

Traducción automática neuronal en atención al cliente: percepción de los estudiantes universitarios de Traducción e Interpretación acerca de la interferencia inter- e intralingüística

Machine translation and customer service: How do translation trainees' perceive inter- and intralinguistic interference

Resumen

Presentamos la continuación de un estudio empírico (Martín de Santa Olalla y Rico en prensa) que analizaba la percepción de la idiomática o naturalidad en un conjunto de oraciones traducidas mediante un motor neuronal del inglés al español, por parte de dos grupos de estudiantes de traducción: el primer grupo sin acceso al texto en inglés y el segundo con acceso al original en inglés y a su traducción en español. Partíamos en este estudio de un corpus de 50 oraciones y un cuestionario con una pregunta en la que los estudiantes tenían que valorar la naturalidad de las oraciones —en una escala de 1 (nada natural) a 5 (muy natural)— y otra en la que se les pedía reformular la oración española. Las respuestas a estas dos preguntas en estos dos grupos nos llevaron a un análisis de carácter cuantitativo que mostraba las diferencias en la percepción de la idiomática y otro cualitativo que mostraba el desempeño en la reformulación de oraciones en una muestra de oraciones. El objetivo de este artículo es profundizar en los fenómenos que originan la falta de idiomática: la interferencia inter- e intralingüística y en las diferentes percepciones de esta en los dos grupos. Para esto, partimos de los datos proporcionados por la aplicación del *t-test*, prueba que nos permite saber si las diferencias son estadísticamente significativas, a las cinco variables en las que se han clasificado las oraciones de nuestro corpus: oraciones idiomáticas, oraciones no idiomáticas por interferencias léxicas, oraciones no idiomáticas por interferencias morfosintácticas, oraciones no idiomáticas por interferencias culturales, oraciones no idiomáticas por incluir rasgos de lo que la industria ha dado en llamar latinoamericano u oraciones no idiomáticas por interferencias latinas. Los resultados muestran, en general, la confianza de los estudiantes en la traducción automática. El dato grueso de nuestro estudio es que los estudiantes dan una nota de 3,4 sobre 5 a las oraciones que hemos considerado no idiomáticas. El análisis cuantitativo detallado nos muestra que los estudiantes dan una nota más alta en naturalidad cuando

Autoría

AURORA MARTÍN DE SANTA

OLALLA SÁNCHEZ

Universidad Complutense de Madrid,
España

aurosant@ucm.es

<https://orcid.org/0000-0002-1880-5048>

Para citar este artículo:

Martín de Santa Olalla, A. (2025). Traducción automática neuronal en atención al cliente: percepción de los estudiantes universitarios de Traducción e Interpretación acerca de la interferencia inter- e intralingüística, *ELUA*, 44, 147-180. <https://doi.org/10.14198/ELUA.29052>

Recibido: 31/12/2024

Aceptado: 26/02/2025

Financiación:

Esta publicación es parte del proyecto de I+D+i ROBOT-TALK PID2022-140897OB-I00 financiado por MCIN/AEI/10.13039/501100011033/ y FEDER/UE

Conflicto de intereses: la autora declara que no hay conflicto de intereses.

© 2025 Aurora Martín de Santa Olalla Sánchez



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

ven el texto en inglés y las diferencias en la percepción de esta característica en los dos grupos son significativas en las cinco variables. Por otra parte, el análisis cualitativo de una pequeña muestra nos deja ver cierta tendencia a la sobrecorrección que es más acusada cuando no tienen acceso al texto en inglés.

Palabras clave:

traducción automática neuronal; atención al cliente; idiomática; interferencia; calidad de la traducción; español.

Abstract

This article is a follow up of an empirical study (Martín de Santa Olalla y Rico in press) which analysed the perception of idiomaticity or naturalness in a set of sentences translated using a neural machine translation engine from English into Spanish by two groups of translation students: the first group without access to the English text and the second with access to the original English text and its Spanish translation. The research was based on a corpus of 50 sentences and a questionnaire with a question in which the students had to evaluate the naturalness of the sentences – on a scale from 1 (not at all natural) to 5 (very natural)– and another in which they were asked to reformulate the Spanish sentence. The responses to these two questions in these two groups led to a quantitative analysis showing differences in perceived idiomaticity and a qualitative analysis showing sentence reformulation performance in a sample of sentences. The objective of this article is to delve into the phenomena that give rise to the lack of idiomaticity: inter- and intralinguistic interference and the different perceptions of this in two groups. For this purpose, we start from the data provided by the application of the t-test, which allows us to know if the differences are statistically significant, to the five variables in which the sentences of our corpus have been classified; namely: idiomatic sentences, non-idiomatic sentences due to lexical interference, non-idiomatic sentences due to morphosyntactic interference, non-idiomatic sentences due to cultural interference, non-idiomatic sentences due to the inclusion of features of what the industry has come to call Latin American Spanish or non-idiomatic sentences due to Latin interference. The results show, in general, students' confidence in machine translation. The main finding of our study is that students give a score of 3.4 out of 5 to the sentences we have considered non-idiomatic. The detailed quantitative analysis shows that students give a higher mark for naturalness when they see the text in English and the differences in the perception of this characteristic in the two groups are significant in all five variables. On the other hand, qualitative analysis of a small sample shows a tendency towards overcorrection which is more significant when they do not have access to the text in English.

Keywords:

neural machine translation; customer service; idiomaticity; interference; translation quality; Spanish.

1. INTRODUCCIÓN

En los últimos años, la investigación en traducción automática (TA) ha retomado el desarrollo de unas líneas de trabajo que tienen como objetivo la definición de las características del texto traducido automáticamente y, en algunos casos, el estudio de la repercusión o la huella que estas dejan en las lenguas traducidas. Estas investigaciones aparecen en muchas ocasiones bajo la palabra clave de *translationese* (*human-translationese*, *machine-translationese* y *post-editedese*). Relacionadas con estos conceptos y estos términos, existe una amplia nómina de publicaciones, como la de Brglez y Vintar (2022), que aborda las diferencias entre la *human-translationese* y la *machine-translationese*; Daems et al. (2017) que, en un estudio exploratorio sobre la *posteditese*,



muestran la incapacidad de un grupo de estudiantes de traducción para distinguir entre textos poseditados para publicar de otros traducidos por humanos y la posibilidad de crear un modelo supervisado de *machine-learning* que pueda distinguir entre el texto traducido por humanos y el texto poseditado; Toral (2019) que califica la *posteditese* como una *translationese* extrema o intensificada; Sánchez-Gijón y Piqué (2019) que, al igual que el citado Toral (2019), advierten sobre la necesidad de abordar la calidad de la traducción más allá de la detección de errores; Sánchez-Gijón (2022) que llama la atención sobre los efectos que, en el medio o en el largo plazo, puede tener la traducción automática sobre los estándares de las lenguas de llegada o Rico (2025) que habla de la imprimación (*priming* en inglés) como la huella que deja la traducción automática en los textos de llegada.

El término *translationese* fue acuñado por Gellerstam (1986, como se citó en Jiménez-Crespo 2023) y se utilizó también en el ámbito de los estudios de traducción basada en corpus (CBTS, *Corpus-Based Translated Studies*, por sus siglas en inglés). Sobre este concepto, Baker (1993) introdujo la noción de los universales de la traducción, características específicas de los textos traducidos independientemente de la combinación lingüística.

Recientemente, Jiménez-Crespo (2023) retoma los argumentos en contra del empleo de este término que se esgrimieron a principios de los 2000 y confirma la validez de estos en contra de su uso en las investigaciones relacionadas con la TA. En su lugar, propone el uso de términos menos marcados como *la lengua de las traducciones* o *las características típicas* o *generales de la lengua traducida*.

Las razones son dos: la primera tiene que ver con la connotación negativa que tiene el término —Jiménez Crespo se refiere aquí, entre otras cosas, al efecto perverso que tuvo la primera utilización de este término en Gellestam (1986) que cuestionaba la profesionalidad del traductor y la calidad del texto traducido. La segunda tiene que ver con el peligro de la sobregeneralización. La sobregeneralización a la que llevó el establecimiento de los universales de la traducción y a la que podríamos llegar en nuestros días al extrapolar resultados y características, que se dan en un determinado dominio, en un género específico, en unas determinadas lenguas y con un motor de traducción concreto, a tendencias generales.

El objetivo de este artículo es profundizar en una de las características de la lengua traducida automáticamente, la idiomaticidad (o mejor dicho la ausencia de ella) en un conjunto específico de géneros textuales (mensajes de correo electrónico, chat o FAQ), relacionados con un ámbito particular, la atención al cliente, en la traducción del inglés al español, con un motor neuronal concreto. Por una parte, nos preguntamos si la percepción de esta falta de idiomaticidad es diferente en dos contextos diferenciados —con o sin acceso al texto en inglés— y en función de las cuatro variables que la originan. A esto dará respuesta el análisis cuantitativo de los datos obtenidos en la primera pregunta de nuestro cuestionario. Además, en la parte cualitativa de nuestro análisis, nos preguntamos en cuál de estos dos contextos obtenemos mejor desempeño en la tarea de corrección o posesición del texto traducido automáticamente.

Siguiendo la terminología y las metáforas de Chesterman (2004), esta investigación es de carácter descriptivo, discurre en carreteras secundarias (*low road*) y va de abajo arriba (*bottom-up*) describiendo, en primer lugar, lo particular, lo que sucede en un ámbito concreto, bajo determinadas condiciones. El objetivo final es sentar las bases que podrían

contribuir a definir las características de los textos traducidos automáticamente en el ámbito de la atención al cliente.

Este trabajo, además, es continuación del iniciado en Martín de Santa Olalla y Rico (en prensa) en el que hacíamos una primera aproximación a la idiomática: su definición y la variabilidad a la que se veía sometida en diferentes contextos, a través de un cuestionario administrado a dos grupos de estudiantes en circunstancias diferentes. La aportación original de este segundo artículo radica en la profundización de las causas que originan la falta de idiomática, a saber: la interferencia interlingüística (interferencia léxica, morfosintáctica y cultural) y la interferencia intralingüística (interferencia latinoamericana¹). Este segundo trabajo aporta, además, nuevos datos cuantitativos que nos confirman la variabilidad a la que se ve sometida la idiomática en diferentes contextos: los dos grupos muestran diferencias estadísticamente significativas en las cinco variables analizadas: oraciones idiomáticas y los cuatro tipos de oraciones no idiomáticas.

En este contexto, el marco teórico que sigue a esta introducción aborda el concepto de idiomática, el concepto genérico de la influencia de las lenguas de partida sobre las lenguas de llegada (interferencia y *shining through*) y el más específico de la imprimación referido a la impronta que dejan el algoritmo o los datos con los que se entrenan los propios sistemas de traducción automática en los textos traducidos (Rico, 2025). Los resultados del trabajo de campo de nuestra investigación corroboran la diferente percepción de lo idiomático en contextos diferentes. Finalmente, las conclusiones ponen de manifiesto la necesidad de profundizar en el estudio de la lengua materna como uno de los pilares de la competencia traductora, proporcionan algunas claves sobre la manera de abordar los trabajos de posesición y advierten sobre la necesidad de medir la calidad a partir de parámetros diferentes de los derivados de la detección de errores.

2. DESARROLLO

El ámbito de la atención al cliente tiene unas características peculiares que vienen determinadas por el tipo de comunicación bilateral que se da entre los llamados *agentes de atención al cliente* y los usuarios o clientes. La función de los agentes es contestar a preguntas o atender a las reclamaciones de los clientes, procesar pedidos y devoluciones y proporcionar información sobre los productos y servicios. En esta tarea, los agentes deben seguir un protocolo específico marcado por la empresa en el que se puede incluir el uso de plantillas, así como de terminología específica que impiden toda variación estilística y que pueden dar como resultado la repetición de estructuras en el proceso comunicativo. En el otro extremo de la cadena de comunicación, el contenido que generan los clientes no sigue ninguna estructura fija, es muy variado e incluye el uso de abreviaturas, emoticonos y expresiones idiomáticas, sin olvidar los errores derivados de escribir en el teclado del móvil (Gonçalves et al. 2022). Desde el punto de vista de la intencionalidad, este contenido es emocionalmente volátil y efímero. Los clientes suelen ponerse

1 El término latinoamericano es el utilizado por la industria para referirse a un hispanoamericano neutro. La distinción latam (latino o latinoamericano) frente a español europeo es simplificadora y podría considerarse poco rigurosa desde el punto de vista académico. La realidad dialectológica del español es, sin duda, mucho más compleja. Prueba de ello son las diferentes propuestas de división del continuo dialectal del español que se han realizado desde hace más de 150 años. Por lo innovador de su aproximación y por la actualidad de su análisis recomendamos al lector interesado la lectura de Moreno y Ueda (2018).

Esta definición de idiomaticidad contrasta, en cierto modo, con la concepción que se da al término en el campo de la fraseología, donde el término cobra una mayor especificidad y se concreta en diferentes definiciones (Fan 2018). Por una parte, Zuluaga (1980, p. 122) se refiere a la idiomaticidad como “el rasgo semántico propio de ciertas construcciones lingüísticas fijas, cuyo sentido no se puede establecer a partir de los significados de sus elementos componentes ni del de su combinación”. De manera complementaria, Corpas (1996, p. 26) se refiere a este concepto como “aquella propiedad semántica que presentan ciertas unidades fraseológicas, por la cual el significado global de dicha unidad no es deducible del significado aislado de cada uno de sus elementos constitutivos”. Mellado (2016) recoge esta idea y la explica en relación con la divisibilidad de la cadena fraseológica y con la autonomía semántica de los constituyentes del fraseologismo, de manera que la idiomaticidad se considera como “una magnitud gradual relacionada con la composicionalidad semántica del significado fraseológico” (Mellado 2016, p. 229). En el ámbito de la traducción automática, los estudios sobre la idiomaticidad son aún incipientes y se han ocupado de cuestiones como el trasvase de locuciones nominales del español al inglés (Ortego y Seghiri 2019), el análisis de la fraseología en el área de la propiedad intelectual e industrial (Alemán 2018), la traducción de paremias y locuciones verbales del ruso al español para la enseñanza de ELE (Ramírez 2022), la definición de métricas automáticas para la evaluación de las expresiones idiomáticas (Baziotis et al 2022) o la creación de un corpus paralelo inglés-alemán de frases idiomáticas para el entrenamiento de motores de traducción automática neuronal (Fadaee et al. 2018). En todo caso, conviene recordar que el tratamiento de las expresiones idiomáticas es aún uno de los grandes desafíos a los que se enfrenta la traducción automática, en primer lugar porque los corpus de entrenamiento no suelen contener este tipo de expresiones en una cantidad suficientemente representativa y, en segundo lugar, porque los modelos de inteligencia artificial tienen dificultades para distinguir entre una expresión idiomática y otra cuyo significado debe interpretarse de modo literal (Parmar y Estrada 2020 y Santing et al. 2022).

Nuestro trabajo orienta lo idiomático hacia el destinatario y el texto meta, a diferencia de la fidelidad, que se orienta hacia la preservación del texto original. Lo idiomático remite a lo que es más genuino, a lo más original, a lo más natural en la lengua de llegada y se refiere a cuestiones léxicas, a cuestiones sintácticas (la pasiva refleja es más genuinamente española que la pasiva) o a cuestiones pragmáticas (el español utiliza en las peticiones fórmulas directas a diferencia del inglés donde predominan las indirectas). Lo idiomático está sujeto a una cierta versatilidad condicionada por variables como la edad o las características socioculturales de los hablantes o incluso el contexto en el que se produce un enunciado, como veremos en esta investigación. Esta es una definición del concepto de idiomaticidad que está en consonancia con el planteamiento que hacen Laver y Mason (2018): “Idiomatic translation n. an approach to translating which, while not sacrificing source-text meaning, seeks to ensure naturalness of expression in the target language” (p. 62).

2.2. El efecto de las lenguas de origen sobre las lenguas de llegada

El efecto de las lenguas de origen sobre las lenguas de llegada ha sido estudiado en el ámbito de la traducción en contextos diferentes con términos distintos. En esta sección nos referiremos a este concepto, con algunas diferencias de matiz, con tres términos diferentes: interferencia (Toury 2012 y Mauranen 2004), *shining through* (Teich 2003) e imprimación (Rico en prensa).

2.2.1. La interferencia

La interferencia o la transferencia negativa es un concepto ampliamente estudiado en el ámbito de la enseñanza de segundas lenguas y también en el de la traducción. Del primero de estos ámbitos, en concreto de Weinreich (1953), viene la definición clásica de la interferencia: “those instances of deviation from the norms of either language which occur in the speech of bilinguals as a result of their familiarity with more than one language” (Weinreich 1953, p. 1). A pesar de que, de la definición anterior, podría inferirse que la interferencia se produce en cualquiera de las lenguas (*either language*), en el ámbito de la enseñanza de las segundas lenguas, tradicionalmente, se ha hablado de la interferencia de la primera lengua, la lengua materna, sobre la segunda. En el ámbito de la traducción, sin embargo, se habla más bien de la influencia de la segunda lengua, la lengua extranjera, la lengua del original, la lengua de la que se traduce, sobre la lengua de llegada, que suele ser la primera lengua del traductor. Nos interesa aquí partir de la reflexión que plantea Mauranen (2004) sobre la definición de la interferencia y su percepción.

En su artículo, Mauranen nos remite a dos publicaciones fundamentales, Baker (1993) y Toury (2012)², con informaciones que podrían considerarse un tanto contradictorias. Efectivamente, Baker (1993) define las *características universales de la traducción* en los siguientes términos: “universal features of translation, that is features with typically occur in translated texts rather than original utterances and which are not the result of interference from specific linguistic systems” (p. 243).

Como vemos Baker (1993) excluye explícitamente la interferencia de las características universales de la traducción.

Por otra parte, sobre la base de las traducciones literarias en diferentes culturas, Toury (2012) identifica dos leyes de traducción. En primer lugar, habla de la ley de estandarización creciente (*law of growing standardization*) “in translation, textual relations obtaining in the original are often modified, sometimes to the point of being totally ignored, in favour of [more] habitual options offered by the target repertoire” (p. 304). En segundo lugar, unas páginas más adelante, habla de la ley de la interferencia (*law of interference*) como una ley fundamental de la traducción “in translation, phenomena pertaining to the make-up of the source text tend to force themselves on the translators and be transferred to the target text” (p. 306).

En esta misma publicación, Toury distingue entre una transferencia positiva (incremento de la frecuencia de características que existen en la lengua de llegada y pueden usarse) o transferencia negativa (desviaciones de la norma que llegan a formar parte del sistema de la lengua). Mauranen (2004) utiliza transferencia —como equivalente a la transferencia positiva de Toury (2012)— e interferencia —como equivalente a la transferencia negativa de Toury (2012)— y sugiere situar estos términos en los dos extremos de un continuo. En un extremo tendríamos una violación flagrante de la norma en la lengua de llegada, un error de traducción, y, en el otro extremo, tendríamos un texto traducido que no podría distinguirse de un texto escrito originalmente en la lengua de llegada. La cuestión es que habría que distinguir en qué punto se rompe esa línea, cuándo la transferencia aceptable se convierte en inaceptable y, sobre todo, quién determina lo que es aceptable y lo que no lo es.

² Mauranen (2004) hacer referencia a Toury (1995), en esta sección nos referiremos a Toury (2012) que es una versión revisada de Toury (1995).



Toury (2012) sugiere que la aceptabilidad de la interferencia está determinada por la aceptación social de la lengua y la cultura de origen en la lengua de llegada “communities differ in terms of their resistance to interference, especially of the ‘negative’ type” (p. 313).

Más adelante pone en relación la aceptación social con nociones como el prestigio y el poder: “tolerance of interference –and hence the endurance of its manifestations- tend to increase when translation is carried out from ‘major’ or highly prestigious language/culture, especially if the target language/culture is ‘minor, ‘weak’ in any sense” (p. 314).

2.2.2. *SL Shining through*

Teich (2003) desarrolla una metodología para el análisis de los textos producidos originalmente en inglés y en alemán y las traducciones inglés-alemán, alemán-inglés a través de corpus comparables y corpus alineados. Parte de la base de que los textos traducidos tienen unas características específicas que los distinguen de los textos producidos originalmente en las lenguas de llegada. Entre estas características específicas recoge dos hipótesis que formula para el par de lengua objeto de su estudio: el inglés y el alemán.

Hypothesis I (formulation 2): SL shining through. In translations from English or German into German or English, respectively, the lexico-grammatical properties of the source language (SL) may shine through, both typological one and registerial ones. This contributes to making English translations (from German) and German translations (from English) different from comparable texts in the same language as the TL.

Hypothesis II (formulation 2): TL normalization. In translations from English or German or German into German or English respectively, adjustments to the properties of the TL are made. The lexico-grammatical means of the TL used may be ‘over-use’, stressing the register features and the typological features of the target language. This contributes to making English translations (from German) and German translation (from English) different from comparable texts in the same language as the TL (Teich 2003, p. 146).

Como el mismo Teich señala, estas dos hipótesis inciden en los conceptos que se ponía de manifiesto en la ley de interferencia (*law of interference*): *Hypothesis I* y a la ley de estandarización creciente (*law of growing standardization*): *Hipótesis II* (ibid., p. 145).

2.2.3. *La imprimación*

Rico (2025) acuña el término *imprimación* para referirse al efecto que causa la traducción automática, la neuronal o la basada en arquitecturas de transformadores, en las lenguas de llegada. Concretamente, Rico se refiere al efecto que tiene “el sesgo que aportan el algoritmo y los datos con los que se entrenan los propios sistemas de traducción” (p. 49). El corpus de análisis de esta investigación está conformado por textos institucionales de la Unión Europea traducidos al español que son representativos de los diferentes hitos tecnológicos en el uso de la traducción automática en la Dirección General de Traducción. Se trata de un estudio longitudinal de carácter cuantitativo que, aunque debería complementarse con otro de carácter cualitativo, podría poner de manifiesto ciertos indicios de imprimación referidos a cuatro aspectos: la diversidad léxica, la densidad léxica, el índice de la longitud del corpus (*length ratio*) y los patrones léxicos.

2.3. Diseño de la investigación y metodología de análisis

2.3.1. Corpus de análisis

El estudio parte de un corpus de 50 oraciones extraídas de comunicaciones reales, totalmente anonimizadas, de diferentes empresas a sus clientes en forma de mensajes de correo electrónico, chats o FAQ. Se trata, en todos los casos, de textos que pertenecen a géneros funcionales — siempre están ligados a un cometido concreto (solucionar una duda, pedir disculpas por algún problema, proporcionar una instrucción)— y son informativos —el objetivo es proporcionar una información—, pero también persuasivos —el objetivo último es mantener una buena relación con el cliente— finalmente, tienen una carga cultural considerable.

Estas oraciones se han traducido automáticamente del inglés al español mediante los motores de traducción automática neuronal de Unbabel, multinacional portuguesa, líder en la utilización de motores de traducción automática neuronal en el ámbito de atención al cliente, gracias a un acuerdo de colaboración con esta empresa. El corpus completo está disponible en acceso a abierto en Martín de Santa Olalla y Rico (2024). Para nuestra investigación hemos agrupado el conjunto de las oraciones del siguiente modo: oraciones idiomáticas y oraciones no idiomáticas³. En primer lugar, el grupo de oraciones idiomáticas comprende las nueve oraciones que se muestran la Tabla 1.

Tabla 1. Oraciones idiomáticas

Original en inglés	Traducción automática al español
<i>Thanks for getting in touch.</i>	Gracias por ponerte en contacto con nosotros.
<i>We're sorry to hear that you are experiencing issues when attempting to access Online features.</i>	Lamentamos que tengas problemas al intentar acceder a las funciones en línea.
<i>We will be happy to help you.</i>	Estaremos encantados de ayudarte.
<i>Open your console's home screen.</i>	Abre la pantalla de inicio de tu consola.
<i>If you have any additional questions or concerns, please do not hesitate to contact us and we will be delighted to assist you.</i>	Si tienes cualquier otra pregunta o consulta, no dudes en ponerte en contacto con nosotros y te ayudaremos con mucho gusto.
<i>I apologize for the inconvenience and look forward to your next message.</i>	Me disculpo por las molestias y espero tu próximo mensaje.
<i>When choosing a new password make sure it is unique to any another account passwords you use.</i>	Al elegir una nueva contraseña, asegúrate de que sea única y no coincida con la de ninguna otra cuenta que uses.
<i>If we can be of any further assistance, please do not hesitate to get in touch.</i>	Si podemos ayudarte en algo más, no dudes en ponerse en contacto con nosotros.
<i>We will need photos of the shipping box.</i>	Necesitaremos las fotos de la caja del envío.

³ Algunas traducciones podrían parecer a priori poco naturales fuera de contexto, pero debe tenerse en cuenta que estos resultados están dentro del estándar de la industria cuando se utiliza TA.

En segundo lugar, el grupo de oraciones no idiomáticas está compuesto por cuarenta y una oraciones traducidas con diferentes problemas relacionados todos ellos con la falta de idiomática. En este caso, hemos hecho una clasificación en cuatro subgrupos que, nos servirán de base para el análisis cuantitativo y cualitativo de una muestra.

Grupo 1. Oraciones no idiomáticas por interferencias interlingüísticas en el léxico. Entre estas, podemos ver ejemplos de cognados (*great/gran; additional/adicional; instant/instantáneas*) o calcos semánticos (*update/actualizar*).

Tabla 2. Oraciones no idiomáticas por interferencias léxicas

Original en inglés	Traducción automática al español
<i>Have a great day Lucía!</i>	¡Qué tengas un gran día, Lucía!
<i>Can I assist with anything additional at the moment?</i>	¿Puedo ayudarte con algo adicional en este momento?
<i>Usually, card transfers are instant but in rare cases, like bank holidays or weekends it gets delayed.</i>	Normalmente, las transferencias son instantáneas , pero en raras ocasiones se retrasan.
<i>I'll update you via email.</i>	Te actualizaré por correo electrónico.

Como han señalado Čulo y Nitzke (2016) o Brglez y Bintar (2022), la traducción automática muestra de manera frecuente cognados como soluciones de traducción, entendiendo como tales las palabras ortográficamente parecidas, frente a otras que son claramente más apropiadas en la lengua de llegada.

Grupo 2. Oraciones no idiomáticas por interferencias interlingüísticas en el plano morfosintáctico. Entre estas, utilización de la pasiva en lugar de la pasiva refleja, utilización del posesivo en español en contextos donde resultaría más natural utilizar el artículo o la traducción de *for you* como *por ti*.

Tabla 3. Oraciones no idiomáticas por interferencias morfosintácticas

Original en inglés	Traducción automática al español
<i>The payment was returned to the sender.</i>	El pago fue devuelto al remitente.
<i>Once we receive this information, we will be able to investigate your issue further.</i>	Una vez que recibamos esta información, podremos investigar tu problema más a fondo.
<i>So that we can get this resolved for you.</i>	Para que podamos resolver esto por ti .

Grupo 3. Oraciones no idiomáticas por interferencias interlingüísticas en el plano cultural. Tenemos, en este caso, oraciones que resultan poco naturales para un hablante de español peninsular. Este tipo de expresiones están muy relacionadas con la manera de entender la atención al cliente en el ámbito anglosajón o más específicamente en los Estados Unidos (Morate Béjar 2018).

Tabla 4. Oraciones no idiomáticas por interferencias culturales

Original en inglés	Traducción automática al español
<i>Be safe, stay healthy and take care.</i>	Estate seguro, mantente saludable y cuídate.
<i>We're so sorry your product arrived in this condition and we apologize for any disappointment caused.</i>	Lamentamos muchos que tu producto hay llegado en esa condición y nos disculpamos por cualquier decepción causada.
<i>May I ask for the order number?</i>	¿ Puedo solicitar el número de envío?
<i>I am happy to arrange for a reshipment.</i>	Estoy feliz de organizar un reenvío.
<i>I've successfully updated the shipping address.</i>	He actualizado con éxito la dirección de envío.

Grupo 4. Oraciones no idiomáticas por efecto de la interferencia intralingüística Al hablar del efecto de la interferencia latina nos referimos a la inclusión de fenómenos léxicos y morfosintácticos que son más propios de lo que la industria ha dado en llamar latam (latinoamericano) en oposición al español europeo. Entre estos fenómenos estaría el empleo de reclamo, en lugar de *reclamación*, *costo* por *coste*; el uso del pretérito indefinido en contextos en los que la variedad centro norte peninsular preferiría el pretérito perfecto compuesto, o el uso de perífrasis durativas de presente (*estés teniendo*) en lugar de formas de presente simple (*tengas*).

Tabla 5. Oraciones no idiomáticas por interferencias latinas

Original en inglés	Traducción automática al español
<i>For any limited warranty claim like a defective product to be replaced, a valid proof of purchase is needed.</i>	Para que se reemplace cualquier reclamo de garantía limitada como un producto defectuoso, se necesita un comprobante de compra válido.
<i>The refund arrived in your account today.</i>	El reembolso llegó a su cuenta hoy.
<i>I am sorry to hear you're having problems with the app.</i>	Lamento que estés teniendo problemas con la aplicación.
<i>I will arrange for a reshipment at no cost.</i>	Organizaré un reenvío sin costo .
<i>After that you tried to order another card and that is why the system did not allow you to pay again.</i>	Después de eso, has intentado pedir otra tarjeta y es por eso que el sistema no te ha permitido pagar de nuevo.

En este caso, la falta de idiomática tiene que ver con la extrañeza con la que puede percibir un español de la variedad centro norte peninsular el uso de léxico o construcciones ajenas a su variedad, del mismo modo que para un mexicano, un colombiano o un peruano resultarían extraños ciertos usos peninsulares, por diferentes a los de su variedad, sin que eso suponga ningún juicio de valor hacia una u otras variedades.

2.3.2. Participantes

Los participantes en el estudio son estudiantes del grado en Traducción e Interpretación de la Universidad Complutense de Madrid, divididos en dos grupos. El grupo 1 está integrado por treinta y cuatro estudiantes de la asignatura “Redacción y composición de textos en español”.

Esta asignatura es la segunda dedicada a la expresión escrita en español y se cursa en el primer cuatrimestre del segundo curso. En ella, se abordan, entre otros contenidos, la corrección gramatical y estilística. Los estudiantes están habituados a corregir textos, pero escritos siempre en español, no traducidos al español. En este momento de sus estudios todavía no han cursado ninguna asignatura de traducción.

El grupo 2 está compuesto por treinta y siete estudiantes de la asignatura “Tecnologías de la traducción asistida por ordenador”, que se cursa en el primer cuatrimestre del tercer año del grado. Esta asignatura es la continuación de otras dos dedicadas a la tecnología, “Tecnologías aplicadas a la traducción y la interpretación I”, que se imparte en primer curso y “Tecnologías aplicadas a la traducción y la interpretación II”, que se imparte en segundo curso. Los estudiantes de este segundo grupo están familiarizados con la traducción automática desde el punto de vista de un usuario avanzado, conocen cómo funcionan los motores de traducción neuronal, han hecho prácticas de evaluación manual y automática y han recibido formación teórica y práctica sobre la posesición.

2.3.3. Instrumentos para la recogida de datos

Para el proceso de recogida de datos se elaboró un cuestionario con dos preguntas aplicadas a las 50 oraciones del corpus. La primera pregunta, “Valora de 1 a 5 la naturalidad de la siguiente oración, siendo 1=nada natural y 5=muy natural”, es la que sirvió de base al estudio cuantitativo. La segunda, “Escribe otra oración que, a tu juicio, resulte más natural” es la que utilizamos en la parte cualitativa de nuestro estudio. Como información contextual, se indicó que la valoración y propuesta de corrección debería basarse en la variedad de español centro norte peninsular y que se utilizaría la segunda persona del singular, variante informal, tú.

En el grupo 1, conformado por estudiantes de la asignatura «Redacción y composición de textos en español», el cuestionario se diseñó como simulación de una tarea real, con la instrucción de revisar las comunicaciones que los agentes del departamento de Atención al Cliente de una empresa multinacional dan al usuario a través de las respuestas en forma de correo electrónico, chats o FAQ. En esta tarea, los participantes solo tenían acceso a las oraciones traducidas automáticamente al español (es decir, no se les dio la oportunidad de ver las oraciones originales en inglés).

En el grupo 2, integrado por estudiantes de la asignatura «Tecnología de la traducción asistida por ordenador», la tarea se planteó en un contexto de posesición con acceso al texto en inglés.

2.3.4. Metodología de análisis

Nuestro estudio recoge datos cuantitativos y cualitativos. En el primer caso, llevamos a cabo tres tipos de análisis con el software de tratamiento estadístico de datos SAS 9.5⁴. A continuación, indicamos el detalle del tipo de análisis realizado:

⁴ La parte estadística de este estudio se llevó a cabo con la inestimable colaboración de Ricardo García, de los Servicios Informáticos de la UCM, a quien agradecemos su tiempo y su ayuda.

- Análisis descriptivo básico con dos tipos de resultados:
 - a. Resultados de las variables categóricas originales, esto es, las que recogen la composición de los dos grupos de estudiantes y las puntuaciones sobre su percepción de la idiomática, de 1 a 5, para cada oración.
 - b. Resultados de las variables numéricas calculadas, es decir, de las puntuaciones de las variables identificadas, a saber: puntuación de las oraciones con interferencias léxicas, puntuación de las oraciones con interferencias morfosintácticas, puntuación de las oraciones con interferencias culturales, puntuación de las oraciones con interferencias latinoamericanas. Para cada una de estas variables se proporciona la siguiente información: muestra útil, en nuestro caso, número de estudiantes (*N*), media (*Mean*), desviación estándar (*StdDev*), valor mínimo (*Min*), mediana (*Median*), cuartiles inferior y superior (*Lower/Upper Quartile*) y valor máximo (*Max*).
- Análisis comparativo de las percepciones de idiomática entre los dos grupos de estudiantes para las cuatro variables descritas en el apartado anterior. A estos resultados se les aplica un *t-test* (o test de la *t* de *Student*) con el fin de conocer si las diferencias son estadísticamente significativas. Este resultado es el que fundamenta las principales conclusiones de nuestro análisis.
- Análisis descriptivo, en forma de tabla, en la que se muestra el perfil de respuesta de cada grupo de estudiantes para cada una de las cincuenta oraciones que componen el corpus. En este estudio también se analiza el resultado de la aplicación del test no paramétrico de la Suma de Rangos de Wilcoxon que nos permitirá ver si las diferencias entre los dos grupos, para cada una de las oraciones, son (o no) significativas.

La recogida de datos se llevó a cabo en dos sesiones consecutivas los días 24 de noviembre de 2022 (grupo 1 de estudiantes) y 29 de noviembre de 2022 (grupo 2 de estudiantes). Cada grupo completó los cuestionarios durante el tiempo de clase de las asignaturas (una hora y cuarenta minutos). Se recogieron un total de 34 respuestas en el grupo 1 y 37 respuestas en el grupo 2.

2.3.5. Resultados y discusión

En este apartado presentamos los resultados. Abordamos, en primer lugar, las respuestas a la primera pregunta sobre la percepción del grado de naturalidad y, en segundo lugar, las respuestas a la segunda pregunta, de carácter abierto en la que se daba a los estudiantes la posibilidad de registrar una opción mejor a la traducción automática.

2.3.5.1. Resultados sobre la percepción de la naturalidad (pregunta 1)

En la siguiente tabla hemos recogido las respuestas de cada uno de los dos grupos de estudiantes con respecto a la percepción de la naturalidad de cada tipo de oración: las idiomáticas, las no idiomáticas interferencias léxicas, las no idiomáticas por interferencias morfosintácticas, las no idiomáticas por interferencias culturales y las no idiomáticas por interferencias latinoamericanas.



Tabla 6. Resultados de la pregunta 1: percepción de la naturalidad

GRUPO	Variable	N	Media	DT	Mín.	Cuartil inferior	Mediana	Cuartil superior	Máx.
Estudiantes grupo 1	percepción de la naturalidad de las oraciones idiomáticas	34	4,05	0,55	2,43	3,75	4,22	4,44	4,78
	percepción de la naturalidad en las oraciones con inteferencias léxicas.	34	3,08	0,67	1,67	2,58	3,19	3,45	4,77
	percepción de la naturalidad en las oraciones con interferencias morfosintácticas.	34	3,61	0,62	2,75	3,20	3,50	4,00	5,00
	percepción de la naturalidad en las oraciones con interferencias culturales.	34	3,00	0,58	2,00	2,54	3,00	3,38	4,69
	percepción de la naturalidad en las oraciones con interferencias latinas.	34	3,15	0,56	1,86	2,80	3,16	3,60	4,00
Estudiantes grupo 2	percepción de la naturalidad de las oraciones idiomáticas	37	4,36	0,36	3,33	4,11	4,44	4,56	5,00
	percepción de la naturalidad en las oraciones con inteferencias léxicas.	37	3,61	0,45	2,73	3,27	3,64	3,91	4,59
	percepción de la naturalidad en las oraciones con interferencias morfosintácticas.	37	3,99	0,48	3,00	3,60	4,00	4,40	4,60
	percepción de la naturalidad en las oraciones con interferencias culturales.	37	3,54	0,56	2,38	3,15	3,62	4,92	4,94
	percepción de la naturalidad en las oraciones con interferencias latinas.	37	3,58	0,46	2,63	3,30	3,50	3,78	4,78

En la tabla, la columna *grupo* proporciona información sobre el grupo de estudiantes al que se refieren las respuestas: grupo de estudiantes 1 y grupo de estudiantes 2. Las respuestas

se refieren a los cinco conjuntos de oraciones, recogidas en la columna *variable*. Las variables que se analizan son puntuaciones promedio por grupos de oraciones: las idiomáticas, las no idiomáticas por interferencias léxicas, las no idiomáticas por interferencias morfosintácticas, las no idiomáticas por interferencias culturales, las no idiomáticas por interferencias latinas. El resto de las columnas muestran los siguientes datos (Bryman 2016, pp. 338-339):

- N: número total de respuestas obtenidas para cada tipo de pregunta.
- media: media aritmética de las respuestas recibidas, esto es, la suma de todos los valores dividida entre el número total de valores.
- DT: desviación estándar o variación en la que las respuestas individuales difieren de la media.
- mín.: puntuación mínima recibida.
- cuartil inferior: valor de la variable tal que la cuarta parte (25 %) de las observaciones son inferiores o iguales a él.
- mediana: punto medio en la distribución de las puntuaciones.
- cuartil superior: valor tal que las tres cuartas partes de las observaciones son inferiores o iguales a él.
- máx.: puntuación máxima recibida.

En la tabla hemos marcado en **negrita** los resultados de la media y la mediana y, en el siguiente gráfico, recogemos estos mismos valores de manera visual.

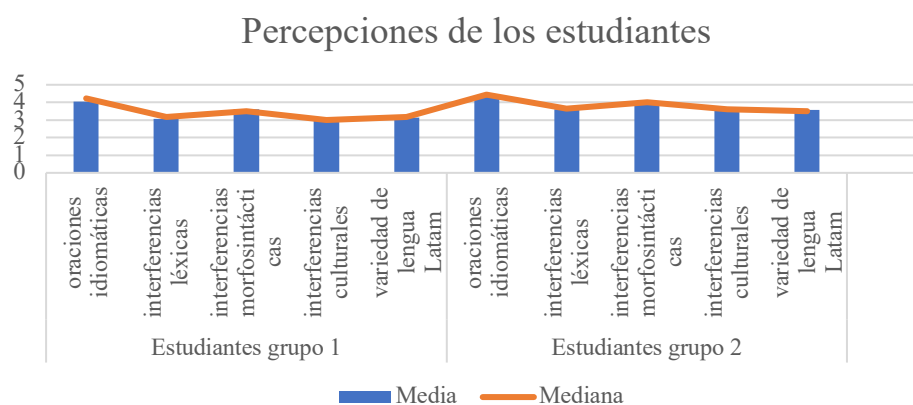


Gráfico 1. Percepción de los estudiantes: media y mediana

Como se infiere de la tabla 1, la puntuación de la naturalidad es mayor, en todas las variables, en el grupo 2. Con el objetivo de analizar si las diferencias que se observan en la descriptiva pueden considerarse estadísticamente significativas, aplicamos un *t-test* (o test de la *t de Student*) a cada una de las cinco variables. En esta prueba, para un nivel de confianza superior al 95 % se deben obtener $p\text{-valores} < \text{Alpha} = 0,05$ (Bhattacharyya y Johnson, 1977). Como se puede apreciar en los resultados de la tabla 2, los valores $p < 0,01$, lo que hace que las diferencias observadas entre los dos grupos pueden considerarse significativas al 99 %.

Tabla 7. Resultados de la aplicación del t-test

VARIABLES	p-valor
Puntuación idiomáticas	<.0,0073
Puntuación de las oraciones no idiomáticas por interferencias léxicas	<.0,0003
Puntuación de las oraciones no idiomáticas por interferencias morfosintácticas	<.0,0062
Puntuación de las oraciones no idiomáticas por interfeferencias culturales	<.0,0002
Puntuación de las oraciones no idiomáticas por interferencias latinas	<.0,0008

El primer análisis de estos datos nos hace pensar que el acceso al texto en inglés inclina a los estudiantes hacia una valoración más positiva de los textos traducidos al español. Esto vendría a confirmar la excesiva confianza de los estudiantes en los resultados de la traducción automática, conclusión a la que se llega también en otros estudios anteriores de carácter experimental a partir de la traducción automática neuronal (véase, por ejemplo, Rico y González (2022), Loock et al (2021)) o estadística (Jolley y Maimone 2015).

2.3.5.2. Resultados sobre la posibilidad de registrar una opción mejor a la de la traducción automática (pregunta 2)

Como apuntamos en el apartado 2.3.3., el cuestionario para la recogida de datos contenía una segunda pregunta en la que se les pedía a los estudiantes que escribiesen otra versión de la oración en español que, a su juicio, resultase más natural. Esta propuesta de corrección debía estar basada en la variedad de español centro norte peninsular y debía utilizar la segunda persona del singular, variante informal tú, para dirigirse al usuario. La tarea tiene relación con la labor de revisión de textos traducidos o, en nuestro caso, sobre la revisión de textos traducidos automáticamente o posesición. Nos interesaba evaluar el impacto que puede tener contar (o no) con el texto original y medir el desempeño de nuestros estudiantes en una tarea —la revisión de textos traducidos— que entraña un alto nivel de dificultad.

Para presentar las respuestas de los estudiantes a esta pregunta, nos centramos aquí en los fenómenos de no idiomaticidad, con la selección de un conjunto de diez oraciones que ilustran los siguientes fenómenos:

- Falta de idiomaticidad por interferencias léxicas.
- Falta de idiomaticidad por interferencias morfosintácticas.
- Falta de idiomaticidad por interferencias culturales.
- Falta de idiomaticidad por interferencias latinas.

Para cada conjunto de estas oraciones, se mostrará:

- El p-valor, resultado de la aplicación del test no paramétrico de la Suma de Rangos de Wilcoxon⁵. Cuanto menor sea el *p-valor* más significativa será la diferencia en los resultados del grupo 1 y del grupo 2.

⁵ Recordemos que este valor nos permite determinar si las diferencias entre los dos grupos de estudiantes son o no significativas.

- El porcentaje de estudiantes que detectan de manera correcta el problema que origina la falta de naturalidad y que, además, proporcionan una oración más natural (columna “desempeño”).
- El porcentaje de estudiantes que sobrecorrige, entendiendo como sobrecorrección el cambio de palabras o estructuras que no plantean ningún problema en la oración original, así como los cambios de estilo innecesarios que, en ocasiones, suponen un cambio sustancial de significado en la oración original (columna “sobrecorrección”).

A continuación, presentamos los resultados agrupados por tipo de oración.

2.3.5.2.1. Falta de idiomática por interferencias léxicas: cognados y calcos semánticos

Tabla 8. Análisis de la oración 1⁶

Núm. oración	Oración en inglés	¿Te gustaría agregar otro producto a tu pedido existente?		
	Traducción automática al español	Desempeño Grupo 1	Desempeño Grupo 2	Sobrecorrección Grupo 2
1	<i>Would you like to add another product to your existing order?</i>			
p-valor		Desempeño Grupo 1	Desempeño Grupo 2	Sobrecorrección Grupo 1
0,0003		47,05 %	37,83 %	17,64 %
			Sobrecorrección Grupo 2	2,7 %

En este primer caso (oración 1), el p-valor es de 0,0003 lo que nos indica que la variación en la percepción de la naturalidad es muy diferente en los dos grupos, hecho que se confirma en los gráficos 2A y 2B.

33 respuestas
Frequency Missing=1

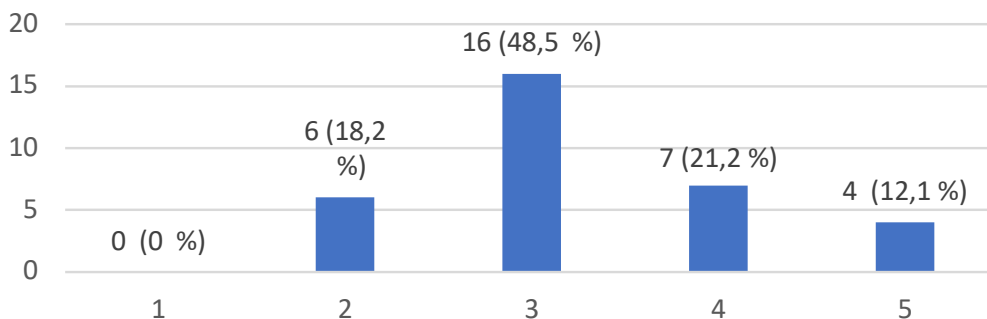


Gráfico 2A Oración 1. Valoración de la naturalidad en el grupo 1

6 El número de oración se refiere a lugar que ocupa en el corpus de estudio.

37 respuestas

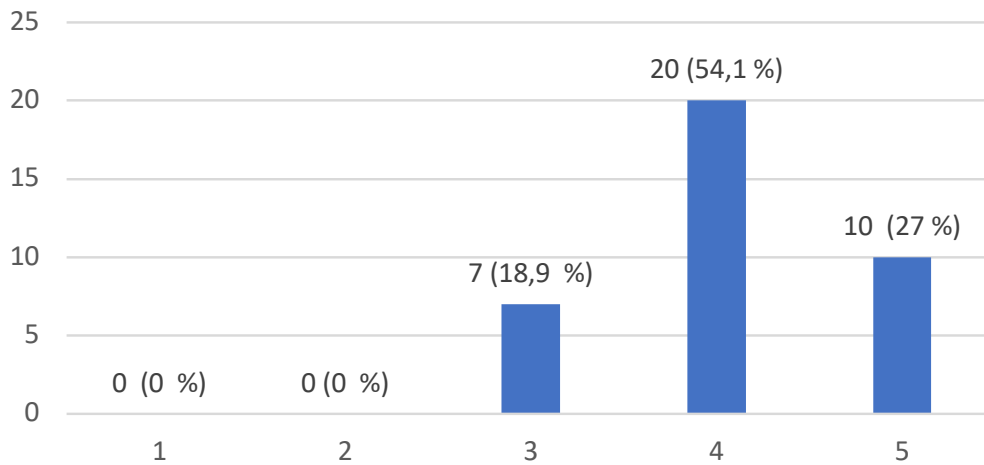


Gráfico 2B. Oración 1. valoración de la naturalidad en el grupo 2

En este caso, la traducción de *existing* como *existente* (cognado) es lo que origina la falta de idiomática. Para evitar este problema, podríamos considerar hasta tres soluciones que podrían resultar más idiomáticas que esta. La primera consistiría en no traducir *existente* (¿Te gustaría agregar otro producto a tu pedido?). Otra posibilidad sería traducir *existing* por *actual* (¿Te gustaría agregar otro producto a tu pedido actual?). Y, la tercera, sería traducir *existing* por *en curso* (¿Te gustaría agregar otro producto a tu pedido en curso?).

Un análisis detallado de las soluciones aportadas por los estudiantes muestra que hay algunas soluciones que podríamos considerar de sobrecorrección que se dan en los dos grupos. Entre estas, la sustitución de *pedido* por *cesta* o *lista* y el cambio de la segunda persona variante informal (*tú*) por la variante formal (*usted*). Recordemos que en las instrucciones se les había pedido explícitamente que utilizarán la variante informal. Como se puede ver en la tabla 3, los estudiantes del grupo 1 sobrecorrigieron más. Además de los cambios ya señalados, cambian *te gustaría* por *deseas* o cambian *otro producto* por *algún otro producto*.

Veamos el siguiente ejemplo.

Tabla 9. Análisis de la oración 29

Núm. oración 29	Oración en inglés	<i>Unfortunately, if you don't have a document like this, then it doesn't seem we have a strong enough case to proceed.</i>		
	Traducción automática al español	Por desgracia, si no tienes un documento como este, no parece que tengamos un caso lo suficientemente fuerte para proceder.		
p-valor	Desempeño Grupo 1	Desempeño Grupo 2	Sobrecorrección Grupo 1	Sobrecorrección Grupo 2
0,0005	0 %	8,8 %	20,58 %	13,51 %

En este segundo caso (oración 29), el p-valor es de 0,0005 lo que nos indica que la variación en la percepción de la naturalidad es muy diferente en los dos grupos, hecho que se confirma en los gráficos 3A y 3B. En el grupo 1, la mayor parte de las respuestas (el 59,4 %) se sitúan entre el 2 y el 3. Mientras que, en el grupo 2, las valoraciones de 3 y 4 son las preferidas para el 80,6 % de las respuestas.

32 respuestas

Frequency Missing=2

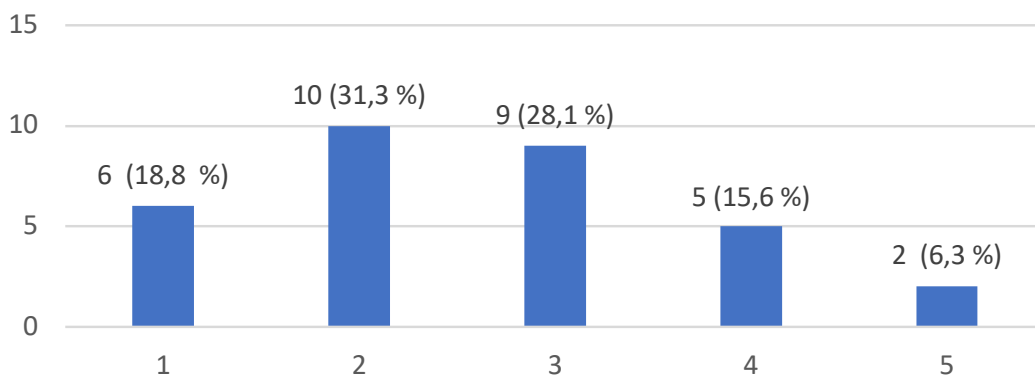


Gráfico 3A Oración 29. Valoración de la naturalidad en el grupo 1

36 respuestas

Frequency Missing=1

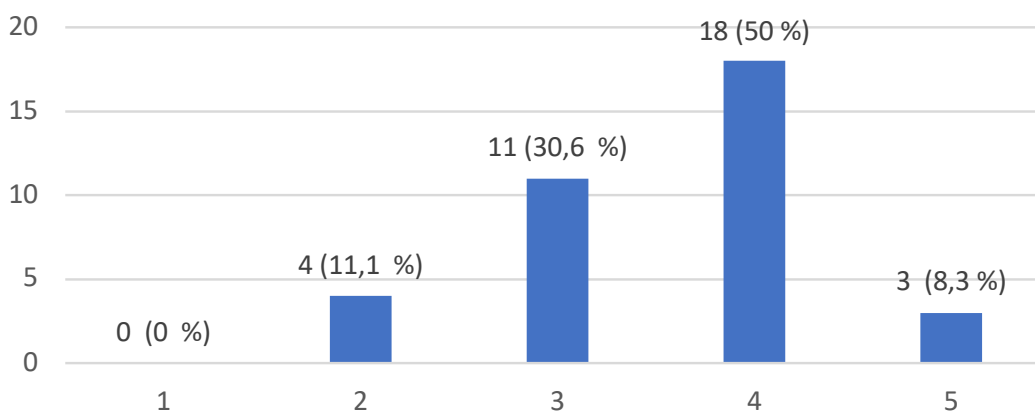


Gráfico 3B. Oración 29. Valoración de la naturalidad en el grupo 2

En esta oración nos encontramos con dos palabras que podrían originar la falta de naturalidad. La primera de ellas es *Unfortunately*, traducido como *por desgracia* y *strong* traducido por *fuerte*. En la traducción de *unfortunately* como *por desgracia* vemos una interferencia de carácter cultural relacionada

con la manera de entender la atención al cliente en el ámbito anglosajón y, más específicamente, en el estadounidense. La traducción de *strong* por fuerte en este contexto nos sitúa ante un caso de calco semántico. Teniendo en cuenta todo esto podríamos considerar traducciones más naturales: *Lamentablemente/desafortunadamente, si no dispones de un documento como este, no parece que tengamos argumentos suficientes/no es un caso lo suficientemente grave/importante, como para proceder.*

En los datos sobre el desempeño hemos considerado como correctas solo las versiones que solucionan los dos problemas que tiene la oración. Sobre el grupo 1, es remarcable el hecho de que, al no estar expuestos al texto en inglés, se aportan soluciones más libres que podrían considerarse adecuadas en este contexto que hemos definido como informal (*Si no dispones de este documento, no podemos ayudarte / Si no dispones de este documento, no podemos hacer nada*). En general, en los dos grupos no se detecta la falta de naturalidad que introduce *por desgracia*. Se ve más el problema de *fuerte* y se aportan soluciones más o menos acertadas (*sólido/importante*).

Veamos el siguiente ejemplo.

Tabla 10. Análisis de la oración 18

Núm. oración	Oración en inglés	<i>I'll update you via email.</i>		
	Traducción automática al español	Te actualizaré por correo electrónico.		
p-valor	Desempeño Grupo 1	Desempeño Grupo 2	Sobrecorrección Grupo 1	Sobrecorrección Grupo 2
0,0278	2,28 %	1,4 %	5,4 %	3,7 %

En este caso, el p-valor (0,0278) nos muestra que sí hay ciertas diferencias entre las valoraciones de la idiomaticidad en los dos grupos. En los gráficos 4A y 4B, podemos ver que, en el grupo 1, el 72,7 % de las respuestas se concentran en los valores 1, 2, 3. En el grupo 2, sin embargo, el 80 % de las respuestas están en la segunda mitad de la escala 3, 4 y 5.

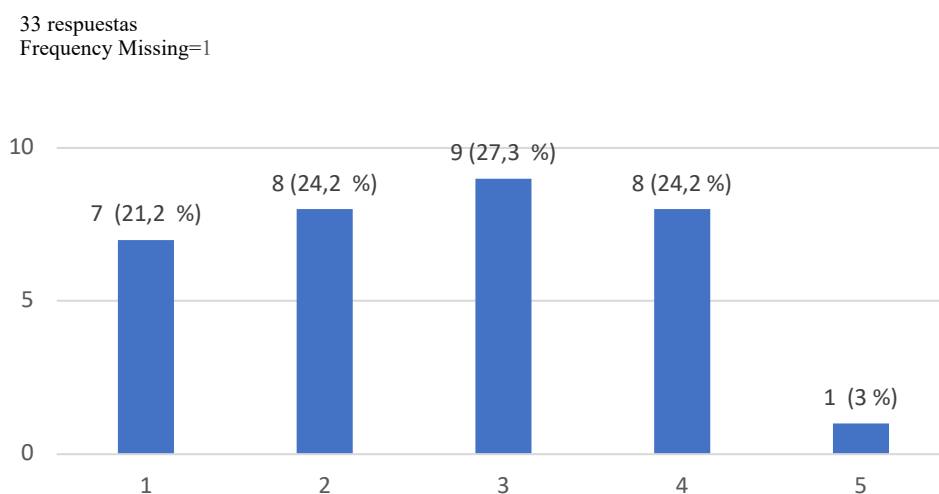


Gráfico 4A. Oración 18. Valoración de la naturalidad en el grupo 1

35 respuestas
Frequency Missing=2

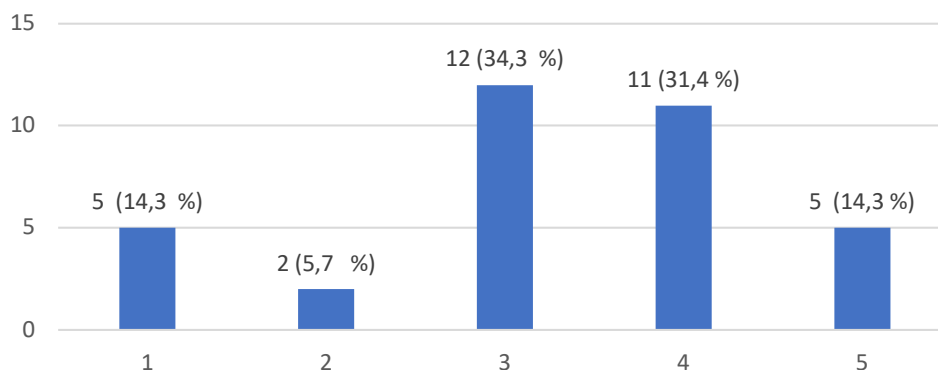


Gráfico 4B. Valoración de la naturalidad en el grupo 2

La traducción de *update* como *actualizar* (calco semántico) es la que origina la falta de idiomática. La solución que podría resultar más natural, en este caso, sería *Te mantendré informado por correo electrónico*. Los ejemplos de sobrecorrección, más numerosos en el grupo 1, tienen que ver con la adición de palabras que podrían ayudar a entender el mensaje (*Te actualizaré la situación por correo electrónico / Te mantendré actualizado por correo electrónico / Contactaremos contigo a través del correo electrónico para ponerte al día de las posibles nuevas actualizaciones*) o la sustitución de otras que podrían resultar innecesarias por el contexto (*Te informaremos de cualquier actualización*).

2.3.5.2.2. Falta de idiomática por interferencias morfosintácticas

Tabla 11. Análisis de la oración 23

Núm. oración 23	Oración en inglés	<i>So that we can get this resolved for you.</i>		
	Traducción automática al español	Para que podamos resolver esto por ti.		
p-valor	Desempeño Grupo 1	Desempeño Grupo 2	Sobrecorrección Grupo 1	Sobrecorrección Grupo 2
0,274	0 %	21,62 %	14,7%	0 %

En este caso, el p-valor (0,274) nos muestra que la valoración de la idiomática es parecida en los dos grupos. Tanto en el grupo 1 como en el grupo 2 la mayor parte de las valoraciones (en torno al 75 %) se sitúan entre 4 y 5.

32 respuestas
Frequency Missing=2

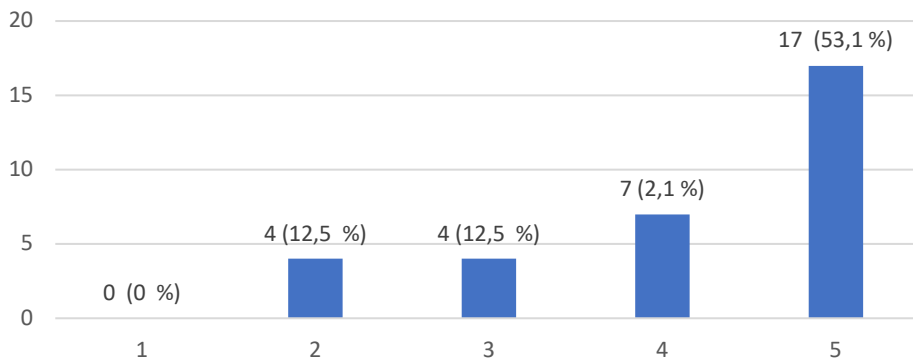


Gráfico 5A. Valoración de la naturalidad en el grupo 1

36 respuestas
Frequency Missing=2

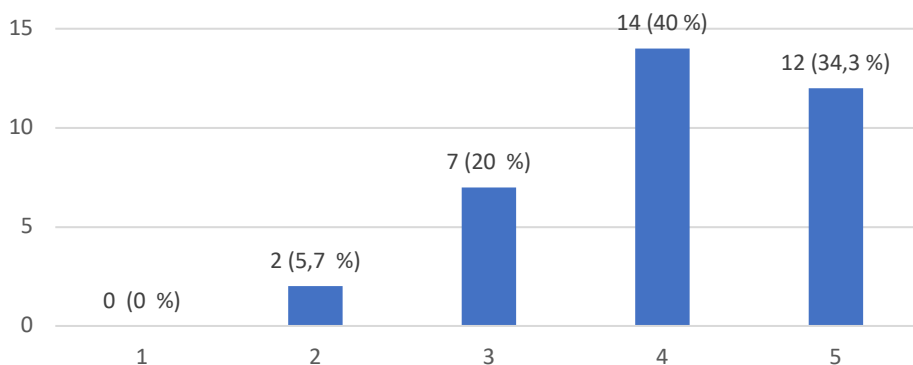


Gráfico 5B. Valoración de la naturalidad en el grupo 2

El problema que origina la falta de idiomática está relacionado con la traducción de *for you* de manera literal *por ti*. En español, lo idiomático, incluso lo correcto en este caso, sería interpretar el *for you* como un objeto indirecto y traducirlo como un enclítico *resolverte*. En este caso, se puede hablar incluso de un cambio de significado “resolver algo por alguien” es resolverlo porque ese alguien es una persona especial mientras que “resolverte” tendría un significado más parecido a “resolverte esto para ti”.

En este caso, el grupo 2, el que tiene acceso al texto en inglés, aporta más soluciones válidas, se identifica más el origen del error y se sustituye la preposición *por* por la preposición *para*. La sobrecorrección del grupo 1 está relacionada con el cambio de *resolver* por *solucionar* y *esto* por *problema*. Cuestiones que nada tienen que ver con el verdadero problema de esta oración.

Veamos el siguiente ejemplo.

Tabla 12. Análisis de la oración 43

Núm. oración 43	Oración en inglés	<i>The payment was returned to the sender.</i>		
	Traducción automática al español	El pago fue devuelto al remitente.		
p-valor	Desempeño Grupo 1	Desempeño Grupo 2	Sobrecorrección Grupo 1	Sobrecorrección Grupo 2
0,0151	20,58 %	8,82 %	8,82%	5,88 %

En este caso, el *p-valor* de 0,0151 nos muestra que sí hay ciertas diferencias en la valoración de la naturalidad de los dos grupos. Como se puede apreciar en los gráficos 6A y 6B. En el grupo 2, prácticamente todas las respuestas están situadas entre el 4 y el 5. En el grupo 1, las respuestas están más repartidas entre el 2 y 5, aunque más de la mitad de las puntuaciones están entre el 4 y 5.

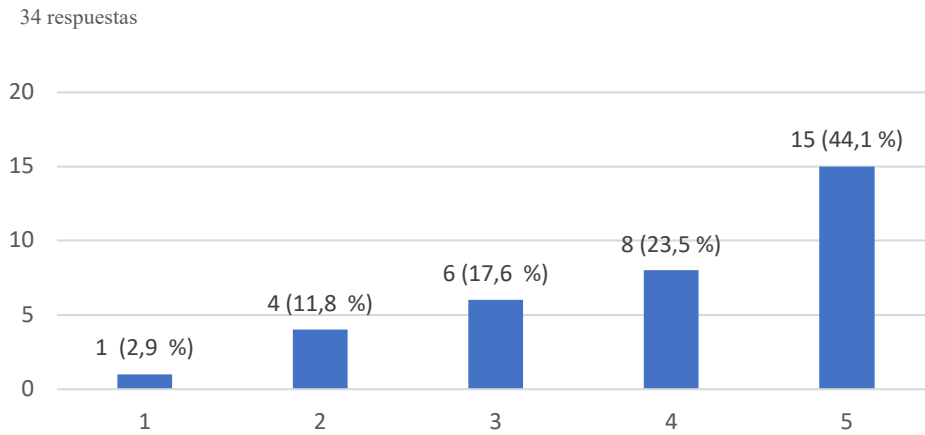


Gráfico 6A. Valoración de la naturalidad en el grupo 1

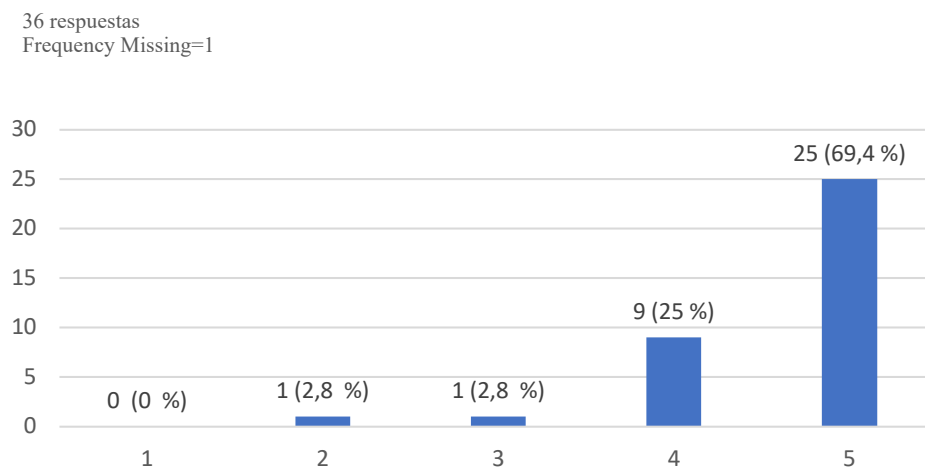


Gráfico 6B. Valoración de la naturalidad en el grupo 2

En esta oración, es la traducción de la pasiva en inglés (*was returned*) por la pasiva en español (*fue devuelto*) lo que causa la falta de idiomática. La solución más idiomática pasaría por la traducción de la pasiva inglesa por una pasiva refleja en español (*El pago se devolvió*). El problema se identifica y corrige de manera significativamente mejor en el grupo 1 que en el grupo 2 (20,58 % de respuestas correctas frente a 8,82 %).

Por otra parte, los casos de sobrecorrección están relacionados con la sustitución del sustantivo *remitente* por el sustantivo *comprador* o *cliente* en el grupo 1 o por la sustitución del pretérito indefinido (*fue devuelto*) por el pretérito perfecto compuesto (*ha sido devuelto*) o por el cambio de *devolver* por *reembolsar* y de nuevo también del pretérito indefinido por el pretérito perfecto compuesto (*El pago ha sido reembolsado*) en el grupo 2.

2.3.5.2.3. Falta de idiomática por interferencias culturales

Tabla 13. Análisis de la oración 14

Núm. oración	Oración en inglés	<i>We're so sorry your product arrived in this condition and we apologize for any disappointment caused.</i>		
	Traducción automática al español	Lamentamos muchos que tu producto hay llegado en esa condición y nos disculpamos por cualquier decepción causada.		
14				
p-valor	Desempeño Grupo 1	Desempeño Grupo 2	Sobrecorrección Grupo 1	Sobrecorrección Grupo 2
0,158	0,30 %	0 %	0,30 %	1,1 %

En este caso, el p-valor (0,158) nos muestra que la valoración de la idiomática tiene ciertas diferencias entre los dos grupos. En el grupo 1, las respuestas se distribuyen entre el 1 y el 5 mientras que el grupo 2 se distribuyen entre el 2 y el 5, con una concentración del 75 % en los dos puntos más altos de la escala.

33 respuestas
Frequency Missing=1

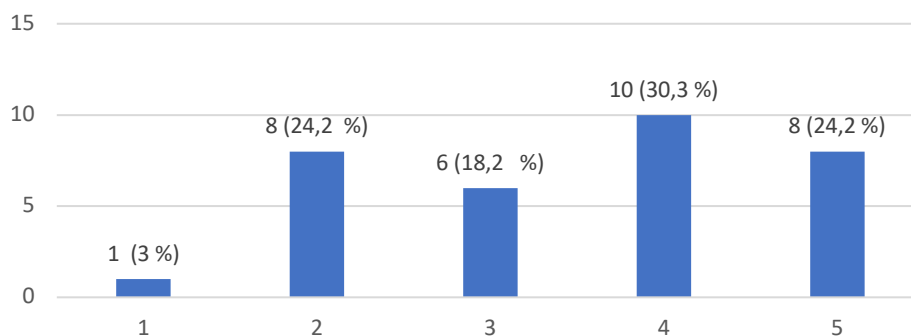


Gráfico 7A. Valoración de la naturalidad en el grupo 1

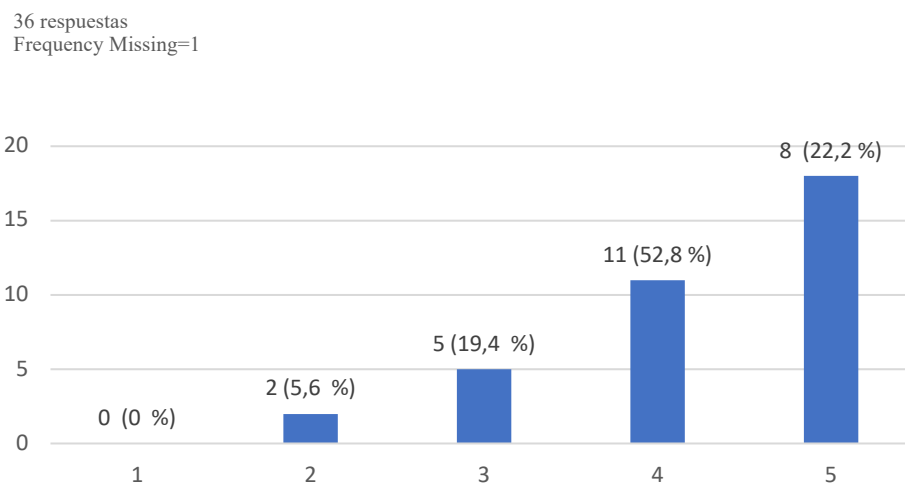


Gráfico 7B. Valoración de la naturalidad en el grupo 2

En esta oración, la traducción de *condition* por *condición* y de *disappointment* por *decepción* originan la falta de idiomática. En el primer caso, como hemos visto en otras ocasiones, la traducción automática opta por una palabra en español con similitud ortográfica con la palabra en inglés. La solución más idiomática sería traducir *condition* por *estado*. El problema que causa la falta de idiomática está en la palabra *dissappointment* que literalmente significa *decepción*. La cuestión es si en España hablaríamos en un caso como este (el producto no llega en buen estado) de *decepción*. Seguramente, en España nos referiríamos a este hecho como *inconveniente* o *problema*.

Si consideramos como soluciones válidas las que corrigen estas dos cuestiones, solo un estudiante del grupo 1 aporta una solución válida. La revisión de las soluciones propuesta muestra que, en general, los estudiantes perciben de manera más clara el problema de la traducción literal de *disappointment* por *decepción*.

En los casos de sobrecorrección se sustituye *nos disculpamos* por *pedimos perdón* o *perdón de antemano* o *lamentamos* por *sentimos*. Tanto los porcentajes de desempeño como los de sobrecorrección son bajos porque, en la mayor parte de los casos, se corrige solo el segundo de los problemas. Es decir, la mayor parte de las propuestas identifican el origen de la falta de idiomática, pero solo de manera parcial.

Veamos el siguiente ejemplo.

Tabla 14. Análisis de la oración 49

49	Núm. oración	Oración en inglés			<i>Be safe, healthy and take care.</i>
		Traducción automática al español			Estate seguro, mantente saludable y cuídate.
	p-valor	Desempeño Grupo 1	Desempeño Grupo 2	Sobrecorrección Grupo 1	Sobrecorrección Grupo 2
	0,0011	23,53 %	10,81 %	35,29 %	32,43 %

El p-valor en ese caso arroja un valor muy bajo (0,0011), lo que se identifica con diferencias significativas en la percepción de la idiomaticidad. Los gráficos 8A y 8B nos muestran que, efectivamente, en el grupo 1, la mitad de los estudiantes dan a esta oración una puntuación de 1 en idiomaticidad. En el grupo 2, sin embargo, solo el 28 % de los estudiantes da una puntuación baja a esta oración, lo que, de nuevo, está relacionado con el hecho de que los estudiantes del grupo 2 tienen acceso al original en inglés.

34 respuestas

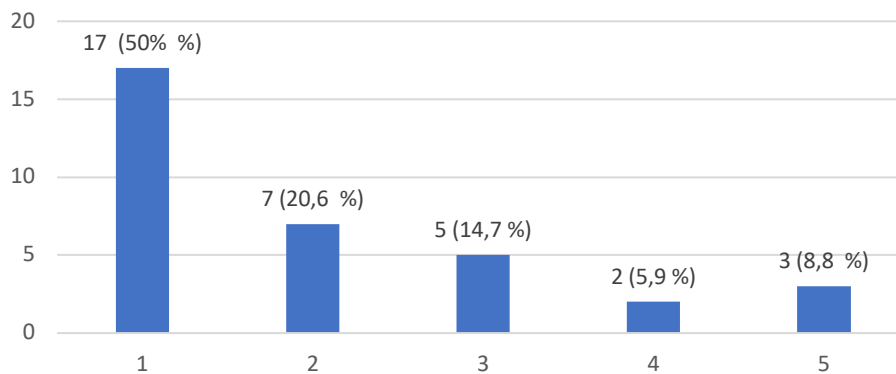


Gráfico 8A. Valoración de la naturalidad en el grupo 1

35 respuestas
Frequency Missing=2

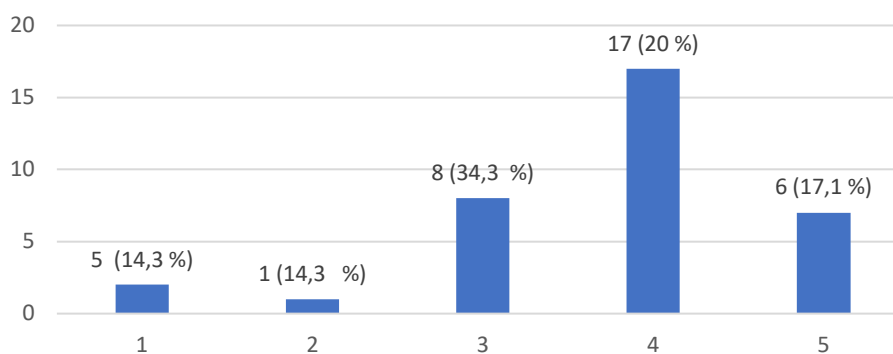


Gráfico 8B. Valoración de la naturalidad en el grupo 2

La falta de idiomaticidad en esta oración está relacionada con la utilización de una fórmula de despedida en inglés cuya traducción literal en español resulta totalmente antinatural. El español resolvería esta alusión a la salud con fórmulas como *cuídate* o *cuídate mucho* que se hicieron habituales en el contexto de la pandemia.

La valoración del desempeño de nuestros estudiantes muestra un porcentaje más alto de respuestas adecuadas en el grupo 1. La ausencia de la oración en inglés hace que se sientan más libres para proporcionar soluciones que no tienen en cuenta el enunciado completo en esta lengua. En este caso, resulta muy llamativa la sobrecorrección que se da en un porcentaje alto tanto en el grupo 1 como en el grupo 2. *Asegúrate, mantén una vida saludable y cuídate, Siéntete seguro, mantente saludables y cuídate, Estate seguro, mantente en forma y cuídate* (en el grupo 1) o *Cuídate y salud, Vete con cuidado, mantente sano y cuídate, Protégete, mantén un estilo de vida saludables y cuídate* (en el grupo 2). Finalmente, entre las soluciones del grupo 1, encontramos oraciones que pierden totalmente la alusión a la salud del original y recurren a las fórmulas de despedida habituales en español (*Un saludo o Muchas gracias, que pases un buen día*).

2.3.5.2.4. Falta de idiomática por interferencias latinoamericanas

Tabla 15. Análisis de la oración 45

Núm. Oración 45	Oración en inglés	<i>I am sorry to hear you're having trouble with the app!</i>		
	Traducción automática al español	Lamento que estés teniendo problemas con la aplicación.		
p-valor	Desempeño Grupo 1	Desempeño Grupo 2	Sobrecorrección Grupo 1	Sobrecorrección Grupo 2
0,5629	2,9 %	0 %	1,25 %	0 %

En este caso, el p-valor (0,5629) nos muestra que la valoración de la idiomática es parecida en los dos grupos. Tanto en el grupo 1 como en el grupo 2 la mayor parte de las valoraciones (87,6 % en el grupo 1 y 97,3 % en el grupo 2) se sitúan entre 4 y 5.

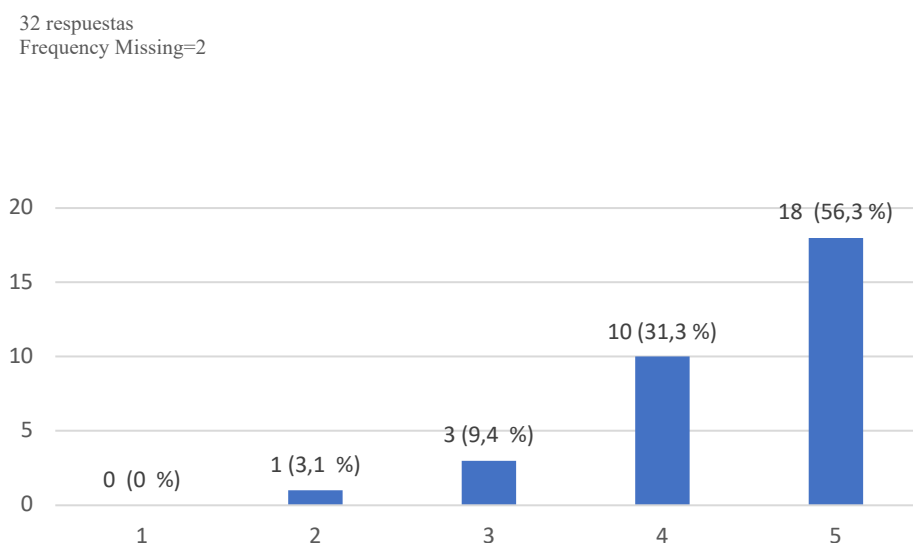


Gráfico 9A. Valoración de la naturalidad en el grupo 1

37 respuestas

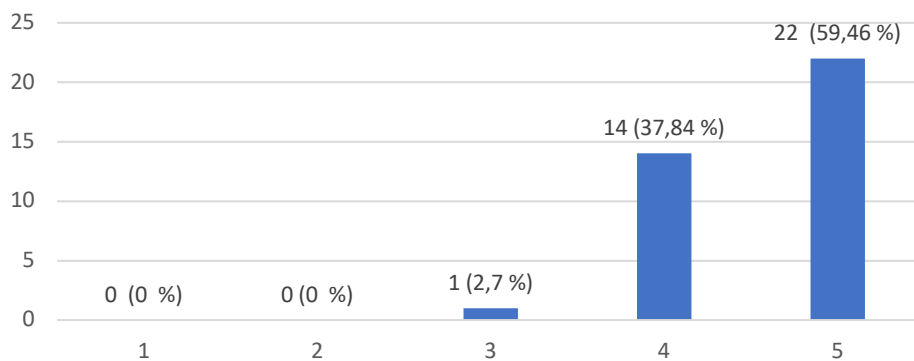


Gráfico 9B. Valoración de la naturalidad en el grupo 2

La falta de idiomática, en este caso, está relacionada con el empleo la perífrasis *estés teniendo* en lugar de *tengas*. Como han señalado García (2012), en América, al aspecto durativo de esta perífrasis se ha sumado otro de carácter puntual que es el que se pone de manifiesto en este ejemplo en el que el español europeo preferiría una forma de presente, en este caso, de subjuntivo (*Lamento que tengas problemas con la aplicación*).

Los bajos porcentajes de desempeño, en los dos grupos, están motivados por el hecho de que, en ambos, la oración se percibe como bastante natural. La sobrecorrección, que se da solo en el grupo 1, tiene que ver con el añadido de palabras que no aportan nada a la solución proporcionada por la traducción automática (*Lamentamos oír que estés teniendo problemas...*) o cambios de orden y de modo en el verbo (*Lamento los problemas que estás teniendo...*) o sustitución de *lamentar* por *sentir*, que ya hemos visto en el análisis de la oración 14, o de *problemas* por *dificultades*.

En el grupo 2, solo dos alumnos corrigen con la apertura y cierre de exclamación, a partir de la puntuación exclamativa del inglés. En ninguno de los dos casos se repara en el problema que podría suponer *estés teniendo*.

2.3.5.3. Discusión

A la vista de los resultados obtenidos, nos parece relevante señalar, en primer lugar, el efecto que tiene en los estudiantes el acceso al texto original en inglés. La posibilidad de tener el texto en inglés les lleva a dar una puntuación significativamente más alta que la que otorgan en aquellos otros casos en los que no tienen el texto en esta lengua. Esto es así en las cinco variables analizadas (oraciones idiomáticas, oraciones no idiomáticas por interferencias léxicas, oraciones no idiomáticas por interferencias morfosintáctias, oraciones no idiomáticas por interferencias culturales, oraciones no idiomáticas por interferencias latinas). De todas ellas, los datos sobre la percepción de la falta de idiomática en oraciones con interferencias léxicas (p-valor <.0,0003)

o culturales ($<p\text{-valor} < 0,0002$) son los que arrojan diferencias más significativas. En el caso de las oraciones con interferencias latinoamericanas, los resultados de nuestro trabajo apuntan a que los estudiantes, quizá por ser un grupo de edad con mayor exposición a los medios de comunicación masiva, perciben este tipo de oraciones como más naturales.

En segundo lugar, nuestro estudio ha revelado la tendencia de los estudiantes a sobre corregir, un fenómeno que no es extraño en la traducción profesional. Cabe destacar que, a la luz de esta pequeña muestra de análisis cualitativo, la tendencia a la sobre corrección es sensiblemente mayor cuando no tienen acceso al texto en inglés. Compárese, a este respecto, los resultados de la celda de sobre corrección en el grupo 1 y 2 en las ocho oraciones de muestra del análisis cualitativo. En siete de los ocho casos, la tendencia a la sobre corrección es mayor en el grupo 1. La sobre corrección es un fenómeno ampliamente estudiado por Nitzke y Gros (2021) que señalan que, en el proceso de posesición, algunos traductores tienden a mejorar innecesariamente el resultado de la traducción automática con el objetivo de conseguir una traducción perfecta, a pesar de que, en las instrucciones de posesición, se les prevenga sobre esto. La sobre corrección no es un error exclusivo de la traducción automática, se da también, por ejemplo, al revisar los resultados que se obtienen de las memorias de traducción y que, incluso, pueden llegar a suponer la revisión del 60 % de coincidencias exactas que, en realidad, no requieren ninguna modificación (Mellinger y Shreve 2016). Según estos autores, la causa de este comportamiento está en que los traductores tienen en mente sus propias traducciones cuando se disponen a trabajar con los resultados de la memoria de traducción y, así, cuando comparan el texto original con el traducido, buscan errores de manera activa, aunque estos no existan, con el fin de acomodar el resultado a su propia traducción ideal. Algo similar ocurre en el proceso de posesición (Nitzke y Gros 2021), cuando los traductores se enfrentan a la traducción automática. El resultado automático no tiene por qué ser necesariamente malo, pero no siempre corresponde a la representación mental que se pueda haber hecho el traductor. Es en este punto donde surge la sobre corrección y se llevan a cabo cambios que no son esenciales.

3. CONCLUSIONES

En este artículo hemos presentado los resultados de un estudio planteado con el objetivo de analizar la percepción de los estudiantes de traducción con respecto a la idiomaticidad en la traducción automática del inglés al español en un corpus de oraciones. El ámbito abordado es la atención al cliente, dada, por una parte, su importancia en el ámbito de la comunicación digital y, por otra, la creciente tendencia a que estos contenidos se traduzcan de manera automática.

En primer lugar, nos parece relevante señalar la excesiva confianza —en nuestro caso en los dos grupos— que tienen los estudiantes en los resultados de la traducción automática. Las puntuaciones, de cada uno de los grupos, para cada una de las variables se reflejan de manera detallada en la Tabla 6. El dato grueso es que los estudiantes dan una nota de 3,4 sobre 5 a las oraciones que hemos considerado no idiomáticas. Está claro que la traducción automática neuronal debe llevar a una reflexión sobre la manera de enseñar a traducir. Cómo hacerlo sería objeto de otro artículo, pero sí se pueden apuntar aquí algunas ideas que abundan en conclusiones a las que han llevado también otras investigaciones.

La primera de estas ideas tiene que ver con el uso de la tecnología y, más concretamente, con el uso de la traducción automática (la neuronal o la basada en arquitecturas transformadoras). Está claro que esta tecnología sitúa al traductor en un escenario diferente más relacionado con la corrección o posesión de textos corregidos. En torno a esta idea, debería articularse, la enseñanza de la tecnología, en consonancia, además, con la definición del área de competencia tecnológica del European Master's in Translation que añade a la utilización de las aplicaciones informáticas más importantes, la necesidad de conocer el funcionamiento básico de la traducción automática y de evaluar el impacto en el proceso traductor y su importancia en el flujo de trabajo (EMT, 2022).

La segunda de estas ideas está relacionada con la primera de las dos metacompetencias que señala Austermuehl (2013): la capacidad de revisar textos en la lengua meta, que suele ser la materna, y la capacidad para documentarse. La enseñanza de la lengua materna en los estudios de traducción requiere de un enfoque específico, muy ligado a las necesidades del traductor. En el contexto de la traducción neuronal o de la traducción basadas en arquitecturas transformadoras requeriría del desarrollo de una conciencia muy clara de lo que marca la especificidad, el carácter único, el genio de las diferentes lenguas.

Como señalan Sánchez-Gijón y Piqué (2020), en los últimos tiempos, hemos visto cómo se llama la atención sobre los sesgos que puede introducir la inteligencia artificial en nuestra manera de percibir la realidad y en nuestra forma de comportarnos. Si nos trasladamos al ámbito lingüístico, reconocer como propios ciertos usos léxicos, sintácticos o modos de hacer que resultan ajenos a nuestra lengua y a nuestra cultura, aceptarlos o, lo que es peor, no llegar a percibir el sesgo que introducen en nuestra manera de percibir la realidad supone, como nos recuerdan Sánchez-Gijón y Piqué (*ibid.*), una pérdida del patrimonio inmaterial de la humanidad que constituyen las lenguas.

Aquí podría contraargumentarse que las lenguas cambian, que lo han hecho en todos los tiempos y que la nuestra podría ser una argumentación purista en exceso. Frente a esto, pensamos que forma parte del trabajo de los lingüistas hacer descripciones atinadas de los estados de lengua y valorar también si lo que aquí estamos describiendo no resultaría en una pérdida de matices, de valores que, al final, nos conducen a una comunicación más limitada con connotaciones culturales que podrían resultar muy ajenas a nuestra forma de percibir la realidad.

Esto trasladado a nuestro ámbito, la atención al cliente, debería llevar a un conocimiento profundo del cliente, de sus peculiaridades lingüísticas y culturales, y a la búsqueda de unas formas de comunicación más empáticas.

La segunda de las conclusiones parte del análisis cuantitativo y cualitativo de los resultados de nuestro trabajo. El análisis cuantitativo muestra las diferencias, estadísticamente significativas, entre el grupo que tiene acceso al texto en inglés frente al que no lo tiene en todas las variables. Esto podría indicar que la posesión se realiza de manera más crítica cuando no se tiene acceso al texto en la lengua original. En esta línea, también podría señalarse que los estudiantes están mucho más acostumbrados a detectar problemas que tienen que ver con la precisión, pero no tanto problemas que tienen que ver con la fluidez, la idiomática, en nuestro caso, hecho que han constatado Loock y Léchauguette (2021).

Por otra parte, la pequeña muestra del análisis cualitativo pone de manifiesto, sin embargo, que los estudiantes proporcionan mejores soluciones y sobrecorrigen menos cuando tienen acceso al texto en inglés. Aquí le corresponde a la industria determinar cuál es manera más eficiente de abordar el trabajo de posesición que, sin duda, será diferente en distintos contextos.

La tercera de las conclusiones tiene que ver con la necesidad de abordar la calidad de la traducción automática con parámetros que van más allá de la detección de errores. Como señalan Sánchez-Gijón y Piqué (2020), en las últimas tres décadas, los requisitos de los clientes en proyectos de traducción y localización multilingüe han dado prioridad a la coherencia léxica, terminológica, sintáctica y discursiva por encima de la naturalidad de la lengua de llegada y la adecuación a diferentes situaciones comunicativas y culturales. Cabría preguntarse si es necesario considerar otra serie de parámetros, como la idiomaticidad, que podrían quedar excluidos de algunos umbrales de calidad.

Finalmente, las investigaciones futuras podrían seguir varias líneas. Se podría, por ejemplo, profundizar en el efecto que tiene la presencia (o ausencia) del texto en la lengua original. En este artículo hemos contrastado las percepciones de la naturalidad y el desempeño al aportar otras soluciones más naturales en dos grupos de dos asignaturas diferentes. En un trabajo futuro, se podrían comparar los resultados de estudiantes de la misma asignatura con un grupo de control con acceso al texto en inglés, que se tomaría como referencia para comprobar y evaluar los efectos de la ausencia del texto en inglés. Otra segunda línea de trabajo sería administrar este mismo cuestionario a un grupo de adultos profesionales de la traducción. La edad podría ser uno de los factores más relevantes en la percepción de lo idiomático. Finalmente, una tercera línea, podría consistir en comparar corpus, de atención al cliente, escritos originalmente en español con otros, como el que aquí presentamos. De esta manera podría estudiarse, a partir de los datos, hasta qué punto han penetrado en el sistema ciertas características que, en principio, podrían ser específicas de los textos traducidos.

Volviendo a la metáfora del principio, transitamos por carreteras secundarias, pero no por ello menos importantes, con conclusiones modestas, aplicables a determinados contextos bajo determinadas circunstancias.

REFERENCIAS BIBLIOGRÁFICAS

- Alemán, V. (2018). Traducción automática y fraseología especializada: el caso de la propiedad intelectual e industrial. En Pino Valero Cuadra, Analía Cuadrado Rey, y Paola Carrión González (coords.) *Nuevas tendencias en traducción: Fraseología, Interpretación, TAV y sus didácticas* (pp. 69-87). Peter Lang.
- Austermuehl, F. (2013). Future (and not-so-future) trends in the teaching of translation technology. *Revista Tradumàtica. Tecnològics de la Traducció*, 11, 326-337. <https://doi.org/10.5565/rev/tradumatica.46>
- Baker, M. 1993. "Corpus linguistics and translation studies: Implications and applications". En M. Baker, G. Francis, & E. Tognini-Bonelli (eds.), *Text and technology: In honour of John Sinclair*, 233–252. John Benjamins. <https://doi.org/10.1075/z.64.15bak>
- Baziotis C., Mathur, P. & Hasler E. (2022). *Automatic Evaluation and Analysis of Idioms in Neural Machine Translation*. <https://doi.org/10.18653/v1/2023.eacl-main.267>

- Bhattacharyya, G. & Johnson, R. A. (1977). *Statistical concepts and methods* (Ser. Wiley series in probability and mathematical statistics). John Wiley & Sons.
- Bryman, A. (2016). *Social research methods* (5^a). Oxford University Press.
- Brglez, M. & Vintar, Š. (2022): Lexical Diversity and Neural Machine Translation, *Information* 13(2). <https://doi.org/10.3390/info13020093>
- Corpas, G. (1996). *Manual de fraseología española*. Gredos.
- Čulo O. y Nitzke, J. (2016): Patterns of Terminological Variation in Post-editing and of Cognate Use in Machine Translation in Contrast to Human Translation. *Proceeding of the 19th Annual Conference of the European Assication of Machine Translation*, 106-114. <https://aclanthology.org/W16-3401/>
- Chesterman, A. (2004). Beyond the particular. En Anna Mauranen y Pekka Kujamäki (eds), *Translation universals: Do they exist?* 33-49. John Benjamins. <https://doi.org/10.1075/btl.48.04che>
- Daems, J., De Clercq, O. y Macken, Lieve (2017). Translationese and posteditese: How comparable is comparable quality? *Linguistica Antverpiensia, New Series: Themes in Translation Studies*, 16, 89–103. <https://doi.org/10.52034/lanstts.v16i0.434>
- EMT (2022). *European Master's in Translation - EMT Competence Framework*. Fecha de consulta: 19 de julio de 2024 <https://commission.europa.eu/news/updated-version-emt-competence-framework-now-available-2022-10-21_en>
- Fadaee, Marzieh A., Bisazza, Arianna y Monz, Chirstof (2018). Examining the Tip of the Iceberg: A Data Set for Idiom Translation. *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*. Fecha de la consulta: 18 de julio de 2024. <https://aclanthology.org/L18-1148.pdf>
- García, A. (2012) Las perífrasis verbales en el español de América. *Analele Universității Ștefan cel Mare*, tomo XVIII, n° 1.
- Gellerstam, M. (1986): Translationese in Swedish novels translated from English. In Lars wollin and Hans Lindquist (eds.), *Translation studies in Scandinavia*. American Psychological Association. *Publications Manual*. American Phychological Association, Washington D.C., pp. 88.95.
- Fan, W. (2018). *Estudio contrastivo de la fraseología china y española desde una perspectiva lingüística*. Comares.
- Gonçalves, M., Buchicchio, M., Stewart, C., Moniz, H. y Lavie, A. (2022). Agent and User-Generated Content and its Impact on Customer Support MT. *Proceedings of the 23rd European Association for Machine Translation EAMT*. Fecha de la consulta: 18 de julio de 2024. <https://aclanthology.org/2022.eamt-1.23.pdf>
- Jolley, J. y Maimone, L. (2015). Free online machine translation: Use and perceptions by Spanish students and instructors. En Aleidine J. Moeller (ed.). *Learn Languages, Explore Cultures, Transform Lives* (pp. 181–200). Robert M. Terry.
- Jimenez-Crespo, M. A. (2023). “Translationese^a (and post-editese?)” no more: on importing fuzzy conceptual tools from Translation Studies in MT research. En *Proceedings of the 24th Annual Conference of the European Association for Machine Translation (EAMT)*. 261-268. Tampere, Finland. European Association for Machine Translation. Fecha de la consulta: 24 de marzo de 2025. <https://aclanthology.org/2023.eamt-1.25>
- Laver, J. y Mason, I. (2018). *A dictionary of translation and interpreting*. Cultural Bridges. Retrieved (2024/12/31) Fecha de consulta. 24 de marzo de 2025. <https://www.cultural-bridges.co.uk/l/dictionary-of-translation-and-interpreting/>
- Loock, R. y Léchaugette, S. (2021). Machine translation literacy and undergraduate students in applied languages: report on an exploratory study. *Revista Tradumàtica. Tecnologies de la Traducció*, 19, 204-225. <https://doi.org/10.5565/rev/tradumatica.281>

- Martín de Santa Olalla, A. y Rico, C. (en prensa): “La traducción automática en el ámbito de la atención al cliente: percepciones de los estudiantes sobre la idiomática”, *Hermeneus, Revista de Traducción e Interpretación*, volumen 27.
- Martín de Santa Olalla, A. y Rico, C. (2024): Corpus-TA_atención al cliente [Data set]. Zenodo <https://doi.org/10.5281/zenodo.10953240>
- Mauranen, A. (2004). Contrasting Language and varieties with translational corpora. *Language in Contrast*, 5 (1), 73-92. <https://doi.org/10.1075/lic.5.1.07mau>
- Mellado, C. (2016). ¿Qué es la idiomática? En: S. Silva (coord.). *Fraseología & Cía: entabulando diálogos reflexivos*. Volumen 2, 229-256. Pontes.
- Mellinger, C. D. y Shreve, G. M. (2016). Match evaluation and over-editing in a translation memory environment. En Ricardo Muñoz Martín (ed.). *Reembedding Translation Process Research.*, (pp. 131-148). John Benjamins. <https://doi.org/10.1075/btl.128.07mel>
- Morate, E. (2018). *Lengua extranjera profesional (inglés) para la gestión administrativa en la relación con el cliente*. Tutor Formación.
- Moreno, F. y Ueda, H. (2018). Cohesion and Particularity in the Spanish Dialect Continuum», *Open Linguistics*, 4, 722–742. <https://doi.org/10.1515/opli-2018-0035>
- Newmark, P. (1988). *Approaches to translation*. Prentice Hall.
- Nitzke, J. y Gros, A. (2021). Preferential changes in revision and post-editing. En M. Koponen, B. Moosop & I. S. Robert (ed.). *Translation, Revision and Post-editing. Industry Practices and Cognitive Processes* (pp. 21- 34). Routledge. <https://doi.org/10.4324/9781003096962-3>
- Ortego, M. T. y Seghiri, M. (2019). La traducción automática de locuciones nominales del español al inglés. A pain in the neck? En C. Carrasco, M. Cantarero y C. Díez (coords.) *Localización: Traducción y sostenibilidad cultural: sustrato, fundamentos y aplicaciones* (pp. 331-342) Ediciones Universidad de Salamanca. <https://doi.org/10.2307/j.ctvrdflts.57>
- Parmar, J. y Estrada, D. (2020): Idiomatic Language Translation and Transfer Learning. Fecha de consulta 24 de marzo de 2025. <https://api.semanticscholar.org/CorpusID:215717832>
- Ramírez, P. (2022). Fraseologismos en el aula de ELE: la problemática de la traducción automática. *Revista Tradumàtica. Tecnologies de la Traducció*, 20, 077-095 <https://doi.org/10.5565/rev/tradumatica.304>
- Rico, C. (2025). Estudio de la imprimación de la traducción automática sobre un corpus de textos del español institucional. *Revista de Humanidades Digitales* 10, 48-72. <https://doi.org/10.5944/rhd.vol.10.2025.41906>
- Sánchez-Gijón, P. y Piqué, R. (2020): “Conseqüències de la traducció automática neuronal sobre les llengües d’arribada”. *Revista Tradumàtica. Tecnologies de la Traducció*, 18, 1-10. <https://doi.org/10.5565/rev/tradumatica.277>
- Sánchez-Gijón, P. (2022): Neural machine translation and the indivisibility of culture and language. *Forum. Revue Internationale d’interprétation et de traduction/International Journal of Interpretation and Translation*, Volumen 20, Issue 2, p. 357-367. <https://doi.org/10.1075/forum.00025.san>
- Santing, A. (2019). Post-editese. An Exacerbated Translations. En *Proceedings of Machine Translation Summit XVII. Research Track*. Dublín, Irlanda. European Association for Machine Translation. 273-281. Fecha de consulta: 18 de julio de 2024. <https://aclanthology.org/W19-6627>
- Santing, L., Sijstermans, R., Anerdi, G. , Jeuris, P. , Ten Thij, M. y Batista, R. (2022). Food for Thought: How can we exploit contextual embeddings in the translation of idiomatic expressions? En *Proceedings of the 3rd Workshop on Figurative Language Processing (FLP)*, (pp.100-110). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2022.flp-1.14>

- Teich, E. (2003): Cross linguistic variation in system and text: a methodology for the investigation of translations and comparable text. Mouton de Gruyter. <https://doi.org/10.1515/9783110896541>
- Toral, A. (2019). Post-editeese. An Exacerbated Translationes. En *Proceedings of Machine Translation Summit XVII. Research Track*. Dublín, Irlanda. European Association for Machine Translation. 273-281. Fecha de consulta: 18 de julio de 2024. <https://aclanthology.org/W19-6627>
- Toury, G. (2012). *Descriptive Translation Studies -and Beyond: Revised Edition* (Original work published 1995). John Benjamins Publishing Company. <https://doi.org/10.1075/btl.100>
- Weinreich, U. (1953) (1974): *Languages in Contact: Findings and Problems*, De Gruyter, Inc., 1974.
- Zuluaga, A. (1980). *Introducción al estudio de las expresiones fijas*. Peter Lang.

