V Encuentro
Migraciones e
internacionalización
del español

OBSERVATORIO
NEBRIJA DEL
ESPAÑOL

La traducción automática en el ámbito migratorio

Celia Rico Pérez
Universidad Complutense de Madrid

ANTONIO DE NEBRIJA
500 AÑOS

FUNDACIÓN
NEBRIJA

La traducción automática en el ámbito migratorio

Celia Rico Pérez
Universidad Complutense de Madrid

© Fundación Antonio de Nebrija
2024

Cualquier forma de reproducción, distribución, comunicación pública o transformación de esta obra solo puede ser realizada con la autorización de sus titulares, salvo excepción prevista por la ley. Diríjase a CEDRO (Centro Español de Derechos Reprográficos) si necesita fotocopiar o escanear algún fragmento de esta obra (www. conlicencia. com; 91 702 19 70 / 93 272 04 45).

Índice

1. **¿Qué es la traducción automática y por qué es relevante en el ámbito migratorio?** **6**

2. **¿Cómo funciona la traducción automática?** **7**

3. **La traducción automática de textos en el ámbito migratorio** **8**

4. **La traducción automática como ayuda en el acceso a la información** **10**

5. **La posedición de textos del ámbito migratorio** **12**

6. **Cuestiones éticas sobre el uso de la traducción automática** **16**

7. **Referencias** **17**

Resumen

Este trabajo se enfoca en cómo la TA, especialmente con los avances en inteligencia artificial, se ha convertido en una herramienta crucial en la traducción profesional, a pesar de no reemplazar completamente al traductor humano. Se discute la naturaleza interdisciplinaria de la TA, su funcionamiento y los desafíos éticos asociados. Además, aborda cómo esta puede ser útil para acceder a información en situaciones migratorias, aunque su uso indiscriminado puede presentar riesgos, especialmente en la traducción de contenidos sensibles o críticos.

1. ¿Qué es la traducción automática y por qué es relevante en el ámbito migratorio?

La traducción automática (TA) es, por su propia definición, aquella que se genera automáticamente sin ningún tipo de intervención humana, es decir, quien traduce es el ordenador, que mediante una serie de algoritmos y entrenamientos previos[1], produce la traducción de un texto original. A pesar de cierta mala fama entre algunos sectores de la traducción, la TA es ya una herramienta útil. Si bien es cierto que esta tecnología no reemplaza al traductor humano, sí ocupa un lugar destacado en el mundo de la traducción profesional (ELIS 2023). Una de las características de la TA es también su interdisciplinariedad, como así indica Quah (2006, 57) en su definición:

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

Los recientes avances en inteligencia artificial han posicionado la traducción automática en el centro del proceso de traducción. Apenas unos pocos años de investigación en motores neuronales primero e inteligencia artificial generativa después, ya han eclipsado alrededor de dos décadas de trabajo de traducción automática estadística, y la industria sin duda se está uniendo a la tendencia (ELIS 2023). Si bien el debate sobre el papel del traductor sigue abierto, muchos proveedores de traducción ya están aprovechando esta tecnología que les permite traducir textos automáticamente sin intervención humana (González Pastor 2023).

Aunque con un componente lingüístico, la TA bebe de disciplinas más técnicas como la lingüística computacional, la estadística o la inteligencia artificial. De hecho, la TA nació gracias a la investigación y el esfuerzo de científicos —ingenieros, lingüistas e informáticos—, que aunaron su trabajo para conseguir un proyecto, sin duda ambicioso donde los haya, como es la traducción entre distintas lenguas sin la intervención humana.

1 Un algoritmo es una serie ordenada de instrucciones de programación que permiten al ordenador realizar una función determinada. En el caso de la traducción automática, el algoritmo se refiere a las instrucciones para que el ordenador traduzca.

Sin ánimo de ser reduccionistas, podemos afirmar que la TA consiste en la traducción de una lengua a otra mediante un programa de software, que es, al fin y al cabo, quien realiza la tarea sin la necesidad de que una persona (el traductor) participe en el proceso. No obstante, este proceso, aparentemente sencillo, implica varias operaciones que van más allá de la simple sustitución de las palabras en una lengua por palabras en otra.

2. ¿Cómo funciona la traducción automática?

Antes de entrar de lleno en el uso de la TA en el ámbito migratorio nos detendremos brevemente en explicar el funcionamiento interno de los programas de TA con el fin de comprender el alcance de estos programas a la hora de generar la traducción de un texto.

A este respecto, conviene recordar que, en cualquier caso, lo cierto es que las máquinas no piensan y, por tanto, las traducciones que genera de manera automática un ordenador no son en ningún caso el resultado de un sistema inteligente sino de una serie de algoritmos previamente programados. Cuando nos encontramos frente a la máquina que traduce no podemos dejar de sentir cierta fascinación al comprobar cómo un ente digital puede producir un texto con apariencia de origen humano. En un intento de explicarnos cómo funciona solemos usar metáforas como que la máquina «aprende», «piensa», «decide», de modo que la humanizamos con el fin de poder comprender algo que nos es desconocido. Esto dota de cierto misterio a esos «cerebros sintéticos» que traducen y que son «indescifrables para el gran público, pero capaces de asumir tareas con un altísimo rendimiento» (Torrijos Caruda, 2022, p. 31). Sin embargo, lo que hay detrás de la máquina y su supuesta inteligencia es, de hecho, un sistema algorítmico complejo de reconocimiento de patrones que procesa cantidades inimaginables de datos, usando para ello una ingente cantidad de recursos informáticos, con un alto coste asociado no solo desde el punto de vista económico sino también medioambiental (Bender et al, 2021).

Una vez procesados los datos, el sistema reproduce de manera probabilística los patrones observados, con distintos grados de fiabilidad y utilidad, pero siempre guiados por los datos de entrenamiento (Bender, 2022). En el caso de la TA los patrones de traducción se extraen a partir de corpus paralelos de millones de palabras y segmentos traducidos con los que se pueden entrenar redes neuronales que, de algún modo, simulan el comportamiento humano al traducir. Así, la calidad de estas traducciones generadas automáticamente depende, por una parte, de los datos que se usen para el entrenamiento y, por otra, de las técnicas de inteligencia artificial empleadas. Sin embargo, conviene recordar que por mucho que la inteligencia artificial pueda reproducir el lenguaje humano lo que en realidad hace no es “comprender” ni “aprender”, sino hacer cálculos y asignar significados a partir de unas instrucciones y ejemplos que ya se han procesado previamente. Además, la “magia” de estos modelos cuando producen lenguaje se debe a los ajustes que se hacen con retroalimentación humana, en la que se recogen puntuaciones sobre la adecuación del contenido que ha generado la máquina para mejorar los resultados automáticos y con ellos, volver a entrenar el sistema. Con este procedimiento se ajusta el comportamiento de la máquina a las preferencias declaradas de un grupo específico de personas, esto es, las personas que etiquetan los resultados (Ouyang et al, 2022). Como vemos, en realidad, la traducción automática es una cuestión de números y el texto que se genera no tiene ninguna intención comunicativa porque, en realidad, la máquina no tiene capacidad para compartir pensamientos de ningún tipo.

En el momento de redactar este informe conviven dos tipos de sistemas de TA basados en inteligencia artificial. Por un lado, los desarrollados a partir de redes neuronales y, por otra, los llamados modelos generativos como ChatGPT. En el primer caso, el modelo basado en redes neuronales artificiales parte de unos datos de entrada que servirán para entrenar la red neuronal, que, a su vez, activa (o no) las neuronas de la fase siguiente (fase de aprendizaje), y que, igualmente, activarán (o no) las neuronas siguientes (fase de resultado). La neurona recibe el estímulo de todas las neuronas que están en la capa anterior (o fuera de su propia capa), y esta misma neurona genera también una salida, es decir, una activación. Por tanto, el proceso de activación de las neuronas artificiales va a depender de la propia activación de las otras neuronas y del llamado peso de las conexiones que se establezcan entre ellas, que puede ser positivo o negativa. De este modo, las redes neuronales artificiales van aprendiendo conforme se van entrenando y asignando el peso adecuado de las conexiones a un resultado acorde a los resultados que se esperan. Este proceso se repite las veces que sea necesario para que la red continúe mejorando los resultados (Sánchez Ramos y Rico Pérez 2020, 24).

Con respecto a los modelos de inteligencia artificial generativa, estos se basan en el procesamiento no secuencial en arquitectura de transformadores. El procesamiento no secuencial constituye uno de los principales avances en TA y ha permitido la mejora cualitativa de los resultados. Esta técnica opera de manera diferente a la arquitectura de redes neuronales recurrentes ya descrita más arriba, que procesa las palabras de una frase como una secuencia y tiene el problema de que le cuesta manejar grandes secuencias de texto, como párrafos largos o ensayos. Por decirlo de algún modo, «cuando llegaban al final de un párrafo, olvidaban lo que había ocurrido al principio» (Markowitz, 2021). La arquitectura de transformadores procesa las palabras de una frase de manera no secuencial y almacena el orden en la que estas se encuentran, no como una estructura ordenada sino como datos con codificaciones de posición. Mediante esta técnica se puede paralelizar y agilizar el procesamiento del texto (Markowitz, 2021). Esta arquitectura está compuesta por un módulo codificador que opera sobre cada palabra y su contexto en la frase de entrada y un segundo módulo descodificador que predice sucesivamente cada palabra de la frase de destino. El algoritmo de aprendizaje utiliza un corpus paralelo para obtener un conjunto de pesos y vectores de atención[2] para el transformador, de forma que los datos de entrenamiento puedan reproducirse y el sistema sea capaz de generalizar más allá de las frases del conjunto de entrenamiento (Pérez-Ortiz, Forcada y Sánchez-Martínez, 2022).

3. La traducción automática de textos en el ámbito migratorio

La traducción de textos del ámbito migratorio implica una situación comunicativa muy interesante. Por una parte, los textos que se manejan en este ámbito destacan por características lingüísticas como las siguientes: la nominalización, la despersonalización, las construcciones formadas por participio+sustantivo, la subordinación reiterada, las construcciones pasivas, el uso de la pasiva refleja y la abundancia de siglas (El-Madkouri Maataoui 2016:131). A esto se añaden las barreras

2 Un vector de atención es una representación vectorial de las palabras que componen una frase de modo que esta se adapta al significado concreto de cada palabra en su contexto. Este mecanismo permite crear un modelo de texto que mira (atiende) a cada una de las palabras de la frase original a la hora de tomar una decisión sobre cómo traducir las palabras de la frase de salida.

culturales que se habrán de sortear para generar un texto aceptable como resultado del proceso de traducción, sobre todo, en términos y expresiones que se refieren a una realidad o a un tipo de documento específico para una administración determinada del Estado.

Por otra parte, la comunicación multilingüe es clave en el contexto migratorio. Debido a las restricciones presupuestarias para la gestión de las necesidades lingüísticas de las personas migrantes (Tesseur, 2017 y 2022), el uso de la traducción automática se ve a menudo como una solución práctica en las diferentes etapas del proceso migratorio. Sin embargo, la utilización indiscriminada de esta tecnología podría suponer una brecha en el flujo comunicativo o incluso, en algún momento, un riesgo importante dependiendo del tipo de contenido a traducir (Nitzke et al, 2019). Por esta razón, es relevante definir criterios que permitan evaluar el impacto que tienen los contenidos traducidos automáticamente como respuesta a las necesidades multilingües de las personas migrantes. Hay aspectos importantes que se refieren, por ejemplo, a los referentes culturales que contienen los textos de este campo y que no tienen un tratamiento adecuado en la traducción automática. En concreto, se trata de los siguientes casos:

a) Documentos oficiales y leyes que no tienen una correspondencia directa en otro idioma.

b) Cuestiones relacionadas con la división geográfica como, por ejemplo, las "provincias" y "comunidades autónomas".

c) Diferentes equivalencias en diferentes idiomas para el mismo concepto. Es el caso, por ejemplo, de la palabra "residencia" y su traducción al francés ("domiciles", "résidence" o "séjour"); o la palabra "ocupación" (empleo), que fuera de contexto puede ser traducida automáticamente con el sentido de "invasión".

d) La traducción de las siglas como, por ejemplo, "Unidad de Grandes Empresas y Colectivos Estratégicos (UGE-CE)" para las que la traducción automática genera una traducción literal en francés "Pôle Grandes Entreprises et Groupements Stratégiques" (CGU-CE).

e) Palabras con un marcado referente cultural como "pareja" que incluyen la posibilidad de estar (o no) casado y que en otras culturas supone necesariamente la necesidad de un matrimonio (como ocurre, por ejemplo, en árabe).

f) La obligación de marcar en los formularios el primer y segundo apellidos.

g) Las siglas en los formularios como, por ejemplo, las referidas al sexo (H y M), las correspondientes al estado civil (S, C, V, D, Sp) o el N.I.E.

No cabe duda de que los desarrollos actuales de los sistemas de traducción automática neuronal dan una calidad muy alta y que los resultados son suficientemente satisfactorios como para poder proporcionar un primer borrador de traducción. En este sentido, es interesante el trabajo de Ricart Vayá y Jordán Enamorado (2022) sobre la aplicación de la traducción automática en contextos de crisis humanitarias. En su estudio, estos autores se centran en criterios de evaluación vinculados con la eficacia en un contexto de necesidad comunicativa inmediata como es el de la comunicación con refugiados ucranianos en España, de manera que "una traducción que podría ser considerada como de baja calidad en un contexto, podría tener una calidad suficiente en otro contexto distinto, dependiendo del propósito del usuario (Ricart Vayá y Jordán Enamorado, 2022, p. 108). Los datos que recogen los autores son fundamentalmente cuantitativos y apuntan a que en más de un 60% de las ocasiones, los resultados de la traducción automática son "perfectos o

casi perfectos" y que en cerca del 30% restante la traducción permite comprender fragmentos de la información original, mientras que alrededor de un 10% de las interacciones son completamente erróneas. Por lo tanto, el uso de la TA permite la comunicación de los refugiados ucranianos con las personas de habla castellana con un porcentaje de éxito elevado. En este punto, conviene mencionar que cuando se trata de evaluar la traducción automática de los referentes culturales no basta con emplear métricas cuantitativas porque este tipo de entidades requiere un análisis que tenga en cuenta también cuestiones cualitativas que, además, podrían suponer una brecha importante en la comunicación. Sin embargo, a pesar de las grandes expectativas de calidad de la traducción automática es necesario tener en cuenta los posibles riesgos que, incluso, podrían tener consecuencias graves. Así lo indican, por ejemplo, Canfora y Ottmann (2020) cuando indican que, entre otros, la traducción automática genera errores tales como omisiones, traducciones erróneas, adiciones o terminología incoherente, que pueden provocar daños, especialmente en dominios críticos para la seguridad. En el caso de los ejemplos que he señalado no se trata claramente de riesgos para la seguridad, aunque sí pueden tener el efecto de romper el flujo de la comunicación y dejar a la persona migrante desasistida ante un texto que aparentemente es fluido y aceptable pero que no consigue los fines comunicativos propuestos.

4. La traducción automática como ayuda en el acceso a la información

Un caso paradigmático de uso de la traducción automática es el acceso a la información a través de Internet. En el estudio realizado por Jiménez-Andrés (2021) se recogen datos de 13 comunidades de personas refugiadas en Grecia, Italia y España y se abordan cuestiones relacionadas con el acceso a la comunicación multilingüe a través de la tecnología así como la accesibilidad de la información disponible. Del estudio se desprende que, a la hora de acceder a la información, estas comunidades se enfrentan no solo a la complejidad de los procedimientos legales sino también a barreras lingüísticas y culturales (Jiménez-Andrés 2021, 81). En un intento por compensar estas estas deficiencias, las ONG que trabajan con personas migrantes recurren a la tecnología (Rico Pérez, 2019, Nurminen y Koponen 2020), del mismo modo que lo hacen también las propias comunidades desplazadas, para quienes supone un elemento clave en el proceso migratorio a la hora de conseguir alimento, refugio, poder comunicarse con la familia u obtener información sobre la ruta por la que se desplazan. Muchas de las narraciones de los participantes revelaron sus dificultades para acceder a información oficial fiable. La dificultad más recurrente era encontrar información en línea sobre los procesos de asilo, la normativa y el estado de su solicitud. Los participantes también plantearon la necesidad de tener información sobre los derechos de los niños, el acceso a los servicios sanitarios y las cuestiones jurídicas.

De los relatos de los participantes que se presentan en el estudio, se desprende que gran parte de la información disponible les resultaba inaccesible. De hecho, la barrera lingüística se menciona de modo sistemático en todas las comunidades como uno de los principales obstáculos para la integración, la búsqueda y la recepción de información. Así se hace hincapié en la escasez de información sobre los procedimientos de asilo en otros idiomas. Por ejemplo, los sitios web sólo están disponibles en las lenguas locales y la información sobre la normativa aplicable y los requisitos

incluye terminología especializada difícil de comprender. Además, los participantes se quejaron de que, incluso cuando se tradujeron algunos sitios web, sólo se tradujeron parcialmente y faltaba información clave. En España, esta deficiencia está muy extendida y no se limita a la información sobre los procedimientos de inmigración, como puede observarse consultando las páginas web de la Administración. En este punto, aunque se utilice traducción automática para trasvasar el texto a otra lengua es necesario tener en cuenta que en el proceso de transferencia de un referente de una cultura a otra, este lleva asociado sus propias connotaciones, esto es, su propia carga cultural específica que, dentro del contexto en el que aparece, puede llegar a plantear un problema cuando en la cultural meta no existe un referente con una carga cultural equivalente. Esta característica de los referentes culturales puede suponer un problema en los textos traducidos automáticamente en función de si ese contexto cultural se ha podido trasvasar o no.

El campo de la traducción automática de los referentes culturales es un área aún poco explorada, quizá porque hasta la llegada de los sistemas basados en inteligencia artificial no se han conseguido resultados de calidad suficiente con los que valiera la pena trabajar. Estos nuevos desarrollos generan, de manera artificial, traducciones que el propio ser humano interpreta como un acto comunicativo válido, dada nuestra innata predisposición a interpretarlos de este modo, es decir, como si la máquina tuviera alguna intencionalidad para transmitir un significado coherente. El problema está en que la máquina no tiene ninguna intención comunicativa por sí misma y que el hecho de que genere textos que se leen de manera fluida no garantiza, a priori, que la traducción sea coherente con el texto original. En el caso de los referentes culturales, precisamente porque remiten a ámbitos específicos conviene explorar la capacidad de la máquina para producir traducciones adecuadas y examinar los resultados mediante una metodología de evaluación que abarque el carácter dinámico de estos referentes a la hora de permitir el trasvase entre dos culturas concretas. Este elemento dinámico hace que los referentes culturales estén ligados estrechamente al contexto y, por lo tanto, cualquier metodología que se emplee para evaluar su traducción automática deberá tener en cuenta este aspecto necesariamente.

Sí es cierto que, dadas las cualidades cada vez más fluidas del texto generado automáticamente, puede dar la impresión de que se crea una situación comunicativa, pero esto es así porque nuestra percepción del texto en lenguaje natural, independientemente de cómo se haya generado, está mediada por nuestra propia competencia lingüística y nuestra predisposición a interpretar los actos comunicativos como si transmitieran un significado y una intención coherentes, lo hagan o no. Así, la comprensión del significado implícito es una ilusión que surge de nuestra singular comprensión humana del lenguaje (Bender *et al* 2021: 616).

El aumento en la fluidez del resultado de la traducción automática modifica la percepción de la adecuación de dicho resultado. Sin embargo, los sistemas de traducción automática pueden producir (y con frecuencia lo hacen) resultados inexactos, aunque fluidos y (de nuevo, aparentemente) coherentes por sí mismos para un consumidor que no ve el texto original o que no puede entenderlo por sí mismo (Martindale y Carpuat 2018). A los grandes modelos lingüísticos de la inteligencia artificial se les puede "pedir" que realicen una serie de tareas de procesamiento del lenguaje natural si se les proporcionan algunos ejemplos de la tarea. No obstante, estos modelos suelen tener comportamientos no deseados tales como inventarse hechos, generar textos sesgados o tóxicos (Ouyang *et al* 2022). La cuestión que surge entonces es la siguiente: ¿qué efecto tiene la traducción automática aplicada a los referentes culturales?

El área de estudio de la traducción automática de los referentes culturales es aún un campo poco explorado, lo cual es natural si pensamos que, hasta la llegada de la traducción automática neuronal, los resultados no tenían suficiente calidad como para que mereciera la pena su análisis. Probablemente, el hecho de que la traducción automática haya pasado casi desapercibida en los estudios de traducción durante décadas tenga su causa en las escasas capacidades de los ordenadores a la hora de procesar idiomas. En realidad, no se puede decir mucho desde el punto de vista de la teoría de la traducción cuando se examinan los textos traducidos por los primeros sistemas basados en reglas. Estos luchaban por producir traducciones precisas en términos sintácticos y, aunque generaban, en el mejor de los casos, traducciones literales palabra por palabra con la ayuda de glosarios y gramáticas compilados manualmente, carecían de la posibilidad de utilizar conocimientos semánticos y pragmáticos. Los desarrollos posteriores aprovecharon toda la capacidad de procesamiento de los ordenadores y convirtieron la traducción en un reto estadístico al alcance del razonamiento matemático. La calidad del texto traducido por la máquina experimentó una auténtica mejora, pero su análisis aún tardó en entrar en la teoría de la traducción. Ha sido con la llegada de la inteligencia artificial cuando por fin ha merecido la pena considerar el texto traducido por una máquina.

5. La posedición de textos del ámbito migratorio

Podemos considerar la posedición como una nueva especialidad relacionada con la TA, ya que es una actividad más de las que ofrece la industria de la traducción. La posedición consiste en la revisión de los textos producidos con traducción automática con el fin de detectar los errores que esta haya podido generar. Este tipo de revisión es diferente de la que se hace con un texto traducido por una persona puesto que el tipo de errores que comete un ordenador cuando traduce son diferentes de los que podría tener un traductor profesional. Es cierto que la TA neuronal da muy buenos resultados, según el tipo de texto que traduzcamos y, precisamente por esta razón, se necesita un buen ojo entrenado en hacer posedición para que no pasen desapercibidos errores importantes, sobre todo de tipo terminológico o fraseológico.

El trabajo de posedición ha generado cierta controversia entre la comunidad traductora, con posturas algo enfrentadas entre detractores y defensores. Para los primeros, la posedición es una tarea a la que no debe someterse un traductor profesional puesto que supone la devaluación de su trabajo. Para los segundos, es un tipo de encargo adicional al de traducción. Cualquiera de las dos opciones es válida siempre que la remuneración que recibamos sea justa con respecto al esfuerzo que realizamos. Así, si tenemos que poseditar un texto de muy baja calidad, que nos va a llevar mucho tiempo, tendremos también que cobrar una cantidad acorde a este trabajo. Por el contrario, si el texto que poseditamos no contiene errores, entonces el precio que cobramos por este trabajo puede ser menor, teniendo en cuenta, claro está, que al hacer esa posedición conviene que pongamos en valor nuestro criterio como expertos por realizar la tarea. En cualquier caso, la discusión sobre las tarifas de posedición sigue suscitando importantes polémicas (Scansani y Mhedhbi, 2020). Es muy importante que tengamos en cuenta que para poder llevar a cabo la posedición es necesario saber qué nivel de calidad se espera del texto revisado. Al utilizar traducción automática no se puede pretender una calidad similar a la que daría un traductor humano a no ser que la intervención que hagamos sobre el texto sea exhaustiva. Con el fin de determinar, precisamente, el

tipo de posedición que conviene hacer en cada caso, se suelen marcar dos tipos de recomendaciones: la posedición rápida y la posedición completa (Sánchez Ramos y Rico Pérez, 2020, 73-91).

La posedición rápida se suele hacer para revisar textos cuyo contenido no tiene mucha relevancia o que son perecederos. Sería el caso, por ejemplo, de mensajes en redes sociales, foros de usuarios o correos electrónicos. El objetivo de la posedición rápida es hacer que el texto sea comprensible, pero sin que haya que hacer grandes cambios para mejorar el estilo. Así, puede parecer que el texto lo ha creado una máquina y contener algún error sintáctico, pero si el mensaje se transmite de manera global, la traducción automática se da por buena. Entre el tipo de modificaciones que se suelen hacer en la posedición rápida, están las siguientes:

a) Corregir términos que no sean correctos.
b) Añadir toda aquella información que se haya omitido.
c) Modificar cualquier error de ortografía.

Por su parte, la posedición completa se lleva a cabo en textos que necesitan un mayor grado de calidad por la importancia de su contenido: informes públicos, contratos, diagnósticos médicos, catálogos y condiciones de servicios, etc. El objetivo de la posedición completa no es solo que la traducción automática sea comprensible sino también que sea adecuada desde el punto de vista estilístico, aunque no llegue a ser tan buena como la que haría un traductor humano.

Veamos a continuación un ejemplo en el que se muestran los dos tipos de posedición en un texto traducido automáticamente. En primer lugar, aparece el texto original, seguido de la traducción automática que ofrece Google Translate. A continuación, muestro el texto revisado con posedición rápida y, por último, la posedición completa. En cada caso, aparece tachado el segmento poseditado (una palabra, una frase o un término) y junto a él, la propuesta de traducción está marcada con un subrayado.

Texto original

> Despite some progress, serious data gaps persist in SDG monitoring. Considerable progress has been made in the availability of internationally comparable data for SDG monitoring: the number of indicators included in the global SDG database increased from 115 in 2016 to 217 in 2022. However, significant data gaps still exist in terms of geographic coverage, timeliness and level of disaggregation, making it difficult to fully comprehend the pace of progress towards the realization of the 2030 Agenda, differences across regions and who is being left behind.
>
> (United Nations, 2022, p. 4)

Traducción automática

> A pesar de algunos avances, persisten importantes lagunas de datos en el seguimiento de los ODS. Se ha logrado un progreso considerable en la disponibilidad de datos internacionalmente comparables para el seguimiento de los ODS: el número de indicadores incluidos en la base de datos mundial de los ODS aumentó de 115 en 2016 a 217 en 2022. Sin embargo, aún existen brechas de datos significativas en términos de cobertura geográfica, oportunidad y nivel de desagregación, lo que dificulta comprender completamente el ritmo del progreso hacia la realización de la Agenda 2030, las diferencias entre regiones y quién se está quedando atrás.

Posedición rápida

A pesar de algunos avances, persisten importantes lagunas ~~de~~ en los datos ~~en~~ para el seguimiento de los ODS. Se ha logrado un progreso considerable en la disponibilidad de datos internacionalmente comparables para el seguimiento de los ODS: el número de indicadores incluidos en la base de datos mundial de los ODS aumentó de 115 en 2016 a 217 en 2022. Sin embargo, aún existen brechas de datos significativas en términos de cobertura geográfica, oportunidad y nivel de desagregación, lo que dificulta el poder comprender completamente el ritmo del progreso ~~hacia la realización de la~~ para llegar a conseguir los objetivos de la Agenda 2030, así como las diferencias entre regiones y quién se está quedando atrás.

Posedición completa

A pesar de algunos avances, persisten importantes lagunas ~~de~~ en los datos ~~en~~ para el seguimiento de los ODS. Se ha logrado un progreso considerable ~~en la disponibilidad de~~ para que haya datos que sean ~~internacionalmente~~ comparables en el ámbito internacional para poder llevar a cabo el seguimiento de los ODS: el número total de indicadores incluidos en la base de datos mundial de los ODS aumentó de 115 en 2016 a 217 en 2022. Sin embargo, aún ~~existen brechas de datos~~ significativas hay lagunas significativas ~~en términos de~~ con respecto a la cobertura geográfica, la oportunidad y el nivel de desagregación, lo que dificulta el poder comprender completamente el ritmo del progreso ~~hacia la realización de la~~ para llegar a conseguir los objetivos de la Agenda 2030, así como las diferencias entre las regiones y quién se está quedando atrás.

Como vemos, en el caso de la posedición rápida se han hecho los mínimos cambios posibles para que el texto resulte comprensible. Así, se han cambiado algunas preposiciones en el siguiente segmento:

> lagunas de datos en el seguimiento

se ha cambiado a

> lagunas ~~de~~ en los datos ~~en~~ para el seguimiento

También se han hecho modificaciones en la siguiente oración:

> lo que dificulta comprender completamente el ritmo del progreso hacia la realización de la Agenda 2030, las diferencias entre regiones y quién se está quedando atrás.

con la siguiente posedición:

> lo que dificulta el poder comprender completamente el ritmo del progreso ~~hacia la realización de la~~ para llegar a conseguir los objetivos de la Agenda 2030, así como las diferencias entre regiones y quién se está quedando atrás.

En el caso de la posedición completa, además de hacer cambios que mejoran la compresión del texto, se han hecho otros que mejoran el estilo, como es el caso del siguiente segmento:

> Se ha logrado un progreso considerable en la disponibilidad de datos internacionalmente comparables para el seguimiento de los ODS.

que se ha poseditado del siguiente modo:

> Se ha logrado un progreso considerable ~~en la disponibilidad de~~ para que haya datos que sean ~~internacionalmente~~ comparables en el ámbito internacional para poder llevar a cabo el seguimiento de los ODS.

En este punto, es conveniente tener en cuenta que la posedición es una tarea que lleva asociado cierto desafío ya que, a menudo, se crea una tensión entre lo que la industria espera en términos de calidad y lo que los traductores están dispuestos a entregar como producto final. Los enfoques convencionales de posedición toman como punto de partida, como hemos visto, la distinción entre posedición rápida y completa, pero la división se difumina cuando se pone en práctica en un proyecto real en el que los traductores suelen encontrar dificultades para diferenciar entre cambios esenciales y cambios de estilo. En el momento en que se diseñaron las primeras guías de posedición, el papel del traductor humano en el proceso de traducción automática se percibía como accesorio, una visión heredada de los primeros días de la investigación cuyo objetivo era lo que entonces se denominó con las siglas FAHQMT (*Fully Automatic High Quality Machine Translation* o traducción automática de alta calidad), (Vieira, 2018, p. 319). A la vista de los últimos avances en traducción automática neuronal y del mayor nivel de calidad de los resultados que estos proporcionan, podemos pensar que la división tradicional de los niveles de posedición deja de ser válida. Esto nos lleva a plantear la posedición como un proceso de dos fases:

Primera fase. Se comprueban los siguientes aspectos:

- que no se ha dejado nada sin traducir.
- que las referencias intertextuales son correctas: números de tablas, sistema de numeración entre las distintas partes del texto, citas, notas a pie de página.
- que no hay contradicciones con el original.
- que se han aplicado de manera correctas las normas ortográficas.
- que la terminología es correcta y que se siguen las indicaciones de los glosarios del proyecto.

Segunda fase. Se corrigen los siguientes errores:

- las palabras o frases que queden en la lengua original (si es necesario).
- las frases sin sentido.
- los errores mecánicos.
- las cuestiones estilísticas sólo cuando la modificación no lleve más tiempo del que llevaría traducir desde cero.

6. Cuestiones éticas sobre el uso de la traducción automática

Para cerrar este informe, dedicaremos este último apartado a la reflexión sobre algunas cuestiones éticas en el uso de la traducción automática.

Una de las principales creencias sobre la traducción automática es que esta va a sustituir a la traducción humana. En realidad, si lo pensamos bien, la transformación de las formas de trabajo es algo que ha ocurrido a lo largo de todos los periodos de la evolución de las sociedades. Desde la Prehistoria, con el uso de la piedra y los metales como herramientas que facilitaban de algún modo el trabajo diario, hasta la Edad Moderna, con la llegada de la revolución industrial, el ser humano siempre ha buscado nuevos métodos y procedimientos que mejorasen su vida. La llegada del ordenador es, desde luego, uno de los grandes desarrollos del siglo XX y ha afectado no solo a los modos de trabajo sino también al ocio y a nuestra vida cotidiana. En este sentido, la traducción no podía quedar atrás. Ya he comentado que los primeros desarrollos en traducción automática sí pretendían originalmente sustituir al traductor humano, pero, con el tiempo, lo que queda claro es que la labor del traductor profesional es siempre necesaria para asegurar la calidad final del texto. Pensemos, por ejemplo, en las consecuencias que podría tener entregar a un paciente su diagnóstico y tratamiento de una enfermedad traducido automáticamente sin que una persona haya comprobado la fidelidad al texto original; o la traducción automática de una sentencia judicial sin que tampoco haya pasado por unas manos profesionales. La traducción automática es, desde luego, una herramienta de ayuda y así es como debemos incorporarla a nuestro trabajo diario.

Junto a la idea de la desaparición del traductor, hay otra creencia también muy popular y es la que se refiere a que la traducción automática es gratuita. Sí es cierto que hay algunas plataformas online que ofrecen traducción automática gratuita, pero, en realidad, este tipo de traducción no es en ningún modo equiparable a la traducción como un servicio profesional. Los portales online de traducción sirven para traducir contenidos que no tienen ninguna relevancia puesto que al usar este tipo de servicios tenemos que ser conscientes de que nadie nos puede asegurar la fidelidad de la traducción con el texto original.

Por último, desde el punto de vista ético debemos tener en cuenta que nuestra obligación como buenos profesionales es usar la traducción automática también de manera profesional. Esto quiere decir que, si decidimos emplearla como herramienta de ayuda para nuestras traducciones, no usaremos tampoco estos portales gratuitos puesto que, entonces, corremos el riesgo de verter en Internet datos personales o confidenciales de nuestros clientes sin tener su autorización para ello. El modo ético de usar la traducción automática es contratar un servicio de pago mensual en alguna de las plataformas de traducción profesional que hay en Internet y asegurarnos de que el tratamiento de nuestros contenidos se hace acorde con la normativa vigente sobre la protección de datos.

7. Referencias

Bender, Emily M., Timnit Gebru, Angelina McMillan-Major, y Shmargaret Shmitchell. 2021. “On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?” *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (FAccT '21). Association for Computing Machinery, New York, NY, USA, 610–623. https://doi.org/10.1145/3442188.3445922

Bender, Emily. 2022. “Look Behind the Curtain: Don’t Be Dazzled by Claims of ‘Artificial Intelligence.” *The Seattle Times*. https://www.seattletimes.com/opinion/look-behind-the-curtain-dont-be-dazzled-by-claims-of-artificial-intelligence/

Canfora, C. & Ottmann, A. (2020). Risks in neural machine translation. *Translation Spaces* 9: 1, 58–77. https://doi.org/10.1075/ts.00021.can

ELIS 2023. *European Language Industry Survey 2023. Trends, expectations and concerns of the European language industry*-. https://elis-survey.org/event/elis-2023-results-analysis/

El-Madkouri Maataoui, Mohamed. 2016. “El lenguaje jurídico-administrativo en el ámbito de la inmigración”. En: Eurrutia Cavero, Mercedes (ed). 2016. *El lenguaje jurídico y administrativo en el ámbito de la extranjería: estudio multilingüe e implicaciones socioculturales*. Berlín, Peter Lang, pp. 127-164.

Ghandour-Demiri, N. (2017). Language and Comprehension Barriers in Greece’s Migration Crisis. *Translators Without Borders*.

González Pastor, D. (Coord) (2023). El impacto de la traducción automática y la posedición en el sector de la traducción en España. Informe de investigación DITAPE 2022. https://hdl.handle.net/10550/85779

Jiménez Andres, M. (2021) Refugee access to information in online and offline environments: results from focus group discussions. *FITISPos-International Journal*, 8(1), 79-95. https://doi.org/10.37536/FITISPos-IJ.2021.8.1.270

Markowitz, D. (2021): «Transformers, Explained: Understand the Model Behind GPT-3, BERT, and T5», Web en línea: https://daleonai.com/transformers-explained

Martindale, Marianna y Marine Carpuat. (2018) “Fluency Over Adequacy: A Pilot Study in Measuring User Trust in Imperfect MT.” En: *Proceedings of the 13thConference of the Association for Machine Translation in the Americas* (Volume 1: Research Track). Association for Machine Translation in the Americas, Boston, MA, pp. 13–25. Versión electrónica: https://www.aclweb.org/anthology/W18-1803

Nitzke, Jean; Silvia Hansen-Schirra & Carmen Canfora. (2019) “Risk management and post-editing competence.” *The Journal of Specialised Translation*, 31, 239-259 https://jostrans.org/issue31/art_nitzke.php

Nurminen, Mary y Maarit Koponen. 2020. Machine translation and fair access to information. *Translation Spaces,* 9(1):150–169.

Pérez-Ortiz, J.A., M. L. Forcada y Felipe Sánchez-Martínez (2022): «How neural machine translation works» en ed. D. Kenny, *Machine translation for everyone: Empowering users in the age of artificial intelligence* (Translation and Multilingual Natural Language Processing 18). Berlin, Language Science Press.

Ouyang, Long. et al. 2022. “Training Language Models to Follow Instructions with Human Feedback.” In *Advances in Neural Information Processing Systems 35: Annual Conference on Neural Information Processing Systems 2022, NeurIPS 2022, New Orleans, LA, USA, November 28 - December 9, 2022*, edited by Sanmi Koyejo et al. https://doi.org/10.48550/arXiv.2203.02155.

Quah, C. K. (2006). *Translation and Technology*. Nueva York, Palgrave Macmillan.

Ricart Vayá, Alicia y Miguel Ángel Jordán Enamorado (2022) “Traducción automática y crisis humanitaria: análisis de la eficacia de Google Translate en la comunicación con refugiados ucranianos en España.” *Revista Tradumàtica.* Tecnologies de la Traducció, 20, pp. 96-114. DOI: https://doi.org/10.5565/rev/tradumatica.306

Rico Pérez, Celia. (2019). Mapping Translation Technology and the Multilingual Needs of NGOs along the Aid Chain. En F. Federici& S. O’Brien (Eds.), *Translation in Cascading Crises*. Routledge .

Sánchez Ramos María del Mar & Celia Rico Pérez. (2020) *Traducción automática. Conceptos clave, procesos de evaluación y técnicas de posedición*. Comares.

Scansani, R. y Mhedhbi, L. (2020). How do LSPs compute MT discounts? Presenting a company's pipeline and its use. *Proceedings of the 22nd Annual Conference of the European Association for Machine Translation*. https://bit.ly/2u1exEO

Tesseur, Wine. (2017) "The translation challenges of INGOs. Professional and non-professional translation at Amnesty International." *Translation Spaces* 6(2), pp. 209–229. DOI: 10.1075/ts.6.2.02tes

Tesseur, Wine. (2022) *Translation as Social Justice: Translation Policies and Practices in Non-Governmental Organisations*. Londres, Routledge.

Torrijos Caruda, Cristina. (2022) "Inteligencia artificial y traducción al español. Proyección, riesgos y responsabilidad." *Puntoycoma*. Boletín de los traductores españoles, 174, pp. 31–40. Versión electrónica: https://www.aieti.eu/noticias/introduce-una-noticia/

United Nations. (2022). *The Sustainable Development Goals Report* 2022. https://unstats.un.org/sdgs/report/2022/

Vieira, Lucas Nunes. (2019) "Post-editing of machine translation." En: O'Hagan M (ed.) *The Routledge handbook of translation and technology.* Londres, Routledge, pp 319-337.

OBSERVATORIO
NEBRIJA DEL
ESPAÑOL

observatorionebrijadelespañol.com

www.nebrija.com
www.fundacionnebrija.org
www.nebrija500.es

FUNDACIÓN
NEBRIJA

ANTONIO DE NEBRIJA
500 AÑOS

GOBIERNO DE ESPAÑA
MINISTERIO
DE ASUNTOS EXTERIORES, UNION EUROPEA
Y COOPERACION