

Nuevas tareas asociadas a la traducción automática y su incorporación al currículo de traducción

Machine translation: new tasks and its incorporation into the translation studies curriculum

Karina Castro San Martín

Universidad de las Américas
Chile

ONOMÁZEIN | Número especial X

Nuevos enfoques y metodologías para la enseñanza de la traducción e interpretación: 114-123

DOI: 10.7764/onomazein.ne10.07

ISSN: 0718-5758



Karina Castro San Martín: Escuela de Traducción e Interpretación, Facultad de Comunicaciones y Artes, Universidad de las Américas, Chile. | E-mail: karina.castro@edu.udla.cl

Fecha de recepción: septiembre de 2021

Fecha de aceptación: marzo de 2022

Número
especial

– X –

Nuevos
enfoques y
metodologías
para la
enseñanza de
la traducción e
interpretación

2022

Resumen

La explosiva generación de contenidos, que a su vez requieren ser traducidos, exige herramientas que produzcan traducciones rápidas y de buena calidad. Los proveedores de servicios lingüísticos han asumido este desafío y han adoptado tecnologías aplicadas a la traducción automática, lo que, consecuentemente, obliga a los traductores a especializarse en este ámbito. Sin embargo, las instituciones de formación de traductores parecen no estar al mismo ritmo de esta tendencia. Por esto, el presente estudio busca describir el estado de la TA y sus usos, tareas y aplicaciones didácticas, con el fin de fomentar la discusión sobre su incorporación o perfeccionamiento en el currículo de traducción.

Palabras clave: traducción automática; preedición; posedición; formación de traductores.

Abstract

The massive generation of content, which needs to be translated, demands tools that produce fast and good quality translations. LSPs have responded to this challenge by adopting MT-based technologies, which, in turn, requires translators to specialize in this field. However, translator training institutions seem not to be keeping pace with this trend. Therefore, this study intends to describe the state of MT and its uses, tasks and didactic applications, in order to stimulate the discussion about its introduction or improvement in the translation curriculum.

Keywords: machine translation; preediting; postediting; translator training.

1. Introducción

Hace 10 años, CSA Research, una entidad investigadora especializada en la industria de los servicios lingüísticos, ya recomendaba a las empresas del rubro adoptar una estrategia de traducción automática o TA para enfrentar la explosiva generación de contenido (CSA Research, 2011), como la que vemos hoy en la industria del entretenimiento y la masificación de las plataformas de *streaming* de series y películas. Los proveedores de servicios lingüísticos parecen haber hecho caso a esta recomendación, ya que, en un estudio reciente de la misma entidad, se concluyó que la adopción de la TA había aumentado un 51 % desde el 2019 (CSA Research, 2021) y esto concuerda con una encuesta de la EUATC o European Union Association of Translation Companies del año pasado, en la que la TA y la posesión se posicionan como los nuevos servicios más populares entre proveedores de servicios lingüísticos de 45 países (EUATC, 2020). Inclusive, hoy no solo existen empresas que incorporan dichos procesos a sus operaciones, sino que también algunas proveen motores personalizados de TA, como, por ejemplo, Pairaphrase (Pairaphrase, s. f.).

Ante este panorama, es válido preguntarse qué están haciendo las instituciones de formación de traductores para afrontar el progreso y expansión de la TA. Cid-Leal y otros (2019) estudiaron la presencia de la TA en los planes de estudio de universidades españolas a nivel de pregrado y posgrado. Si bien se detectó la presencia de asignaturas y perfiles asociados al uso de tecnología aplicada la traducción, la TA solo se contemplaba de manera marginal en la mayoría de los planes de estudios analizados.

Aunque aún falta llevar a cabo investigaciones de esta naturaleza en el contexto formativo local y, posiblemente, regional, en instancias de intercambio y perfeccionamiento docente, como las Jornadas Nacionales de Enseñanza de la Traducción y la Interpretación¹, se ha evidenciado solo una breve discusión en torno a la TA y su incorporación en el currículo de traducción.

Es por esto que en el presente artículo se busca, por una parte, describir el estado actual de la TA y algunos de sus usos y tareas asociadas y, por otra parte, plantear actividades didácticas, con el fin de instalar la discusión en torno a la TA y su incorporación o perfeccionamiento en el currículo de traducción.

1 Las Jornadas Nacionales de Enseñanza de la Traducción y la Interpretación son una instancia anual que reúne a docentes de todo Chile y el extranjero, quienes exponen sobre temáticas en torno a la enseñanza e investigación en traducción e interpretación. Cada jornada contempla también una charla magistral, a cargo de un profesional o investigador destacado. La versión más reciente (IX) se realizó en agosto de este año y la institución organizadora fue la Pontificia Universidad Católica de Chile. El programa de la última versión se puede encontrar en el siguiente enlace: http://letras.uc.cl/letras/jornadasfortich/images/Programa_IX_Jornadas_Traduccion_final.pdf.

2. Estado de la TA

Antes de incorporar o actualizar la TA en el currículo de traducción, es importante estar al tanto de sus mejoras más recientes. Cuando hablamos de TA, inmediatamente la asociamos a la empresa Google y su motor de traducción, aunque otras grandes empresas, tales como Microsoft, Amazon y Facebook (Synced, 2017), y otras emergentes, como Deepl, también se han mantenido a la vanguardia en TA. Sin embargo, no es casualidad que se establezca dicha asociación, ya que Google ha sido una empresa pionera en la implementación, investigación y divulgación de tecnología aplicada a la TA (Synced, 2017).

En el año 2016, Google incorporó a su motor de TA el modelo de redes neuronales (Turovsky, 2016; Wu y otros, 2016), lo que permitió la traducción de oraciones completas (distanciándose de la traducción a nivel de palabra o frase de los modelos basados en reglas y estadísticos) y el aprendizaje en base a ejemplos a lo largo del tiempo (si bien el modelo antecesor, la TA estadística, ya permitía entrenar motores de traducción a base de ejemplos provenientes de corpus bilingües, este aprendizaje se optimiza en el modelo de redes neuronales, al requerir menos datos de entrenamiento [Synced, 2017]).

Desde el año 2017 en adelante, Google complementó su tecnología de TA con el modelo de redes *Transformer*, el que permite procesar secuencias de texto de entrada más extensas (ej., oraciones más largas, las cuales antes no podían procesarse de forma eficiente [Synced, 2017]) e incorpora un *bloque atencional*, el que analiza toda la secuencia de entrada y establece relaciones relevantes entre las palabras y grupos de palabras de dicha secuencia (Codificando Bits, 2020). Para mayor información sobre el funcionamiento de una red *Transformer*, véase Codificando Bits (2020) y Vaswani y otros (2017).

Tales mejoras sin duda representan un avance considerable en el desarrollo de la TA durante la última década y, aunque se han planteado investigaciones en las que se establece que la TA ha alcanzado una calidad similar a la humana o la llamada *paridad humana* (Hassan y otros, 2018), también existe una serie de factores que condicionan dicha calidad, tales como el tamaño y confiabilidad de los datos de entrenamiento, el tipo de texto y el campo temático, la calidad del texto fuente y los pares de idiomas de trabajo (Acolad, 2019). Debido a esto, aún se requiere la intervención humana antes y después de la traslación automática de textos, así como también en otras modalidades de traducción, tales como la traducción texto-imagen o en la traducción de lenguas con bajos recursos digitales (Zhang y Zong, 2020), entre otras áreas, que se discutirán en la siguiente sección.

3. Nuevas tareas asociadas a la TA y algunas propuestas didácticas para su incorporación al currículo de traducción

Luego de revisar el estado actual de la TA, procederemos a explorar algunos de los usos y tareas asociados a esta, los cuales requieren de distintos grados de intervención humana, en

distintos momentos del proceso traductor. La descripción de cada uso o tarea se complementa con propuestas didácticas para su implementación en el currículo de traducción.

3.1. Preedición y posesición

La tarea de preedición “consiste, a grandes rasgos, en modificar el contenido en lengua original antes de ser traducido de forma automática para facilitar el proceso y mejorar su resultado; de esta manera se reduce la carga de posesición” (Cid-Leal y otros, 2019: 189). En otras palabras, se edita el texto en lengua fuente para minimizar los errores que pueda producir la TA y optimizar el trabajo del poseedor. Para llevar a cabo dicha edición, se deben implementar distintas estrategias o reglas. Mercader y Sánchez (2016) proponen ocho reglas, las cuales se definieron a partir del análisis de errores generados en la TA. Algunas de estas se describen en la siguiente tabla:

TABLA 1

Ejemplos de reglas de preedición (Mercader y Sánchez, 2016)

INCLUIR ARTÍCULOS	DEFINIR LA POSICIÓN DE LAS COMAS
But the ruthless competition between them has helped drive a dynamic industry with *a broadening international appeal	Smith says the devil is in the details and that the EPA appears to be (...). The conservation officers check on them twice a day, * but they don't have a vet, (...).

Miyata y Fujita (2021), tras implementar el proceso de preedición con 100 oraciones del japonés al inglés, chino y coreano, definen una tipología de 39 operaciones de edición, entre las que destacan el *uso de sinónimos*, para mejorar la selección léxica en la TA, y la *adición de información* a partir del contexto inferido por el editor humano.

La posesición, por su parte, corresponde a la corrección y edición del producto de la TA. Se distingue entre la *posesición completa*, que busca lograr una calidad similar a la humana, y la *posesición parcial*, cuyo fin es generar un texto lo suficientemente comprensible para el destinatario (Cid-Leal y otros, 2019: 189). No se ahondará sobre este proceso aquí, ya que en la literatura abundan las investigaciones en torno a este (véase, por ejemplo, González y Rico, 2021).

3.1.1. Propuesta didáctica

La implementación de las tareas de preedición o posesición, sin duda, debe partir con una descripción de sus fundamentos y las reglas que gobiernan cada proceso. En el caso de la preedición, el docente estará a cargo de definir los textos de trabajo, un proceso que implica corroborar la pertinencia de las reglas a aplicar (dado que la calidad de la TA mejora constantemente, algunas de las reglas planteadas en las investigaciones citadas podrían no aplicar-

se en todos los casos). Si bien el docente puede optar por realizar actividades en las que se implementen reglas ya definidas, también puede orientarlas hacia el análisis de errores de traducción generados en la TA, para así definir las reglas a aplicar (de hecho, este fue el enfoque adoptado en las investigaciones citadas).

Para la posesición, puede que el proceso de selección de textos sea menos complejo (aunque, de todas formas, debería corroborarse que la TA genera errores lo suficientemente interesantes y complejos de analizar). Las actividades en base a este proceso podrían abordar la diferenciación entre posesición completa o parcial y la evaluación de la calidad de TA entre distintos motores o sus implicancias éticas y profesionales, tales como la importancia de la intervención humana para garantizar la calidad del producto final y la fijación de tarifas de posesición (González y Rico, 2021: 5-6).

Estos procesos también podrían implementarse en la clase de gramática contrastiva, como una forma de explicitar las diferencias entre las lenguas de trabajo de los traductores en formación.

Finalmente, el análisis de los errores que se identifican en los procesos de preedición y posesición puede fomentar la discusión sobre los errores producidos por los estudiantes, dado que podrían establecerse similitudes (por ejemplo, la investigadora a cargo de este estudio ha detectado en su práctica docente que la ausencia de artículos es una de las faltas más recurrentes y este es uno de los aspectos que se abordan en la preedición).

3.2. Traducción a nivel de documento

A pesar de que las redes neuronales *Transformer* son capaces de procesar secuencias de texto más extensas, los motores de TA siguen limitándose al nivel de frase y oraciones y a la información contextual dentro de ellas y no a lo largo del texto (Synced, 2017; Martínez, 2019). Dicha limitación afecta la coherencia y cohesión a lo largo del texto y genera errores de correferencia, inconsistencia léxica de términos polisémicos y de concordancia (Zhang y Zong, 2020; Martínez, 2019), tal como se ilustra en el siguiente ejemplo propuesto por Martínez (2019):

TABLA 2

Ejemplo de error de concordancia a nivel de documento (Martínez, 2019)

Maria won the first prize. **She** was very happy. **The winner** was the youngest one.

Maria ganó el primer premio. **Ella** estaba muy feliz. **El ganador** fue el más joven.

Si bien existen investigaciones que exploran la traducción automática neuronal a nivel de documento o *docNMT*, en la que se emplean métodos como el aprovechamiento de oraciones próximas, estas aún se encuentran en un nivel exploratorio (Zhang y Zong, 2020).

3.2.1. Propuesta didáctica

La capacidad de análisis contextual y la habilidad para garantizar la coherencia y cohesión de los textos son competencias que seguirán, posiblemente por varios años, distinguiendo a los traductores humanos de los automáticos. Por lo tanto, las actividades enfocadas en la traducción a nivel de documento podrían incluir la posesión a este nivel y la reflexión en torno a la intervención humana como garante de la calidad final del texto y cómo esto puede extenderse a la negociación con clientes (González y Rico, 2021: 5). Junto con esto, las debilidades de la TA en este ámbito pueden servir como base para actividades que busquen enfatizar y reforzar la importancia de la documentación y análisis contextual de los textos.

3.3. Traducción de lenguas con bajos recursos digitales y traducción multimodal

Una lengua de bajos recursos digitales es “aquella que no posee una vasta cantidad de textos publicados y disponibles de manera digital, puede ser porque no hay una cantidad grande de hablantes o no hay producción escrita” (Gutiérrez y otros, 2015). Dichas lenguas suponen un desafío para los motores de TA, ya que estos requieren de bitextos digitales para entrenarse y, si estos son escasos o inexistentes, la calidad de la traducción será deficiente.

La traducción multimodal puede incluir la presencia de texto, junto con otro elemento que aporte significado, como una imagen (traducción de texto-imagen), el cambio de medio (traducción de voz a texto, mediante el uso de *software* de reconocimiento de voz automático) y la simultaneidad (interpretación simultánea automática) (Zhang y Zong, 2020: 12). En el caso de la traducción de texto-imagen, por ejemplo, la imagen podría aportar información útil, como en el caso de la oración *He went to the bank*, acompañada de la imagen de un hombre aproximándose a un río (Zhang y Zong, 2020: 4), lo que deja en claro que se trata de la orilla de un río y no el edificio de un banco o asiento.

3.3.1. Propuestas didácticas

Dadas las limitaciones de las lenguas con bajos recursos digitales, las actividades asociadas a estas podrían desarrollarse como proyectos a mediano y largo plazo, en los que primero se definan las lenguas de trabajo, se implementen métodos para obtener textos en esas lenguas y se generen bitextos (mediante la digitalización y alineación de textos), para luego entrenar y configurar un motor de traducción de TA. Si bien esto puede parecer intimidante, existen recursos en línea que cuentan con todo lo necesario para llevar a cabo este proceso (en Sendra, 2020, se describe la implementación de un motor de TA, usando la plataforma Google Collaboratory) y, con la debida preparación o con la asistencia de un experto, cualquier traductor podría personalizar un motor de TA.

En el caso de la traducción multimodal, las actividades podrían enfocarse en la preparación o evaluación de materiales para la traducción imagen-texto (por ejemplo, definir si una

imagen efectivamente aporta información útil y cómo esta favorece o facilita la tarea de traducción) o analizar la calidad del producto de la traducción de voz a texto o la interpretación simultánea automáticas.

4. Discusión

Lo anteriormente expuesto deja en claro que la TA es una herramienta que nos ofrece amplias posibilidades de investigación y se presenta como un complemento a la labor traductora, en lugar de una amenaza (dadas sus deficiencias y dependencia de la intervención humana). Esto, sumado a las tendencias y necesidades del mercado, hace imperante que los planteles educativos definan perfiles de egreso y eventualmente asignaturas en los planes de estudio en los que se incorpore o refuerce la práctica de la TA. Cid-Leal y otros (2019) plantean que existe una cierta reticencia entre los profesionales hacia las tareas asociadas a la TA o derechamente preocupación en relación con esta, pero lo anterior se debe a la falta de formación y desconocimiento. Este último punto nuevamente refuerza la necesidad de incorporar la TA al currículo de traducción y valida la realización de estudios como los del presente artículo, en los que se busca ofrecer un panorama general de la TA y visibilizar cómo el traductor interactúa con ella.

5. Bibliografía citada

ACOLAD, 2019: “Neural Machine Translation: Everything You Need to Know” [<https://blog.acolad.com/neural-machine-translation>, fecha de consulta: 24 de septiembre de 2021].

CID-LEAL, Pilar, María ESPÍN-GARCÍA y Marisa PRESAS, 2019: “Traducción automática y posesición: Perfiles y competencias en los programas de formación de traductores”, *MonTI* 11, 187-214.

CODIFICANDO BITS, 2020: “Redes Transformer (... o el fin de las Redes Recurrentes)” [<https://www.codificandobits.com/blog/redes-transformer/>, fecha de consulta: 24 de septiembre de 2021].

CSA RESEARCH, 2011: “As Content Volume Explodes, Machine Translation Becomes an Inevitable Part of Global Content Strategy” [<https://csa-research.com/Blogs-Events/CSA-in-the-Media/Press-Releases/rise-of-machine-translation>, fecha de consulta: 24 de septiembre de 2021].

CSA RESEARCH, 2021: “Building a Comprehensive View of Machine Translation’s Potential” [<https://csa-research.com/Blogs-Events/Blog/Building-a-Comprehensive-View-of-Machine-Translations-Potential>, fecha de consulta: 24 de septiembre de 2021].

EUATC, 2020: “EUROPEAN LANGUAGE INDUSTRY SURVEY 2020 BEFORE & AFTER COVID-19” [https://ec.europa.eu/info/sites/default/files/2020_language_industry_survey_report.pdf, fecha de consulta: 24 de septiembre de 2021].

GONZÁLEZ, Diana, y Celia RICO, 2021: “POSEDITrad: la traducción automática y la posesión para la formación de traductores e intérpretes”, *RIDU* 15.

GUTIÉRREZ, Ximena, Elena VILCHIS y Rocío CERBÓN, 2015: “Recopilación de un corpus paralelo electrónico para una lengua minoritaria: el caso del español-náhuatl” [https://www.researchgate.net/publication/312895849_Recopilacion_de_un_corpus_paralelo_electronico_para_una_lengua_minoritaria_el_caso_del_espanol-nahuatl], fecha de consulta: 24 de septiembre de 2021].

HASSAN, Hany, Anthony AUE, Chang CHEN, Vishal CHOWDHARY, Jonathan CLARK, Christian FEDERMANN, Xuedong HUANG, Marcin JUNCZYS-DOWMUNT, William LEWIS, Mu LI, Shujie LIU, Tie-Yan LIU, Renqian LUO, Arul MENEZES, Tao QIN, Frank SEIDE, Xu TAN, Fei TIAN, Lijun WU, Shuangzhi WU, Yingce XIA, Dongdong ZHANG, Zhirui ZHANG y Ming ZHOU, 2018: “Achieving Human Parity on Automatic Chinese to English News Translation” [<https://arxiv.org/pdf/1803.05567.pdf>], fecha de consulta: 24 de septiembre de 2021].

MARTÍNEZ, Eva, 2019: *Document-Level Machine Translation: Ensuring Translational Consistency of Non-Local Phenomena*. Tesis doctoral, Universidad Universitat Politècnica de Catalunya.

MERCADER, Julia, y Felipe SÁNCHEZ, 2016: “Analysis of translation errors and evaluation of preediting rules for the translation of English news texts into Spanish with Lucy LT”, *Tradumàtica* 14, 172-186.

MIYATA, Rei, y Atsushi FUJITA, 2021: “Understanding Pre-Editing for Black-Box Neural Machine Translation” [<https://arxiv.org/pdf/2102.02955.pdf>], fecha de consulta: 24 de septiembre de 2021].

PAIRAPHRASE, s. f.: “Training Pairaphrase” [<https://www.pairaphrase.com/how-it-works/training-pairaphrase/>], fecha de consulta: 24 de septiembre de 2021].

SENDRA, Jaime, 2020: “TRANSFORMER para la Traducción de Texto” [<https://jaimesendraberenguer.medium.com/transformer-para-la-traducci%C3%B3n-de-texto-91c6d57d375d>], fecha de consulta: 24 de septiembre de 2021].

SYNCED, 2017: “History and Frontier of the Neural Machine Translation” [<https://medium.com/syncedreview/history-and-frontier-of-the-neural-machine-translation-dc981d25422d>], fecha de consulta: 24 de septiembre de 2021].

TUROVSKY, Barak, 2016: “Found in translation: More accurate, fluent sentences in Google Translate” [<https://www.blog.google/products/translate/found-translation-more-accurate-fluent-sentences-google-translate/>], fecha de consulta: 24 de septiembre de 2021].

VASWANI, Ashish, Noam SHAZEER, Niki PARMAR, Jakob USZKOREIT, Llion JONES, Aidan GOMEZ, Tukas KAISER y Illia POLOSUKHIN, 2017: “Attention is all you need” [<https://arxiv.org/pdf/1706.03762.pdf>], fecha de consulta: 24 de septiembre de 2021].

WU, Yonghui, Mike SCHUSTER, Zhifeng CHEN, Quoc V. LE, Mohammad NOROUZI, Wolfgang MACHEREY, Maxim KRIKUN, Yuan CAO, Qin GAO, Klaus MACHEREY, Jeff KLINGNER, Apurva SHAH, Melvin JOHNSON, Xiaobing LIU, TUKASZ KAISER, Stephan GOUWS, Yoshikiyo KATO, Taku KUDO, Hideto KAZAWA, Keith STEVENS, George KURIAN, Nishant PATIL, Wei WANG, Cliff YOUNG, Jason SMITH, Jason RIESA, Alex RUDNICK, Oriol VINYALS, Greg CORRADO, Macduff HUGHES y Jeffrey DEAN, 2016: “Google’s Neural Machine Translation System: Bridging the Gap between Human and Machine Translation” [<https://arxiv.org/pdf/1609.08144.pdf>, fecha de consulta: 24 de septiembre de 2021].

ZHANG, Jiajun, y Chengqing ZONG, 2020: “Neural Machine Translation: Challenges, Progress and Future” [<https://arxiv.org/abs/2004.05809>, fecha de consulta: 24 de septiembre de 2021].