la información se refiere, su forma de «comprender» los textos no ha cambiado demasiado. Como indica Kenny (2022, p. 39), un sistema NMT puede traducir una expresión idiomática correctamente, pero no lo hará por una decisión consciente del contexto y sus implicaciones, sino porque el motor infiere de los textos de entrenamiento que es la opción más probable basándose en lo que ya conoce. Esto conlleva que un fallo en el reconocimiento de los elementos implicados podría afectar a toda la oración que contiene la EMV, generando situaciones de obstáculos de comunicación derivados de la presencia de estos elementos fraseológicos. De ahí la importancia de hacer un uso adecuado de cara al progreso de la TA neuronal y, por supuesto, del progreso de la TA de las EMV.

Una vez delimitado el objeto de estudio en este trabajo (las EMV) y definidos los STA, procedemos a describir la metodología que vamos a emplear.

2. Los «textos de entrenamiento» se refieren a todos los textos que han servido para entrenar al sistema NMT en algún momento, bien durante la fase de creación («corpus de entrenamiento») o la de uso y de los cuales el sistema ha tenido la oportunidad de aprender.

3. Metodología

Para detectar los patrones de comportamiento que los STA presentan en la actualidad con respecto a las EMV en el par de lenguas de español-inglés, en primer lugar, delimitamos la muestra de análisis basándonos en la compilación y explotación de un corpus monolingüe, a continuación, seleccionamos los STA de los que vamos a obtener la TA de las EMV y, por último, explicamos el procedimiento empleado para la recogida y análisis de los datos.

3.1. La selección de la muestra de análisis

Para obtener ejemplos reales, acotamos el corpus a los folletos turísticos sobre gastronomía. Los folletos turísticos son un género textual específico dentro del ámbito del turismo (Calvi, 2010, 2019) diseñados para promocionar las ofertas culinarias de una región o establecimiento. Estos folletos tienen un propósito dual: informar y persuadir al lector. Por un lado, informan de los detalles sobre la gastronomía local, ofrecen una descripción de los platos típicos, de los ingredientes autóctonos y de las técnicas culinarias tradicionales. Al mismo tiempo, buscan convencer al lector de visitar el destino promocionado, enfatizando la calidad y singularidad de su oferta gastronómica.

> Una publicación distribuida gratuitamente, la mayoría de las veces en forma de desplegable o cuadernillo y, entre sus funciones, destacan la de informar, promover una imagen y seducir y atraer al turista. En los folletos, solemos encontrar partes descriptivas, expositivas y argumentativas combinadas con partes prácticas y elementos gráficos. Estos últimos elementos (fotografías) activan la imaginación del receptor y provocan su interés por el producto, servicio o destino turístico. (Soto, 2014, p. 66)

Los folletos turísticos que versan sobre la gastronomía no solo pretenden informar y describir las principales atracciones culinarias de un determinado lugar, sino que también buscan incitar al receptor a fin de persuadirlo y, para conseguirlo, el emisor necesita conocer las características psicológicas y lingüísticas de los destinatarios. Por tanto, el folleto adquiere las propieda-

des de un anuncio publicitario que a va a ser consumido por un público receptor interesado en viajar y conocer nuevos lugares e, incluso, podríamos señalar que se trata de un texto con un acto estratégico en el que el emisor impone su opinión al lector, que es el turista gastronómico.

Además, el diseño visual es un aspecto crucial de estos folletos. Se utilizan imágenes atractivas de platos y ambientes gastronómicos para captar la atención del lector, junto con una tipografía y una paleta de colores que refuerzan la identidad cultural del destino. Este enfoque visual busca no solo informar, sino también evocar una respuesta emocional en el lector, incentivándolo a vivir la experiencia culinaria presentada.

Una vez delimitado el género textual, procedemos a compilar el corpus siguiendo un protocolo similar al empleado en estudios previos (Seghiri, 2017; Ortego, 2024a, 2024b; Sánchez, 2022; Peñuelas, 2024a, 2024b; entre otros). La compilación se llevó a cabo siguiendo cuatro fases:

- Búsqueda de textos redactados por empresas u organismos que proporcionen servicios turísticos que se centrasen en la vertiente gastronómica de estos y que se ciñesen a la tipología textual a la que se ha acotado el estudio. Para este trabajo solo se tuvieron en cuenta aquellos textos que aparecían alojados en las páginas web como un documento independiente.
- Descarga, de manera manual, de los textos seleccionados en su formato original (PDF).
- Conversión de los textos a archivos TXT UTF8 para que estos pudiesen ser procesados por herramientas de gestión de corpus, utilizando AntFileConverter (Anthony, 2020).
- Archivo de los textos. En este caso, dado que se trata de un corpus monolingüe, solo se separaron los textos de acuerdo con su formato (PDF o TXT).

A cada texto se le asignó una denominación única basada en el siguiente estándar: identificador numérico según orden de descarga + identificador de la temática (en este caso «GT», como abreviatura de «gastroturismo») + identificador de origen del texto (abreviatura de la provincia o comunidad autónoma para la que se ha escrito) + fecha de descarga (aaaammdd). De este modo, el archivo 001GTSO20230518, por ejemplo, el primer documento del corpus, contiene información relevante a la provincia de Soria y se descargó el 18 de mayo de 2023.

El resultado es un corpus monolingüe virtual compuesto de 30 folletos sobre el turismo gastronómico en España –con un número total de palabras (casos o *tokens*) de 93.185 y 16.569 palabras únicas (tipos o *types*)– catalogado y estructurado para su explotación.

A continuación, el corpus se explota utilizando dos programas de gestión de corpus –Sketch Engine (Kilgarriff et al., 2004) y AntConc (Anthony, 2020)– que permitieron analizar y extraer la información pertinente de manera rápida y eficaz. A pesar de que los dos programas presentan herramientas y funcionamientos similares, se optó por utilizar ambas herramientas simplemente por una cuestión de preferencia con respecto a la interfaz de cada una de las herramientas.

El procedimiento de extracción de las EMV se inició utilizando la opción «WordList», una de las herramientas básicas de AntConc que permite generar un listado con las palabras únicas, o tipos, del corpus ordenadas según su frecuencia de aparición. Para asegurar que los resultados eran lingüísticamente relevantes, se utilizó una *stoplist* para que el programa ignorase las palabras de menor relevancia como pueden ser artículos, preposiciones o conjunciones. Los resultados se revisaron uno a uno con la herramienta «N-grams» de Sketch Engine, la cual permite filtrar y anidar los resultados, las posibles combinaciones de cada término, prestando siempre atención a la frecuencia de este.

Este proceso de revisión reveló lo común que resulta utilizar expresiones construidas en torno al término «punto», el cual aparece representado en el corpus 44 veces, en el campo del turismo gastronómico español. Las EMV detectadas en el corpus que incluyen «punto» presentan características que las hacen interesantes para su análisis:

- Algunas de ellas presentan cierto grado de idiomaticidad, lo que implica que: a) su significado no se puede inferir directamente del significado de sus componentes sino a través del conocimiento sociocultural de la lengua; y b) es probable que existan diferencias sustanciales con sus equivalentes en lengua inglesa.
- Su composición gramatical representa distintos patrones (sust + sust, sust + prep + sust, sust + adj, etc.), lo que nos lleva a considerar las posibles diferencias en los resultados.

No obstante, se detectó que cinco de las repeticiones no guardan relación con el turismo gastronómico en ninguna de sus vertientes, sino que tenían un trasfondo histórico o geográfico, por lo que en esta ocasión no se tendrán en cuenta. Además, otra de las oraciones en las que aparecía «punto» quedó descartada puesto que, aunque hace referencia al punto de cocción, parte de la EMV se omitía por sobreentenderse en la situación («Se prepara a la parrilla y se sirve generalmente poco hecho. Este punto hay a quien le produce cierto rechazo por dar la impresión de venir ensangrentado»).

Al tener en cuenta estas consideraciones, se decidió proceder al análisis utilizando como base las siguientes EMV únicas, que se dan en un total de 38 oraciones (tabla 1).

Tabla 1. Listado de las EMV únicas a analizar.

Expresiones multiverbales	Frec.
a punto de nieve	1
aguja de hacer punto	1
en tu punto	1
estar a punto	3
punto caramelo	1
punto de congelación	1
punto de encuentro	1
punto de hebra fina	1
punto de referencia gastronómica	1
punto de sal	7
punto deseado de cocción	1
punto fuerte	15
punto idóneo de cocción	1
punto óptimo de maduración	1
tener su punto	1
un punto picante	1

Una vez acotada la muestra de análisis procedemos a delimitar los sistemas de traducción automática.

3.2. La selección de los sistemas de TA

Para seleccionar los STA que forman parte de este estudio se han tenido en cuenta dos puntos básicos: por un lado, se tuvo en cuenta el hecho de que, a pesar de que dos motores pueden pertenecer a un mismo paradigma, todos ellos se habrán entrenado con un corpus diferente (Bowker, 2023, p. 101) por lo que los resultados obtenidos serán distintos; por el otro, se buscó que fuesen representativos de los sistemas que un usuario cualquiera pudiese utilizar, es decir, que sean lo suficientemente conocidos como para que el usuario considere recurrir a ellos. Por supuesto, la popularidad varía dependiendo del par de lenguas, pues no todos los motores traducen a todas lenguas y, en algunos casos, un par de lenguas concreto solo está disponible a través de una lengua intermedia.

Analizando literatura que trata los STA más populares o hace alusión a ellos (Carré et al., 2022; Rivera-Trigueros, 2022; Bowker, 2023) y contrastando los listados hallados, hemos limitado la muestra a cuatro traductores automáticos: DeepL (NMT), Google Translate (NMT), Microsoft Translator (HMT) y Yandex Translate (SMT en el par de lenguas ES-EN).[3] Los cuatro motores de traducción dan la opción de traducir directamente del español al inglés (estadounidense) sin hacer uso de lenguas pivote y consideramos que, dada su variedad, conforman una base sólida de cara a comparar los pros y contras de cada uno de ellos.

Además, todos ellos se caracterizan por ser gratuitos, estar disponibles en línea y ofrecer la posibilidad de traducir imágenes en su versión aplicación de móvil y archivos completos en su versión web, hechos que podrían potenciar su uso en un contexto turístico.

3.3. La recogida y el análisis de los datos

Una vez determinadas las 16 EMV únicas que conforman el estudio, con sus 38 ocurrencias totales, procedimos a traducirlas uti-

3. El último comunicado de la empresa con respecto a la arquitectura de su STA (Yandex, 2017) indica que se iba a comenzar a utilizar un modelo híbrido, pero únicamente en el par de lenguas inglés-ruso; no hay indicaciones de que esto haya cambiado, por lo que se considerará un traductor automático que funciona con una arquitectura completamente estadística.

lizando los STA arriba indicados. En este respecto es necesario tener en cuenta dos cuestiones:

- Las EMV no se tradujeron de manera aislada, sino que se tradujeron junto sus contextos para incrementar las posibilidades de una traducción apropiada. Como mencionábamos anteriormente, se trata de expresiones convencionalizadas mediante su uso y que, por lo tanto, evolucionan a medida que pasa el tiempo y cambia la situación sociocultural en la que se dan; esto implica que, en algunos casos, podemos encontrar expresiones que presentan varios significados en función del contexto en el que se dan.
- Dado el ámbito en constante cambio en el que nos encontramos, cabe notar que todas las traducciones se obtuvieron en mayo de 2023.

El resultado fueron 152 segmentos en inglés que procedimos a analizar manualmente. El análisis se focalizó en la detección de errores en las traducciones de las EMV de manera estricta, por lo que no se contabilizaron en el total de errores aquellas palabras ajenas a las expresiones –como puede ser el caso en «[tu] punto fuerte», donde el «tu» no se considera parte de la expresión– o errores que parecen surgir a raíz de la presencia de una EMV en la oración que no ha sido interpretada de manera correcta.

Todos los errores hallados se categorizaron siguiendo una modificación propia del sistema de evaluación humana de traducciones propuesto por Ortiz Boix (2016, pp. 63-64), el cual se basa en la precisión y fluidez de las traducciones. Esta clasificación de errores tiene como objetivo analizar los fallos hallados en traducciones automáticas de documentales, por lo que no todas las categorías propuestas se ajustan a las necesidades de nuestro estudio, motivo por el que incorporamos seis nuevas: *gramatical person*, *tenses*, *meaning change*, *wrong suggestion*, *spelling* y *repetition*.

No obstante, a pesar de la diferencia de objetivos de ambos estudios y la incompatibilidad de ciertas categorías con las expresiones multiverbales, decidimos no prescindir de ninguna de las propuestas por Ortiz Boix (2016, pp. 63-64), pues consideramos que podrían ser útiles de cara a catalogar y analizar los errores hallados fuera de las EMV en estudios futuros. Por lo tanto, el sistema de evaluación de traducciones utilizado en este estudio consta de 27 categorías contando las adiciones, 15 de ellas referidas a la precisión y 12 a la fluidez (tabla 2).

Tabla 2. Sistema de categorización de errores empleado en el análisis.

Categorías de precisión	Categorías de fluidez
Terminology	Spelling
Mistranslation	Capitalization
Overly Literal	Diacritics
False Friend	Variant
Should not have been translated	Typography
Date/time	Punctuation
Unit conversion	Unpaired elements
Number	
Entity	Grammar
Grammatical person	Morphology
Tenses	Part of speech
Meaning change	Agreement
Omission	Word order
Addition	Function words
Untranslated	Unintelligible
Wrong suggestion	Repetition

Consideramos que todos los errores detectados tenían la misma relevancia desde el punto de vista de la traducción, por lo que todos ellos se contabilizaron de manera individual para permitirnos detectar los patrones de comportamiento de cada STA de una manera más precisa. En otras palabras: una misma EMV podía contener más de un error.

Una vez expuesta la metodología empleada en este trabajo, procedemos a mostrar los resultados de su aplicación en la muestra de estudio seleccionada.

4. Análisis y resultados

De las 38 oraciones objeto de estudio, 21 de ellas (55,263 %) no presentaron errores en ninguna de sus traducciones, lo que supone que los 73 errores hallados se concentraron en las 17 oraciones restantes. Desde el punto de vista de las traducciones, tan solo 53 de los 152 segmentos analizados presentaron errores, es

decir, no se observaron problemas en la traducción del 65,132 % de las EMV que conforman la muestra de estudio (ver tabla 1).

La metodología utilizada a lo largo de este trabajo nos permite analizar estos datos desde dos perspectivas diferentes. Por un lado, analizar los datos por tipo de error permite comprender las tendencias generales que la TA presenta a día de hoy a la hora de traducir del español al inglés, lo que a su vez nos posibilita determinar qué carencias lingüísticas y comunicativas pueden desprenderse de su uso; por el otro, analizar los datos de cada STA de manera aislada permite comprender los patrones de traducción propios de cada uno de ellos, facilitando así a la redacción el empleo de lenguajes controlados para ser traducidos con un STA concreto o la preedición de los mismos. A continuación, ilustramos los resultados obtenidos desde estas dos perspectivas.

4.1. Resultados por categorización de errores

Los 73 errores hallados durante la fase de análisis se corresponden con 10 de las categorías contempladas, de las cuales siete corresponden a categorías de precisión (*terminology*, *overly literal*, *tenses*, *omission*, *addition*, *untranslated* y *meaning change*) y tan solo tres a categorías de fluidez (*function words*, *agreement* y *part of speech*). Además, la distribución de los errores muestra una clara predominancia en las categorías de precisión, pues estas representan el 94,521 % de los errores, con un total de 69. En contraste, los problemas de fluidez se reducen a un error en *part of speech* (1,370 %), uno en *agreement* (1,370 %) y dos en *function words* (2,740 %).

Esta distribución tan irregular de los errores, sin embargo, es esperable. Aunque una de las características básicas de las expresiones multiverbales es el hecho de que están compuestas por al menos dos palabras, lo cierto es que suelen ser unidades lingüísticas bastante breves y, en nuestra muestra de16 EMV únicas, las más largas están conformadas por solo cuatro palabras. En consecuencia, es comprensible que la presencia de errores de fluidez sea limitada ya que la simplicidad estructural de las expresiones dificulta la aparición de este tipo de errores.

No obstante, este no es el único caso de distribución irregular hallado en los resultados del estudio, pues, entre los errores de precisión, observamos una clara tendencia a la traducción demasiado literal de las expresiones multiverbales. Este error se ha contabilizado en un total de 39 ocasiones, por lo que representa un 56,522 % de los errores de precisión y un 53,425 % de los errores totales. Las cifras registradas en esta categoría son muy superiores a las halladas en el resto, puesto que la segunda categoría más numerosa es *terminology*, que consta de 13 ejemplos en la muestra de estudio, y la tercera, *meaning change*, tan solo cuenta con 6 (tabla 3).

Tabla 3. Índice de errores por categoría de precisión.

	N.º errores	% errores precisión	% errores totales
Overly Literal	39	56,522 %	53,425 %
Terminology	13	18,841 %	17,808 %
Meaning change	6	8,696 %	8,219 %
Tenses	5	7,246 %	6,849 %
Omission	4	5,797 %	5,479 %
Addition	1	1,449 %	1,370 %
Untranslated	1	1,449 %	1,370 %

Resulta muy significativo que esta categoría de error por sí sola suponga más de la mitad de los errores tanto de precisión como de los totales, puesto que se trata de un tipo de error que resulta incompatible con las características intrínsecas de las expresiones multiverbales. Esto nos lleva a preguntarnos hasta qué punto se relacionan estos errores con los hallados en el resto de las categorías y qué impacto han podido tener en estas, principalmente, en los errores por terminología y en los de falso sentido. Este interés está motivado por dos razones: la primera de ellas responde a la vinculación existente entre estas tipologías, pues una traducción literal puede afectar directamente a la precisión terminológica y, en ocasiones, alterar el significado último del texto traducido; la segunda es el hecho de que la combinación de estas categorías asciende al 75,362 % de los errores de precisión y al 71,233 % de los totales. Esta prevalencia sugiere que los problemas de traducción literal y de terminología están intrínsecamente ligados.

Al analizar los datos compilados, observamos que de los 13 errores de terminología que se han detectado, 11 (84,615 %) se

dan en EMV en las que también hay un problema de traducción demasiado literal. Las dos únicas excepciones las hallamos en el STA neuronal DeepL en relación con la expresión «punto de sal» (tabla 4). A diferencia de lo ocurrido en el resto de los sistemas, DeepL no ha optado por traducir «punto» literalmente como *point*, sin embargo, las opciones proporcionadas por este motor tampoco se adecúan al contexto en el que nos encontramos: por un lado, *salt content*, si bien sigue guardando relación con comida, es un término propio de los valores nutricionales que se pueden encontrar en el empaquetado de determinados productos; por el contrario, *salt level*, es un término que se da en el contexto de piscinas u otros cuerpos de agua.

Tabla 4. Muestra de traducción de la EMV «punto de sal» proporcionada por DeepL.

a.	Damos **punto de sal**, colamos y reservamos.	
	DL	Add **salt**, strain and reserve.
b.	Cuando la trucha esté ya cocida y comprobado el **punto de sal**, hay que sacar los trozos, para ponerlos sobre el pan y seguidamente verter el caldo por encima.	
	DL	When the trout is cooked and **the salt** has been checked, remove the pieces and place them on the bread and then pour the broth over them.
c.	Ponemos las vieiras **a punto de sal** y las pasamos por el polvo de cecina formándoles como una piel, las cortamos y las disponemos en el plato junto con el pan de broa que, previamente, habremos cortado en finas láminas y puesto crujiente.	
	DL	**Salt** the scallops and pass them through the jerky powder to form a skin, cut them and place them on the plate together with the broa bread, previously cut into thin slices and crispy.

No obstante, como podemos observar en el apartado C de la tabla anterior, se trata de una EMV que este sistema ha sido capaz de tratar manera apropiada en otras ocasiones, sin caer en cuestiones de traducción demasiado literal al optar por una forma de expresión típica del día a día y que puede encontrarse en numerosas recetas y blogs de cocina.

Esta tendencia a la traducción literal, por lo tanto, ha supuesto numerosos casos de errores terminológicos, un patrón de comportamiento que hemos detectado en todos los STA del estudio. Sin embargo, esto parece no replicarse en los falsos sentidos hallados en la muestra de estudio. Los seis errores que conforman la categoría *meaning change* parecen guardar una mayor relación con los cambios de tiempo verbal y con la adición innecesaria de contenido a la EMV traducida.

La adición se ha registrado en una única traducción proporcionada por Google Translate, donde se le ha añadido un carácter cuantitativo a la expresión «[se le añade el] punto de sal», que en el original denota gusto o preferencia, al traducirla como «[add] a bit of salt». En cuanto a los cambios de significado a causa de una alteración de los tiempos verbales, hemos registrado cinco, es decir, que ninguno de los seis errores en la categoría *meaning change* se asocia a cuestiones de traducción literal o de terminología. Estos cinco errores se han dado en las traducciones de la EMV «estar a punto», donde cabe destacar que Microsoft Translator ha sido el único que ha mostrado una traducción apropiada, si bien se considera la posibilidad de que esto se deba a una carencia en el corpus, pues la alteración de los indicadores temporales del contexto no parecía modificar los resultados de observados en la EMV.

El hecho de que los errores detectados en las categorías *terminology*, *overly literal* y *meaning change* se correspondan con el 79,452 % de los errores totales nos llevó a considerar si la longitud de las EMV podría estar relacionada. Como mencionábamos anteriormente, esta tendencia en las traducciones apunta a un problema con los algoritmos de identificación y segmentación del texto de origen, por lo que el número de componentes de una EMV podría estar directamente relacionada con estos fallos. No obstante, los errores detectados no siguen un patrón que indique que una extensión mayor de la EMV afecte sobremanera a la capacidad de procesamiento de los STA.

Esto nos llevó a cuestionar el impacto que podía tener la frecuencia de uso general de una EMV en la traducción. Sin embargo, al comparar los resultados obtenidos en algunas de las expresiones más comunes que componen la muestra de análisis como pueden ser «punto fuerte» y «estar a punto», hemos observado que, a pesar de su alta frecuencia, ambas se sitúan en puntos opuestos de la escala de errores por ocurrencia. De esta manera, «punto fuerte», que se daba en 15 de las oraciones traducidas, ha sido traducida correctamente en todas las ocasiones, mientras que «estar a punto» presentaba una incidencia de 3,333 errores/ocurrencia al concentrar diez fallos en tan solo tres oraciones.

Una vez presentados los datos recopilados por categoría de error, procedemos a explicarlos desde la perspectiva del rendimiento general de cada uno de los STA empleados en este trabajo.

4.2. Resultados por sistemas de traducción automática

Si bien el número total de errores cometidos por los cuatro sistemas estudiados es bastante similar (DeepL, 17; Google Translate, 16; Microsoft Translator, 19; Yandex.Translate, 21), lo cierto es que existen grandes diferencias en cuanto al rendimiento mostrado por cada uno de ellos a la hora de traducir las EMV contextualizadas propuestas.

La primera diferencia la encontramos en el número de tipologías de error que cada STA abarca. Desde esta perspectiva, observamos que, aunque Microsoft Translator es uno de los sistemas que más errores comete, tan solo registra tres tipos de error: terminológicos, por traducción literal y por palabras funcionales. Esto contrasta con los seis tipos de error hallados en Google Translate y los siete de DeepL y Yandex.Translate. Cabe mencionar que, en la muestra de análisis seleccionada, Google Translate ha cometido cinco errores asociados con la categoría *wrong suggestion*, siendo el único en cometer este tipo de error, si bien estos no se han tenido en cuenta para el conteo final de errores por considerar que afectaban a la oración completa y no a la EMV de manera estricta.

En cuanto a errores específicos, es notable que Google Translate no comete errores asociados con la fluidez, pero es el único que ha presentado problemas con la adición de información. Por su parte, DeepL es el único que presenta errores en dos categorías de fluidez, *part of speech* y *agreement*, y el único que comete este tipo de error; mientras que Yandex.Translate es el único que ha dejado parte de una EMV sin traducir. En este sentido, Microsoft Translate no solo destaca por no presentar una gran variedad de tipos de error, sino que, además, es el único en cometer un tipo concreto de error.

A pesar de esto, este motor es el que presenta un rendimiento más irregular, lo que nos lleva a la segunda diferencia: la cantidad de errores cometidos por categoría. Desde esta perspectiva, cabe destacar los 15 errores registrados en *overly literal* por parte de Microsoft Translator. Este pico negativo denota un problema a la hora de trasvasar estructuras entre lenguas por parte del sistema, algo que ha sucedido en el 39,473 % de las oraciones de este estudio. Se trata de un porcentaje muy alto de incidencia, lo que contrasta con los 4 errores (10,526 %) cometidos por DeepL en esta categoría, lo que lo sitúa como el STA que tiene una menor tendencia a este tipo de error (hecho que hemos corroborado al observar que, además, este es el sistema que más a menudo intenta simplificar las estructuras o alejarse del original) (figura 1).

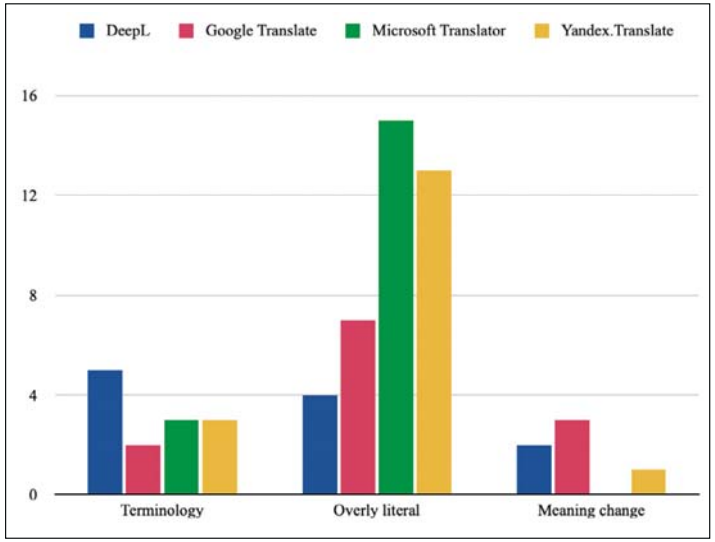

Figura 1. Distribución de los errores más comunes según los STA.

Al centrar el análisis en los errores más comunes que mencionábamos anteriormente, detectamos, no obstante, que esta es la categoría con la diferencia más. Los datos extraídos indican que la media de error de esta categoría se sitúa en los 9,75 errores por sistema, 6,5 puntos más que en el caso de *terminology* y 8,25 en el caso de *meaning change*. Como cabría esperar, dada la tendencia hacia estas tres categorías, las cuales constituyen el 79,452 % de todos los errores, la media de error en el resto se sitúa entre el 0,25 y el 1,25.

Tras haber descrito los resultados recabados desde las dos perspectivas consideradas para la fase de análisis de este trabajo, procedemos a presentar las conclusiones.

5. Conclusiones

La proliferación de los sistemas de traducción automática disponibles en línea de manera gratuita nos ha llevado a considerar hasta qué punto su uso podía afectar a la comunicación y en qué formas podría hacerlo, particularmente en lo referido a las expresiones multiverbales, pues se consideran unidades complejas de traducir y extremadamente comunes en los lenguajes naturales. En este trabajo hemos analizado esta incidencia usando para ello textos del turismo gastronómico, un ámbito que se caracteriza por la abundancia de EMV que utiliza, y cuatro de los sistemas de traducción automática más populares del mercado.

Los datos obtenidos del análisis han mostrado una clara tendencia hacia la traducción literal de las expresiones, un hecho que resulta de por sí incompatible con las características intrínsecas de las EMV, ya que estas presentan con regularidad cierta idiomaticidad, por lo que sus significados no pueden deducirse a partir de la suma de los significados de sus componentes. Esta tendencia a la traducción literal afecta a la precisión terminológica de las EMV traducidas, un resultado sin duda ligado a la estrecha relación existente entre ambas tipologías de error, aunque, en este caso concreto, no se ha asociado a cambios de significado a pesar de estar estos también estrechamente ligados a las categorías previamente mencionadas.

Además, hemos observado que estos errores no están relacionados con la frecuencia de uso general de una EMV específica. En este respecto, hemos detectado que una EMV común, como podría ser el caso de «punto fuerte», que aparecía recogida 15 veces en nuestro corpus, ha sido traducida de manera apropiada en todos los casos, mientras que expresiones igualmente comunes como «estar a punto» presentaban hasta 10 errores en tan solo tres ocurrencias. Los datos apuntan a que este patrón de errores está relacionado con problemas durante la fase de fragmentación del texto de entrada.

No obstante, los indicios apuntan a que la longitud de las EMV no es un factor determinante en la aparición de errores a causa de problemas en la identificación y fragmentación que terminan por dividir incorrectamente las expresiones. Dado que este comportamiento se ha dado en EMV de distintas longitudes, todo apuntaría a que el problema radica en la naturaleza idiomática y compleja de las expresiones.

En el futuro, sería fundamental analizar más a fondo la relación entre los patrones de comportamiento destacados de los distintos motores de traducción aquí estudiados y varios factores lingüísticos y operacionales. En primer lugar, es necesario examinar cómo la estructura gramatical de las EMV influye en la incidencia de errores específicos, pues determinar si ciertos patrones gramaticales son más propensos a errores podría orientar el desarrollo de algoritmos más precisos y eficaces.

Asimismo, sería necesario profundizar en las carencias características de cada motor, como pueden ser las repeticiones o sugerencias incorrectas que mencionábamos en el apartado anterior, pues comprender estas deficiencias específicas permitirá personalizar las mejoras en cada sistema, aumentando su eficiencia y precisión en la traducción de EMV. Además, sería necesario comprobar cómo la densidad terminológica de las oraciones de origen afecta a la traducción de las EMV, ya que los datos parecen indicar que la complejidad terminológica de las oraciones afecta a la capacidad de los STA para procesar y traducir adecuadamente las EMV.

Por supuesto, sería necesario ampliar la investigación a un mayor número de EMV y aplicar el análisis a otros campos del saber, lo que nos permitirá evaluar y comprender en mayor profundidad cómo las expresiones multiverbales son procesadas por los STA actuales y cómo continuar refinando estos de cara a eliminar las barreras lingüísticas con mayor eficacia.

6. Agradecimientos

El presente trabajo se ha realizado en el marco de los proyectos nacionales de I+D titulados *Multi-lingual and Multi-domain Adaptation for the Optimisation of the VIP system* (VIP II) (PID2020-112818GB-I00), coordinado por la Dra. Gloria Corpas Pastor (Universidad de Málaga) y *App para entrenar en posedición de traducción automática neuronal mediante la gamificación en entornos profesionales* (GAMETRAPP) (TED2021-129789B-I00), coordinado por la Dra. Cristina Toledo Báez (Universidad de Málaga).

7. Referencias bibliográficas

Álvarez Jurado, M. (2020). Adquisición y transmisión del conocimiento experto a través de la traducción de las guías turísticas de arquitectura. *Onomázein*, (NE VII), 1-17. https://doi.org/10.7764/onomazein.ne7.01

Anthony, L. (2020). *AntConc (Versión 3.5.9) [Programa de ordenador]*. Waseda University. https://www.laurenceanthony.net/software/antconc/

Austermühl, F. y Kortenbuck, A. (2012). A translator's sword of Damocles? An introduction to machine translation. En F. Austermühl (Ed.), *Electronic tools for translation* (3ª ed.) (pp. 153-176). Routledge.

Biber, D., Johansson, S., Leech, G., Conrad, S. y Finegan, E. (1999). *Grammar of spoken and written English*. Pearson Education Limited.

Bowker, L. (2023). *De-mystifying Translation: Introducing Translation to Non-translators*. Routledge. https://doi.org/10.4324/9781003217718

Bowker, L. y Buitrago Ciro, J. (2019). *Machine translation and global research: towards improved machine translation literacy in the scholarly community*. Emerald Publishing.

Brown, P., Cocke, J., della Pietra, S., della Pietra, V., Jelinek, F., Mercer, R. y Roossin, P. (1988). A statistical approach to language translation. *Proceedings of the 12th conference on Computational linguistics*, *1*, 71-76.

Brown, P., Cocke, J., della Pietra, S., della Pietra, V., Jelinek, F., Lafferty, J. D., Mercer, R. y Roossin, P. (1990). A statistical approach to language translation. *Computational linguistics*, *16*(2), 79-85.

Calvi, M. V. (2010). Los géneros discursivos en la lengua del turismo. *Ibérica: Revista de la Asociación Europea de Lenguas para Fines Específicos (AELFE)*, *19*, 9-32.

Calvi, M. V. (2019). Géneros discursivos, diversidad cultural y traducción de textos turísticos. *Altre Modernità: Rivista di studi letterari e culturali*, *21*, 69-86. https://doi.org/10.13130/2035-7680/11633

Carré, A., Kenny, D., Rossi, C., Sánchez-Gijón, P. y Torres-Hostench, O. (2022). Machine translation for language learners. En D. Kenny (Ed.), *Machine translation for everyone: Empowering users in the age of artificial intelligence* (pp. 187-207). Language Science Press. https://doi.org/10.5281/zenodo.6760024

Corpas Pastor, G. (2013). Detección, descripción y contraste de las unidades fraseológicas mediante tecnologías lingüísticas. En I. Olza Moreno y E. Manero Richard (Coords.), *Fraseopragmática* (pp. 335-374). Frank & Time.

Costa-jussà, M. R. y Fonollosa, J. A. (2015). Latest trends in hybrid machine translation and its applications. *Computer Speech and Language*, *32*(1), 3-10.

Hernández Mogollón, J. M., Di-Clemente, E. y López Guzmán, T. (2015). El turismo gastronómico como experiencia cultural. El caso práctico de la ciudad de Cáceres. *Boletín de la Asociación de Geógrafos Españoles*, *68*, 407-425.

ISO (2017). *18587 International Standard. Translation services – Post-editing machine translation output* – Requirement. Recuperado el 18 de diciembre de 2021 en https://www.iso.org/standard/62970.html

Jackendoff, R. (1977). *The architecture of the Language Faculty*. MIT Press.

Kenny, D. (2022). Human and machine translation. En D. Kenny (Ed.), *Machine translation for everyone: empowering users in the age of artificial intelligence* (pp. 23-49). Language Science Press.

Kilgarriff, A., Rychlý, P., Smrž, P. y Tugwell, D. (2004). The Sketch Engine. *Proceedings of the 11th EURALEX International Congress*, 105-116. http://www.sketchengine.eu

Koehn, P.; Och, F. J. y Marcu, D. (2003). Statistical Phrase-Based Translation. *Proceedings of the 2003 Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics*, 127-133.

Microsoft (s.f.). *Languages*. Recuperado el 2 de septiembre de 2023 en https://www.microsoft.com/en-us/translator/languages/

Ministerio de Industria, Comercio y Turismo (2023, 2 de febrero). Datos de Frontur y Egatur del INE [Nota de prensa]. Recuperado el 4 de febrero de 2024 en http://www.mincotur.gob.es/es-es/GabinetePrensa/NotasPrensa/2023/Paginas/En-2022-visitaron-España-71,6-millones-de-turistas-internacionales-que-realizaron-un-gasto-de-87.061-millones-de-euros.aspx

Mitkov, R., Seretan, V., Corpas Pastor, G. y Monti, J. (2018). Multiword units in machine translation and translation technology. En R. Mitkov, V. Seretan, G. Corpas Pastor y J. Monti (eds.), *Multiword units in machine translation and translation technology* (pp. 1-38). John Benjamins.

Moorkens, J. (2022). Ethics and machine translation. En D. Kenny (Ed.), *Machine translation for everyone: empowering users in the age of artificial intelligence* (pp. 121-140). Language Science Press.

Oliver, A. (2016). *Herramientas Tecnológicas para Traductores*. Editorial UOC. https://elibro.net/es/lc/uva/titulos/58451

Organización Mundial del Turismo. (2023). *World tourism barometer – May 2023 (excerpt)*, 21(2). Recuperado el 24 de mayo de 2024 de

https://webunwto.s3.eu-west-1.amazonaws.com/s3fs-public/2023-05/UNWTO_Barom23_02_May_EXCERPT_final.pdf
Organización Mundial del Turismo y Basque Culinary Center. (2019). *Guía para el desarrollo del turismo gastronómico*. OMT. https://doi.org/10.18111/9789284420995
Ortego Antón, M. T. (2019). *La terminología del sector agroalimentario (español-inglés) en los estudios contrastivos y de traducción especializada basados en corpus: los embutidos*. Peter Lang.
Ortego Antón, M. T. (2024a). The Design of Torrezno TRAD: the Semiautomatic Spanish-English Writing and Translation Aid Tool. En I. Peñuelas Gil, I. y M. T. Ortego Antón, (Eds.), *Interpreting and Translation for Agri-food Professionals in the Global Marketplace* (pp. 69-84). DeGruyter. https://doi.org/10.1515/9783111101729-004
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. https://doi.org/10.5007/2175-7968.2024.e94647
Ortiz Boix, C. (2016). *Implementing Machine Translation and Post-Editing to the Translation of Wildlife Documentaries through Voice-over and Off-screen Dubbing* [Tesis doctoral]. Universidad Autónoma de Barcelona. http://hdl.handle.net/10803/400020
Penadés Martínez, I. (2015). *Para un diccionario de locuciones. De la lingüística teórica a la fraseografía práctica.* Universidad de Alcalá.
Peñuelas Gil, I. (2024a). From corpora to a semi-automatic tool: the use of model lines for the development of TorreznoTRAD. En I. Peñuelas Gil, I. y M. T. Ortego Antón (Eds.), *Interpreting and Translation for Agri-food Professionals in the Global Marketplace* (pp. 105-118). DeGruyter. https://doi.org/10.1515/9783111101729-006
Peñuelas Gil, I. (2024b). *Estudio contrastivo del tratamiento de las expresiones multiverbales del turismo gastronómico en los sistemas de traducción automática del español al inglés* [Tesis doctoral]. Universidad de Valladolid.
Pérez Blanco, M. y Izquierdo, M. (2021). Developing a corpus-informed tool for Spanish professionals writing specialized texts in English. En J. Lavid-López, C. Maíz-Arévalo y J. R. Zamorano-Mansilla (Eds.), *Corpora in Translation and Contrastive Research in the Digital Age* (pp. 147-173). John Benjamins.
Quah, C. K. (2006). *Translation and Technology*. Palgrave Macmillan.
Rivera-Trigueros, I. (2022). Machine translation systems and quality assessment: a systematic review. *Lang Resources & Evaluation*, *56*, 593-619. http://doi.org/10.1007/s10579-021-09537-5

Sag, I. A., Baldwin, T., Bond, F., Copestake, A. y Flickinger, D. (2002). Multiword expressions: A pain in the neck for NLP. *Proceedings of the third international conference on intelligent text processing and computational linguistics (CICLING 2002)*, 1–15.
Sánchez Carnicer, J. (2022). *Traducción y discapacidad. Un estudio comparado de la terminología inglés-español en la prensa escrita*. Peter Lang. https://doi.org/10.3726/b19567
Sánchez Ramos, M. M. y Rico Pérez, C. (2020). *Traducción automática: Conceptos clave, procesos de evaluación y técnicas de posedición*. Comares.
Seghiri, M. (2017). Metodología de elaboración de un glosario bilingüe y bidireccional (inglés-español/español-inglés) basado en corpus para la traducción de manuales de instrucciones de televisores. *Babel*, *63*(1), 43-64. https://doi.org/10.1075/babel.63.1.04seg
Soto Almela, J. (2014). *Los términos culturales en el ámbito turístico español-inglés: traducción, manipulación y recepción real en usuarios anglófonos*. Tesis doctoral. Universidad de Murcia. Recuperado el 13 de junio de 2022 de http://hdl.handle.net/10201/41596
Vieira, L. N. (2020). Machine Translation in the News: A Framing Analysis of the Written Press. *Translation Spaces*, *9*(1), 98-122. https://doi.org/10.1075/ts.00023.nun
Yamada, K. y Knight, K. (2001). A syntax-based statistical translation model. *Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics*, 523-530.
Yandex (2017). Supported languages. *Yandex Support*. Recuperado el 2 de septiembre de 2023 de https://yandex.com/support/translate-mobile/supported-langs.html
Zens, R., Och, F. J. y Ney, H. (2002). Phrase-based statistical machine translation. En M. Jarke, J. Koehler y G. Lakemeyer (Eds.), *KI 2002: Advances in Artificial Intelligence: 25th Annual German Conference on AI, KI 2002 Aachen, Germany, September 16-20, 2002 Proceedings 2479* (pp. 18-32). Springer.