



UNIVERSITAT DE VIC
UNIVERSITAT CENTRAL
DE CATALUNYA

LA PRODUCTIVIDAD DE LA TRADUCCIÓN ASISTIDA Y AUTOMÁTICA EN LOS TEXTOS DEL ÁREA DE ESTADÍSTICA: UNA EVALUACIÓN DE LA TRADUCCIÓN DEL INGLÉS AL ESPAÑOL

Autora: Margalina, Vasilica-Maria

Tutor: Canovas Mendez, Marcos

Curso: 2022-2023

Trabajo Fin de Master

Máster en Traducción Especializada

Universitat de Vic-Universitat Central de Catalunya

Fecha de depósito: 6 de septiembre de 2023

AGRADECIMIENTOS

Este Trabajo de Fin de Master no hubiera sido posible sin la ayuda de muchas personas a las que me gustaría mostrar mi agradecimiento.

En primer lugar, me gustaría agradecer al Prof. Dr. Christian M. Ringle de la Universidad de Hamburgo, Alemania, por facilitarme los textos para la elaboración de este trabajo.

En segundo lugar, quiero agradecer a mi tutor Marcos Canovas por sus valiosas recomendaciones en la realización de este trabajo. Además, agradezco a todos los profesores del Master de Traducción Especializada de la Universidad de Vic por los conocimientos compartidos y la guía recibida durante el master.

Por último, me gustaría agradecerle a mi familia por el apoyo incondicional y la paciencia.

Índice de contenido

1.	Introducción.....	9
1.1.	Justificación.....	9
1.1.	Objetivos	10
2.	Marco teórico	11
2.1.	Aclaración de los conceptos de traducción asistida y traducción automática	11
2.2.	La productividad en el proceso de traducción con el uso de las TIC a la traducción	13
2.2.1	El proceso de posesición en la traducción automática y TAO	13
2.2.2.	Revisión de la literatura sobre la productividad en la traducción en la era digital	16
3.	Metodología	30
4.	Resultados.....	36
5.	Discusión de los resultados.....	45
6.	Conclusiones.....	48
	Referencias	50
7.	Anexos.....	57
7.1.	Anexo 1.....	57
7.2	Anexo 2.....	58
7.3	Anexo 3.....	59

Índice de cuadros, tablas y figuras

CUADROS

Cuadro 1. Estudios sobre la productividad en el proceso de traducción ...	17
Cuadro 2. Revisión de estudios que han analizado el esfuerzo temporal ..	19
Cuadro 3. Revisión de estudios que han analizado el esfuerzo técnico.....	24
Cuadro 4. Revisión de estudios que han analizado el esfuerzo cognitivo ..	28
Cuadro 5. Descripción de la muestra de textos y segmentos seleccionados	31
Cuadro 6. Variables analizadas.....	32
Cuadro 7. <i>Tipologías de los errores MQM utilizados en el análisis</i>	34
Cuadro 8. <i>Ficha clasificación de errores y cálculo OQS_TA</i>	57
Cuadro 9. Segmentos traducción automática y fuzzy matches	59

TABLAS

Tabla 1. Estadísticos descriptivos.....	36
Tabla 2. Estadísticos descriptivos de Prod_TA en función del nivel de BLEU	37
Tabla 3. <i>Tipos y niveles de los errores</i>	38
Tabla 4. Resultados de la prueba Mann-Whitney H1	39
Tabla 5. <i>Resultados prueba U de Mann-Whitney H2 y H3</i>	40
Tabla 6. <i>Estadísticos descriptivos de la productividad en función de la presencia de errores del ámbito de la estadística</i>	40
Tabla 7. Estadísticos descriptivos por texto fuente de los segmentos	41
Tabla 8. Resultado de las correlaciones de Spearman.....	42

Tabla 9. Análisis descriptivo de las fuzzy matches de la memoria de traducción.....	44
--	----

Tabla 10. Resultados Shapiro-Wilk.....	58
---	----

FIGURAS

Figura 1. <i>Frecuencia de los errores</i>	37
---	----

RESUMEN

El presente trabajo de investigación tiene como objetivo evaluar la productividad de la posesición traducción asistida y la traducción automática del inglés al español de textos de estadística. Se realizó un análisis descriptivo de los niveles de productividad de la posesición de la traducción automática y de la posesición de la traducción asistida con la traducción automática integrada de 191 segmentos obtenidos de cuatro casos de estudio de un libro de estadística. Así mismo, se compararon los niveles de productividad de la posesición en función del nivel de calidad medido con una métrica automática (BLEU) y una métrica de evaluación humana (OQS), calculada aplicando el método MQM. El sistema utilizado para la traducción automática y la traducción asistida es Trados Studio. El método MQM permitió también analizar el tipo de errores y su nivel de gravedad. La prueba U de Mann-Whitney mostro que existen diferencias significativas en la productividad de la posesición de la traducción automática y la productividad de la posesición de la traducción asistida con traducción automática integrada. Además, la existencia de errores en la traducción de la terminología influye en los niveles de productividad en la posesición realizada con los dos procedimientos. Adicionalmente, se encontraron correlaciones significativas entre el tamaño del segmento y la calidad de la traducción automática con la productividad en el caso de ambos procedimientos.

Palabras clave: BELU, calidad, posesición, traducción asistida, traducción automática

ABSTRACT

The objective of this research is to evaluate the productivity of post-editing of computer-assisted and machine translation of statistical texts from English to Spanish. A descriptive analysis of the productivity levels of machine translation post-editing and computer-assisted translation with integrated machine translation post-editing was performed on 191 segments obtained from four case studies of a statistic book. Furthermore, the productivity levels of the post-editing process were compared according to the quality level measured with an automatic metric (BLEU) and an human evaluation metric (OQS), calculated by applying the MQM method. Trados Studio was the system used for machine translation and computer-assisted translation. The MQM method was also used to analyze the type of errors and their severity levels. The Mann-Whitney U test showed that there are significant differences in the productivity of machine translation post-editing when comparing to the productivity of machine translation post-editing with integrated machine translation. In addition, the existence of errors in the translation of terminology influences the productivity levels in the post-editing performed with the two procedures. Significant correlations were also found between segment size and machine translation quality with productivity in the case of both procedures of post-editing.

Keywords: BLEU, computer-assisted translation, machine translation, post-editing, quality

1. Introducción

La traducción automática junto a la presión del precio representan desafíos mayores para las empresas europeas, según European Language Industry Survey (EUATC, 2022). Pero a la vez la traducción automática y la posesición están vistas como oportunidades para poder hacer frente a los desafíos y para mejorar la productividad. A la vez, se observa un aumento del interés en la literatura por investigar el efecto de los nuevos avances en la traducción automática traducción asistida sobre la productividad del traductor. Por tanto, el número estudios de la traducción automática, definida como un subcampo de la lingüística computacional o procesamiento del lenguaje natural que investiga el uso del software para traducir texto o voz de un lenguaje natural a otro (Qun y Xiaojun, 2023), ha aumentado de manera considerable en la última década. En este sentido, la presente investigación representa un aporte al campo de investigación de la traducción automática.

1.1. Justificación

Los avances en tecnología y en inteligencia artificial han contribuido al rápido desarrollo de los sistemas de traducción automática, su empleo convirtiéndose en “algo tan esencial como habitual” (López Pereira, 2019, p. 2). Los nuevos sistemas basados en inteligencias artificial y en redes neuronales han logrado producir traducciones más fluidas y naturales, los cuales deberían aumentar la productividad de la posesición (Rivera-Trigueros, 2022). Los buenos resultados de estas nuevas tecnologías han provocado afirmaciones de que la calidad de la traducción de estos sistemas es la misma que la traducción humana o predicciones relacionadas con el fin de la traducción humana como profesión (Castilho *et al.*, 2018). No obstante, lejos de reemplazar a la traducción humana, hoy en día la traducción automática ha sido incorporada a los tradicionales sistemas de traducción asistida por ordenador (TAO), siendo utilizada juntamente con las memorias de traducción y los glosarios de términos. A la vez, ha aumentado la

importancia del proceso de posesición para obtener traducciones de alta calidad (Do Carmo y Moorkens, 2020).

Así mismo, los resultados de los estudios indican que, aunque se ha producido una mejora sustancial en la traducción automática con las nuevas tecnologías, los resultados en términos de calidad difieren entre pares de idiomas y áreas de especialidad (Wang, Wu, He, Huang y Ward Church, 2022). Sin embargo, existen muy pocos estudios que haya investigado el efecto sobre la productividad de la integración de las nuevas tecnologías de TA en los sistemas TAO. La mayoría de los estudios comparan las recomendaciones de la TA con las coincidencias *fuzzy matches* de las memorias de traducción, sin tener en cuenta de que hoy en día estas sugerencias se muestran los sistemas de traducción de manera simultánea.

Las métricas utilizadas para medir la productividad y la calidad de la traducción representan otro aspecto relevante en la investigación en el campo de la traducción automática, sobre el cuál no se ha llegado todavía a un consenso en la literatura científica (Stasimioti y Sonsoni, 2020; Rivera-Trigueros, 2022). Pero si hay un aspecto sobre el cual coinciden varios autores: la evaluación de la traducción automática requiere tanto de métricas automáticas como de evaluaciones humanas siempre que se necesita un análisis profundo e informativo (Chunyu y Tak-ming, 2015; Gutiérrez-Artacho, Olvera-Lobo y Rivera-Trigueros, 2018). Por esta razón, la presente investigación pretende contribuir a la evaluación de la traducción automática mediante el análisis de su productividad y calidad.

1.1. Objetivos

El objetivo de la investigación es evaluar la productividad de la traducción asistida por ordenador y la traducción automática del inglés al español en los textos del área de estadística.

Adicionalmente, la investigación tiene los siguientes objetivos específicos:

- Evaluar las diferencias de productividad en la posesición en función de los niveles de calidad de la traducción automática.

- Identificar el impacto de los errores en la traducción automática de la terminología estadística en la productividad.
- Analizar las diferencias de productividad de la posesión entre la traducción asistida por ordenador y la traducción automática.

2. Marco teórico

2.1. Aclaración de los conceptos de traducción asistida y traducción automática

El desarrollo de la tecnología de la información de las últimas décadas ha dejado su huella en la traducción, mediante el surgimiento de varias aplicaciones informáticas con diferentes enfoques y para tareas específicas del proceso de traducción. Destacan principalmente dos tipos de aplicaciones: las utilizadas para la traducción asistida y las enfocadas en la traducción automática. Estos dos tipos tecnologías tienen enfoques y resultados diferentes, además de ser utilizados es distintos contextos (Crăciunescu, Gerding-Salas y Stringer-O'Keeffe, 2004). Además, dos nuevos términos han surgido en la traducción que es necesario aclarar: la traducción asistida por ordenador (TAO) y la traducción automática (TA). A pesar de las diferencias, tanto la TAO como la TA tienen el mismo objetivo: mejorar la eficacia de la traducción mediante el uso de la tecnología (Chunyu y Tak-ming, 2015). Otro objetivo de cualquier tipo de traducción (TAO, TA y humana) es conseguir traducciones de alta calidad (Sin-wai, 2015).

Los sistemas TAO iniciaron su desarrollo a principios de los años 90 para responder a la necesidad de empresas e instituciones de traducir un gran volumen de texto en plazos ajustados (Garcia, 2015). En este contexto, era necesario desarrollar herramientas que permitieran reutilizar traducciones revisadas y aplicar sistemáticamente la misma terminología. Esto se consiguió con la TAO, que representa un proceso complejo que se lleva a cabo a través de herramientas específicas y adaptadas a las necesidades del traductor para asistirle de manera eficiente durante su trabajo (Crăciunescu *et al.*, 2004; Oliver, 2016). Los recursos más utilizados en los sistemas TAO

son las memorias de traducción, las herramientas de alineación de textos y los glosarios (Fernández-Rodríguez, 2010; Oliver, 2016). Estos sistemas también pueden incluir funciones para la gestión de proyectos, corrección ortográfica, recuento de palabras y líneas y la posibilidad de generar diferentes tipos de archivos (Melby y Wright, 2015).

La reutilización de traducciones mediante memorias de traducción representa la característica más reconocida y apreciada de la TAO (Jaworski, 2013). Las memorias de traducción son bases de datos en los cuales se almacenan segmentos previamente traducidos del mismo texto o de otros textos para su futura reutilización (Crăciunescu *et al.*, 2004). Normalmente, un sistema TAO registra un segmento de la lengua de origen (generalmente una oración) combinado con un segmento de la lengua meta; es decir, pares bilingües. Durante la traducción, el sistema detecta un segmento idéntico o similar (denominado *fuzzy matches*) y lo sugiere para la nueva traducción (Screen, 2017). Estos segmentos pueden provenir de textos traducidos previamente por el mismo traductor o por otro profesional (Oliver, 2016). Para generar memorias traducción a partir de segmentos traducidos previamente sin una aplicación informática de traducción asistida, los sistemas TAO incorporan herramientas de alineación de textos. En el proceso de alineación, los documentos paralelos (uno en la lengua fuente y el otro en la lengua meta) se emparejan, segmentan y codifican para su importación en la memoria de traducción designada (García, 2015).

Los glosarios, junto a los diccionarios y las bases de datos terminológicos representan funciones de los sistemas TAO que permiten la gestión terminológica, utilizándose con el objetivo de automatizar y facilitar el proceso de consulta (Fernández-Rodríguez, 2010). Durante el proceso de traducción de una frase, el traductor recibe coincidencias de palabras o frases provenientes de estas herramientas. Los glosarios se dividen en dos categorías: los diccionarios incorporados y los glosarios creados por el usuario (Jaworski, 2013).

American Translator Association (s.f.) define la traducción automática como el uso del software automatizado que traduce un texto sin la intervención humana. Los sistemas de traducción automática se han basado

tradicionalmente en los enfoques basados en reglas y los apoyados en corpus, pero en los últimos años han surgido nuevos desarrollos con enfoques mixtos o basados en redes neuronales (Rivera-Trigueros, 2022), que han logrado producir traducciones para algunos pares de idiomas y dominios con muy alta adecuación y fluidez en particular (Wang *et al.*, 2022).

Por lo tanto, la principal diferencia entre los dos conceptos está representada por el grado de la intervención humana en el proceso de traducción. Esta diferencia está también relacionada con el nivel de automatización del proceso de traducción. En este sentido, el objetivo de la traducción automática era, hasta hace unos años, que el proceso completo sea realizado por un sistema informático; mientras que, en la traducción asistida por ordenador, el traductor selecciona entre diferentes herramientas que automatizan fases del proceso, lo que le ofrece el control del proceso de traducción (Fernández-Rodríguez, 2010). No obstante, la TA ya está disponible en la mayoría de las herramientas TAO y desempeña un rol cada vez más importante en el proceso de traducción (Vieira, Alonso y Bywood, 2019). Hoy en día, las sugerencias de la TA se añaden a las típicas coincidencias exactas (*exact matches*) y parciales (*fuzzy matches*) de la memoria de la traducción (Caldwell, Castilho, O'Brien y Mitchell, 2016). A continuación, un traductor humano se encarga de realizar la posesición de las sugerencias provenientes de la TA y de la memoria de traducción (Jaworski, 2013). Por tanto, la integración entre los dos tipos de aplicaciones ha aumentado, lo que ha hecho que delimitación tradicional entre los dos conceptos sea cada día más difusa.

2.2. La productividad en el proceso de traducción con el uso de las TIC a la traducción

2.2.1 El proceso de posesición en la traducción automática y TAO

El proceso de traducción se lleva a cabo en dos etapas: una primera etapa, que tiene como resultado un borrador o versión preliminar del texto en la lengua meta; y la etapa de revisión, durante la cual se pueden realizar desde pequeños ajustes hasta cambios radicales al texto (Crăciunescu *et al.*, 2004).

No obstante, la automatización de los procesos de traducción ha llevado al surgimiento de una nueva actividad: la posesición (Mercader-Alarcón y Sánchez-Martínez, 2016). El aumento en la incorporación de la posesición en muchos procesos de trabajo de traducción se produjo con el aumento considerable de la calidad de los resultados de la traducción automática partir del surgimiento de la TA basada en redes neuronales un (Álvarez, Oliver y Badia, 2020).

La posesición se puede definir como la edición, modificación o corrección de la versión preliminar del texto resultada de la traducción automática (Allen, 2003, p. 296). Se puede hablar de posesición "ligera" o "completa" en función de si el objetivo es obtener una traducción de difusión o de asimilación (O'Brien, 2022). Estos dos niveles de posesición se basan en los propuestos por Allen (2003), quien explica que la posesición "ligera" representa una posesición "mínima", es decir, con muy pocos cambios realizados en los resultados de la TA. Por tanto, cada uno de estos dos niveles de posesición requiere de un esfuerzo diferente (Ortiz-Boix y Matamala, 2016).

Según la definición de posesición, parecería que la traducción automática hubiera reemplazado la primera etapa del proceso de traducción y la posesición hubiera tomado el lugar de la etapa de revisión (Do Carmo y Moorkens, 2020). Sin embargo, Do Carmo y Moorkens (2020) argumentan que la posesición no puede ser considerada una revisión porque las dos actividades tienen objetivos diferentes. El único objetivo de la revisión es eliminar errores de una traducción finalizada, condición que no se cumple en el caso de la posesición porque la traducción automática produce solo un resultado o una sugerencia y no una versión final de la traducción, la cual está bajo la responsabilidad del poseedor. Es más, muchos profesionales han afirmado que, después de poseitar algunos segmentos provenientes de la TA, borran los segmentos restantes y traducen desde cero (Parra y Arcedillo, 2015).

A pesar de estas declaraciones, el uso de un resultado de la TA y su posterior posesición por parte de un traductor se ha convertido en una práctica común en el caso de algunos sectores y para ciertos pares de idiomas (Statismioti y Sosoni, 2020). Hoy en día, la posesición se ha consolidado como un servicio

ofrecido al cliente por parte de las empresas de traducción, el cual cuenta hasta con su propia norma internacional, la ISO 18587, que fue creada en el año 2017 (Vieira *et al.*, 2019). La mejora de las técnicas de traducción automática junto a la incorporación de los sistemas TA en las herramientas TAO fueron los factores principales que llevaron al aumento del número de empresas que incluyen la posesición entre sus procesos de traducción (Mercader-Alarcón y Sánchez-Martínez, 2016). Hoy en día, las sugerencias de la traducción automática son utilizadas en el proceso de decisión junto a las coincidencias, los términos y las concordancias de las memorias de traducción (Do Carmo y Moorkens, 2020). Es más, en algunos contextos, la memoria de traducción y la TA se pueden combinar, de modo que el contenido de la MT se puede utilizar para perfeccionar el resultado de la TA, lo que aumenta el grado de integración de las distintas funciones y tecnologías a disposición de los traductores (Vieira *et al.*, 2019).

A pesar del uso cada vez más extendido de la TA por la industria de la traducción, muchos traductores desconfían que los sistemas de TA puedan producir resultados de una calidad que pueda contribuir al aumento de su productividad (Sánchez-Gijón, Moorkens y Way, 2019). La calidad de la TA no es la única razón por la cual los traductores desconfían de la posesición. Los efectos de la TA sobre el mercado de la traducción (Vieira *et al.*, 2019) y sobre la profesión del traductor (Statismioti y Sosoni, 2020) se encuentran también entre los factores de desconfianza o rechazo. Así mismo, según el estudio llevado a cabo por Moorkens y O'Brien (2015), los traductores se quejan de que la actividad de posesición implica menos creatividad y menos oportunidades para la mejora de la calidad en comparación con la traducción desde cero.

No obstante, las razones que hay detrás de la resistencia podrían no corresponder con la realidad. Algunos estudios muestran que los niveles de calidad de los segmentos MT y los resultados de la TA son similares (Gueberof, 2009; Sánchez-Gijón *et al.*, 2019). Otros estudios han encontrado que la traducción obtenida mediante la posesición puede ser de la misma calidad e incluso superior a la que resulta de la traducción desde cero (Toral, Wieling y Way, 2018). Pero hay que tener en cuenta que la calidad no es el único factor que puede afectar la productividad del proceso de traducción,

incluyendo la etapa de posesición. En la siguiente sección se presentan los resultados de varios estudios al respecto.

2.2.2. Revisión de la literatura sobre la productividad en la traducción en la era digital

Ante la presión del incremento del contenido a traducir y de la necesidad de reducir costes a la que se ha enfrentado en los últimos años, la industria de la traducción se ha enfocado en el aumento de la productividad (Statismioti y Sosoni, 2020). En la era digital, la productividad en la traducción se consigue mediante el uso de la tecnología, la traducción colectiva, el reciclaje de las traducciones, la reutilización de traducciones, la competencia profesional, la búsqueda de beneficios, el ahorro de mano de obra y el ahorro de costes (Sin-way, 2015). Así mismo, cada vez más empresas incluyen la posesición en sus procesos de traducción porque, en general, esta actividad aumenta la productividad y los beneficios (Mercader-Alarcón y Martínez-Sánchez, 2016).

La productividad al usar la traducción automática ha recibido mucha atención en la literatura desde muchas perspectivas diferentes, aunque la mayoría de los estudios se han enfocado en la traducción de más contenidos en menos tiempo y con menos esfuerzo (Briva-Iglesias, O'Brien y Cowan, 2023). Varios estudios han tenido como objetivo medir el esfuerzo de posesición para evaluar la productividad de la traducción automática (Moorkens, Toral, Castilho y Way, 2018). Según Krings (2001), el esfuerzo de posesición es una combinación del esfuerzo temporal, cognitivo y técnico. No obstante, no todos los estudios usan este enfoque tridimensional en su análisis de productividad y tampoco han comparado la productividad entre los mismos tipos de traducción, tal como se puede observar en el Cuadro 1. En este sentido, se pueden identificar cuatro grupos principales de estudios: un primer grupo, que compara la productividad entre la TAO y la TA; un segundo grupo, que compara la TA con la traducción manual; un tercer grupo, que compara los tres tipos de traducción; y, por último, están los estudios que comparan la productividad entre diferentes tipos de TA, principalmente la basada en redes neuronales y la estadística.

Cuadro 1. Estudios sobre la productividad en el proceso de traducción

Autor(es)	Lenguas	Tipo de traducción	Participantes	Dimensión productividad
Guerberof (2009)	EN>ES	TAO/TA	8 traductores profesionales	esfuerzo temporal
Federico, Cattelan y Trombetti (2012)	EN>IT EN>DE	TA/TAO+TA	12 traductores profesionales	esfuerzo temporal, esfuerzo técnico
Guerberof (2012)	EN>ES	TAO/TA/Traducción manual	24 traductores y 3 revisores	esfuerzo temporal
Aranberri, Labaka, Diaz de Iarazza y Sarasola (2014)	EN>EU	TA/Traducción manual	6 traductores profesionales y 6 usuarios finales	esfuerzo temporal
Gaspari, Toral, Naskar, Groves y Way (2014)	EN>DE EN>NL	TA/Traducción manual	20 profesionales de los medios de comunicación	esfuerzo temporal
Texeira (2014)	EN>ES	TAO/TA/Traducción manual	10 traductores	esfuerzo temporal, esfuerzo cognitivo
Vieira (2014)	FR>EN	TA/TA	13 sujetos	Esfuerzo cognitivo
Parra y Arcedillo (2015)	EN>ES	TAO//TA/traducción manual	8 traductores profesionales	Esfuerzo temporal
Fernández-Torné y Matamala (2016)	EN>CA	TA/Traducción manual/creatividad	14 estudiantes nativos en catalán	esfuerzo temporal, esfuerzo técnico, esfuerzo cognitivo.
Screen (2017)	EN>CY	TAO/TA	8 traductores profesionales y 2 revisores	esfuerzo temporal
Castilho <i>et al.</i> (2018)	EN>DE EN>PT EN>RU EN>EL	TAS/TARN	3 traductores profesionales, 4 en el caso del griego (EL)	esfuerzo temporal, esfuerzo técnico
Toral <i>et al.</i> (2018)	EN>CA	TAFS/TARN/Traducción manual	6 traductores profesionales	esfuerzo temporal, esfuerzo técnico, esfuerzo cognitivo
Toledo-Báez (2018)	EN><ES	TA/TA	Estudiantes de traducción (no se especifica el número)	esfuerzo temporal
Carl y Toledo Baéz (2019)	EN>ES EN>ZH	TA/Traducción manual	48 estudiantes de traducción chinos y 32 españoles; 16 estudiantes de traducción chinos y 8 traductores profesionales españoles para la revisión	esfuerzo temporal
Läubli <i>et al.</i> (2019)	DE>FR DE>IT	TAO/TARN	4 traductores profesionales	esfuerzo temporal

			(dos por cada lengua meta); 2 expertos para la evaluación de la calidad	
Sánchez-Gijón <i>et al.</i> (2019)	EN>ES	TAO/TARN	8 profesionales de la traducción con una experiencia de entre 2 y 32 años	esfuerzo temporal, esfuerzo técnico
Álvarez <i>et al.</i> (2020)	EN>ES	TARN/TARN	4 traductores profesionales	esfuerzo temporal, esfuerzo técnico
Stasimioti y Sosoni (2020)	EN>HE	TA/Traducción manual	12 traductores profesionales	esfuerzo temporal, esfuerzo técnico, esfuerzo cognitivo
Álvarez, Oliver y Badia (2021)	EN>ES	TAFS/TARN	29 estudiantes y traductores participaron en la encuesta; 4 traductores profesionales participaron en el experimento	esfuerzo temporal, esfuerzo técnico

Hay que mencionar también el estudio de Federico *et al.* (2012), en el cual se compara la traducción automática (TA) con la traducción que integra la traducción automática con la traducción asistida (TAO+TA). Teniendo en cuenta que hoy en día la mayoría de los sistemas de traducción TAO suelen integrar aplicaciones TA, en la presente investigación se plantea la siguiente hipótesis:

H1: Existen diferencias significativas en la productividad de la posesición de la traducción automática (TA) y la productividad de la posesición de la traducción realizada con TAO+TA

Esfuerzo temporal

El esfuerzo temporal es la medida utilizada con más frecuencia en los estudios sobre productividad, aunque con diferentes conceptualizaciones. La primera se basa en la definición de Krings (2001), para quien el esfuerzo temporal representa el tiempo dedicado a la posesición. En los estudios que usan esta conceptualización, la medición consiste en el recuento de los minutos (Stasimioti y Sosoni, 2020) o segundos (Screen, 2017; Álvarez *et al.*, 2021) de la traducción o posesición del texto, la frase o los segmentos (véase Cuadro 2). Otros estudios conceptualizan el esfuerzo técnico como velocidad de traducción o de posesición. En este grupo de estudios, se mide el número de palabras poseído por hora (Aranberri *et al.*, 2014), minutos (Guerberof,

2009, 2012) o segundos (Gaspari *et al.*, 2014; Castilho *et al.*, 2018), así como la media de palabras procesadas por el traductor en una hora (Federico *et al.*, 2012; Toral *et al.*, 2018) o los segundos por cada cien palabras (Texeira, 2014).

Cuadro 2. Revisión de estudios que han analizado el esfuerzo temporal

Autores	Medidas	Resultados
Guerberof (2009)	Número de palabras por minuto	La velocidad de procesamiento es mayor para TA frente a la TAO, no obstante, hay grandes variaciones por individuos.
Federico <i>et al.</i> (2012)	La media de palabras procesadas por el traductor en una hora	La velocidad depende de la calidad de la sugerencia realizada por TAO y TA y el rendimiento del traductor. Al integrar TA en el TAO se gana en productividad.
Guerberof (2012)	Número de palabras por minuto	La velocidad de traducción es mayor en presencia de <i>fuzzy matches</i> y TA, que sin estas sugerencias. Existe un nivel similar de productividad entre TA y <i>fuzzy matches</i> entre 85 y 94%. La ganancia en productividad varía por cada traductor. La media del aumento de productividad es del 32% en el caso de <i>fuzzy matches</i> y del 37% por TA. Los traductores perciben menos esfuerzo temporal que la posesición de los textos traducidos por personas. La velocidad de los traductores y su experiencia no afecta el esfuerzo temporal.
Aranberri <i>et al.</i> (2014)	palabras traducidas por hora	Los traductores han obtenido un aumento de productividad del 17,66% al utilizar TA y los usuarios finales del 12,43%. El aumento varía en función del texto fuente. Los usuarios tienen mejor percepción sobre el uso y la productividad de la TA que los traductores. Los textos que obtienen mayor calidad en la TA alcanzan un mayor aumento de la productividad.
Gaspari <i>et al.</i> (2014)	palabras traducidas por segundo	Se obtuvo un aumento de productividad de 3,5% para EN>DE y de 13,6% para NL>EN. En la traducción DE<EN y EN>NL la productividad fue menor frente a la traducción desde cero. Las percepciones de los traductores son sesgadas y la mayoría prefieren la traducción manual.

Texeira (2014)	segundos por cada 100 palabras	En general, el esfuerzo temporal es menor al traducir con <i>fuzzy matches</i> y TA que desde cero. Las percepciones de los traductores coinciden con los resultados cuantitativos.
Parra y Arcedillo (2015)	número de palabras por hora	La productividad aumenta con TA (24,09%) y con <i>fuzzy matches</i> entre 75% y 84% (22,52%). En media, la productividad de la posesición TA es mayor que la posesición de <i>fuzzy matches</i> entre 75% y 84%.
Fernández-Torné y Matamala (2016)	duración en segundos de la tarea (creación, traducción o posesición)	El esfuerzo temporal de la posesición es menor en el caso de la creación o traducción, pero sin diferencias significativas.
Screen (2017)	segundos por frase	<i>Fuzzy matches</i> de mínimo 70% y <i>exact matches</i> , así como TA, hacen más rápido el proceso de traducción. El aumento en productividad no disminuye la calidad.
Castilho <i>et al.</i> (2018)	número de palabras por segundo	NMT mostró mejores rendimientos, con la excepción del ruso. Sin embargo, la diferencia no es significativa en ninguno de los idiomas.
Toral <i>et al.</i> (2018)	número de palabras por hora	La traducción TAFS reduce el esfuerzo temporal en un 18%, mientras que la TARN lo hace en un 36%. El tamaño del segmento fuente influye en el esfuerzo temporal.
Toledo-Báez (2018)	el tiempo empleado por cada post-editor para editar un número medio de palabras procesadas en un tiempo determinado	No hay un impacto significativo de la formación y la direccionalidad en la productividad.
Carl y Toledo Baéz (2019)	duración de la posesición	Todas las categorías de errores tienen un impacto significativo en la duración de la posesición.
Läubli <i>et al.</i> (2019)	número de palabras por hora	El uso de TARN aumentó en 59,74% la productividad en DE>FR y en 9,26% en DE>IT y conduce a niveles de calidad similares o ligeramente mejores.
Sánchez-Gijón <i>et al.</i> (2019)	Tiempo empleado por cada participante en la posesición de cada segmento	Los segmentos TARN muestran una mayor variabilidad en el tiempo necesario para la posesición que los segmentos TAO. La posesición TAO requiere más tiempo en comparación con los segmentos TA. La percepción de la productividad de los poseedores coinciden con sus métricas de rendimiento.

Álvarez <i>et al.</i> (2020)	tiempo de posesición en segundos normalizado por el número de tokens del segmento	El tiempo es una buena medida del esfuerzo, pero no siempre corresponde con las otras métricas de esfuerzo y varía mucho entre los traductores. HTER y HBLUE no están correlacionados con el esfuerzo temporal.
Stasimioti y Sosoni (2020)	duración en minutos de la posesición y de la traducción manual	En media, se ahorra un 19,8% de tiempo en la posesición de segmentos TARN. La experiencia en posesición afecta el tiempo necesario para la posesición.
Álvarez <i>et al.</i> (2021)	segundos por segmento	La media del esfuerzo temporal de la TARN es menor frente a TAFS, pero sin diferencias significativas.

Además de las diferencias en los enfoques de conceptualización y medición del esfuerzo temporal, no todos los estudios analizan las diferencias de productividad entre los mismos tipos de traducción. Varios estudios se han enfocado en comparar el esfuerzo temporal entre la posesición de resultados de traducción de la TAO y los resultados obtenidos con la TA. En concreto, Läubli *et al.* (2019) hallaron un aumento del 59,74% para la TA del alemán al francés y del 9,26% para la TA del alemán al italiano. Otros estudios también confirman que, en términos de tiempo, se obtiene en media mayor productividad con la TA en comparación con la TAO (Guerberof, 2009; Federico *et al.*, 2012; Sánchez-Gijón *et al.*, 2019). Así mismo, Screen (2017) mostró que los niveles de productividad son similares entre las TA y los *exact matches* o los *fuzzy matches* de mínimo 70%. No obstante, hay ciertos factores que influyen en estas diferencias, como la experiencia del traductor (Guerberof, 2009), rendimiento del traductor y la calidad de las sugerencias (Federico *et al.*, 2012).

Además, los traductores suelen ahorrar tiempo con la TA también en comparación con la traducción desde cero, según los resultados de los estudios llevados a cabo por Guerberof (2012), Aranberri *et al.* (2014), Gaspari *et al.* (2014), Teixeira (2014), Toral *et al.* (2018) y Stasimioti y Sosoni (2020). Es más, los resultados del estudio Guerberof (2012) mostraron que el aumento de productividad es del 37% en la traducción del inglés al español. Toral *et al.* (2018) encontraron un aumento de productividad similar, del 36%, para la TARN, pero la mejora varía en función del modelo TA utilizando, porque en el caso de la TAFS el incremento es menor, del 23%. Así mismo,

Stamsiotti y Sonsoni (2020) encontraron que se ahorra en media un 19,8% de tiempo con la TA en comparación con la traducción manual del inglés al griego, aunque este ahorro es afectado por la experiencia en posesión del traductor. No solo la experiencia en posesión es importante, sino también su experiencia como traductor. Aranberrí *et al.* (2014) determinaron un aumento de productividad al traducir del inglés al euskera del 17,66% para los traductores profesionales y del 12,43% para los usuarios finales; no obstante, el aumento de productividad varía en función de la calidad del resultado de la TA y la familiaridad del traductor con el área de especialidad del texto. Por su parte, Fernández-Torné y Matamala (2016) encontraron que, aunque el tiempo de posesión con la TA es menor, las diferencias con la traducción desde cero no son significativas; mientras que Gaspari *et al.* (2014) hallaron que los resultados dependen de los pares de idiomas analizados. Los resultados de la investigación realizada por estos autores indican un aumento de productividad del 3,5% para la TA del inglés al alemán y del 13,6% para la TA del holandés al inglés; pero en el caso de la TA alemán-inglés e inglés-holandés, la posesión tomó más tiempo que la traducción desde cero.

Por supuesto, no solo la TA puede contribuir a ahorrar el tiempo de los traductores, sino también la TAO. Concretamente, Guerberof (2012) que los aumentos de productividad de la TA y los *fuzzy matches* entre el 85% y el 94% son similares al comparar con la traducción desde cero del español al inglés, con un aumento de productividad del 32% en el caso de la TAO. Los resultados de este estudio son confirmados por Parra y Arcedillo (2015), quienes encontraron niveles de productividad similares, del 24,09% para la TA y del 22,52% para la TAO, tomando en cuenta el mismo par de idiomas que en el estudio de Guerberof (2012), pero con *fuzzy matches* de entre el 75% y el 84%. Por su parte, Texeira (2014) también corrobora que la TAO alcanza mayores niveles de productividad que la traducción desde cero.

Otro grupo de estudios comparan el esfuerzo temporal de dos tipos de traducción automática: la estadística (TAFS) y la basada en redes neuronales (TARN). No obstante, aunque todos los estudios encontraron que con la TARN se consigue una mayor reducción del esfuerzo temporal, los resultados son diferentes. En el caso del estudio de Torral *et al.* (2018), se consigue casi el

doble de productividad con la TARN en comparación con la TAFS, aunque la reducción del esfuerzo temporal varia en función del tamaño del segmento fuente. En cambio, aunque en los estudios Castilho *et al.* (2018) y Álvarez *et al.* (2021) se encontró ligeras reducciones del esfuerzo temporal con la TARN, las diferencias no son significativas.

Por último, hay que mencionar algunos los estudios que se enfocaron en estudiar algunos factores que pueden afectar el esfuerzo temporal en el caso de la TA. Por ejemplo, Toledo-Báez *et al.* (2018) mostraron que la formación del traductor y la direccionalidad (traducir al idioma nativo y viceversa) no afecta la productividad. Por su parte, Carl y Toledo Báez (2019) hallaron que todos los tipos de errores están correlacionados con el esfuerzo temporal; mientras que Álvarez *et al.* (2020) destacaron que el esfuerzo temporal varia mucho entre los traductores.

Al analizar los resultados de estos estudios con los mencionados previamente, se puede concluir que hay muchos factores que influyen en el esfuerzo temporal que varían desde factores individuales relacionados con el traductor (formación y experiencia), los pares de idiomas, el motor de traducción automática y la calidad de los resultados de la TA. No obstante, de todos los factores analizados en relación con el esfuerzo temporal, el único abordado desde una perspectiva doble es la calidad. Por una parte, se analiza el efecto de la calidad de las sugerencias TA o de la TAO en el esfuerzo temporal, con resultados que no so concluyentes. Por ejemplo, Federico *et al.* (2012) y Aranberri *et al.* (2014) encontraron indicios que la velocidad de la posesición de los segmentos TA o TAO dependen de la calidad; mientras que, Álvarez *et al.* (2020) no encontraron una correlación significativa. Por otra parte, Screen (2017) analizó el efecto de la reducción del esfuerzo temporal en la calidad del texto final traducido, los resultados de sus estudios indicando que no existe un efecto negativo del aumento de la productividad en la calidad. Ha que mencionar que estas investigaciones utilizan medidas diferentes de calidad. BLUE es la única medida que se repite en los estudios llevados a cabo por Aranberri *et al.* (2014) y Álvarez *et al.* (2020). Otras medidas automáticas utilizadas son TER, perplejidad y "fuera de vocabulario" (Aranberri *et al.*, 2014), NIST, RIBES y WER (Álvarez *et al.*, 2020). Por su parte, Screen (2017) utilizó el marco LISA para evaluar la calidad.

Hay que destacar también los estudios que también evaluaron las percepciones de los traductores sobre el esfuerzo temporal. Los primeros estudios muestran que los traductores tienen una percepción negativa sobre la TA y la posesición, además consideran que la posesición requiere más tiempo y que los resultados de la TA son de peor calidad que la traducción manual (Aranberri *et al.*, 2014; Gaspari *et al.*, 2014). Sin embargo, estas percepciones no se corresponden con la realidad, ya que los resultados cuantitativos, tanto para la calidad como del esfuerzo temporal de la posesición, muestran mejores resultados (Gaspari *et al.*, 2014). En cambio, los resultados de otros estudios muestran que los resultados de las métricas automáticas (p. ej. BLUE o el tiempo de la posesición) coinciden con las percepciones de los traductores (Texeira, 2014; Sánchez-Gijón *et al.*, 2019).

Esfuerzo técnico

El esfuerzo técnico se refiere a las “operaciones puramente técnicas” (Krings, 2001, pág. 179). Por tanto, normalmente, su medición consiste en el conteo del número de pulsaciones en el teclado (Castilho *et al.*, 2018; Toral *et al.*, 2018; Stamsioti y Sosoni, 2020; Álvarez *et al.*, 2021). Algunos autores han usado, sin embargo, otras medidas, como el porcentaje de palabras cambiadas (Federico *et al.*, 2012), acciones del ratón, cambios del teclado al ratón, el total de transiciones de ventana (Fernández-Torné y Matamala, 2016), HTER, HBLEU (Álvarez *et al.*, 2020) y distancia de edición (Sánchez-Gijón *et al.*, 2019; Álvarez *et al.*, 2020). El Cuadro 3 presenta en más detalles estas medidas y los resultados de los estudios.

Cuadro 3. Revisión de estudios que han analizado el esfuerzo técnico

Autores	Medidas	Resultados
Federico <i>et al.</i> (2012)	porcentaje de palabras cambiadas por el traductor en las sugerencias TAO	El esfuerzo disminuye significativamente cuando las coincidencias TAO son complementadas con coincidencias TA.

Fernández-Torné y Matamala (2016)	acciones del teclado (incluido el total de tipos de caracteres y otras pulsaciones), acciones del ratón (incluidos clics, movimientos y desplazamientos), cambios del teclado al ratón, y el total de transiciones de ventana.	El esfuerzo técnico es menor con la posesición, pero no hay diferencias significativas con la traducción desde cero o la traducción creativa.
Castilho <i>et al.</i> (2018)	Pulsaciones por segmento	Aunque es esfuerzo técnico de posesición es menor con la TARN, las diferencias con la TAFS no son significativas
Toral <i>et al.</i> (2018)	número de pulsaciones de teclado para realizar la traducción final	Hay un reducción del esfuerzo técnico del 23% en el caso de TARN y del 9% para la TAFS. El tamaño del texto fuente influye en el esfuerzo técnico. Existen diferencias en el tipo de teclas que se pulsan entre la TA y la traducción desde cero.
Sánchez-Gijón <i>et al.</i> (2019)	distancia de edición entre los segmentos TAO y los segmentos TA	La distancia de edición de los segmentos TA es significativamente menor en comparación con los segmentos TAO.
Álvarez <i>et al.</i> (2020)	pulsaciones, HBLEU, HTER y distancia de edición	El sistema TARN transformador requiere menos esfuerzo técnico que el seq2seq. Existe una alta correlación entre las pulsaciones y el esfuerzo temporal.
Stasimioti y Sosoni (2020)	número de pulsaciones de teclado (inserciones y supresiones)	El esfuerzo técnico es significativamente mayor en la traducción manual que en la posesición. En la traducción es mucho más alto el número de inserciones; mientras que en la posesición, hay un mayor número de supresiones.
Álvarez <i>et al.</i> (2021)	número de pulsaciones por segmento	Existe una diferencia considerable en el esfuerzo técnico entre la TARN y la TAFS, siendo menos en el primero tipo de traducción.

Los resultados de los estudios del primer grupo, en el cual se compara la productividad de la TA con la TAO, muestran que el esfuerzo técnico

disminuye significativamente con la TA (Federico *et al.*, 2012; Sánchez-Gijón *et al.*, 2019). En el estudio de Federico *et al.* (2012) se registró un descenso del esfuerzo técnico de más del 50% tanto para la traducción del inglés al alemán como del inglés al italiano cuando las sugerencias de la TAO eran complementadas con las sugerencias de la TA. Por su parte, Sánchez-Gijón *et al.* (2019) encontraron que las diferencias no son significativas para los segmentos entre 10 y 19 palabras, aunque el esfuerzo técnico para la posesición de la TA si es significativamente menor que la posesición de los segmentos que representan *fuzzy matches* de entre 80% y 90% de la TAO.

Otro grupo de estudios es el que analiza las diferencias entre el esfuerzo técnico de la TA con el requerido por la traducción desde cero. En este caso también se encontró que con la TA el esfuerzo técnico es menor (Fernández-Torné y Matamala, 2016; Stasimioti y Sosoni, 2020). No obstante, en el caso del estudio llevado a cabo por Fernández-Torné y Matamala (2016) las diferencias no son significativas; mientras que, Stasimioti y Sosoni (2020) hallaron que el esfuerzo técnico de la traducción manual es significativamente mayor que la posesición. En este último estudio también se encontró que en el caso de la traducción manual el número de inserciones es significativamente más alto; en cambio, la posesición requiere un número más alto de supresiones. Así mismo, los resultados indican que el esfuerzo técnico es menor para los traductores con experiencia en posesición.

Por último, hay que mencionar el grupo de estudios que compara el esfuerzo técnico entre los distintos tipos de traducción automática, principalmente entre la traducción estadística (TAFS) y la basada en redes neuronales (TARN). En este caso, los resultados de las investigaciones son similares con las del anterior grupo. Es decir, en ambos estudios, que conforman el grupo, se encontró que el esfuerzo técnico es menor en el caso de la TARN, pero en el caso de la investigación realizada por Castilho *et al.* (2018) las diferencias con TAFS no son significativas; mientras que, en el estudio llevado a cabo por Álvarez *et al.* (2021) la ganancia en productividad si es considerable. No obstante, no todos los sistemas basados en TARN obtienen los mismo resultados, según Álvarez *et al.* (2020), quienes encontraron un que un sistema TARN transformador requiere menos esfuerzo técnico que un sistema

seq2seq. Además, en este mismo estudio se encontró una correlación muy alta entre el esfuerzo técnico y el esfuerzo temporal.

Esfuerzo cognitivo

El esfuerzo cognitivo se define como “el alcance y tipo de procesos cognitivos que deben activarse para subsanar una deficiencia en el resultado de la TA” (Krings, 2001, pág. 179). Como se puede observar, la definición de Krings (2001) se refiere exclusivamente a la traducción automática, pero el concepto se aplica incluso a la posesión de los segmentos provenientes de la memoria de traducción (Guerberof, 2009). A diferencia del esfuerzo temporal y el esfuerzo técnico, esta dimensión de la productividad no se puede medir a través de la observación directa y, por tanto, hay que utilizar un *proxy* (Ortiz-Boix y Matamal, 2016).

En general, los *proxys* usados en los estudios se han obtenido a través de protocolos de seguimiento ocular o de pensamiento en voz alta (Álvarez *et al.*, 2020). No obstante, en los estudios más recientes se utilizan con más frecuencia los protocolos de seguimiento ocular (*eye-tracking*), tal como indican las medidas utilizadas en los estudios presentados en el Cuadro 4. En el caso de estos protocolos, un equipo no intrusivo registra los movimientos y fijaciones oculares (O’Brien, 2011). Por tanto, los datos colectados con la ayuda de esta herramienta incluyen el recuento de fijaciones, la duración media de la fijación (Vieira, 2014; Stamsimioti y Sosoni, 2020) y el número total de miradas (Stamsimioti y Sosoni). Así mismo, en las investigaciones más recientes, se utilizan varios indicadores calculados a partir de las pausas, como el tiempo total de la pausa, la media del tiempo de las pausas, número total de pausas, la ratio del tiempo total de las pausas entre el tiempo total de traducción y el ratio del las pausas entre el número de palabras (Ortiz-Boix y Matamal, 2016; Toral *et al.*, 2014). Hay que mencionar también el estudio de Texeira (2014) que emplea como medidas del esfuerzo cognitivo: el número de pulsaciones en determinadas teclas y el número total de caracteres del texto (incluidos los espacios). Así mismo, destaca el estudio de Vieira (2014), en el cual, además del seguimiento ocular, se aplica una encuesta con preguntas con respuestas de escala Likert de 9 puntos para medir el esfuerzo cognitivo percibido.

Cuadro 4. Revisión de estudios que han analizado el esfuerzo cognitivo

Autores	Medidas	Resultados
Texeira (2014)	el número de pulsaciones en las teclas pertinentes y el número total de caracteres del texto meta final (incluidos los espacios)	Hay un mayor número de ediciones en el caso de la traducción desde cero que en la TA.
Vieira (2014)	duración media de fijación, recuento de fijaciones, esfuerzo cognitivo percibido	Existe una correlación significativa entre METEOR y el esfuerzo cognitivo. Los efectos de las características lingüísticas del texto fuente son menores, con un efecto mayor en la percepción del esfuerzo entre los poseedores con nivel bajo de competencias en el idioma fuente. Se observó una relación significativa entre la capacidad de memoria de trabajo de los sujetos y la productividad.
Fernández-Torné y Matamala (2016)	tiempo total de pausa, tiempo medio de pausa, número de pausas, ratio de las pausas entre el número de palabras	Existe un menor esfuerzo cognitivo en el caso de la posesición, pero la diferencia con la traducción manual y la creativa no es significativa.
Toral <i>et al.</i> (2018)	número de pausas, media de la duración de la pausa, ratio del tiempo total dedicado a las pausas dividido por el tiempo total de la traducción	Hay un aumento de alrededor del 65% del esfuerzo cognitivo en la posesición en comparación con la traducción desde cero.
Stasimioti y Sosoni (2020)	recuento de fijaciones, duración de las fijaciones en segundos, número total de miradas	El recuento de fijaciones y la duración de las fijaciones y el número total de miradas son mayores en el caso de la traducción manual en comparación con la posesición. El esfuerzo cognitivo es menor para los traductores con experiencia en posesición. En el caso de la posesición la mayor atención se enfocó en el área de texto meta; mientras que en el caso de la traducción manual en el área del texto fuente.

La mayoría de los estudios que analizan el esfuerzo cognitivo se enfocan en la comparación entre la TA y la traducción manual. A diferencia de los resultados del análisis del esfuerzo temporal y el esfuerzo técnico, no existe un consenso en los resultados. Stasimioti y Sosoni (2020) encontraron que el número de fijaciones y su duración, así como el número de miradas, son

menores en el caso de la posesición; es decir, la TA requiere un esfuerzo cognitivo menor que la traducción desde cero. Así mismo, los resultados muestran que los traductores no se fijan en las mismas áreas, registrándose un mayor número y duración de las fijaciones en el área del texto fuente en el caso de la traducción desde cero y en el área del texto meta para la TA. Por su parte, Fernández-Torné y Matamala (2016) hallaron también una disminución del esfuerzo cognitivo en el caso de la posesición, pero la diferencia con la traducción manual no es significativa. En cambio, Toral *et al.* (2018) encontraron el resultado opuesto; es decir, que el esfuerzo cognitivo en la posesición es mayor que en la traducción desde cero, con un incremento de alrededor del 65%.

Hay que destacar el estudio de Texeira (2004) en el que se analizó si la integración de la TA en la TAO disminuye el esfuerzo cognitivo, los resultados corroborando su planteamiento, al necesitar la traducción con sugerencias del sistema TA y de la memoria de traducción menos ediciones. No obstante, si se comparan las medidas utilizadas con las empleadas para medir el esfuerzo técnico (véase el Cuadro 3), se observa una gran semejanza. Los mismos autores concluyen que las medidas utilizadas no son las adecuadas para medir el esfuerzo cognitivo.

Existen también estudios como el de Vieira (2014), en el cual el enfoque se pone en la identificación de los factores que afectan el esfuerzo cognitivo en la posesición de la TA del francés al inglés. Los resultados muestran que la calidad del texto (medida con el índice METEOR) tiene un efecto negativo sobre el esfuerzo cognitivo, mientras que la longitud de las frases tiene un efecto positivo. Hay que mencionar que se encontró también que el esfuerzo cognitivo percibido por los participantes al experimento varía en función del nivel de francés. Otra característica individual que afecta la productividad es la capacidad de memoria de los participantes, resultados en el que no se observó efectos significativos de la edad de los traductores. Respecto a los factores que afectan el esfuerzo cognitivo, hay que añadir que Stasimioti y Sosoni (2020) hallaron que el esfuerzo cognitivo es menor para los traductores con experiencia en posesición.

Considerando los resultados de los estudios revisados anteriormente se plantean las siguientes hipótesis:

H2: La productividad de la TA varía en función de la existencia de errores de traducción de la terminología del área estadística.

H3. La productividad de la TAO integrada con TA varia en función de la existencia de errores de traducción de la terminología del área estadística.

3. Metodología

La población del presente estudio consiste en los textos de estadística en inglés (Estado Unidos), como lengua fuente, y en español como lengua meta. Para el análisis, se seleccionaron cuatro textos que corresponden a una actualización de los casos de estudio de los capítulos del 2 a 5 del libro *A Primer on Partial Least Squares Structural Equation Modelling*, de Hair et al. (2022). El libro explica la aplicación de la técnica estadística de mínimos cuadrados parciales (PLS-SEM) con la ayuda del software SmartPLS.

La actualización de los casos de estudio es el resultado de las prácticas del máster de Traducción Especializada de la Universidad de Vic realizadas por la autora de la presente investigación en Fördeverein Industrielles Managment e.V., una fundación perteneciente a la Universidad Técnica de Hamburgo, Alemania, que participa en el desarrollo del software SmartPLS. Los casos de estudio han sido publicados online¹ por los autores y pueden ser descargados de manera gratuita. Los cuatro casos de estudio tienen un total de 12.337 palabras. No obstante, se precedió a borrar del texto las indicaciones para el uso del software, quedando solo el texto que explica el método y los resultados estadísticos, que suma en total 4.431 palabras. Por tanto, la muestra final para la presente investigación es de 191 segmentos, tal como se muestra en el Cuadro 5.

¹ <https://www.pls-sem.net/>

Cuadro 5. Descripción de la muestra de textos y segmentos seleccionados

Nº Texto	Nº de palabras	Nº de segmentos
Texto 1	1285	59
Texto 2	316	15
Texto 3	952	40
Texto 4	1878	77
Total	4.431	191

La variable objetivo del análisis es la productividad del proceso de posesición utilizando dos tipos de procedimiento: uno en el que solo se posedita la traducción del segmento resultante de la traducción automática y otro en el que además de la traducción automática se hizo uso de herramientas TAO, como una base terminológica con 82 términos relacionados con la técnica estadística PLS-SEM y la memoria de traducción resultante de la traducción de los casos. Para todo el proceso de posesición se utilizó Trados Studio 2022 como herramienta TAO, así como la tecnología de traducción automática neuronal Language Weaver, integrada en la aplicación. Se selecciono Trados Studio para la presente investigación porque es la herramienta TAO más utilizada por las empresas europeas, según el estudio realizado por EUATC (2022).

En primer lugar, se creó un proyecto para la traducción automática de los cuatro casos. A continuación, se procedió a la posesición de los textos traducidos con Language Weaver, cronometrando el proceso por cada segmento. El tiempo obtenido en segundos se utilizó como medida de productividad de la traducción automática.

A continuación, se creó un proyecto independiente para cada texto. En el primer paso, se creó el proyecto de traducción en Trados Studio para el Texto 1, en el cual se adjunto la base terminológica y se creo una memoria de traducción. Se tradujo automáticamente el proyecto con la ayuda del Language Weaver y se poseditaron los segmentos cronometrándose la duración del proceso (en segundos) para medir la variable Prod_TA_TAO. Al finalizar, este proceso, se procedió a la alineación de los segmentos traducidos, los cuales se guardaron en la memoria de traducción creada. En el segundo paso, se creo un nuevo proyecto para el Texto 2, al cual se cargo la misma base terminológica y la misma memoria de traducción del proyecto anterior. Después, se procedió a traducir utilizando el motor de traducción

automática y las concordancias de la memoria, para finalizar con la posesición, la actualización de la memoria y la medición de la variable Prod_TA_TAO. El mismo proceso se repitió para los dos textos restantes. Los datos obtenidos de las mediciones de Prod_TA y Prod_TA_TAO por cada segmento se usaron para calcular la diferencia de productividad, así como su equivalente en porcentajes.

Cuadro 6. Variables analizadas

Variable		Medida	Autor(es)
Productividad	Productividad traducción automática (Prod_TA) Productividad traducción automática + TAO (Prod_TA_TAO) Diferencia entre Prod_TA_TAO y Prod_TA	Duración en segundos de la posesición del segmento TA Duración en segundos de la posesición del segmento TA+TAO Diferencia entre duración de la posesición del segmento TA+TAO y la duración del segmento TA ($DIFF = Prod_TA_TAO - Prod_TA$) La diferencia entre las dos duraciones medida en porcentaje (DIFF %)	Gaspari <i>et al.</i> (2014), Screen (2017) y Carl y Toledo Baéz (2019); Sánchez-Gijón <i>et al.</i> (2019)
calidad (BLEU_TA y BLEU_TM)		BLEU	Aranberri <i>et al.</i> (2014), Gaspari <i>et al.</i> (2014), Para y Arcedillo (2015), Castilho <i>et al.</i> (2018), Álvarez <i>et al.</i> (2020)
Calidad global de la traducción automática (OQS_TA)		Puntuación global de calidad obtenida aplicando el método MQM	Lommel (2018), MQM (2023)
Longitud del segmento		Número de caracteres (con espacios incluidos)	Federico <i>et al.</i> (2012), Vieira (2014), Álvarez <i>et al.</i> (2020) y Álvarez <i>et al.</i> (2021)
Errores estadística (ET)		Variable ficticia binaria, la cual recibe el valor 1 cuando se encontró un error de traducción relacionado con el ámbito de la estadística y 0 en el caso contrario.	

Bilingual Evaluation Understudy (BLEU) se utilizó como medida de calidad porque es la métrica más popular para evaluar los resultados de la traducción automática (Rivera-Trigueros, 2022). BLEU es una métrica de precisión que utiliza las unidades lingüísticas indivisibles (*n*-grams), calculando un índice en función del número máximo de apariciones de los *n*-grams en el texto de referencia (Gutiérrez-Artacho *et al.*, 2018). Además, tal como se puede observar en el Cuadro 6, es utilizada con frecuencia en los estudios que

analizan la relación entre la calidad y la productividad de la traducción automática.

Para calcular el BLEU, se utilizó la herramienta ofrecida por la empresa TILDE². Como textos de referencia se utilizó las traducciones de los casos realizadas durante las mismas prácticas por la autora. Para la revisión de las traducciones se consultó el libro *Manual de Partial Least Squares Structural Equation Modeling (PLS-SEM)*, que es la traducción al español de la segunda edición del libro de los textos fuente, que fue realizada por los profesores Castillo Apraiz, Cepeda Carrión y Roldán (2019) y que tiene una alta similitud con la versión actual de los textos.

No obstante, una medida automática como BLUE es con frecuencia insuficiente para evaluar la traducción automática y, por esta razón, se recomienda realizar también una evaluación humana para una mejor interpretación de los resultados (Gutiérrez-Artacho *et al.*, 2018). El uso de un texto de referencia para el cálculo del BLUE es la principal limitación de esta medida porque difícilmente va a existir una traducción única correcta entre tantas traducciones posibles para un texto fuente (Chunyu y Tak-ming, 2015).

El método considerado más fiable para la evaluación humana de la calidad de la traducción es el basado en una calificación de los errores, porque la identificación de los errores es, en general, más objetiva y coherente entre los evaluadores que la valoración de la calidad de la traducción (Chunyu y Tak-ming, 2015). En la presente investigación, se completó la evaluación de la calidad con el análisis de los errores de traducción con la ayuda de una clasificación de los errores adaptada de MQM (*multidimensional quality metrics*). Este método de evaluación fue creado por DFKI y TAUS en el marco del proyecto QT21 financiado por la Unión Europea, el cual tenía como objetivo integrar los dos modelos desarrollados por las dos entidades y, por tanto, estandarizar las tipologías de los errores utilizadas en la evaluación cualitativa de la calidad de la traducción (Lommel, 2018).

² El cálculo de BLEU se realizó en la aplicación ofrecida por la empresa TILDE en la siguiente dirección web <https://www.letsmt.eu/Bleu.aspx>

El marco MQM es utilizado con frecuencia en los estudios de traducción porque permite crear métricas personalizadas basadas en las necesidades de los investigadores y sus categorías de error se pueden usar tanto para evaluar el texto en su conjunto o frase por frase (Mariana, Cox y Melbry, 2015). En el presente estudio se aplicó el marco con un enfoque analítico, es decir, se realizó una clasificación de los errores para cada segmento resultado de la traducción automática en parte. Para el análisis, se seleccionaron las siguientes categorías de errores: *terminology* (terminología), *accuracy* (precisión), *linguistic convention* (convenciones lingüísticas), *style* (estilo) y *locale convention* (convención local). Dentro de las primeras dos categorías se incluyeron subcategorías relacionados con la traducción de la terminología del área de estadística. En el siguiente cuadro se detallan y definen las categorías utilizadas para el análisis en base a las definiciones ofrecidas por el MQM Committee (2023).

Cuadro 7. *Tipologías de los errores MQM utilizados en el análisis*

Categoría	Definición
Terminología	Los errores que se producen cuando un término no se ajusta a las normas terminológicas del ámbito o cuando un término del texto meta no es el equivalente normativo correcto del término correspondiente del texto fuente.
Término erróneo estadística	
Término erróneo otro	
Precisión	Errores que ocurren cuando el texto meta no se corresponde con precisión con el contenido proposicional del texto fuente, debido a la distorsión, omisión o adición al mensaje.
Traducción incorrecta estadística	
Traducción incorrecta estadística	
Omisión estadística	
Omisión otro	
Sin traducir estadística	
Sin traducir otro	Errores relacionados con la calidad lingüística del texto, incluidos problemas de gramática, ortografía y puntuación.
Convenciones lingüísticas	
<i>Gramática</i>	
<i>Genero</i>	
<i>Puntuación</i>	
<i>Ortografía</i>	Errores en un texto que no se derivan de las faltas gramaticales sino de la desviación de las guías de estilo de la organización o muestran un estilo lingüístico inapropiado. En el presente estudio se evaluó el uso de un registro formal.
Estilo	
Registro	Errores que ocurren cuando la traducción incumple los requisitos del contenido o formatos específicos de la lengua local para los elementos datos. En la presente investigación se evaluó si el resultado de la traducción automática cumple con el formato de número utilizado en España.
Convención local	
Formato de número	

A cada tipo de error se le asigna un nivel de gravedad. El marco MQM establece cuatro niveles: *neutro* (existe una solución diferente, pero no se debe asignar una penalización), *menor* (el error tiene un impacto limitado en la calidad del contenido), *mayor* (el error afecta significativamente la comprensión, lectura o uso del texto) y *crítico* (el contenido no es adecuado para el objetivo o representa un riesgo físico, financiero o de reputación). A cada nivel de error se le asigna un valor: 0, 5, 10 y 25, respectivamente. Estos valores son utilizados posteriormente para el cálculo de varias métricas: PWPT (*Per-Word Penalty Total*), ONPT (*Overall Normed Penalty Total*) y OQS (*Overall Quality Score*). En el presente estudio, se realizó un análisis descriptivo de las categorías de los errores y su nivel de gravedad. Además, la métrica OQS se empleó en el análisis descriptivo, pero también en el análisis correlacional. En el Cuadro 8 del Anexo 1 se presenta la ficha del marco MQM con las categorías de errores utilizadas, los niveles de gravedad y las fórmulas de cálculo de las métricas, llenada con los datos de uno de los segmentos utilizados en el análisis.

Otra variable utilizada con frecuencia en los estudios en relación con la productividad es la longitud de los segmentos, al estar vinculada con las dos dimensiones de la productividad: esfuerzo temporal y esfuerzo cognitivo. Por tanto, también se incluyó en el presente estudio. Esta variable se midió calculando la totalidad de los caracteres, incluidos los espacios en blanco.

Para el análisis de las variables se aplicaron varios estadísticos. En primer lugar, se realizó un análisis descriptivo de todas las variables del estudio con la ayuda de estadísticos como la media, la mediana, la desviación estándar, el mínimo y el máximo. El objetivo es obtener una visión general de las variables, así como identificar diferencias en función de la métrica BLEU o del texto fuente. Además, se aplicó la prueba Mann-Whitney para determinar si existen diferencias en la productividad entre los dos tipos analizados en la presente investigación: la posesión de la traducción automática (TA) y la posesión de la traducción que integra TA y herramientas TAO (bases terminológicas y memoria de traducción). La prueba se aplicó para comparar la productividad de la muestra completa de segmentos, así como los segmentos con errores vinculados al ámbito de la estadística y los segmentos sin este tipo de errores.

Se hizo también un análisis de correlación para identificar las relaciones entre la longitud del segmento y la productividad, así como la relación entre la calidad y la productividad, y entre la longitud del segmento fuente y la calidad. Por último, se realizó un análisis descriptivo y cualitativo de los *fuzzy matches* obtenidos durante el proceso de traducción, así como una comparación de la calidad y productividad con los resultados de la traducción automática.

4. Resultados

Se inicia el análisis presentando los resultados de la evaluación de los estadísticos descriptivos de las variables tamaño del segmento, calidad de la traducción automática (BLEU_TA), calidad multidimensional de la traducción automática (OQS_TA), productividad de la posesición con traducción automática (Prod_TA) y productividad de la con traducción asistida y traducción automática integrada (Prod_TA_TAO). La Tabla 1 presenta los resultados del análisis general de la muestra de 191 segmentos.

Tabla 1. Estadísticos descriptivos

Variable	Media	Mediana	Desviación estándar	Mínimo	Máximo
Tamaño segmento	150,46	139,00	68,66	36	424
BLEU_TA	58,86	60,66	19,56	8,24	100,00
OQS_TA	73,64	90,00	45,10	-200	100
Prod_TA	40,43	37,00	22,65	5,00	120,00
Prod_TA_TAO	35,57	28,00	18,53	6,00	123,00
DIF	-7,84	-5,00	16,95	-95,00	47,00
DIF%	-6,06	-17,39	62,99	-79,17	522,22

Como se puede observar, la media de los segmentos es de alrededor de 150 caracteres (con espacios incluidos), con un mínimo de 36 caracteres y un máximo de 424. El valor de la medida de calidad BLEU supera el límite mínimo de 45 puntos constatados como el nivel a partir del cual se produce un aumento de productividad en las tareas de posesición realizadas por los traductores profesionales (Parra y Arcedillo, 2015). La media de la variable Prod_TA en función del nivel del BLEU (Tabla 2) muestra que, en general, este es también el caso del presente estudio, con la excepción de los niveles entre 67-74, en los cuales, se dedicó un tiempo ligeramente superior a la traducción que en los segmentos con puntuaciones inferiores a 45 (45,26

segundos vs. 44,12 segundos). No obstante, habría que realizar un análisis detallado de los segmentos con menor productividad para identificar las causas.

Tabla 2. Estadísticos descriptivos de Prod_TA en función del nivel de BLEU

BLEU	N	%	Media	Mediana	Desv. típ.	Mínimo	Máximo
Menos de 45	45	23,6	44,12	41,00	21,38	15,00	101,00
45-54	25	13,1	43,96	35,00	27,05	9,00	120,00
55-64	45	23,6	39,96	37,00	18,62	12,00	86,00
65-74	39	20,4	45,26	44,00	23,56	9,00	108,00
75-84	20	10,5	32,20	26,00	20,59	6,00	74,00
85-94	8	4,2	39,00	35,50	26,72	10,00	93,00
95-100	1	0,5	21,00	21,00		21,00	21,00
100	8	4,2	12,25	9,50	8,55	5,00	30,00
Total	191	100%					

Los estadísticos descriptivos de la Tabla 1 indican también que el nivel de calidad basado en los errores (OQS_TA), con una media de 73,64. Por tanto, en el caso de un alto número de segmentos se registra un número reducido de errores. Concretamente, el 39,79% de los segmentos han obtenido una puntuación superior al 95%, es decir, sin errores o con pocos errores menores, seguido por un 16,75%, con una puntuación entre el 85 y 95, lo que también indica una baja frecuencia de los errores.

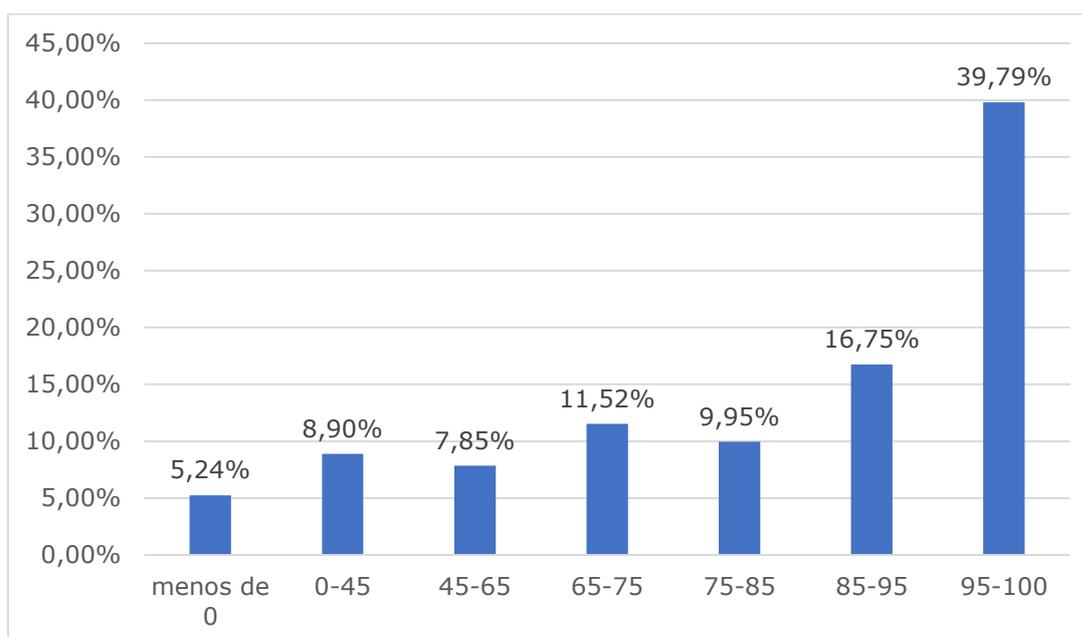


Figura 1. Frecuencia de los errores

En total, se identificaron 416 errores, de los cuales, una buena parte están relacionadas con la precisión (40,63%), la terminología (30,05%) y las convenciones lingüísticas (22,36%). Como se puede ver en la Tabla 3, la mayor parte de los errores son neutrales (108) o menores (195), es decir, sin un impacto en el uso y la comprensión del texto.

Tabla 3. *Tipos y niveles de los errores*

Tipo de error	Nivel de gravedad				Total	%
	Neutro	Menor	Mayor	Crítico		
Terminología	24	83	7	11	125	30,05%
Terminología estadística	19	83	7	11	120	28,85%
Terminología otro	5	0	0	0	5	1,20%
Precisión	31	46	83	9	169	40,63%
Traducción incorrecta estadística	6	13	47	4	70	16,83%
Traducción incorrecta otro	17	31	24	0	72	17,31%
Omisión estadística	1	0	2	0	3	0,72%
Omisión otro	1	1	0	0	2	0,48%
Sin traducir estadística	6	1	10	4	21	5,05%
Sin traducir otro	0	0	0	1	1	0,24%
Convenciones lingüísticas	52	38	2	1	93	22,36%
Gramática	35	36	1	1	73	17,55%
Género	0	0	1	0	1	0,24%
Puntuación	9	2	0	0	11	2,64%
Ortografía	8	0	0	0	8	1,92%
Estilo	0	2	0	0	2	0,48%
Registro	0	2	0	0	2	0,48%
Convención local	1	26	0	0	27	6,49%
Formato de número	1	26	0	0	27	6,49%
Total	108	195	92	21	416	100%

Además, se observa que los errores relacionados con la traducción de los términos estadísticos, sea debido al uso de un término erróneo (28,85%), la traducción incorrecta de los términos y expresiones de la estadística (16,83%), la cantidad de términos sin traducir (5,05%) o la omisión (0,72%), están entre los más frecuentes. Del total 191 segmentos, el 59,16% (113 segmentos) registraron algún error en el uso de los términos o la precisión de la traducción vinculada a la estadística. No obstante, la traducción automática también tuvo como resultado la traducción incorrecta de otros términos que no están relacionados con la estadística (17,31%) y faltas gramaticales (17,55%), como el uso incorrecto de las preposiciones o el orden erróneo de

las unidades de sintaxis en la frase, entre otros. En el caso de los formatos de número, se cambió correctamente el formato de número del inglés estadounidense al formato europeo utilizado en España. Pero en la traducción automática también se cambió el formato de los números de los títulos correspondientes a las figuras y las tablas.

Las medias de las variables Prod_Tao_TA y Prod_TA (Tabla 1) indican un aumento de productividad en el proceso de posesición cuando se utilizan conjuntamente la traducción automática con otras herramientas de la TAO, como la memoria o una base de datos terminológica. Se necesitaron en media 33,57 segundos para poseer un segmento proveniente del motor de traducción automática con la ayuda de la memoria y la base de términos, en comparación con 40,43 segundos cuando no se contó con estas herramientas. En media, se obtuvo un aumento de productividad de 7,84 segundos, equivalente a 6,06%. El mayor aumento de productividad en un segmento es de 79,17% (mínimo en la Tabla 2 en DIFF%), aunque existieron también segmentos en los cuales el uso de las herramientas TAO disminuyeron la productividad, siendo la mayor pérdida de productividad de 522,22% (máximo en la Tabla 3 en DIFF%). Además, esta diferencia es significativa, según los resultados de la prueba Mann-Whitney (Tabla 4). Por tanto, se valida la hipótesis H1.

Tabla 4. Resultados de la prueba Mann-Whitney H1

	Hipótesis nula	Sig.	Decisión
H1	La distribución de Prod_PE es la misma entre las categorías herramientas (TA vs. TA+TAO)	0,00*	Rechazar la hipótesis nula

*El nivel de significancia es de 0,05.

Además, la frecuencia de errores en la traducción de la terminología estadística tiene un impacto en la productividad de la posesición de la traducción automática (Prod_PE_TA) y también en la productividad de la posesición de la traducción TAO integrada con TA (Prod_PE_TA_TAO). Según los resultados de la prueba U de Mann-Withney, existen diferencias significativas en los niveles de productividad en la posesición de los segmentos con errores en la traducción automática de los términos de

estadística y los segmentos en los que no se han observado estos tipos de errores (Tabla 5). Por tanto, se validan las hipótesis H2 y H3.

Tabla 5. Resultados prueba U de Mann-Whitney H2 y H3

	Hipótesis nula	Sig.	Decisión
H2	La distribución de Prod_PE_TA es la misma entre los segmentos con errores en la traducción relacionada con la estadística y los segmentos sin errores.	0,00*	Rechazar la hipótesis nula
H3	La distribución de Prod_PE_TA_TAO es la misma entre los segmentos con errores en la traducción relacionada con la estadística y los segmentos sin errores.	0,00*	Rechazar la hipótesis nula

*El nivel de significancia es de 0,05.

Concretamente, la media del tiempo de posesición de la traducción automática de los segmentos sin errores vinculados a la terminología estadística es de 32,32 segundos; mientras que, la media de posesición de los segmentos con este tipo de errores es de 46,03 segundos (Tabla 6). Es decir, la productividad en la posesición de los segmentos sin errores de traducción relacionados con la estadística es un 70,21% mayor.

Tabla 6. Estadísticos descriptivos de la productividad en función de la presencia de errores del ámbito de la estadística

Variable	ET	Media	Mediana	Desviación típica	Mínimo	Máximo	N
Prod_PE_TA	SI	46,03	41,00	23,46	8,00	120	N
	NO	32,32	28,00	18,80	5,00	86,00	113
Prod_PE_TA_TAO	SI	37,70	31,00	20,21	8,00	123,00	78
	NO	25,15	25,00	12,61	6,00	59,00	113

Los mismo ocurre en el caso de la posesición de los segmentos con traducción TAO integrada con TA, en la cual la media de productividad de los segmentos sin errores relacionados con la estadística es de solo 25,15 segundos frente a los 37,70 segundos que son necesarios en media para poseer los segmentos con este tipo de errores. La diferencia de productividad es de 66,71%. Es decir, la presencia de errores relacionados con la estadística influye también en la productividad de la posesición de los segmentos con traducción TAO integrada con TA. No obstante, se puede observar que la media de productividad es menor para Prod_PE_TA_TAO tanto en los segmentos con errores del ámbito de la estadística como los segmentos sin

este tipo de errores. Es más, se obtiene una ganancia media de productividad del 81,90% en los segmentos con este tipo de errores y del 77,81% en el caso contrario.

En el análisis se tuvieron en cuenta segmentos de cuatro textos, que, aunque provienen del mismo libro, pueden tener algunas diferencias (por ejemplo, un número de caracteres diferente por segmento, estructuras de frases y un vocabulario diferente). Por tanto, se debe analizar si esas diferencias pueden determinar también resultados diferentes en los niveles de calidad y de productividad. En Tabla 7 se presentan los resultados de los estadísticos descriptivos de las variables analizadas en función del texto.

Tabla 7. Estadísticos descriptivos por texto fuente de los segmentos

Variable	Texto	Media	Mediana	Desviación típica	Mínimo	Máximo	N
Tamaño segmento	Texto 1	143,39	134,00	68,45	36	424	59
	Texto 2	133,47	140,00	49,99	39,00	223,00	15
	Texto 3	151,23	143,23	61,99	54,00	295,00	40
	Texto 4	158,00	143,00	74,90	64,00	397,00	77
BLEU_TA	Texto 1	59,77	60,91	17,39	8,24	100,00	59
	Texto 2	59,54	59,05	22,41	10,13	100,00	15
	Texto 3	55,33	61,97	22,30	9,24	94,88	40
	Texto 4	59,87	59,59	19,23	10,30	100,00	77
OQS_TA	Texto 1	83,71	94,12	20,27	27,59	100,00	59
	Texto 2	93,32	95,24	7,65	77,78	100,00	15
	Texto 3	44,83	65,69	68,10	-200,10	100,00	40
	Texto 4	77,06	91,67	42,15	-188,46	100,00	77
Prod_TA	Texto 1	41,83	40,00	25,99	5,00	108,00	59
	Texto 2	34,47	29,00	15,51	11,00	60,00	15
	Texto 3	41,68	36,00	22,52	8,00	120,00	40
	Texto 4	39,87	36,00	21,29	6,00	104,00	77
Prod_TA_TAO	Texto 1	32,15	29,00	21,03	8,00	123,00	59
	Texto 2	30,13	29,00	13,47	9,00	52,00	15
	Texto 3	30,95	24,00	17,49	9,00	75,00	40
	Texto 4	34,22	29,00	18,03	6,00	103,00	77
DIF_P 40	Texto 1	-9,68	-8,00	17,74	-53,00	47,00	59
	Texto 2	-4,33	-2,00	15,58	-34,00	23,00	15
	Texto 3	-10,68	-8,50	20,63	-95,00	29,00	40
	Texto 4	-5,65	-4,00	14,16	-51,00	35,00	77
DIF %	Texto 1	-5,67	-26,53	78,89	-72,41	522,22	59
	Texto 2	-5,66	17,24	40,23	-59,65	79,31	15
	Texto 3	-15,08	-21,28	49,48	-79,17	181,25	40
	Texto 4	-1,74	-9,30	58,59	-68,42	388,89	77

Como se puede observar, aunque existen algunas diferencias en cuanto a la media del tamaño de los segmentos entre los textos, estas no son muy grandes. Los segmentos del Texto 2 tienen en media un menor número de caracteres (133,47), mientras que el Texto 4 alcanzan la mayor media (158,00); es decir, una diferencia de 24,53 de caracteres. No obstante, destaca el hecho que a pesar de que sus segmentos tienen en media el mayor número de segmentos, alcanzan la mayor media en los puntos BLEU (59,87). Por tanto, a primera vista no parece que la longitud de los segmentos afecte la calidad de la traducción automática, ya que los segmentos del Texto 4 alcanzan la mayor media en cuanto a la puntuación BLEU (59,87). En cambio, aunque la métrica OQS coincide con el BLUE en cuanto al texto con menor calidad (Texto 3), la mayor media y mediana se alcanza en el Texto 2 (93,32 y 95,24) respectivamente. A pesar del resultado en calidad, el uso de las herramientas TAO solo ha aportado en media un aumento del 1,74% de la productividad en la posesición. Destaca que el mayor aumento de la productividad al combinar TA con las herramientas TAO se obtuvo en el Texto 3 (15,08%), el que obtuvo la menor puntuación BLEU (55,33) y OQS (44,83), siendo el segundo texto con segmentos de mayor tamaño (151,23). En consecuencia, hay que analizar la relación que existe entre estas variables. A continuación, se presentan los resultados de las correlaciones del coeficiente de Spearman (Tabla 8).

Tabla 8. Resultado de las correlaciones de Spearman

	Tamaño segmento	BLEU_TA	OQS_TA	Prod_TA	Prod_TA_TAO	DIF
Tamaño segmento	1,000	0,127	-0,179*	0,574**	0,478**	-0,249**
BLEU_TA	0,127	1,000	0,338*	-0,220**	-0,224**	0,049
OQS_TA	-0,179*	0,338*	1,000	-0,401**	-0,409**	0,053
Prod_TA	0,574**	-0,220**	-0,401**	1,000	0,645**	-0,593**
Prod_TA_TAO	0,478**	-0,224**	-0,409**	0,645**	1,000	0,146*
DIF	-0,249**	0,049	0,053	-0,593**	0,146*	1,000

*La correlación es significativa a un nivel de 0,01 (dos colas)

**La correlación es significativa a un nivel de 0,05 (dos colas)

Se aplicó el coeficiente de correlación de Spearman, porque los resultados de la prueba Shapiro-Wilk, que alcanzó una significación menor a 0,05 en el caso

de todas las variables, mostraron que las variables no tienen una distribución normal (Véase Tabla 10, Anexo 2). Por tanto, hay que utilizar un coeficiente de correlación no paramétrico, como es el caso del coeficiente Spearman.

A priori, parecería que no existe una relación significativa entre el tamaño del segmento en la lengua fuente (inglés) y la calidad del segmento traducción con un motor de traducción automática en la lengua meta en español (español) en el caso de los textos de estadística. El coeficiente de correlación es de solo 0,127, mostrando una correlación débil y sin significación. En cambio, la correlación con la medición multidimensional de la calidad de la traducción automática (OQS_TA) sí arroja una correlación significativa y negativa (-0,179), es decir, los segmentos de mayor longitud suelen ser los que registran un mayor número de errores o de mayor nivel. Además, existe una correlación significativa entre el tamaño del segmento y la productividad. El coeficiente de Spearman de 0,574 indica una correlación positiva considerable (Mondragón Barrera, 2014).

Destaca la correlación significativa que existe entre la métrica BLEU del segmento traducido con el motor de traducción automática y la productividad. Los coeficientes son de -0,220 en caso de la traducción automática (Prod_TA) y de -0,224 cuando la posesición se realiza también con herramientas TAO (Prod_TA_TAO). Hay que recordar que la productividad se midió en segundos por segmento, por tanto, el coeficiente indica que la correlación es negativa y media en ambos casos. Es decir, mayores niveles de calidad están relacionados con menores tiempos de traducción, por tanto, con una mayor productividad. La correlación es significativa y negativa también con OQS_TA (-0,401 y -0,409, respectivamente) y ligeramente mayor en comparación con la correlación con el BLUE. Por tanto, se confirma la relación entre la calidad de la traducción y la productividad. Hay que destacar también la correlación positiva y media de 0,338 entre las dos métricas BLUE y OQS_TA, lo que representa un indicio del rendimiento medio de las dos métricas de calidad.

No es de extrañar que exista también una correlación positiva significativa entre Prod_TA y Prod_TA_TAO (coeficiente Spearman de 0,645) porque en ambos casos se utiliza el mismo motor de traducción automática y, si se tarda más en traducir un segmento TA, lo más probable se tarde más aún cuando

se utilicen también herramientas TAO. Esto se confirma por la correlación considerable (-0,593) que existe en el esfuerzo temporal realizado en los dos tipos de correlaciones. Hay que mencionar que el resultado negativo de la correlación es el resultado de la medición de la variable como la diferencia entre Prod_TA_TAO y Prod_TA. Por tanto, en realidad, la relación es positiva entre ambas variables, al igual que la relación con el tamaño del segmento. El resultado indica que cuando la duración de posesión es alta en el segmento TA, es menor la duración de la posesión cuando también se utilizan herramientas TAO.

La memoria de traducción generada al traducir los casos de estudio del libro no contiene suficiente información para generar coincidencias (*fuzzy matches*) en el texto de mínimo 75%, que es el que se encontró que aumenta la productividad en estudios previos (Para y Arcedillo, 2015). No obstante, en la posesión del Texto 3 y el Texto 4, se configuró un nivel mínimo de coincidencias del 50%, lo que resultó en nueve *fuzzy matches* de diferentes niveles. A continuación, se realizó un análisis descriptivo de la calidad y la productividad de la posesión de estas *fuzzy matches*, así como una comparación con los segmentos en la lengua meta resultadas de la traducción automática (Tabla 9).

Tabla 9. Análisis descriptivo de las *fuzzy matches* de la memoria de traducción

Texto	Nº seg.	BLEU_TA	Prod_TA	% FMS	BLEU_TM	Prod_TM	DIF	DIF %
Texto 3	3	29,26	61	58%	17,69	29	-32	-52,46
	6	46,62	120	66%	31,61	32	-88	-73,33
	18	43,01	20	68%	20,95	45	25	125,00
	39	76	23	51%	10,88	42	19	82,61
	7	58,4	23	72%	53%	43	20	86,96
Text 4	8	71,85	97	60%	44,64	232	135	139,18
	22	45,4	59	60%	48,00	94	35	59,32
	62	100	8	51%	10,18	201	193	2412,5
	78	79,68	8	51%	4,84	34	26	325

Como se puede observar en la Tabla 9, solo en el caso del segmento 8 del Texto 3, la concordancia de la memoria de traducción (FMS) obtuvo una puntuación BLEU (48,00) superior a la alcanzada por el segmento traducido con el motor de traducción automática (45,4). Sin embargo, la diferencia es

demasiado pequeña para que pueda tener algún efecto en la productividad. Es más, se registró un aumento del esfuerzo temporal del 59,32%. En este caso, se trata de un *FMS* del 60%, pero incluso en el segmento 7, que tiene un nivel de *FMS* del 72% no se observó un aumento de productividad, todo lo contrario, se requirió aumentar en un 86,96% el esfuerzo temporal. No obstante, hay también algunas excepciones, en el caso de los segmentos 3 y 6, en los cuales, aunque la puntuación BLEU fue menor, se consiguió mejorar la productividad en un 52,46% y 73,33%, respectivamente.

Un análisis más detallado de los segmentos (62 y 78), muestra que algunas concordancias, aunque alcanzan un *fuzzy match* del 51%, los niveles del BLEU son muy bajos de solo 10,18 y 4,84, respectivamente. Esto demuestra que el BLEU es una métrica bastante buena para medir la calidad de las traducciones automáticas o de las *fuzzy matches*. Por ejemplo, el segmento con el menor resultado BLEU (4,84) es el siguiente "PLS-SEM permite probar estadísticamente la significación de la relación hipotetizada (Capítulo 6).", mientras que, la sugerencia del motor de traducción automática es "Así, podemos proceder a la evaluación del modelo estructural (Capítulo 6)". Claramente, la segunda es una traducción mucho mejor del segmento fuente "Thus, we can proceed with the evaluation of the structural model (Chapter 6)". Los nueve segmentos, tanto en la lengua fuente, como en sus respectivas traducciones (traducción automática y *fuzzy matches*) están publicados en el Cuadro 9 del Anexo 3.

5. Discusión de los resultados

Los resultados de la presente investigación demostraron que la traducción automática de los textos de estadística puede producir niveles de calidad superiores al mínimo requerido para conseguir un aumento de productividad en el proceso de posesición. Concretamente, los segmentos analizados obtuvieron un resultado del BLUE superior al mínimo de 45 indicado por el estudio de Parra y Arcedillo (2015). Además, el uso de la traducción automática junto a las herramientas TAO, como el glosario y la memoria de traducción, contribuyen a una mejora significativa de la productividad. En este sentido, se confirma el resultado del estudio de Federico *et al.* (2012), en el cual se demostró que al integrar TAO y TA se gana en productividad.

Así mismo, en general, en los segmentos con mayor nivel de calidad también obtienen mayores niveles de productividad, tal como se encontró también en la investigación realizada por Aranberri *et al.* (2014).

Previos estudios hallaron también que la memoria de traducción mejora la productividad en presencia de *fuzzy matches* de 70% (Screen, 2017), el nivel de productividad de la posesición con la TA es similar al de los *fuzzy matches* entre 85% y 94% (Guerberof, 2012) o incluso superior frente a *fuzzy matches* entre 75% y 84% (Parra y Arcedillo, 2015). El análisis de los *fuzzy matches* de la memoria de traducción utilizada en el presente estudio no mostró una relación entre el nivel de *fuzzy match* y la productividad. Tampoco se identificó una diferencia significativa entre la sugerencia de la traducción automática y las coincidencias de la memoria de traducción. No obstante, hay que tener en cuenta de que solo se dispuso de 9 *fuzzy matches* para el análisis, lo que representa una limitación relevante para poder evaluar el impacto del nivel de *fuzzy matches* en la productividad.

Según Arranberri *et al.* (2014), el aumento de productividad varía en función del texto. Los resultados del análisis confirman que el texto influye en la variación de la productividad, porque el texto que obtuvo la mejor puntuación tanto en el BLEU como en el OQS registró la mejor media de productividad; mientras que, el texto con el peor resultado en ambas métricas requirió mayor esfuerzo temporal durante el proceso de posesición. Pero destaca el hecho de que el texto con los niveles de calidad más bajos obtuvo la mayor ganancia de productividad al momento de utilizar la TA integrada en la TAO. Además, hay que mencionar que el tamaño de los segmentos de los textos también varía, lo que tiene un impacto en la ventaja obtenida con la traducción automática o el uso de la TAO. Sin embargo, al contrario de la relación negativa encontrada en la investigación llevada a cabo por Toral *et al.* (2018), en el presente estudio se obtuvo una correlación positiva. Es decir, la ganancia en productividad, tanto en la TA como en la traducción TAO con TA integrada, es mayor en el caso de los segmentos de mayor tamaño.

Otro factor que hay que mencionar es el tipo de errores cometidos por la traducción automática, que tienen un impacto significativo sobre la duración de la posesición (Carl y Toledo-Báez, 2019). Los resultados de la presente

investigación muestran que cuando la traducción automática comete errores de traducción relacionados con el área de estadística, como el uso del término erróneo, la traducción incorrecta de los términos de estadística, la omisión y la falta de traducción de términos estadísticos, influye de forma significativa y negativa en la productividad. Este tipo de errores representan una proporción significativa de los errores cometidos por la traducción automática en el presente estudio. El uso de un glosario y de una memoria de traducción junto a la traducción automática resulta en unas ganancias substanciales de productividad, en estos casos. Por tanto, la capacidad del software de traducción automática de traducir correctamente los términos específicos de la especialidad del texto traducido influye en la productividad de la posesición. No obstante, el uso de estas herramientas tuvo un aporte relevante en la productividad de la posesición también en el caso de los segmentos sin este tipo de errores. Sería interesante evaluar la variación de la productividad al utilizar otra aplicación de traducción automática y analizar también si existen diferencias en los tipos, la cantidad y el nivel de gravedad de los errores.

En los estudios empíricos se encontró que el aumento de productividad varía en función de las características y la experiencia del traductor con la posesición, el software utilizado y la especialidad (Aranberri *et al.*, 2014; Toledo Báez *et al.*, 2018; Álvarez *et al.*, 2020). La posesición y la evaluación de los errores fueron realizados por una sola persona con un alto nivel de conocimiento de la técnica estadística explicada en los textos, lo que representa una limitación de la presente investigación. Para poder evaluar el impacto de estos factores en los resultados, se deberían utilizar datos obtenidos de la posesición y evaluación de los errores realizada por otros traductores con características diferentes y con otro nivel de experiencia.

Otra limitación de la investigación es que se utilizó solo un enfoque unidimensional de medición de la productividad: el esfuerzo temporal. Aunque se trata de la medida más utilizada en los estudios empíricos, varias investigaciones han medido también el esfuerzo técnico y el esfuerzo cognitivo (Fernández-Torné y Matamala, 2016; Toral *et al.*, 2018; Stasimioti y Sosoni, 2020). El enfoque multidimensional permitiría evaluar el impacto

de las herramientas en cada tipo de esfuerzo e identificar como contribuyen en la mejora de la productividad.

Otro aspecto que hay que discutir es el uso de las medidas de calidad de la traducción. En la presente investigación se ha utilizado una métrica automática (BLEU) y otra resultada de la evaluación humana (OQS). Todas las métricas tienen limitaciones y un cierto componente subjetivo (Gutiérrez-Artacho, Olvera-Lobo y Rivera-Trigueros, 2018). Para evaluar la fiabilidad de las métricas de calidad, los estudios empíricos utilizan la correlación con el objetivo de medir la consistencia entre los dos tipos de métricas (Chunyu y Tak-ming, 2015). La métrica BLEU es la más popular justo porque en previas investigaciones mostró una alta correlación con la evaluación humana de la traducción automática (Rivera-Trigueros, 2022). El coeficiente de Spearman fue utilizado en el presente estudio con este propósito. La correlación positiva y significativa entre el BLUE y OQS indica la existencia de consistencia. Sin embargo, el valor de la correlación no es suficientemente alto para indicar una alta fiabilidad. Habría que ampliar el estudio a muestras de segmentos más grandes y con datos obtenidos de varios traductores para poder evaluar mejor la fiabilidad de estas dos métricas.

6. Conclusiones

El objetivo de la presente investigación fue evaluar la productividad de la traducción asistida y automática del inglés al español en el caso de los textos de estadística. Los resultados mostraron que los traductores pueden utilizar las herramientas TA y TAO para mejorar su productividad, aunque la validación de la H1 muestra que existen diferencias entre los dos tipos de posesición. Concretamente, los mejores resultados se obtienen cuando se utiliza un software de traducción que integra ambos tipos de herramientas. Por tanto, la integración de TA y TAO permite reducir el esfuerzo temporal de la posesición. En el caso de la presente investigación se utilizó el software de traducción Trados Studio y el motor de traducción automática Language Weaver, que mostraron niveles aceptables de calidad en el área de la estadística.

El estudio aporta a la investigación sobre la traducción automática y la traducción asistida mediante la propuesta de uso de dos medidas de calidad para la traducción automática y la traducción asistida: una medida automática (BLEU) y otra medida obtenida mediante la evaluación humana (OQS). Concretamente, se proponen categorías de errores específicamente relacionados con la traducción de términos de estadística y categorías específicas para las partes del texto que no están específicamente relacionadas con la especialidad. Los valores asignados al nivel de gravedad de los errores permiten calcular una puntuación global de calidad para cada segmento del análisis. Esta puntuación puede ser utilizada posteriormente para un análisis más detallado de la eficiencia en la traducción automática de un determinado software de traducción automática y traducción asistida para una especialidad en concreto y para una mejor interpretación del efecto sobre productividad.

La inclusión de categorías vinculadas a la traducción de los términos de estadística en la evaluación humana de la calidad del texto permitió también identificar la proporción del texto que contenía este tipo de errores. Además, la validación de H1 y H2 mostró que existen diferencias significativas en la productividad de la posesición cuando en la traducción automática surgen errores que pertenecen a estas categorías. En concreto, cuando ocurren errores en la traducción de términos de la estadística, el nivel de productividad es menor.

Así mismo, los resultados demuestran que las características del texto fuente afectan tanto la calidad de la traducción automática como la duración del proceso de posesición. No obstante, el estudio debería ser aplicado a textos de otras metodologías estadísticas, obtener evaluaciones de calidad de otros traductores y analizar otros softwares de traducción automática y asistida para identificar otros factores que pueden afectar la productividad del proceso de posesición, como las características de los traductores, entre otros. Además, la medición de la productividad debería incluir también el esfuerzo técnico y el esfuerzo cognitivo.

Referencias

- Allen, J. (2003). Post-editing. In H., Soomers (ed.). *Computers and Translation: A translator's guide*, 35, 297-318. Benjamins Translation Library. <https://doi.org/10.1075/btl.35.19all>
- Álvarez, S., Oliver, A., y Badia, T. (2020). Quantitative Analysis of Post-Editing Effort Indicators for NMT. In Proceedings of 22nd Annual Conference of the European Association of Machine Translation (pp. 411-420).
- Álvarez, S., Oliver, A. y Badia, T. (2021). Comparing NMT y PBSMT for post-editing in-domain formal texts: A case study. In Tra&Co Group (ed.). *Translation, interpreting and cognition: The way out of the box*, 33-46. Berlin: Language Science Press. <https://doi.org/10.5281/zenodo.4545031>
- American Translator Association. (s.f.). *What is machine translation?* <https://www.atanet.org/client-assistance/machine-translation/>
- Arranberri, N., Labaka, G., Diaz de Ilarazza, y Sarasola (2014). Comparison of post-editing productivity between professional translators and lay users. En O'Brien, S., Simard, M., y Specia, L. (eds.). Third Workshop on Post-Editing Technology and Practice. The 11th Conference of the Association for Machine Translation in the Americas, Vancouver, October 22-26, 2014.
- Briva-Iglesias, V., O'Brien, S. y Cowan, B.R. (2023). The impact of the traditional and interactive post-editing on Machine Translation User Experience, quality and productivity. *Translation, Cognition & Behavior*, 1-27. <https://doi.org/10.1075/tcb.00077.bri>
- Cadwell, P., Castilho, S., O'Brien, S. y Mitchell, L. (2016). Human factors in machine translation and postediting among institutional translators. *Translation Spaces*, 5(2), 222-243. <https://doi.org/10.1075/ts.5.2.04cad>

- Carl, M. y Toledo Baéz, M.C. (2019). Machine translation errors and the translation process: a study across different languages. *Journal of Specialised Translation*, 31, 107-132.
- Castilho, S., Moorkens, J., Gaspari, F., Sennrich, R., Way, A., y Georgakopoulou, P. (2018). Evaluating MT for Massive Open Online Courses. A multifaceted comparison between PBMST and NMT systems. *Machine Translation*, 32(3), 255-278.
- Chunyu, K. y Tak-ming, B.W. (2015). Evaluation in machine translation and computer-aided translation. In C. Sin-wai (Ed.). *The Routledge Encyclopedia of Translation Technology* (1st ed., pp. 213-236). Routledge.
- Crăciunescu, O., Gerding-Salas, C., y Stringer-O’Keeffe, S. (2004). Machine Translation and Computer-Assisted Translation: a New Way of Translating? *Translation Journal*, 8(3).
<https://translationjournal.net/journal/29computers.htm>
- Do Carmo, F. y Moorkens, J. (2020). Differentiating, editing, post-editing, and revision. In *Translation Revision and Post-editing* (1ed), pp. 35-49. Routledge.
- EUATC. (2022). *European Language Industry Survey 2022*. ELIS Research.
https://fit-europe-rc.org/wp-content/uploads/2022/03/ELIS-2022_survey_results_final_report.pdf
- Federico, M., Cattelan, A. y Trombetti, M. (2012). Measuring Using Productivity in Machine Translation Enhanced Computer Assisted Translation. En Proceedings of the 10th Conference of the Association for Machine Translation of the Americas 2012.
- Fernández-Rodríguez, M. (2010). Evolución de la traducción asistida por ordenador. De las herramientas de apoyo a las memorias de traducción. *Sendeban*, 21, 201-230.

- García, I. (2015). Computer-aided translation. In C. Sin-wai (Ed.). *The Routledge Encyclopedia of Translation Technology* (1st ed., pp. 68-87). Routledge.
- Gaspari, F., Toral, A., Naskar, S., Groves, D. Way, A. (2014). Perception vs Reality: Measuring Machine Translation Post-Editing Productivity. En O'Brien, S., Simard, M., y Specia, L. (eds.). *Third Workshop on Post-Editing Technology and Practice*. The 11th Conference of the Association for Machine Translation in the Americas, Vancouver, October 22-26, 2014.
- Guerberof, A. (2009). Productivity and quality in MT post-editing. In *Beyond Translation Memories: New Tools for Translators Workshop*. MTSummit, Ottawa, Canada. <https://aclanthology.org/2009.mtsummit-btm.7.pdf>
- Gutiérrez-Artacho, J., Olvera-Lobo, M-D., y Rivera-Trigueros, I. (2018). Human Post-editing in Hybrid Machine Translation Systems: Automatic and Manual Analysis and Evaluation. In Rocha, Á., Adeli, H., Reis, L.P., Costanzo, S. (eds.). *Trends and Advances in Information Systems and Technologies. WorldCIST'18 2018*. Advances in Intelligent Systems and Computing, vol. 745. Springer. https://doi.org/10.1007/978-3-319-77703-0_26
- Hair, J.F., Hult, G.M.T., Ringle, C.M., & Sarstedt, M. (2022). *A Primer on Partial Least Squares Structural Equation Modelin (PLS-SEM)*, 3rd ed. Thousand Oaks, CA: Sage.
- Jaworski, R. (2013). Anubis – Speeding Up Computer-Aided Translation. In Przepiórkowski, A., Piaseki, M., Jassem, K., Fuglewicz, O. (eds.). *Computational Linguistics. Studies in Computational Intelligence*, vol. 458. Springer, Berlin. https://doi.org/10.1007/978-3-642-34399-5_14
- Krings, H.P. (2001). *Repairing Texts: Empirical Investigations of Machine Translation Post-Editing Processes*. Ohio: Kent State University Press.

- Läubli, S., Amrhein, C., Düggelein, P., Gonzalez, B., Zwahlen, A. y Volk, M. (2019). Post-editing Productivity with Neural Machine Translation: An Empirical Assessment of Speed and Quality in the Banking and Finance Domain. En *Proceedings of Machine Translation Summit VII: Research Track*, pp. 267-272. Dublin, Irlanda. European Association for Machine Translation.
- Lommel, A. (2018). Metrics for Translation Quality Assessment: A Case for Standardising Error Typologies. In J., Moorkens, S., Castilho, F., Gaspari y Doherty, S. (eds.). *Translation Quality Assessment. Machine Translation: Technologies and Applications*, 1. Springer, Cham. https://doi.org/10.1007/978-3-319-91241-7_6
- López Pereira, A. (2019). Traducción automática neuronal y traducción automática estadística. *Revista Tradumàtica. Technologies de la Tradució*, 17, 1-19. <https://doi.org/10.5565/rev/tradumatica.235>
- Mariana, V., Cox, T. y Melby, A. (2015). The Multidimensional Quality Metrics (MQM) Framework: a new framework for translation quality assessment. *The Journal of Specialized Translation*, 23, 137-161.
- Melby, A.K. y Wright, S.E. (2015). Translation memory. In C. Sin-wai (Ed.). *The Routledge Encyclopedia of Translation Technology* (1st ed., pp. 662-677). Routledge.
- Mercader-Alarcón, J. y Sánchez-Martínez, G. (2016). Analysis of translation errors and evaluation of pre-editing rules for the translation of English news texts into Spanish with Lucy LT. *Revista Tradumàtica*, 14, 172-186. <https://doi.org/10.5565/rev/tradumatica.164>
- Mondragón Barrera, M.A. (2014). Uso de la correlación de Spearman en un estudio de intervención en fisioterapia. *Movimiento científico*, 8(1), 98-100.
- Moorkens, J. y O'Brien, S. (2015). Post-editing Evaluations: Trade-offs between Novice and Professional Participants. In I.D., El-Kahlout, M., Özkan, F., Sánchez-Martínez y G., Ramírez-Sánchez, F., Hollowood y

A., Way. (eds.). *Proceedings of the 18th Annual Conference of the European Association for Machine Translation*, 75-81.

Moorkens, J., Toral, A., Castilho, S., y Way, A. (2018). Translators' Perceptions of Literary Post-editing using Statistical and Neural Machine Translation. *Translation Spaces*, 7(2), 240-262. <https://doi.org/10.1075/ts.18014.moo>

MQM Committee. (2023). *MQM (Multidimensional Quality Metrics). The MQM Error Typology*. <https://themqm.org/error-types-2/typology/>

O'Brien, S. (2011). Towards predicting post-editing productivity. *Machine Translation*, 25(3), 197-215. <https://doi.org/10.1007/s10590-011-9096-7>

O'Brien, S. (2022). How to deal with errors in machine translation: Post-editing. In D., Kenny.(ed.). *Machine translation for everyone: Empowering users in the age of artificial intelligence*, 105-120. Berlin: Language Science.

Oliver, A. (2016). *Herramientas tecnológicas para traductores*. Barcelona: Editorial UOC.

Parra, C. y Arcedillo, M. (2015). *Machine translation evaluation made fuzzier: A study on post-editing productivity and evaluation metrics in commercial settings*. Proceedings of the MT Summit XV, Research Track, pp. 131-144. Miami (Florida), 30 octubre – 3 noviembre.

Qun, L. & Xiaojun, Z. (2023). Machine Translation. In C. Sin-wai (Ed.). *The Routledge Encyclopedia of Translation Technology* (2nd ed., pp. 105-119). Routledge.

Rivera-Trigueros, I. (2022). Machine translation systems and quality assessment: a systematic review. *Lang Resources & Evaluation*, 56, 593-619.

- Sánchez-Gijón, P., Moorkens, J., y Way, A. (2019). Post-Editing Neural Machines Translation versus Translation Memory Segments. *Machine Translation*, 33, 31-59. <https://doi.org/10.1007/s10590-019-09232-x>
- Screen, B. (2017). Productivity and quality when editing machine translation and translation memory outputs: an empirical analysis of English to Welsh translation. *Studia Celtica Posnaniensia*, 2(1), 119-142. <https://doi.org/10.1515/scp-2017-0007>
- Sin-wai, C. (2015). The development of translation technology 1967-2013. In C. Sin-wai (Ed.). *The Routledge Encyclopedia of Translation Technology* (1st ed., pp. 3-31). Routledge.
- Stasimioti, M. y Sosoni, V. (2020). Translation vs Post-editing of NMT Output: Measuring effort in the English-Greek language pair. 1st Workshop on Post-Editing in Modern-Day Translation. *Proceedings of the 14th Conference of the Association for Machine Translation in the Americas*. (pp. 109-124), October 6-9.
- Texeira, C. (2014). Perceived vs. measured performance in the post-editing of suggestions from machine translation and translation memories. En O'Brien, S., Simard, M., y Specia, L. (eds.). *Third Workshop on Post-Editing Technology and Practice*. The 11th Conference of the Association for Machine Translation in the Americas, Vancouver, October 22-26, 2014.
- Toledo Baéz, M.C. (2018). Machine Translation and Post-editing: Impact of Training and Directionality on Quality and Productivity. *Revista Tradumàtica. Tecnologies de la Traducció*, 16, 24-34. <https://doi.org/10.5565/rev/tradumatica.215>
- Toral, A., Wieling, M., y Way, A. (2018). Post-editing Effort of a Novel With Statistical and Neural Machine. *Frontiers in Digital Humanities*, 5(9), 1-9. <https://doi.org/10.3389/fdigh.2018.00009>
- Vieira, L.N. (2014). Indices of Cognitive Effort in Machine Translation Post-Editing. *Machine-Translation*, 28(3), 187-216.

Wang, H., Wu, H., He, Z., Huang, L. y Ward Church, K. (2022). Progress in Machine Translation. *Enigneering*, 18, 143-153.
<https://doi.org/10.1016/j.eng.2021.03.023>

7. Anexos

7.1. Anexo 1

Cuadro 8. Ficha clasificación de errores y cálculo OQS_TA

MQM Scorecard: Full MQM-Core Error Typology with 4 Severity Levels								
Segment No.	Source text	Trados TA Text	Postedited Text					
136	As shown in Exhibit A5.7, you see the standardized outer weights for the formative measurement models (e.g., QUAL), standardized outer loadings for the reflective measurement models (e.g., CUSL), and a 1.000 for the relationship between the CUSA construct and its single-item measure.	Como se muestra en el documento adjunto A5,7, puede ver los pesos exteriores estandarizados para los modelos de medición formativa (por ejemplo, QUAL), las cargas exteriores estandarizadas para los modelos de medición reflectiva (por ejemplo, CUSL) y un 1,000 para la relación entre el constructo CUSA y su medida de un solo elemento.	Como se muestra en la Figura A5.7, usted ve los pesos externos estandarizados para los modelos de medida formativos (p. ej., QUAL), las cargas externas estandarizadas para los modelos de medida reflectivos (p. ej., CUSL) y un 1,000 para la relación entre el constructo CUSA y su único ítem de medida.					
		Error Severity Levels:	Neutral	Minor	Major	Critical	Error Type Penalty Total	
		Severity Multipliers:	0	1	5	25		
ET Nos	Error	Error Types	Error Counts				ET Weights	ETPTs
1		Terminology	0	0	0	0	1,0	0,0
1.1	pesos exteriores, modelos de medición, cargas exteriores, modelos de medición, elemento.	Wrong statistic term	1	4	0	0	1,0	4,0
1.2		Wrong term other	0	0	0	0	1,0	0,0
2		Accuracy	0	0	0	0	1,0	0,0
2.1.2		Mistranslation statistic	0	0	0	0	1,0	0,0
2.1.3	documento adjunto	Mistranslation other	0	0	1	0	1,0	5,0
2.2.1		Addition statistic	0	0	0	0	1,0	0,0
2.2.2		Addition other	0	0	0	0	1,0	0,0
2.3.1		Omission statistic	0	0	0	0	1,0	0,0
2.4.1		Omission other	0	0	0	0	1,0	0,0
2.5.1		Untranslated statistic	0	0	0	0	1,0	0,0
2.5.2		Untranslated other	0	0	0	0	1,0	0,0
3		Linguistic conventions (was Fluency)	0	0	0	0	1,0	0,0
3.1		Grammar	0	0	0	0	1,0	0,0
3.1.?		Gender	0	0	0	0	1,0	0,0
		Punctuation	0	0	0	0	1,0	0,0

3.3		Spelling	0	0	0	0	1,0	0,0
4		Style	0	0	0	0	1,0	0,0
4.1		Register	0	0	0	0	1,0	0,0
5		Locale convention	0	0	0	0	1,0	0,0
5.1	A 5,7	Number format	0	1	0	0	1,0	1,0
					Absolute Penalty Total (APT):			10,00
		Evaluation Word Count (EWC):	52		Per-Word Penalty Total (PWPT):			0,1923
		Reference Word Count (RWC):	1000		Overall Normed Penalty Total (ONPT):			192,31
		Penalty Scaler (PS):	1,00		Overall Quality Score (OQS):			80,77
		Max. Score Value (MSV):	100,00					
					Overall Quality Fraction (OQF):			0,81
		ONPT=PWPT × RWC × PS						
		OQF=1 - (ONPT / RWC)						
		OQS = OQF × MSV						

7.2 Anexo 2

Tabla 10. Resultados Shapiro-Wilk

Variables	Estadístico	gl	Sig.
Tamaño segmento	0,922	191	0,000
BLEU_TA	0,985	191	0,042
CM_TA	0,581	191	0,000
Prod_TA	0,944	191	0,000
Prod_TA_TAO	0,888	191	0,000
DIF	0,953	191	0,000
DIF%	0,636	191	0,000

7.3 Anexo 3

Cuadro 9. Segmentos traducción automática y fuzzy matches

Nº Texto	Nº seg.	Fuente	Segmento meta TA	segmento FMS
Texto 3	3	Exhibit A4.2 displays the results table for the outer loadings.	El Anexo A4,2 muestra la tabla de resultados para las cargas exteriores.	La Figura A4.2 muestra el informe de resultados para los coeficientes path en formato de matriz.
	6	Exhibit A4.3 shows the internal consistency reliability values in matrix format.	El Anexo A4,3 muestra los valores de confiabilidad de consistencia interna en formato de matriz.	La Figura A4.3 muestra el informe de resultados para los coeficientes path en formato de matriz.
	18	Exhibit A4.6 shows the HTMT values for all pairs of constructs in a matrix format.	El Anexo A4,6 muestra los valores HTMT para todos los pares de construcciones en un formato de matriz.	La Figura A4.6 muestra el informe de Akaike para los coeficientes path en formato de matriz.
	39	Exhibit A4.10 summarizes the results of the reflective measurement model assessment.	El Anexo A4,10 resume los resultados de la evaluación del modelo de medición reflexiva.	La Figura A4.10 muestra el informe de resultados para los coeficientes path en formato de matriz.
Texto 4	7	Exhibit A5.1 shows the constructs and their relationships, which represent the extended structural model for our PLS-SEM example in the remaining chapters of the text.	El Anexo A5,1 muestra los constructos y sus relaciones, que representan el modelo estructural extendido para nuestro ejemplo PLS-SEM en los capítulos restantes del texto.	La Figura A5.1 muestra los constructos y sus relaciones, los cuales representan el modelo estructural para el estudio de caso de PLS-SEM.

	8	<p>To summarize, the extended corporate reputation model has three main theoretical components: (1) the target constructs of interest (namely, CUSA and CUSL); (2) the two corporate reputation dimensions, COMP and LIKE, that represent key determinants of the target constructs; and (3) the four exogenous driver constructs (i.e., ATTR, CSOR, PERF, and QUAL) of the two corporate reputation dimensions.</p>	<p>En resumen, el modelo de reputación corporativa extendida tiene tres componentes teóricos principales: (1) los constructos objetivo de interés (a saber, CUSA y CUSL); (2) las dos dimensiones de reputación corporativa, COMP y SIMILARES, que representan los determinantes clave de los constructos objetivo; y (3) los cuatro constructos exógenos (es decir, ATTR, CSOR, PERF y QUAL) de las dos dimensiones de reputación corporativa.</p>	<p>En resumen, el modelo de reputación corporativa simple tiene dos componentes teóricos principales: (1) los constructos dependientes de interés, a saber, CUSA y CUSL (constructos endógenos), y (2) las dos dimensiones de reputación corporativa COMP y LIKE (constructos exógenos), que representan los determinantes clave de los constructos dependientes.</p>
	22	<p>Since the number of missing values is relatively small (i.e., less than 5% missing values per indicator; Chapter 2), we use mean value replacement instead of casewise or pairwise deletion to treat the missing values.</p>	<p>Dado que el número de valores faltantes es relativamente pequeño (es decir, menos del 5 % de los valores faltantes por indicador; Capítulo 2), utilizamos el reemplazo de valores medios en lugar de la eliminación por casewise o por pares para tratar los valores faltantes.</p>	<p>De nuevo, si el número de valores perdidos por indicador en su conjunto de datos es relativamente pequeño (es decir, menos del 5 % de valores perdidos por indicador; 2), se recomienda reemplazar por el valor de la media en lugar de la eliminación de casos para tratar los valores perdidos al ejecutar.</p>

	62	For example, the critical values for significance levels of 1% ($\alpha = 0.01$) and 5% ($\alpha = 0.05$) probability of error are 2.57 and 1.96 (two-tailed), respectively.	Por ejemplo, los valores críticos para los niveles de significación de 1 % ($\alpha = 0,01$) y 5 % ($\alpha = 0,05$) de probabilidad de error son 2,57 y 1,96 (dos colas), respectivamente.	Por ejemplo, el límite inferior y superior del intervalo de confianza del 95 % del HTMT para CUSA y COMP son 2,57 y 1,96, respectivamente.
	78	Thus, we can proceed with the evaluation of the structural model (Chapter 6).	Así, podemos proceder a la evaluación del modelo estructural (Capítulo 6).	PLS-SEM permite probar estadísticamente la significación de la relación hipotetizada (Capítulo 6).