

ISSN: 1579-9794

**La traducción automática de los referentes culturales.  
Propuesta de una metodología de evaluación aplicada a  
textos del ámbito migratorio**

**Cultural references and machine translation: a  
methodology for evaluation. The case of administrative  
texts in the area of migration**

CELIA RICO PÉREZ  
celrico@ucm.es  
Universidad Complutense de Madrid

Fecha de recepción: 20/02/2023

Fecha de aceptación: 25/07/2023

**Resumen:** Los referentes culturales constituyen uno de los grandes desafíos para la traducción automática (TA). Pese a existir numerosos estudios *ad hoc* acerca de esta herramienta, la magnitud de casos en los que se puede aplicar excede los límites de las investigaciones actuales. La TA de las manifestaciones culturales propias de una lengua se enmarca dentro los campos poco explorados hasta el momento. En consecuencia, se muestra la necesidad de revisar los métodos de evaluación que a menudo se emplean para determinar la validez de las traducciones hechas por la máquina y aplicarlo al caso de estos referentes, con especial incidencia en el ámbito migratorio. En esta línea, el artículo presenta una propuesta metodológica para la evaluación basada fundamentalmente en datos cualitativos sobre la fluidez, la precisión y la aceptabilidad de las traducciones automáticas, a los que se añaden datos cuantitativos sobre la percepción de estos mismos criterios. En primer lugar, se contextualizan los referentes culturales en el marco de la TA y la inteligencia artificial. A continuación, se presentan los diferentes métodos de evaluación de evaluación automática y manual, con la definición de una metodología específica para la evaluación de los referentes culturales. Con el fin de ilustrar cómo se puede llevar a la práctica esta metodología, se muestra un caso de evaluación de referentes culturales mediante un estudio exploratorio llevado a cabo con textos administrativos del ámbito migratorio.

**Palabras clave:** Referentes culturales, Traducción automática, Evaluación de la traducción, Textos administrativos, Migración

**Abstract:** Cultural references constitute one of the great challenges for machine translation (MT). Despite the existence of numerous *ad hoc* studies on this technology, the number of cases where it can be applied exceeds the limits of current research. MT of a language's own cultural manifestations is one of the fields that have been little explored so far. Consequently, there is a need to review the evaluation methods often used to determine the validity of the texts produced by MT, with a focus on cultural referents on the field of migration. In this line, the article presents a methodological proposal for MT evaluation which is based primarily on the following qualitative data: fluency, accuracy, and acceptability. To these, quantitative data is added on the perception of these same criteria. First, the cultural references are contextualised in the framework of MT and artificial intelligence. Then, the different evaluation methods of automatic and manual evaluation are presented, with the definition of a specific methodology for the evaluation of cultural references. To illustrate this methodology from a practical point of view, a case of evaluation of cultural references is shown by means of an exploratory study carried out with administrative texts in the field of migration.

**Keywords:** Cultural references, Machine translation, Translation evaluation, Administrative texts, Migration

#### INTRODUCCIÓN

La catalogación de los ámbitos a los que pueden referirse los referentes culturales está ligada a una serie de contextos concretos como son el medio natural, el patrimonio cultural, la cultura social o la cultura lingüística (cf. Molina, 2006). En el proceso de transferencia de un referente de una cultura a otra, este lleva asociado sus propias connotaciones, en otras palabras, 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 de forma automática en función de si ese contexto cultural se ha podido trasvasar o no. Si tomamos como ejemplo un referente de la cultura social como puede ser el uso del tratamiento de cortesía *usted*, observaremos que su traducción automática plantea problemas de equivalencia que solo pueden resolverse por medio del contexto. Veamos los dos ejemplos siguientes, ambos traducidos automáticamente con el programa de traducción DeepL:

Texto original	Traducción automática
1) ¿Quiere usted un poco más de vino?	<i>Do you want some more wine?</i>

2) ¿Quiere usted un poco más de vino, Sra. García?	<i>Would you like some more wine, Mrs. Garcia?</i>
--	--

**Tabla 1. Ejemplos de TA DeepL**

Fuente. Elaboración propia

En el primer caso, la carga cultural asociada a *usted* queda anulada en la traducción automática, puesto que no hay un contexto específico que permita a DeepL resolver la equivalencia de otro modo que no sea utilizando el pronombre *you*; mientras que, en el segundo caso, al añadir la referencia contextual *Sra. García*, DeepL cuenta con la información suficiente como para generar una equivalencia con un nivel de cortesía similar en inglés.

El campo de la traducción automática de referentes culturales es un área aún carente de estudios, quizá porque hasta la llegada de los sistemas basados en inteligencia artificial no se habían conseguido resultados con suficiente calidad que justificaran debidamente su estudio. Estos nuevos desarrollos artificiales son capaces de generar traducciones que el propio ser humano interpreta como un acto comunicativo válido. En parte, debido a nuestra innata predisposición a interpretarlos de este modo, es decir, como si la máquina tuviera la intención de transmitir un significado coherente. El problema reside en que la máquina no tiene ninguna intención comunicativa por sí misma y que el hecho de que genere textos que se lean de manera fluida no garantiza, *a priori*, que la traducción se corresponda 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 dinamizador hace que los referentes culturales estén ligados estrechamente al contexto y, por lo tanto, cualquier metodología que se emplee para evaluar la traducción automática de estos deberá considerar en todo momento este aspecto.

Por otra parte, parece necesario que desde los Estudios de Traducción se reflexione acerca de la validez de las categorías de evaluación que tradicionalmente se emplean con el fin de adaptarlas al análisis de la traducción automática de los referentes culturales. Pongamos por caso, estrategias típicamente humanas como pueden ser la domesticación o la extranjerización (Venuti, 1995). Hacerlo de este modo conferiría, a la máquina de una intencionalidad similar a la que se espera de la persona que traduce, lo que, como ya se ha mencionado con anterioridad es hasta cierto punto inverosímil. Además, atribuirle al programa de traducción esa capacidad humana inherente contribuye a restar agencia al propio proceso humano de

traducción en el sentido que apunta Vieira (2019) al analizar los diferentes grados de control que se pueden dar en la interacción de la persona con la traducción automática.

En este artículo se plantea, en primer lugar, una reflexión sobre el lugar que pueden ocupar los referentes culturales en el marco de la traducción automática y la inteligencia artificial, para proseguir con una revisión de los diferentes modelos que se emplean en la evaluación de la traducción automática: los modelos automáticos y los manuales (Sánchez Ramos y Rico Pérez, 2020, p. 34-48). A continuación, se proponen tres criterios de evaluación que parecen más acordes con la problemática que nos encontramos: la adecuación, la fluidez y la aceptabilidad. Finalmente, con el fin de ilustrar la puesta en práctica de la metodología de evaluación que se plantea, se muestra su aplicación en el caso concreto de los textos administrativos en el campo de las migraciones. Tras la discusión de los resultados obtenidos, el artículo concluye con la exposición de por qué no es posible aplicar criterios universales para la evaluación de la traducción automática. Referentes culturales, traducción automática e inteligencia artificial

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 que aparenta tener un origen humano. En un intento por descifrar los entresijos de su funcionamiento recurrimos a recursos como la prosopopeya para afirmar 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» (cf. Torrijos, 2022, p. 31). Sin embargo, lo que hay detrás de la máquina y su supuesta inteligencia no deja de ser un sistema algorítmico complejo de reconocimiento conformado por patrones, que procesa cantidades inimaginables de datos y usa para ello una ingente cantidad de recursos informáticos, con un alto coste asociado no solo desde el punto de vista económico, sino también medioambiental (cf. Bender *et al.*, 2021)<sup>1</sup>. 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 (cf. Bender,

---

<sup>1</sup> Para que podamos hacernos una idea, se estima que el procesamiento de un transformador GPT-3 como el que ha desarrollado OpenAI para #ChatGPT requiere una infraestructura eléctrica y de computación por valor de varios millones de dólares, de manera que esta solo puede activarse en los servidores de un *gigante* de la informática como es el caso de Microsoft. En cierto modo, esta situación no hace más que consolidar la actuales *dictaduras* en la infoesfera (Vetere, 2021).

2022). En el caso de la traducción automática, los patrones de traducción se extraen a partir de un compendio de corpus paralelos con millones de palabras y de segmentos traducidos gracias a los que se pueden entrenar las redes neuronales que, de algún modo, simulan el comportamiento humano al traducir. Por ende, la calidad de estas traducciones generadas automáticamente dependería de dos factores: los datos que se usen para el entrenamiento y 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. De ahí que la *magia* de estos modelos cuando producen lenguaje se deba a los ajustes que se hacen con retroalimentación humana, en los que se recogen datos acerca de las puntuaciones sobre la adecuación del contenido que ha generado la máquina para así poder 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 (cf. 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, la máquina *per se* no tiene capacidad para compartir pensamientos de ningún tipo. Sí es cierto, sin embargo, que, dadas las cualidades cada vez más fluidas del texto generado automáticamente, puede dar la impresión de que se crea una situación comunicativa. Esto es así porque nuestra percepción del texto, 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, con independencia de si, efectivamente, lo transmiten. Así, la comprensión del significado implícito es una ilusión que surge de nuestra singular comprensión humana del lenguaje (cf. Bender *et al.*, 2021, p. 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 resultados inexactos, aunque fluidos y coherentes por sí mismos para un consumidor que no ve el texto original o que no puede entenderlo por sí mismo (cf. Martindale y Carpuat, 2018). A los grandes modelos lingüísticos de 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 (cf. Ouyang *et al.*, 2022). La cuestión que

surge entonces es la siguiente: qué efecto tiene la traducción automática aplicada a los referentes culturales.

#### 1. MÉTODOS DE EVALUACIÓN EN TRADUCCIÓN AUTOMÁTICA

Los métodos de evaluación en traducción automática se suelen clasificar en dos tipos: la evaluación *automática*, es decir, la realizada por el propio software sin intervención humana, y la evaluación *manual*, realizada por profesionales de la traducción (Sánchez Ramos y Rico Pérez, 2020, p. 34-48). Estos dos tipos se emplean en diferentes contextos de evaluación tales como la comparación de dos o más sistemas para determinar cuál ofrece mejores resultados en un determinado par de idiomas, la comparación de las diferentes versiones de un mismo programa para analizar qué mejoras se han realizado en cada una de ellas o la evaluación de los resultados que da un sistema con textos de un campo de especialidad determinado. En el caso de las métricas de evaluación automática, estas comparan las traducciones obtenidas por el programa con una traducción de referencia realizada por un traductor profesional. Aunque una correspondencia exacta no es posible, se considera que la mejor traducción es aquella que más se acerca a la traducción de referencia. La proximidad entre la traducción de referencia y la que ha generado el programa se calcula de forma distinta según las métricas y de ellas, BLEU (Bilingual Evaluation Understudy)<sup>2</sup> (cf. Papineni *et al.*, 2002) es la que se usa con más frecuencia. Para realizar los cálculos de evaluación de la calidad se utilizan tres parámetros, siempre relacionados con el grado de coincidencia léxica entre la traducción automática y la traducción de referencia (cf. Sánchez y Rico, 2020, p. 31-54):

- a) El número de palabras comunes entre la traducción de referencia y la generada de forma automática.
- b) La longitud del texto traducido automáticamente (calculado en el número total de palabras).
- c) La longitud de la traducción de referencia (calculada en el número total de palabras).

A partir de estos tres parámetros, se pueden establecer las siguientes relaciones:

1. La relación entre el número de palabras comunes (a) y la longitud del texto traducido automáticamente (b). Con esta relación se calcula la *precisión*, esto es, cuantas palabras del texto traducido automáticamente no están en la traducción de referencia. La puntuación que se obtiene penaliza la

---

<sup>2</sup> Bilingual Evaluation Understudy) o Bilingual Evaluation Understudy (BLEU).

generación de palabras adicionales que no aparezcan en la traducción de referencia.

2. La relación entre el número de palabras comunes (a) y la longitud de la traducción de referencia (c). Con esta relación se calcula la *exhaustividad*, es decir, la cantidad de palabras de la traducción de referencia que no aparecen en la traducción automática.
3. La calidad de la traducción resultante del sistema de TA es asignada con un valor en una escala del 1 al 100, de manera que cuanto más alta sea la puntuación, mayor será la calidad de la traducción.

Como vemos, este tipo de métricas automáticas miden únicamente el grado de similitud entre una traducción automática y un documento de referencia contrastado en la lengua meta, pero no se centran, en ningún caso, en la expresión adecuada del significado del texto origen por lo que, en realidad, carecen de validez en cuanto a la medida de la calidad semántica de la traducción. Esta es una de las grandes desventajas de las métricas automáticas y, aunque es cierto que sirven para dar una estimación de la calidad en grandes volúmenes de texto, conviene atender también a los resultados que se obtienen de la evaluación de oraciones individuales. Este es, precisamente, uno de los puntos más controvertidos de BLEU y se debe a que, al calcular la calidad de una traducción de acuerdo con una traducción de referencia, no se tiene en cuenta que cabe la posibilidad de que un texto traducido sea bueno sin ser necesariamente idéntico al de referencia.

Con respecto a la evaluación manual de la traducción automática, esta se lleva a cabo de *modo analítico*, es decir, identificando errores uno por uno, con categorías típicas como errores terminológicos, ambigüedad léxica, omisiones, concordancia o errores de puntuación. Generalmente, se utilizan plantillas correctoras que establecen una tipología de errores con una serie de valores de penalización para cada uno de ellos, con el problema añadido de que no son valores estándar, puesto que cada métrica asigna un peso diferente a las distintas tipologías de error. Este tipo de evaluación nos da una imagen pormenorizada de qué problemas se han identificado en el texto traducido de manera automática. A la hora de aplicar las métricas de evaluación manual, los evaluadores primero identifican el error en función de una tabla previa de clasificación y luego asignan una puntuación según una escala determinada, de forma que al final se obtiene una puntuación que determina la calidad del texto meta. Algunas de las métricas de evaluación manual más utilizadas son SAEJ2450 (cf. SAE 2001), el modelo MQM (cf. Almaghout y Specia, 2013) y el modelo TAUS (cf. Valli, 2015).

La bibliografía especializada en la evaluación de la traducción automática es extensa y abarca diversas disciplinas como la informática, la lingüística computacional, la estadística o la inteligencia artificial (ver, por ejemplo, Way, 2018). En los Estudios de Traducción, la evaluación de la traducción automática ha hecho su entrada en los últimos años, coincidiendo precisamente con la implantación casi generalizada de este tipo de programas, y se ha centrado principalmente en la posesión (cf. Koponen *et al.*, 2021), adoptando –y adaptando, en la medida de lo posible– los métodos de las disciplinas citadas. Ahora bien, los Estudios de Traducción no son ajenos, en modo alguno, a la evaluación y esta es, precisamente, un área de trabajo de gran tradición (cf. Conde, 2022). Por otra parte, en la evaluación de la traducción automática no pueden aplicarse los mismos criterios que en la evaluación de la traducción humana, puesto que son actividades diferentes y con objetivos complementarios (cf. Mihalache, 2021). Hacerlo de otro modo, supone poner en pie de igualdad los resultados de una y otra, lo que de alguna manera lleva intrínseca la devaluación del factor humano e impide una adecuada traslación de los referentes culturales en tanto que «naciones específicas-culturales de un país o de un ámbito cultural, unidades de comunicación que poseen una estructura semántica y pragmática compleja» (cf. Luque, 2009). La cuestión es saber hasta qué punto estos métodos sirven realmente para evaluar los referentes culturales traducidos de manera automática.

## 2. CRITERIOS PARA LA EVALUACIÓN DE LOS REFERENTES TRADUCIDOS AUTOMÁTICAMENTE

La evaluación de los referentes culturales traducidos mediante traducción automática puede considerarse un nuevo campo de estudio. Hasta ahora, el análisis ha seguido la metodología propuesta fundamentalmente por Molina (2006), quien define los referentes culturales (*culturemas*) como elementos dinámicos que existen precisamente como la consecuencia de un trasvase cultural entre dos culturas concretas, de manera que al entrar en contacto a través de la traducción se puede provocar un problema de comprensión. Por lo tanto, los elementos culturales no existen fuera de un contexto, puesto que su actuación como tal depende del contexto en el que aparezcan (cf. Molina 2006, pp. 78 -79). Este es un punto clave a la hora de definir una metodología de evaluación de los referentes culturales traducidos automáticamente porque, como hemos podido comprobar en el apartado anterior, el contexto no es precisamente un aspecto que se tenga en cuenta en los criterios que se suelen utilizar. La evaluación de los elementos culturales en el resultado de la traducción automática requiere un análisis cualitativo para el que las métricas automáticas, como BLEU, no son adecuadas. Tampoco resultan aptas las métricas manuales, puesto que usan



escalas analíticas que incluyen anotaciones de errores y el cálculo de la proporción de estos errores con respecto al número total de palabras del texto traducido. Por consiguiente, son evaluaciones que no aportan datos relevantes para el caso de los referentes culturales porque no recogen información concreta sobre la esencia dinámica del referente cultural.

La propuesta de criterios que se presenta en este artículo toma como punto de partida el eje funcionalista, que permite tener en cuenta el uso final del texto en una situación o contexto determinado de modo que para que una traducción sea buena debe *funcionar* (cf. Conde, 2022). A partir de este marco de trabajo se proponen tres criterios básicos para la evaluación de los referentes culturales, siguiendo las definiciones de Castilho *et al.* (2018, pp. 18, -20):

- a) Adecuación: grado en que la traducción transfiere el significado de la unidad de la lengua de partida a la de llegada.
- b) Fluidez: grado en que la traducción sigue las reglas de la lengua de llegada.
- c) Aceptabilidad: grado en que el texto traducido satisface las necesidades de los potenciales usuarios.

Adicionalmente, y como se muestra en el siguiente apartado, la aplicación de estos tres criterios permite obtener datos cualitativos que revelan información significativa.

### 2.1 *Aplicación práctica de los criterios: el caso de textos del ámbito migratorio*

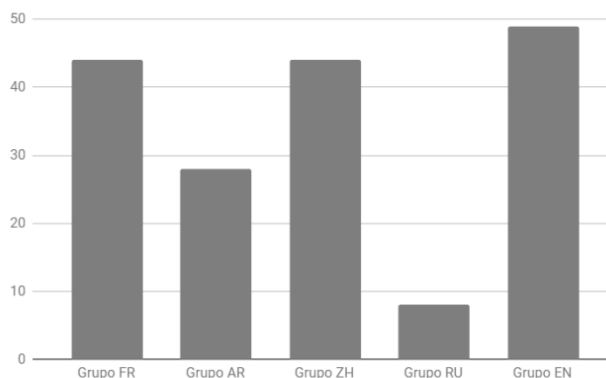
La forma en que se comunica la administración pública con las personas migrantes presenta un ámbito de situación comunicativa muy interesante en el que analizar la traducción automática de los referentes culturales. Por una parte, los textos que se manejan en este ámbito destacan por características lingüísticas como 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, 2016, p. 131). A esto se añaden las barreras 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 (cf. Tesseur, 2017 y

2022), el uso de traducción automática se ve a menudo como una solución práctica en las diferentes etapas de la cadena de ayuda (cf. Rico, 2020). Sin embargo, la utilización indiscriminada de esta tecnología podría suponer una brecha en el flujo de trabajo comunicativo o incluso, en algún momento, un riesgo significativo, todo dependiendo del tipo de contenido que se vaya a traducir (cf. Nitzke *et al.*, 2019). Por esta razón, es importante conocer el impacto que tienen los contenidos traducidos automáticamente como respuesta a las necesidades multilingües de las personas migrantes. En este sentido, con el objetivo de mostrar una aplicación práctica de los criterios de evaluación que se proponen en este artículo, en los siguientes apartados se presenta un estudio exploratorio sobre la evaluación de los referentes culturales traducidos automáticamente en textos del ámbito migratorio.

### 2.1.1 Características de los participantes en el estudio

El estudio exploratorio se llevó a cabo entre el 17 y el 19 de enero de 2022 con un grupo de 176 estudiantes de la asignatura Técnicas y Recursos para la Traducción e Interpretación en los Servicios Públicos en el Máster de Comunicación Intercultural, Interpretación y Traducción en los Servicios Públicos de la Universidad de Alcalá. Este grupo de estudiantes tenía como combinación lingüística ES-AR, FR, RU, ZH, EN-ES con la siguiente distribución por grupo de idioma (Gráfico 1):



**Gráfico 1. Distribución del grupo de lenguas**

Fuente. Elaboración propia

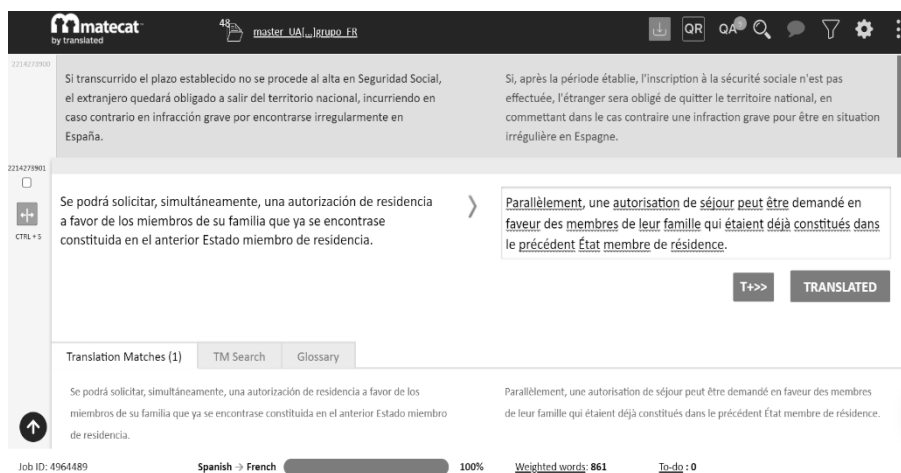
### 2.1.2 Instrumento

Para la recogida de datos se creó un cuestionario a través de un formulario de Google con dos tipos de preguntas para evaluar la calidad del resultado de la traducción automática según los criterios de adecuación,

fluidez y aceptabilidad (en el Anexo I pueden verse las preguntas que se hicieron). El primer grupo de preguntas recogía datos cuantitativos mediante una escala Likert de 5 puntos en la que los participantes indicaban el grado de acuerdo con una serie de afirmaciones relativas a la calidad del texto traducido automáticamente y que se referían de manera explícita a los tres criterios citados. Así, por ejemplo, para recoger la información sobre el criterio de *adecuación*, los participantes debían indicar su grado de acuerdo con la siguiente afirmación: «la traducción automática del texto con el que he trabajado traslada de manera adecuada el significado del texto original». El segundo grupo de preguntas se centraba en datos cualitativos, con preguntas abiertas sobre la calidad de la traducción automática. En este sentido, se pedía expresamente a los participantes que dieran su opinión sobre el resultado de la traducción automática y que indicaran algún ejemplo problemático, en el caso de que lo hubiera, con su correspondiente explicación. Para el criterio de *adecuación*, por ejemplo, la pregunta se formulaba de la siguiente manera: «¿hay algún caso en el que la traducción automática del texto con el que has trabajado no haya trasladado de manera adecuada el significado del texto original? Indica en qué caso y explícalo».

### 2.1.3 Corpus de análisis

El corpus de análisis estaba compuesto por textos del Portal de Migraciones del Ministerio de Inclusión, Seguridad Social y Migraciones de España. Se trata, principalmente, de contenidos informativos sobre trámites y procedimientos con la Administración tales como la contratación de trabajadores extranjeros, reagrupación, estancia por estudios o autorizaciones de diversa índole. Asimismo, el corpus de análisis contenía formularios y modelos de solicitud de autorización de estancia, de residencia o de trabajo. Todos los textos se tradujeron en las combinaciones de idiomas que estudiaban los participantes encuestados con la herramienta MyMemory, el sistema de traducción automática proporcionado por la plataforma de traducción MateCat que combina tanto la tecnología de Google Translator como el motor de traducción de Microsoft Translator. Los participantes tenían acceso a los textos traducidos directamente en la plataforma de traducción MateCat, de manera que podían consultar simultáneamente el texto original y su traducción. En la Imagen 1 se puede ver un ejemplo.

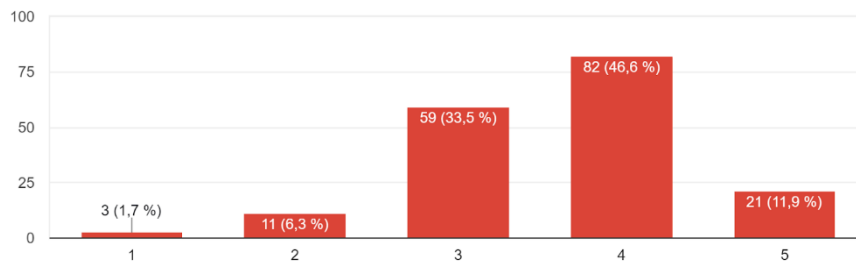


**Imagen 1. Ejemplo de texto original con su traducción automática en la plataforma MateCat**  
Fuente. Elaboración propia

#### 2.1.4 Resultados

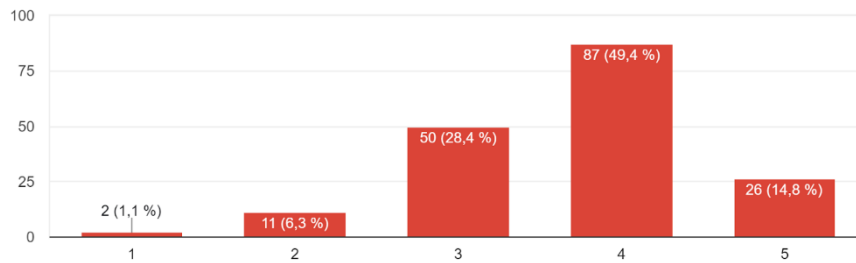
Como puede observarse en los Gráficos 2 a 4, la recogida de datos cuantitativos muestra que los resultados en las tres categorías de evaluación (adecuación, fluidez y aceptabilidad) se agrupan, principalmente, entre las puntuaciones 3 y 4, con predominio de esta última. En el caso de la *adecuación*, los resultados muestran que el 33,5 % de los participantes asigna una puntuación de 3 al texto traducido automáticamente y que el 46,6 % asigna una puntuación de 4, es decir, que están de acuerdo (puntuación 3) o muy de acuerdo (puntuación 4) en que la traducción automática es adecuada para el fin comunicativo propuesto. Para el criterio de *fluidez* el resultado es muy similar y se divide entre el 28,4 % para la puntuación de 3 y el 49,4 % para la puntuación de 4. Por último, el criterio *aceptabilidad* se distribuye entre el 39,8 % para la puntuación de 3 y el 40,9 % para la puntuación de 4.

1. Adecuación de la traducción automática. La traducción automática del texto con el que he trabajado traslada de manera adecuada el significado y ha necesitado poca (o ninguna) posesición.  
176 respuestas



**Gráfico 2. Adecuación de la traducción automática**  
Fuente. Elaboración propia

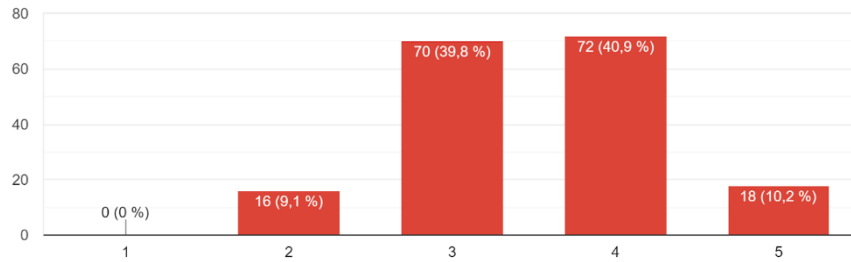
2. Fluidez de la traducción. La traducción automática del texto con el que he trabajado se lea, en su conjunto, de manera fluida con respecto a las normas de la lengua de llegada.  
176 respuestas



**Gráfico 3. Fluidez de la traducción automática**  
Fuente. Elaboración propia

3. Aceptabilidad de la traducción. ¿Hasta qué punto crees que la traducción automática del texto con el que has trabajado resulta aceptable desde el punto de vista lingüístico?

176 respuestas



**Gráfico 4. Aceptabilidad de la traducción automática**

Fuente. Elaboración propia

A la vista de estos datos cuantitativos, observamos que para los participantes en el estudio la traducción automática cumple, en la mayoría de los casos, con el criterio de trasladar de manera adecuada el significado del texto original, ser aceptable para que una persona migrante pueda conocer los trámites administrativos en cada caso y transmitir el sentido de forma fluida con respecto a las normas de la lengua de llegada. Por lo tanto, se puede concluir que la traducción automática de los textos objeto del estudio, cumplen con la finalidad comunicativa básica.

Ahora bien, más allá del análisis cuantitativo, resulta interesante detenerse en el análisis de los datos cualitativos pues nos permiten una observación directa del resultado de la traducción automática y sus consecuencias a la hora de establecer una comunicación fluida. Si bien es cierto que, de manera general, los participantes muestran su satisfacción con la traducción automática, debemos ser conscientes de que lo hacen con ciertas precauciones al respecto. A continuación, se recogen cuatro comentarios que, a modo de ejemplo, sirven para mostrar el tipo de precauciones que señalan los participantes :

1. Comentario: «Cada vez hay menos errores por parte del traductor automático, pero el traductor sigue siendo muy útil en el proceso de traducción: el registro cultural y la gramática no son perfectamente dominados por los programas de traducción automática».

2. Comentario: «Al tratarse de formularios la traducción automática ha sido muy útil y a la vez no [...] debido a la cantidad de siglas e iniciales empleadas que necesitaban de la modificación completa del segmento».
3. Comentario: «El resultado de la traducción automática es bastante aceptable. No obstante, se debe prestar atención a los elementos extratextuales, como los números o las etiquetas debido a los fallos que puede originar la herramienta».
4. Comentario: «Quisiera poner de relieve que la traducción automática de los segmentos resultó ser mejor de lo que podía pensar antes de empezar la actividad, y a excepción de expresiones más complejas por su estilo o sus referentes culturales, debo admitir que MateCat tradujo muy bien el texto, y me ha sorprendido bastante».

Los datos cualitativos recogidos por el experimento contienen mucha más información de la que aquí se muestra, puesto que recogen la opinión de un grupo de 176 participantes. En este sentido, es necesario apuntar que el análisis de todos estos datos necesita un estudio exhaustivo que, en todo caso, excede los objetivos del presente trabajo. Los cuatro comentarios que aquí se señalan sirven, no obstante, para ilustrar el hecho de que a pesar de que la traducción automática se percibe en un rango medio-alto con respecto a los criterios de adecuación, fluidez y aceptabilidad, algunos referentes culturales no se han traducido correctamente. En concreto, se trata de los siguientes casos:


- a) Documentos oficiales y leyes que no tienen una correspondencia directa en otro idioma (es el caso, por ejemplo, de una Ley orgánica).
- b) La división geográfica en *provincias* y *comunidades autónomas*.
- c) Diferentes equivalencias para el mismo concepto: *residencia* («domiciles», «résidence» o «séjour»); *ocupación* (empleo), traducido automáticamente con el sentido de 'invasión'.
- d) La traducción de las siglas como, por ejemplo, *Unidad de Grandes Empresas y Colectivos Estratégicos (UGE-CE)*<sup>3</sup> para

---

<sup>3</sup> En el texto original, la Unidad de Grandes Empresas y Colectivos Estratégicos (UGE-CE) se refiere al organismo constituido en el año 2007 para dar respuesta ágil y asesoramiento experto a las necesidades planteadas por empresas y organismos que requieren traer a España personal no comunitario de especiales características (Secretaría de Estado de Migraciones, 2022)

las que la traducción automática genera una traducción literal «Pôle Grandes Entreprises et Groupements Stratégiques (CGU-CE)».

- e) Palabras con un marcado referente cultural como *pareja* que incluyen la posibilidad de estar casado o no y que en otras culturas incluye necesariamente la necesidad de un matrimonio (así ocurre, por ejemplo, en árabe).
- f) La obligación de marcar en los formularios el primer y el segundo apellido.
- 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 NIE (Imagen 2).

	<p>EX-00</p> <p><b>Solicitud de autorización de estancia y prórrogas (LO 4/2000 y RD 557/2011)</b></p>	<p>Espacios para sellos de registro</p>
--	--	---

1) DATOS DEL EXTRANJERO/A	
PASAPORTE _____	N.I.E. _____
1er Apellido _____	2º Apellido _____
Nombre _____	Sexo <sup>(1)</sup> H <input type="checkbox"/> M <input type="checkbox"/>
Fecha de nacimiento <sup>(2)</sup> ____/____/____	Lugar _____ País _____
Nacionalidad _____	Estado civil <sup>(3)</sup> S <input type="checkbox"/> C <input type="checkbox"/> V <input type="checkbox"/> D <input type="checkbox"/> Sp <input type="checkbox"/>
Nombre del padre _____	Nombre de la madre _____
Domicilio en España _____	Nº _____ Piso _____
Localidad _____	C.P. _____ Provincia _____
Teléfono móvil _____	E-mail _____
Representante legal, en su caso _____	DNI/NIE/PAS _____ Título <sup>(4)</sup> _____

**Imagen 2.** Formulario para la solicitud de autorización de estancia y prórrogas  
Fuente. Portal de Migraciones del Ministerio de Inclusión, Seguridad Social y Migraciones

### 2.1.5 Discusión

No cabe duda de que los desarrollos actuales de los sistemas de traducción automática neuronal cuentan con una calidad muy alta y que, como se ha mostrado, los resultados son suficientemente satisfactorios como para poder proporcionar una buena comunicación desde el punto de vista de los parámetros propuestos. En este sentido, los resultados de este estudio



están en consonancia con trabajos similares como el que proponen Ricart y Jordán (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 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 (cf. Ricart y Jordán, 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 cerca del 30 % restante la traducción permite comprender fragmentos de la información original, mientras que alrededor de un 10 % de las interacciones se consideran completamente erróneas. Por lo tanto, el uso de la TA permite la interacción 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, como ya se apuntaba al principio de este artículo, este tipo de entidades requiere de un análisis que tenga en cuenta también cuestiones cualitativas que, además, podrían suponer una brecha importante en la comunicación, como demuestran los ejemplos que aparecen en los resultados. No obstante, a pesar de las grandes expectativas de calidad de la traducción automática es necesario tener en cuenta los posibles riesgos que podrían llegar a tener consecuencias graves. Así lo afirman, por ejemplo, Canfora y Ottmann (2020) cuando indican que la traducción automática genera errores tales como omisiones, traducciones erróneas, adiciones o terminología incoherente, entre otros, que pueden provocar daños, especialmente en dominios críticos para la seguridad. En el caso de los referentes culturales que he señalado en este experimento, no se trata claramente de riesgos para la seguridad, aunque sí pueden provocar una ruptura en el flujo de la comunicación y dejar a la persona migrante desasistida ante un texto que tiene una apariencia fluida y aceptable, pero que no consigue los fines comunicativos propuestos que la situación requiere.

#### CONCLUSIONES

El trabajo que he presentado aquí parte de la propuesta de que la evaluación de los referentes culturales en los textos traducidos automáticamente requiere el uso de métricas que atiendan a la naturaleza dinámica de estos elementos que, en la traducción, están a caballo entre dos culturas. Desde este punto de vista, parece que los métodos de evaluación que tradicionalmente se han venido aplicando en la investigación de la

traducción automática no son los más efectivos, puesto que no permiten recoger información sobre el contexto en el que se inscribe el elemento cultural. Como se ha mostrado, las métricas automáticas contrastan la traducción generada automáticamente con una traducción previa que se usa como referencia, tomando como punto de comparación el número de palabras comunes y la longitud del texto. Por su parte, las métricas manuales consisten, de forma esencial, en la identificación de unos errores a los que se les asigna un valor de penalización. Así pues, resulta evidente que ambos tipos de métricas impiden una adecuada evaluación de la riqueza de matices contextuales asociados a los referentes culturales. En este sentido, tal como apunta Candel (2022, p. 117), no podemos decir que existan escalas universales de valoración de la calidad, de modo que la evaluación de la calidad del resultado de la traducción automática suele depender de factores como la finalidad del texto o el valor que se dé a la inmediatez de la traducción. En el caso de la evaluación de los referentes culturales la propuesta que se ha planteado en este trabajo toma como punto de partida el eje funcionalista, de manera que la calidad de la traducción automática se evalúa a partir del uso final del texto en un contexto determinado. Los criterios propuestos son la adecuación, la fluidez y la aceptabilidad.

Con el fin de ilustrar cómo se pueden llevar a la práctica estos criterios, en este artículo se ha presentado un experimento de evaluación en el caso concreto de los textos administrativos del ámbito migratorio. Esta investigación se diseñó como un enfoque de método mixto, recopilando datos cualitativos y cuantitativos sobre las percepciones del resultado de la traducción automática. El análisis de los resultados ha revelado que los contenidos traducidos de forma automática se consideran, en términos generales, adecuados, aceptables y fluidos para los fines de la comunicación de la administración con las personas, aunque algunos conceptos culturales clave presenten traducciones erróneas y requirieran eventualmente de la intervención de un traductor profesional. Este experimento, aunque se ha realizado a pequeña escala, ha servido para mostrar de manera sencilla cómo se puede llevar a cabo la evaluación de los referentes culturales en los textos traducidos de manera automática.

#### REFERENCIAS BIBLIOGRÁFICAS

- Almaghout, H. y Specia, L. (2-6 de septiembre 2013). A CCG-based quality estimation metric for statistical machine translation [póster]. MT Summit XIV, Niza, Francia. <https://aclanthology.org/2013.mtsummit-posters.4.pdf>
- Bender, E. M. (11 de mayo de 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/>
- Bender, E. M., Gebru, T., McMillan-Major, A., y Shmitchell, S. (2021). On the dangers of stochastic parrots Can Language Models Be Too Big? [comunicación]. Proceedings of the 2021 ACM Conference on Fairness, Accountability and Transparency, (pp. 610–623). Association for Computing Machinery. <https://doi.org/10.1145/3442188.3445922>
- Candel-Mora, M. A. (2022). Fine-tuning machine translation quality-rating scales for new digital genres: The case of user-generated content. *ELUA Estudios de Lingüística Universidad de Alicante*, 38, 117-136. <https://doi.org/10.14198/elua.21900>
- Canfora, C. y Ottmann, A. (2020). Risks in neural machine translation. *Translation Spaces*, 9(1), 58–77. <https://doi.org/10.1075/ts.00021.can>
- Castilho, S.; Doherty, S.; Gaspari, F. y Moorkens, J. (2018). Approaches to Human and Machine Translation Quality Assessment. En J. Moorkens, S. Castilho, F. Gaspari y S. Doherty (Eds.), *Translation Quality Assessment from Principles to Practice* (pp. 9-38). Springer International. <https://doi.org/10.1613/jair.1.12007>.
- Conde Ruano, J. T. (2022). Calidad. En *Enciclopedia de traducción e interpretación (ENTI)*. [https://www.aieti.eu/enti/quality\\_SPA/](https://www.aieti.eu/enti/quality_SPA/)
- El-Madkouri Maataoui, M. (2016). El discurso del lenguaje jurídico-administrativo español: análisis y perspectivas. En M. Eurrutia Cabrero (Coord.). *El lenguaje jurídico y administrativo en el ámbito de la extranjería: Estudio multilingüe e implicaciones socioculturales* (pp. 127-164). Peter Lang.
- Google Translator. <https://translate.google.es/>
- Koponen, M., Mossop, B. Robert, I. S. y Scocchera, G. (Eds.) (2021) *Translation, revision and post-editing*. Routledge.
- Luque Nadal, L. (2009) Los culturemas: ¿unidades lingüísticas, ideológicas o culturales? *Language Design*, 11, 93-120.
- Martindale, M. y Carpuat, M. (2018). Fluency Over Adequacy: A Pilot Study in Measuring User Trust in Imperfect MT. En C. Cherry y G. Neubig (Eds.) *Proceedings of the 13th Conference of the Association for Machine Translation in the Americas (Volume 1: Research Track)*, (pp. 13-25). Association for Machine Translation in the Americas, <https://www.aclweb.org/anthology/W18-1803>

- Microsoft Translator. <https://translate.google.es/>
- Mihalache, I. (2021). Human and Non-Human Crossover: Translators Partnering with Digital Tools. En R. Desjardins, C. Larssonneur y P. Lacour (Eds.) *When Translation Goes Digital. Case Studies and Critical Reflections* (pp. 19-44). Palgrave Macmillan.
- Molina, L. (2006). *El otoño del pingüino: análisis descriptivo de la traducción de los culturemas*. Publicaciones de la Universidad Jaime I.
- MyMemory. <https://guides.matecat.com/my>
- Nitzke, J.; Hansen-Schirra, S. y Canfora C. (2019). Risk management and post-editing competence. *JoSTrans. The Journal of Specialised Translation*, 31, 239-259.
- Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C. L., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, S., Schulman, J., Hilton, J., Kelton, F., Miller, L., Simens, M., Askell, A., Welinder, P., Christiano, P., Leike, J., Lowe, R. (2022). Training language models to follow instructions with human feedback. <https://doi.org/10.48550/arXiv.2203.02155>
- Papineni, K., Roukos, S., Ward T., y Zhu, W. (julio 2002). BLEU: a method for automatic evaluation of machine translation. [Presentación de comunicación]. ACL-2002: 40th Annual meeting of the Association for Computational Linguistics (pp 311–318).
- Portal de Migraciones del Ministerio de Inclusión, Seguridad Social y Migraciones. <https://www.inclusion.gob.es/web/migraciones/home>
- Ricart Vayá, A. y Jordán Enamorado, M. A. (2022). Traducción automática y crisis humanitaria: análisis de la eficacia de Google Translate en la comunicación con refugiados ucranianos en España. *Revista Tradumática*, 20, 96-114. <https://doi.org/10.5565/rev/tradumatica.306>
- Rico Pérez, C. (2020). Translation technologies for the aid-chain. En F. M. Federici y S. O'Brien (Eds.) *Translation in Cascading Crises* (pp. 112-131). Routledge.
- SAE (2001). SAEJ2450 Translation quality metrics. [http://www.apex-translations.com/documents/sae\\_j2450.pdf](http://www.apex-translations.com/documents/sae_j2450.pdf)
- Sánchez Ramos, M. M. y Rico Pérez, C. (2020). Traducción automática. Conceptos clave, procesos de evaluación y técnicas de posesición. Comares.

- Secretaría de Estado de Migraciones. (2022). Unidad de grandes empresas y colectivos estratégicos. <https://www.inclusion.gob.es/web/unidadgrandesempresas/sobre-nosotros>
- Tesseur, W. (2017). The translation challenges of INGOs. Professional and non-professional translation at Amnesty International. *Translation Spaces*, 6(2), 209–229. <https://doi.org/10.1075/ts.6.2.02tes>
- Tesseur, W. (2022). *Translation as Social Justice: Translation Policies and Practices in Non-Governmental Organisations* (1.ª ed.). Routledge. <https://doi.org/10.4324/9781003125822>
- Torrijos Caruda, C. (2022). Inteligencia artificial y traducción al español. Proyección, riesgos y responsabilidad. *Puntoycoma*, 174, 31–40. <https://www.aieti.eu/noticias/introduce-una-noticia/>
- Valli, P. (26-27 de noviembre de 2015). The TAUS Quality Dashboard [Presentación de comunicación]. Proceedings of the 37th Conference Translating and the Computer, (pp. 127–136). <https://aclanthology.org/2015.tc-1.17.pdf>
- Venuti, L. (1995): *The Translator's Invisibility*, Routledge.
- Vetere, G. (2021). *Textnology. Imminent. Research Report*. <https://imminent.translated.com/textnology>
- Vieira, L. N. (2019). Post-editing of machine translation. O'Hagan, M. (Ed.), *The Routledge handbook of translation and technology* (1.ª ed.), pp. 319-337). Routledge. <https://doi.org/10.4324/9781315311258>
- Way, A. (2018). Quality Expectations of Machine Translation. Moorkens, J., Castilho, S., Gaspari, F., y Doherty, S. (Eds.), *Translation Quality Assessment from Principles to Practice* (pp. 159-178). Springer International. <https://doi.org/10.1613/jair.1.12007>

ANEXO I. LISTADO DE PREGUNTAS PARA LA RECOGIDA DE DATOS MEDIANTE EL CUESTIONARIO

1) Preguntas para la recogida de datos cuantitativos sobre la calidad del texto traducido automáticamente

a) Pregunta sobre la adecuación de la traducción automática:

*Expresa, en una escala de 1 a 5, tu acuerdo/desacuerdo con la siguiente afirmación: «la traducción automática del texto con el que he trabajado traslada de manera adecuada el significado del texto original».*

1      2      3      4      5

Totalmente en desacuerdo                        Totalmente de acuerdo

---

**Imagen 3. Escala Likert correspondiente a la pregunta sobre la adecuación de la traducción automática**

Fuente. Cuestionario sobre la evaluación de la traducción automática de elaboración propia

b) Pregunta sobre la fluidez de la traducción automática:

*Expresa, en una escala de 1 a 5, tu acuerdo/desacuerdo con la siguiente afirmación: «la traducción automática del texto con el que he trabajado se lee, en su conjunto, de manera fluida con respecto a las normas de la lengua de llegada».*

1      2      3      4      5

Totalmente en desacuerdo                        Totalmente de acuerdo

---

**Imagen 4. Escala Likert correspondiente a la pregunta sobre la fluidez de la traducción automática**

Fuente. Cuestionario sobre la evaluación de la traducción automática de elaboración propia

c) Pregunta sobre la aceptabilidad de la traducción automática:

*Expresa, en una escala de 1 a 5, tu acuerdo/desacuerdo con la siguiente afirmación: «la traducción automática del texto con el que he trabajado resulta aceptable para que una persona migrante conozca los diferentes trámites administrativos y pueda rellenar las diferentes solicitudes con esa información».*

1      2      3      4      5

Totalmente en desacuerdo                        Totalmente de acuerdo

---

**Imagen 5 . Escala Likert correspondiente a la pregunta sobre la aceptabilidad de la traducción automática**

Fuente. Cuestionario sobre la evaluación de la traducción automática de elaboración propia

2) Preguntas para la recogida de datos cualitativos sobre la calidad del texto traducido automáticamente

d) Pregunta sobre la adecuación de la traducción automática:

*¿Hay algún caso en el que la traducción automática del texto con el que has trabajado no haya trasladado de manera adecuada el significado del texto original? Indica en qué caso y explícalo con detalle.*

e) Pregunta sobre la fluidez de la traducción automática:

*¿Hay algún caso en el que la traducción automática no se lea de manera fluida? Indica en qué caso y explícalo con detalle.*

f) Pregunta sobre la aceptabilidad de la traducción automática:

*¿Hay algún caso en el que la traducción automática del texto con el que has trabajado no sea aceptable? Indica en qué caso y explícalo con detalle*