

Traducción automática y crisis humanitaria: análisis de la eficacia de Google Translate en la comunicación con refugiados ucranianos en España



Alicia Ricart Vayá
Miguel Ángel Jordán Enamorado



Alicia Ricart Vayá
IULMA. Universitat de València
alicia.ricart@uv.es;
ORCID:
[0000-0002-1523-3814](https://orcid.org/0000-0002-1523-3814)



Miguel Ángel Jordán Enamorado
IULMA. Universitat de València
miguel.jordan@uv.es;
ORCID:
[0000-0003-4668-1537](https://orcid.org/0000-0003-4668-1537)

Resumen

El objetivo de este trabajo es evaluar la eficacia de la traducción automática en un contexto de emergencia humanitaria. Para ello, se realizaron entrevistas con refugiados ucranianos utilizando Google Translate. Los resultados de nuestro análisis muestran la eficacia de dicha herramienta y señalan los aspectos que se han de mejorar.

Palabras clave: traducción automática, Google Translate, comunicación, fluidez, inteligibilidad.

Resum

L'objectiu d'aquest treball és avaluar l'eficàcia de la traducció automàtica en un context d'emergència humanitària. Per això, es van realitzar entrevistes amb refugiats ucraïnesos utilitzant Google Translate. Els resultats de la nostra anàlisi mostren l'eficàcia d'aquesta eina i assenyalen els aspectes que cal millorar.

Paraules clau: traducció automàtica, Google Translate, comunicació, fluïdesa, intel·ligibilitat.

Abstract

The aim of this paper is to evaluate the effectiveness of machine translation in a humanitarian emergency context. To this end, interviews with Ukrainian refugees were conducted using Google Translate. The results of our analysis show the effectiveness of this tool and point out the aspects that need to be improved.

Keywords: machine translation, Google Translate, communication, fluency, intelligibility.

1. Introducción

El conflicto bélico entre Rusia y Ucrania, que comenzó durante el mes de febrero de 2022, ha dado lugar a la mayor crisis de refugiados desde la Segunda Guerra Mundial. Según los datos ofrecidos por las Naciones Unidas con fecha de junio de 2022, más de ocho millones de personas han abandonado Ucrania desde el inicio de la guerra, de las cuales, casi ciento veinticinco mil han encontrado refugio en España. Según los datos ofrecidos por el gobierno español a principios de abril de 2022, el 23,65% de los

Revista Tradumàtica. Número 20 (deseembre 2022).

Cita recomanada: RICART VAYÁ, ALICIA; JORDÁN ENAMORADO, MIGUEL ÁNGEL (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, 096-114.
<<https://doi.org/10.5565/rev/tradumatica.306>>



<http://revistes.uab.cat/tradumatica>

ISSN 1578-7559

refugiados ucranianos en España que habían tramitado un permiso de protección temporal son menores de 18 años y el 67,58% son mujeres (Ministerio de Interior, 2022). Esta situación ha conllevado un ingente despliegue humanitario en todos los niveles de la sociedad, tanto por parte de instituciones públicas y privadas como de familias e individuos que han colaborado en la acogida y atención de las personas que escapaban de la guerra. La ayuda ofrecida por individuos e instituciones ha tenido como finalidad cubrir no solo las necesidades básicas de los refugiados, ofreciéndoles un lugar en el que hospedarse, alimentos y ropa, sino que también se han desarrollado programas de formación en idiomas (Instituto Cervantes, 2022), escolarización de menores (Casanova, 2022) e inserción laboral (Helguera, 2022).

Como es lógico, en todo este proceso ha sido necesaria la comunicación con las personas atendidas, que en muchos casos desconocían el idioma del país que las acogía. Para lograr esta comunicación, se ha contado con la colaboración de traductores profesionales y de voluntarios con conocimientos de las lenguas requeridas. Sin embargo, en muchas ocasiones no ha sido posible contar con esta asistencia, por lo que se ha recurrido al uso de herramientas digitales de traducción automática.

Durante los últimos años, se han publicado numerosos trabajos centrados en el análisis de la traducción automática (en adelante TA) y, en concreto, en la calidad de sus resultados (Castilho *et al.*, 2018; Freitag *et al.*, 2021; Chatzikoumi, 2020). La mayoría de estos trabajos han abordado esta tarea desde el punto de vista de la traducción profesional, es decir, han analizado hasta qué punto la TA puede ser útil en el proceso de traducción de un texto de manera que agilice o incluso pueda llegar a sustituir la labor del traductor (Rico, 2017; Canfora y Ottmann, 2020). Existen también diversos estudios sobre la aplicabilidad de la TA en contextos de crisis (Lewis *et al.*, 2011; Translators Without Borders, 2016; Caldwell *et al.*, 2019; Federici y O'Brien, 2019; O'Brien, 2019). Sin embargo, en la mayoría de estos trabajos los usuarios de la TA son profesionales de la traducción. A día de hoy todavía son pocos los trabajos que han investigado la utilidad de la TA en situaciones de crisis humanitaria en las que la mayoría de los usuarios de la TA son personas sin una especial formación en lenguas ni en traducción, que se ven obligadas a recurrir a estas herramientas por una necesidad comunicativa, en la que no importa tanto la corrección del texto traducido como la posibilidad de comunicarse con diversos interlocutores.

En el presente artículo explicaremos la investigación que hemos realizado con el objetivo de evaluar la eficacia de la TA por medio de la herramienta Google Translate (en adelante GT) en una situación de crisis humanitaria. Para llevar a cabo dicha investigación, contamos con la colaboración del Centro de Idiomas de la Universitat de València, por medio del cual pudimos mantener entrevistas individuales de manera presencial con un grupo de treinta y dos refugiados ucranianos de edades comprendidas entre los 25 y los 55 años que asisten a clases de español en dicha institución.

Con la finalidad de contextualizar los resultados de dicha investigación, ofreceremos en primer lugar un marco teórico sobre la evaluación de la calidad de la TA. Posteriormente, explicaremos la metodología seguida en nuestro análisis y los resultados

a los que dio lugar. Por último, ofreceremos nuestras conclusiones y señalaremos algunas líneas de investigación para futuros estudios.

2. Marco teórico

2.1. Problemas y ventajas de la traducción automática

La TA ha evolucionado enormemente en las últimas décadas, dando lugar a una sucesión de sistemas que han ido mejorando la calidad de los resultados (Castilho *et al.*, 2017). Sin embargo, a pesar de los grandes avances logrados, es evidente que los textos resultantes de este tipo de traducciones todavía son mejorables. En este apartado, mencionaremos brevemente tanto los problemas de la TA como las ventajas que ofrece a sus usuarios. Al hablar de problemas en la TA, nos referimos a aquellos conflictos que se presentan al utilizar estas herramientas con cualquier tipo de textos. No abordaremos, por tanto, los problemas específicos de los distintos sistemas de traducción ni de los diferentes motores de TAN.

En su estudio sobre la TA, Ramlow (2008) analiza los problemas de orden lingüístico que surgen de este proceso y establece una clasificación con cuatro grupos principales: problemas de orden morfológico, léxico, sintáctico y aquellos que tienen que ver con la referencia pronominal. Existen además otros tipos de problemas en la TA, algunos son formales, como aquellos que tienen que ver con la puntuación; también se han de tener en cuenta los posibles errores existentes en el texto de origen, que pueden ser detectados por un traductor humano, pero que no serán tratados como tales por el traductor automático. Otro problema que puede surgir es la traducción de palabras que no deben ser traducidas como, por ejemplo, algunos términos técnicos, nombres propios, siglas o anglicismos (Mendoza, 2017). Habrá de tenerse presente, además, la posible repetición de los mismos errores a lo largo del texto, el empobrecimiento de la expresión y la falta de creatividad, ya que los programas de TA no son capaces de introducir modificaciones retóricas ni ofrecer soluciones creativas ante los diversos problemas de traducción (Navío, 2003).

A pesar de los posibles problemas de la TA, no hay duda de que también ofrece un gran número de ventajas a sus usuarios que mencionaremos brevemente a continuación. La cualidad más llamativa de la TA es su rapidez, puesto que ofrece sus resultados de manera prácticamente inmediata. Además, facilita la comprensión de lenguas desconocidas o de las que no se posee un conocimiento suficiente. La TA favorece por tanto la productividad, puesto que agiliza procesos que de otro modo requerirían una mayor dedicación de tiempo y de recursos humanos. La TA facilita la autonomía de aquellos particulares o empresas sin conocimientos de lenguas extranjeras que necesiten disponer de una información en otra lengua. Los programas de TA suelen ser sencillos de utilizar, de manera que son accesibles para personas con pocos conocimientos informáticos (Mendoza, 2017; Castilho, 2017; Oliver, 2020).

Como es lógico, es necesario tener en cuenta la finalidad de la traducción a la hora de decidir si la TA es el procedimiento adecuado para realizar dicho trabajo. De manera

generalizada, se pueden clasificar los fines de la traducción en tres grupos. Por una parte, la traducción para la difusión, es decir, la que se lleva a cabo cuando el texto resultante será publicado en cualquier tipo de soporte o plataforma. Por otra parte, se encuentra la traducción para la asimilación, cuya finalidad es acceder a la información recogida en el texto. En tercer lugar, se puede hablar de la traducción para la comunicación interpersonal por medio de textos escritos, ya sean correos electrónicos, servicios de mensajería digital, redes sociales, etc. De estos tres grupos, el primero es el único en el que no es suficiente el uso de la TA para lograr un resultado adecuado. Sin embargo, en los otros dos, la TA puede servir para lograr los objetivos de dicha traducción (Hutchins, 2001).

2.2. Evaluación de la calidad de la traducción automática

La traducción es un proceso complejo que incluye elementos cognitivos, lingüísticos, sociales y culturales. Por esta razón, no resulta sencillo definir los parámetros con los que evaluar los resultados de dicho proceso. Debido a su importancia, no es de extrañar que la evaluación de la calidad de la traducción (*translation quality assessment*. TQA) haya sido un tema de debate dentro de los estudios de traducción y, en particular, en el campo de la traducción automática (Hutchins y Somers 1992; Arnold *et al.* 1994; White 2003; Hamon *et al.* 2007; Castilho *et al.*, 2017). Por esta razón, antes de abordar este tema, es necesario definir qué se entiende por calidad de la traducción, puesto que este concepto se puede analizar desde perspectivas diferentes.

Koby *et al.* (2014) afirman que una traducción de calidad ha de cumplir los requisitos de fidelidad y fluidez exigidos tanto por la audiencia como por la finalidad de dicha traducción y ha de cumplir todas las condiciones acordadas entre quien la solicita y quien la ofrece, teniendo en cuenta las necesidades del usuario final. En el caso de la TA, el consumidor final puede ser un usuario de Internet o un traductor profesional (Specia *et al.*, 2010), de manera que el resultado esperado por dicho consumidor puede ser muy diferente al de una traducción humana. Según la teoría de skopos (Reiss y Vermeer, 2014) la traducción es un acto con un propósito específico, cuyo resultado es el texto meta. El propósito que gobierna el proceso de traducción es definido por el solicitante y la traducción debería funcionar de manera que se cumpla dicho propósito. Según House (2014), para que una traducción sea correcta, debe existir una completa equivalencia semántica y pragmática entre el texto de origen y el texto meta.

A lo largo de los últimos años se han diseñado diferentes técnicas de evaluación de los diversos tipos de TA. En la literatura internacional, estas técnicas se han clasificado en automáticas o humanas. Las primeras se refieren a las evaluaciones realizadas por herramientas informáticas, mientras que las segundas son llevadas a cabo por personas (Han *et al.* 2021). En el presente trabajo nos centraremos en la evaluación humana y, en concreto, en aquella en la que los evaluadores expresan un juicio directo sobre la calidad de la traducción. Las diversas investigaciones sobre la evaluación humana de la traducción han dado lugar a taxonomías diferentes en las que los mismos criterios reciben denominaciones variadas. Castilho *et al.* (2018) ofrecen la siguiente clasificación en la que se recogen las principales propuestas anteriores:

- a) Adecuación: también llamado precisión o fidelidad en algunos estudios sobre la TA, es un principio consolidado en la evaluación de los resultados de la TA (Koehn, 2009). Este criterio evalúa hasta qué punto el resultado de la TA transmite el significado del texto de origen (TO).
- b) Fluidez: se evalúa si el texto meta (TM) sigue las reglas de la lengua meta (LM).
- c) Legibilidad: se refiere a la facilidad con la que un texto puede ser leído por una o varias personas. Para evaluarlo se siguen diferentes pautas que se suelen basar en rasgos lingüísticos como la frecuencia de uso de determinadas palabras, la longitud de las frases, etc.
- d) Comprensibilidad: este criterio se aplica para evaluar cómo de comprensible es un texto para el lector. Mientras que en el criterio de legibilidad depende de manera inherente del texto, la comprensibilidad puede variar en función del lector, es decir, dependiendo de su edad, formación o conocimiento del tema tratado. Al no existir un criterio unificado de evaluación de este principio, en el campo de la TA se suele recurrir a la valoración por medio de escalas Likert (Matas, 2018), ejercicios de cloze testing, es decir, aquellos en los que hay que completar algunos huecos en un texto con la palabra adecuada, o pruebas de memoria en las que se evalúa la cantidad de información que el lector recuerda tras leer el resultado de la traducción.
- e) Aceptabilidad: este término se aplica a diferentes campos relacionados con la lingüística y la traducción. En el contexto de la TA se refiere al grado en el que el resultado de la traducción responde a las expectativas y necesidades de los lectores o usuarios. Según Castilho (2016) los usuarios encontrarán la traducción más aceptable cuando responda a sus expectativas a pesar de los errores que pueda contener y, por el contrario, la encontrarán menos aceptable cuando los errores de la TA afecten en alguna medida a la finalidad del texto resultante.
- f) Ranking: este criterio se suele aplicar al análisis contrastivo de los resultados de diferentes sistemas de TA al traducir un mismo texto.
- g) Usabilidad: este término originado en contextos digitales se suele aplicar al grado de dificultad con el que los usuarios de diferentes entornos digitales son capaces de interactuar con ellos. Aplicado a la TA, este término ha sido definido de maneras diversas. En respuesta a dicha inconsistencia, Doherty y O'Brien (2014) proponen que se adopte la definición de un estándar internacional de calidad, la ISO/TR 16982, que define la usabilidad como “la medida en la que un producto puede ser utilizado por usuarios específicos para lograr unos fines específicos con efectividad, eficiencia y satisfacción en un contexto de uso específico” (International Organization for Standardisation, 2002).
- h) Actuación: por medio de este criterio se evalúa el modo en el que los usuarios utilizan un producto o servicio en el que hay un texto traducido por medio de TA. Se puede evaluar por medio de parámetros objetivos, como el tiempo empleado en la lectura del texto o la interacción digital con dicho texto, o subjetivos, es decir, recogiendo la opinión de los usuarios.

A los criterios aquí recogidos se les ha de añadir el criterio de inteligibilidad, es decir, el grado de facilidad con el que un lector puede leer y entender una traducción. Este criterio engloba los que la lista recién mencionada denomina “legibilidad” y “comprensibilidad”. Tal y como se explicó anteriormente, el criterio de legibilidad se centra en el texto mientras que el de comprensibilidad lo hace en el lector. El criterio de inteligibilidad no realiza esta diferenciación, muchas veces difícil de llevar a cabo, y valora la facilidad de comprensión del texto de un modo general (Fiederer y O'Brien, 2009). La inteligibilidad es uno de los criterios fundamentales en todo proceso de

evaluación de la traducción (Van Slype, 1979; Roturier, 2006; Bawa y Kumar, 2021). Arnold *et al.* (1994: 169) afirman que el criterio de inteligibilidad es “un modo tradicional de evaluar la calidad de la traducción”. La inteligibilidad puede verse afectada por errores gramaticales, confusiones en la traducción y palabras no traducidas. Según estos autores, un alto grado de inteligibilidad suele implicar un alto grado de adecuación o fidelidad. Dabbadie *et al.* (2002: 14) también afirman que la inteligibilidad es una de las métricas más frecuentes en la evaluación humana de la TA.

Como se puede apreciar, los límites entre algunos criterios son difusos y dependen en gran medida del contexto y de la finalidad de la traducción. Por esta razón, a la hora de evaluar la calidad de una traducción concreta, deberán de ser las personas responsables de dicha tarea quienes elijan los criterios que mejor se ajusten a las circunstancias.

3. Objetivos

El objetivo principal de este trabajo es evaluar la utilidad de la TA por medio del análisis de la eficacia de GT en un contexto de crisis humanitaria, como la derivada de la guerra entre Rusia y Ucrania, en el que personas, en ocasiones sin conocimientos de lenguas extranjeras, se ven obligadas a interactuar con personas e instituciones de otros países. Nuestra investigación se centra, por tanto, en el uso de la TA por personas ajenas al campo de la traducción en contextos reales en los que prima la función comunicativa del lenguaje. Por esta razón, los criterios de evaluación de la traducción en los que nos centraremos son aquellos que estudian los factores que están directamente vinculados con la eficacia en un contexto de necesidad comunicativa inmediata. Concretamente, nos hemos centrado en la fluidez y la inteligibilidad. Como es lógico, los criterios de fluidez e inteligibilidad están conectados entre sí, puesto que, como se dijo en el apartado anterior, los errores gramaticales pueden dificultar la comprensión, de modo que la alta o baja calidad de la fluidez redundará en la inteligibilidad. Sin embargo, cada criterio enfatiza unos aspectos distintos, de manera que, al evaluar ambos, se obtiene un resultado más detallado de la calidad de la TA.

Para llevar a cabo nuestra investigación, se utilizó la herramienta de reconocimiento de voz de GT, de manera que se pudo comprobar su grado de eficacia. Por esta razón, en este trabajo ofreceremos también algunos datos complementarios de dicha herramienta, que influyen en la eficacia de la comunicación y que serán objeto de futuras investigaciones.

4. Metodología

Como ya se ha dicho con anterioridad, este trabajo pretende analizar la eficacia de la TA en contextos ajenos al mundo de la traducción profesional, en los que se lleva a cabo una interacción comunicativa inmediata. Nuestra intención es, en concreto, comprobar hasta qué punto el uso de una herramienta de TA, como es el caso de GT, puede facilitar la comunicación inmediata entre personas de cualquier ámbito de la

sociedad en una situación de emergencia humanitaria, como la que se produjo durante los primeros meses de la guerra entre Rusia y Ucrania, en los que personas de contextos muy diferentes entraron en contacto con refugiados ucranianos, es decir, en una situación en la que la finalidad comunicativa está relacionada con las necesidades básicas de los usuarios de dichas herramientas. Por esta razón, al diseñar el procedimiento de la recolección y análisis de datos, se intentó que ambas fases estuvieran conectadas con el contexto real en el que estas comunicaciones tienen lugar.

La primera fase de esta investigación consistió en la recolección de datos para nuestro análisis. Para esta labor, contamos con la colaboración del Centro de Idiomas de la Universitat de València, que durante los meses posteriores al inicio del conflicto bélico entre Rusia y Ucrania ofreció clases de español gratuitas para personas de origen ucraniano. Al plantearnos la posibilidad de realizar esta investigación, tuvimos en cuenta la situación tan delicada en la que se encontraban las personas que formarían parte de nuestro estudio. Por esta razón, antes de llevarla a cabo se realizaron diversas entrevistas con los responsables del Centro de Idiomas de la Universitat de València, quienes, a su vez, explicaron la actividad a los refugiados ucranianos que asisten a sus cursos de castellano y pidieron voluntarios para participar en nuestra investigación. Finalmente, el grupo que formó parte de nuestra investigación estuvo compuesto por un total de treinta y dos personas de edades comprendidas entre 25 y 55 años, de los cuales veintinueve eran mujeres y tan solo tres hombres. Este hecho se debe a que la mayoría de hombres ucranianos han permanecido en su país para luchar en la guerra.

En primer lugar, se pidió a los participantes que respondieran a un breve cuestionario con las siguientes preguntas:

1. ¿Qué otros idiomas conoces además del ucraniano?
2. ¿Qué nivel tienes en esos idiomas?
3. ¿Habías utilizado algún traductor automático antes de salir de Ucrania?
4. ¿Has utilizado algún traductor automático desde que llegaste a España? ¿Cuáles?

La finalidad de este cuestionario era obtener una visión general de su contexto comunicativo, es decir, de los conocimientos y herramientas de los que disponían para comunicarse en su día a día con las personas de su entorno. Posteriormente, dentro de la fase de recogida de datos para el análisis, se mantuvieron entrevistas individuales de alrededor de diez minutos de duración con ayuda de la herramienta de reconocimiento de voz de GT (de ahora en adelante GT) en su versión para PC, que fueron grabadas con la aplicación Xbox, que permite la grabación de la pantalla del ordenador y del sonido externo. Durante las entrevistas, los investigadores hablaban en castellano y las personas entrevistadas lo hacían en ucraniano. Se utilizó la herramienta de reconocimiento de voz para agilizar el proceso comunicativo. El contenido de estas entrevistas giró en torno a la información personal de los participantes y a las experiencias vividas desde su llegada a España. Con la finalidad de recrear el contexto de una conversación de la vida ordinaria, no se utilizó un cuestionario predefinido, sino que las preguntas variaron en función de la información aportada por cada participante, de manera que se establecieron diálogos que pretendían asemejarse en cierta medida a los que, presumiblemente, los refugiados mantienen con las personas de su entorno. Por medio de estas conversaciones, se pudo confeccionar un corpus de textos traducidos

de ucraniano a castellano por medio de la TA, que, como se explicará más adelante, fue objeto de evaluación.

El criterio de fidelidad, también denominado adecuación, que evalúa hasta qué punto el resultado de la TA transmite el significado del texto de origen, no es objeto directo de nuestro análisis. Sin embargo, es evidente que la fidelidad es un criterio fundamental en la evaluación de la TA, por esta razón, se tuvo en cuenta durante el proceso de recogida de datos. Para ello, se contó con la colaboración de dos personas de origen ucraniano con un alto conocimiento de castellano que estuvieron presentes durante las entrevistas, a las que se les solicitó que supervisaran la fidelidad de la traducción de ucraniano a castellano realizada por GT, señalando aquellas traducciones en las que el contenido semántico del texto meta era significativamente distinto al del texto original.

Además, para garantizar la fiabilidad de nuestro análisis, durante las entrevistas se pidió a los participantes que, después de responder oralmente, comprobaran si el texto que aparecía en la ventana en ucraniano de GT recogía fielmente lo que habían dicho y, en caso de no ser así, que volvieran a decirlo o que reformularan dicha información. Las intervenciones en las que el reconocimiento de voz no había funcionado correctamente no se recogieron en los textos que fueron evaluados. Una vez finalizadas las entrevistas, se transcribieron los textos traducidos de ucraniano a castellano resultantes de la TA, que se dividieron en frases para su posterior evaluación según los criterios de fluidez y legibilidad. En total, el corpus de análisis estuvo compuesto por 487 frases.

La segunda fase de nuestra investigación consistió en la evaluación de los resultados de la TA. Con la finalidad de que dicha evaluación reflejara el contexto real del uso de la TA, se seleccionó a un total de ocho evaluadores ajenos al mundo de la traducción y de la especialización en lenguas extranjeras. Los evaluadores elegidos tenían entre 25 y 65 años y habían cursado estudios superiores. La evaluación de todas las categorías se llevó a cabo por medio de una escala de 1 a 4, en la que cada puntuación respondía a un grado de calidad definido, como se puede apreciar en las Tablas 1 y 2. Ambas escalas fueron diseñadas de acuerdo con el modelo propuesto por la Translation Automation User Society (TAUS, 2013).

Fluidez	Categoría	Descripción
4	Perfecta	Se refiere a un texto sin ningún tipo de error
3	Buena	Se refiere a un texto que se lee fluidamente, aunque haya algunos errores poco relevantes
2	Poco fluida	Se refiere a un texto que está mal escrito y es difícil de entender
1	Incomprensible	Se refiere a un texto que está mal escrito y no se comprende

Tabla 1. Escala de evaluación de la fluidez

Inteligibilidad	Categoría	Descripción
4	Todo	Se entiende todo el significado de la traducción
3	La mayoría	Se entiende casi todo el significado de la traducción
2	Poco	Se entienden fragmentos del significado de la traducción
1	Nada	No se entiende nada del significado de la traducción

Tabla 2. Escala de evaluación de la inteligibilidad

Existe una amplia propuesta de herramientas de evaluación (Callison-Burch *et al.* 2007; Przybicki *et al.* 2009; Graham *et al.*, 2013; Chatzikoumi, 2020; Freitag *et al.*, 2021). En nuestro trabajo decidimos seguir la propuesta desarrollada por la Translation Automation User Society (TAUS) en 2011 por la claridad y simplicidad de dichas escalas, ya que, como se indicó anteriormente, las personas encargadas de la evaluación no formaban parte del mundo de la lingüística ni de la traducción profesional.

En el siguiente apartado se muestran los resultados del proceso de evaluación de los criterios de fidelidad e inteligibilidad. Para la representación gráfica de dichos resultados, nos hemos inspirado en la propuesta realizada por Kim y Colominas (2020: 38-39) puesto que algunos de los apartados de su investigación guardan cierta semejanza con el presente trabajo tanto en los criterios evaluados como en la metodología de evaluación de los resultados de la TA.

5. Resultados

5.1. Resultados de la evaluación de la fluidez y la inteligibilidad

Tal como se mencionó en el apartado de la metodología, aunque los criterios de calidad de la TA que son objeto de análisis en este trabajo son los de fluidez e inteligibilidad, durante el proceso de recogida de datos se tuvo también en cuenta el criterio de fidelidad, debido a su relevancia para la calidad de la traducción. La evaluación de este criterio fue llevada a cabo por dos personas de origen ucraniano con un alto nivel de conocimiento de castellano. Al no tratarse de uno de los objetivos de nuestro trabajo, la evaluación de este criterio no siguió la metodología empleada para los criterios de fluidez e inteligibilidad. En esta ocasión, se solicitó a los evaluadores que señalaran aquellas frases en las que el significado del texto original no se mantuviera en absoluto en el texto meta. Como resultado de esta evaluación, se concluyó que el 9,3% de las frases traducidas no mantenía nada del significado del texto original.

Para la representación gráfica de los resultados ofrecidos por los evaluadores, nos hemos inspirado en la propuesta realizada por Kim y Colominas (2020: 38-39). En la Tabla 3 se muestra la puntuación media que cada evaluador asignó a los resultados de GT según los criterios de fluidez e inteligibilidad. Como se puede apreciar, la media

total de ambos criterios es muy cercana, aunque el criterio de inteligibilidad recibió una evaluación un 3,13% superior al de fluidez.

	E1	E2	E3	E4	E5	E6	E7	E8	Media
Fluidez	2,85	2,57	2,65	2,78	2,74	2,99	2,81	2,68	2,75
Porcentaje	71,25 %	64,25 %	66,25 %	69,50 %	68,50 %	74,75 %	70,25 %	67%	68,7 %
Intelig.	3,09	2,54	2,68	3,13	2,64	3,22	2,99	2,82	2,88
Porcentaje	77,25 %	63,50 %	67%	78,25 %	66%	80,50 %	74,75 %	70,50 %	72%

Tabla 3. Resultados por evaluador (E) de los criterios de fluidez e inteligibilidad

Aunque las puntuaciones asignadas por los evaluadores varían, todas ellas asignan a las traducciones un valor superior al 60% tanto en la fluidez como en la inteligibilidad, con una puntuación media cercana al 70% en el primer criterio y algo superior a este porcentaje en el segundo.

En los Gráficos 1 y 2, se muestran con más detalle el porcentaje de puntuaciones asignadas por cada evaluador a los criterios de fluidez e inteligibilidad.

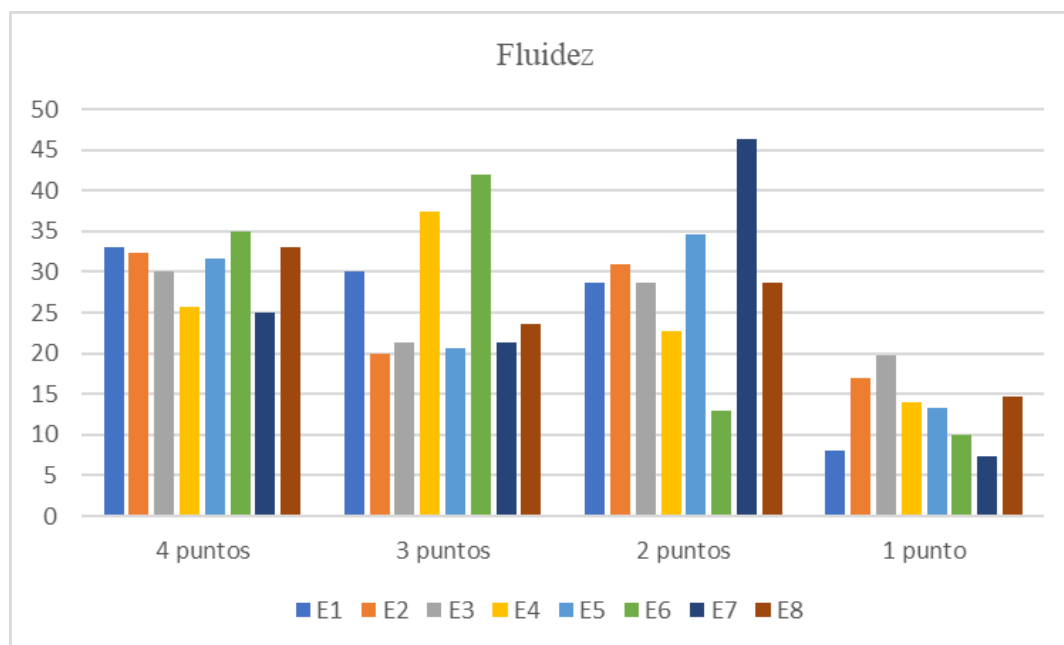


Gráfico 1. Distribución de los resultados del criterio de fluidez

En el Gráfico 1, se puede apreciar cierta homogeneidad tanto en el porcentaje de frases traducidas que fueron evaluadas con la máxima calificación como en el de aquellas que recibieron la mínima. Sin embargo, existe una mayor diversidad en aquellas que recibieron 2 o 3 puntos. En las Tablas 2 y 3 se ofrecen algunos ejemplos de los

textos traducidos de ucraniano a castellano por GT valorados por los evaluadores según el criterio de fluidez (Tabla 2) y de inteligibilidad (Tabla 3).

	E1	E2	E3	E4	E5	E6	E7	E8
Y trabajo como especialista en exportación, los pedidos se envían a Alemania y Polonia, pero actualmente no hay trabajo.	3	4	4	3	3	4	3	4
Llegué a Volyn hace un mes y fue muy aterrador porque no tenía donde vivir pero ahora vivo con mi amigo del trabajo y fue muy difícil pelear en español.	3	3	2	3	3	2	3	3
Creo que es un poco diferente porque hace mucho calor en España y la ropa es más abierta en Ucrania, pero para los hombres, ¿cuál es la diferencia?	2	2	2	3	3	2	2	3
Y vine aquí solo, mi familia se quedó en Ucrania y no podía irse.	4	4	4	4	4	4	4	4
Cada día no es como otro así que cada día es algo nuevo y tengo muchas impresiones porque veo algo nuevo cada día algo interesante y no puedo decir eso porque hay muchas emociones intensas.	1	1	2	2	1	1	2	1

Tabla 2. Ejemplos de frases evaluadas según el criterio de fluidez

En el Gráfico 2 se puede ver que el bloque con mayores porcentajes es el de aquellas frases que fueron evaluadas con la máxima puntuación. También se puede apreciar que los porcentajes de frases evaluadas con 1 punto está por debajo del 20% en todos los casos e incluso por debajo del 10% en los resultados de tres evaluadores.

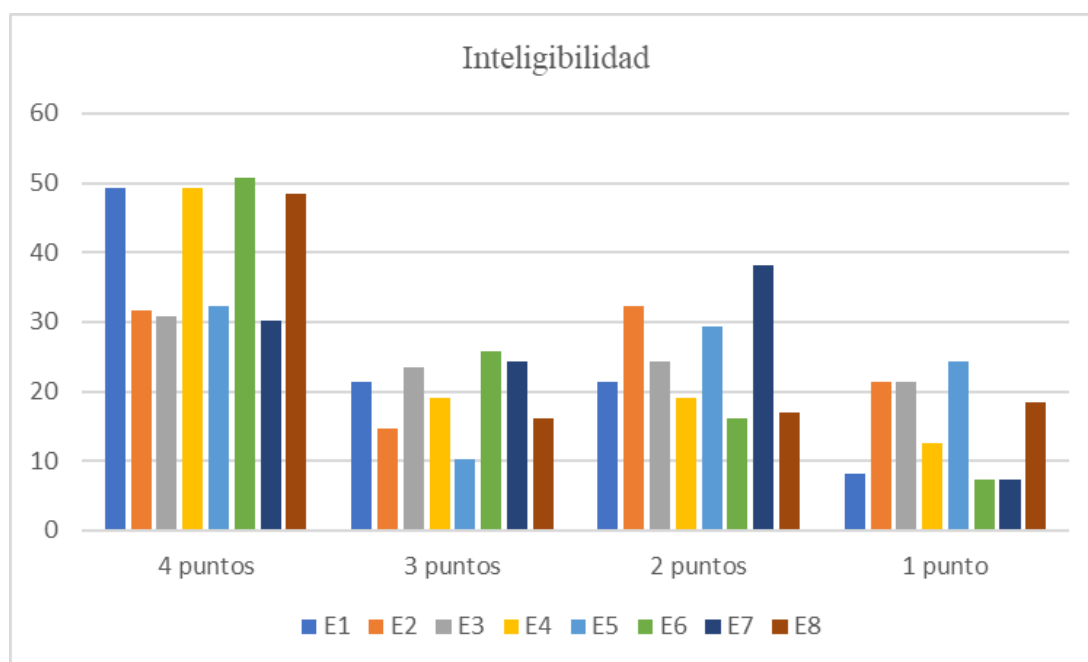


Gráfico 2. Distribución de los resultados del criterio de inteligibilidad

	E1	E2	E3	E4	E5	E6	E7	E8
Y trabajo como especialista en exportación, los pedidos se envían a Alemania y Polonia, pero actualmente no hay trabajo.	4	4	4	4	4	4	4	4
Llegué a Volyn hace un mes y fue muy aterrador porque no tenía donde vivir pero ahora vivo con mi amigo del trabajo y fue muy difícil pelear en español.	3	3	3	4	3	4	3	3
Creo que es un poco diferente porque hace mucho calor en España y la ropa es más abierta en Ucrania, pero para los hombres, ¿cuál es la diferencia?	3	4	3	3	3	4	3	3
Y vine aquí solo, mi familia se quedó en Ucrania y no podía irse.	4	4	4	4	4	4	4	4
Cada día no es como otro así que cada día es algo nuevo y tengo muchas impresiones porque veo algo nuevo cada día algo interesante y no puedo decir eso porque hay muchas emociones intensas.	2	2	3	2	2	3	3	2

Tabla 3. Ejemplos de frases evaluadas según el criterio de inteligibilidad

5.2. Otros datos relevantes del proceso de TA

En este apartado ofreceremos algunos datos sobre la calidad de la TA al utilizar la herramienta de voz de GT, puesto que influyen en el proceso de comunicación. Sin embargo, en este artículo no profundizaremos en este tema, puesto que será objeto de análisis de futuras investigaciones.

Tal como se explicó en el apartado 4, durante las entrevistas se solicitó a los participantes que revisaran las transcripciones realizadas por el sistema de reconocimiento de voz de GT para comprobar si eran correctas. Aunque los errores que se produjeron no afectan a nuestros resultados, ya que esas frases no se incluyeron en las evaluaciones, nos parece relevante señalar que en el 15% de las intervenciones se detectaron errores de transcripción.

Cada vez que se dio un error de este tipo, se solicitó al participante que repitiera lo que había dicho y, ante la persistencia de algunos de estos errores, los participantes tuvieron que reformular sus palabras utilizando estructuras más sencillas. Algunos de los errores que se detectaron fueron, por una parte, la incapacidad de GT para reconocer que se estaba formulando una pregunta. También se detectó que esta herramienta se equivocaba con frecuencia cuando las intervenciones contenían varias oraciones, puesto que, de manera habitual, no se detectaba el final de una oración y el principio de la siguiente, lo que daba lugar a errores en la traducción. Como consecuencia de este hecho, algunos de los participantes modificaron su manera de hablar durante las entrevistas, pasando de intervenciones largas con todo tipo de oraciones a otras más breves buscando estructuras sintácticas más sencillas.

Un error de traducción recurrente en las entrevistas fue la asignación equivocada del género, ya que GT traducía frecuentemente con género masculino, sustantivos y adjetivos que debían ser femeninos. También se detectó un uso inconsistente de los signos de puntuación y de la acentuación de ciertas palabras. Sin embargo, a pesar de los errores señalados, en todas las entrevistas fue posible mantener un diálogo fluido en el que se intercambió información sobre temas variados.

6. Discusión

En el marco teórico de este artículo se ofreció una breve explicación sobre la evaluación de la calidad de la TA y se señalaron algunos de los criterios principales que se suelen tener en cuenta al llevar a cabo esta tarea. Existen numerosas propuestas de metodologías y taxonomías para realizar estos análisis y no hay duda de que es un tema que sigue en desarrollo. Sin embargo, uno de los puntos comunes de las diferentes propuestas es que, a la hora de evaluar la calidad de una traducción, ya sea humana o automática, un factor determinante es su finalidad, es decir, el propósito para el cual la ha solicitado el usuario final de dicha traducción. Por lo tanto, se puede afirmar 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.

En el contexto de una persona que se ha visto obligada a abandonar su país de origen y buscar refugio en otro país, cuyo idioma desconoce, no hay duda de que la finalidad de la traducción es práctica y comunicativa, es decir, lo importante para esa persona es ser capaz de comunicar y comprender la información más relevante para su supervivencia, en una primera fase, y para lograr cierto bienestar, en una fase posterior.

Según los resultados del cuestionario que respondieron los participantes en nuestro estudio, solo un 20% de ellos tenía un nivel de inglés suficiente para poder comunicarse con fluidez, y tan solo un 6% sabía algo de castellano antes de llegar a España. Por esta razón, es lógico que todos los participantes afirmaran haber hecho uso de la TA desde que abandonaron su país, aunque solo un 28% la había usado en diferentes ocasiones cuando todavía estaba en Ucrania.

Teniendo en cuenta este contexto, los resultados de la evaluación de los criterios de fluidez e inteligibilidad y los datos adicionales ofrecidos en el apartado 5.2, podemos afirmar que, a pesar de que todavía existe un amplio margen de mejora, la TA es una herramienta de gran utilidad en un contexto de emergencia humanitaria, como el provocado por la guerra entre Rusia y Ucrania. Si se combinan los datos de la evaluación de la fluidez y la inteligibilidad, se puede concluir que en más de un 60% de las ocasiones, los resultados de la traducción realizada por GT son perfectos o casi perfectos, cerca del 30% restante dichos resultados permiten comprender fragmentos de la información original, mientras que alrededor de un 10% de las interacciones serán 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.

Tal como se señaló en la sección de resultados, la puntuación media de la inteligibilidad fue ligeramente superior a la de la fluidez. Estos resultados se corresponden con el hecho de que habitualmente el lector de una traducción será capaz de inferir el significado a pesar de que contenga algunos errores gramaticales sin demasiada relevancia.

Estas cifras resultarían insuficientes en un contexto en el que se pretende que la TA ofrezca un resultado lo más cercano posible al de un traductor profesional. Sin embargo, se ha de tener en cuenta que, en un contexto de interacción comunicativa inmediata, en el que los interlocutores se encuentran presentes, existen factores adicionales que refuerzan la comunicación, como es la capacidad de inferencia del interlocutor, el lenguaje no verbal, la posibilidad de repetir o reformular lo que se ha dicho, el contexto, etc. De manera que las carencias de la TA se podrían compensar con estos factores y probablemente no impedirían que se llevara a cabo una comunicación efectiva.

La TA es, por tanto, una herramienta efectiva que permitirá una comunicación básica entre personas que hablan diferentes idiomas. Sin embargo, las limitaciones mencionadas en el apartado 5.2 afectan al proceso comunicativo cuando este implica interacciones de larga duración o de contenidos de mayor complejidad. El uso de la TA implica un protocolo comunicativo distinto del natural, con un proceso que ralentiza la interacción de los participantes, puesto que cada mensaje ha de ser reconocido por la aplicación de TA, traducido y leído o escuchado por el otro participante que, a su vez, responderá

del mismo modo. Tal como se dijo, el porcentaje de error del reconocimiento de voz fue de un 15% durante las entrevistas realizadas en nuestro estudio. Aunque este dato no es extrapolable al resto de interacciones, es representativo de que, con cierta frecuencia, será necesario repetir o reformular lo que se ha dicho. Además, el uso de la TA obliga a simplificar el lenguaje y las estructuras y a acortar los mensajes para permitir que la aplicación lo pueda traducir correctamente, de manera que es necesaria cierta capacidad de reformulación en los participantes para lograr una mayor efectividad del proceso comunicativo que, además, también requerirá habilidades de inferencia en el interlocutor que le permitan comprender aquellas traducciones con un nivel medio-bajo de fluidez e inteligibilidad. Por todas estas razones, se puede afirmar que el uso de la TA durante un periodo de tiempo prolongado o con temáticas complejas exige un esfuerzo adicional por parte de los hablantes que dificulta el proceso de comunicación.

Se puede afirmar, por lo tanto, que la TA dista aún mucho de poder equipararse a la interpretación simultánea o incluso a la interpretación consecutiva, al menos en la traducción entre el castellano y el ucraniano. Para lograr una comunicación más fluida, GT debería mejorar las limitaciones señaladas en el párrafo anterior, permitiendo interacciones más naturales que exijan un esfuerzo menor de los participantes.

7. Conclusión

El desarrollo y el proceso de mejora de los motores de TA es un tema de gran actualidad, debido al auge de estas herramientas en diversos ámbitos de la sociedad. A pesar de las reticencias iniciales de algunos sectores de la traducción profesional, a día de hoy los motores de TA se han convertido en un instrumento que puede facilitar la labor del traductor. Por esta razón, desde hace años se están llevando a cabo numerosas investigaciones orientadas a evaluar la calidad de la TA. Sin embargo, la mayoría de esas investigaciones se llevan a cabo en contextos de traducción profesional.

En el presente artículo hemos ofrecido un primer acercamiento a la evaluación de la TA, en concreto, del traductor GT, en un contexto diferente al que acabamos de señalar, en el que los usuarios son personas ajenas al mundo de la traducción y utilizan la TA como herramienta con una finalidad comunicativa básica. Los datos obtenidos por medio de nuestro análisis de los criterios de fluidez e inteligibilidad nos permiten afirmar que la TA es una herramienta de gran utilidad en un contexto de emergencia humanitaria, como es el provocado por la guerra entre Rusia y Ucrania, puesto que permite que un alto porcentaje de la información sea comprensible para los interlocutores. Sin embargo, todavía es necesario implementar ciertas mejoras en las aplicaciones de TA para lograr un proceso comunicativo más fluido.

Nuestro estudio es, por tanto, una contribución innovadora al campo de la evaluación de la calidad de la TA. Somos conscientes, sin embargo, de las limitaciones de este trabajo, que no pretende extraer conclusiones generales sobre el uso de la TA en contextos reales. Este trabajo ha de ser completado con futuras investigaciones en las que se incluyan otros traductores automáticos, otros idiomas, corpus con un número

más amplio de textos, y en los que se analice con más detalle todo el proceso comunicativo que incluye el uso de la TA.

BIBLIOGRAFÍA

- Alarcón Navío, E. (2003). Traducción automática versus traducción humana: tipología de errores. In: Actas del I Congreso Internacional de la Asociación Ibérica de Estudios de Traducción e Interpretación: Granada, 12-14 de Febrero de 2003. Granada: AIETI, pp. 721-738.
- Arnold, D.; et al. (1994). Machine Translation: An Introductory Guide. Oxford: NCC Blackwell.
- Bawa, S.; Kumar, M. (2021). A comprehensive survey on machine translation for English, Hindi and Sanskrit languages. Journal of Ambient Intelligence and Humanized Computing. <<https://doi.org/10.1007/s12652-021-03479-0>>. [Accessed: 20220715].
- Cadwell, P.; O'Brien, S.; DeLuca, E. (2019). More than tweets: A critical reflection on developing and testing crisis machine translation technology. Translation Spaces, v. 8, n. 2, pp. 300-333. <<https://doi.org/10.1075/ts.19018.cad>>. [Accessed: 20220715].
- Callison-Burch, C.; Fordyce, C. S.; Koehn, P.; Monz, C.; Schroeder, J. (2007). (Meta-) evaluation of machine translation. In: Proceedings of the Second Workshop on Statistical Machine Translation, Prague, June 2007, pp. 136-158. <<https://doi.org/10.3115/1626355.1626373>>. [Accessed: 20220711].
- Canfora, C.; Ottmann, A. (2020). Risks in neural machine translation. Translation Spaces, v. 9, n. 1, pp. 58-77. <<https://doi.org/10.1075/ts.00021.can>>. [Accessed: 20220711].
- Casanova, M. A. (2022). Cómo escolarizamos a los refugiados ucranianos. El Correo (abril). <<https://www.elcorreo.com/internacional/escolarizamos-refugiados-ucranianos-20220405102947-ntrc.html>>. [Accessed: 20220510].
- Castilho, S. (2016) Measuring acceptability of machine translated enterprise content. PhD thesis. School of Applied Language and Intercultural Studies, Dublin City University. <[https://doras.dcu.ie/21342/1/Castilho\(2016\)_Acceptability_of_Machine_Translated_Enterprise_Content_PhDThesis.pdf](https://doras.dcu.ie/21342/1/Castilho(2016)_Acceptability_of_Machine_Translated_Enterprise_Content_PhDThesis.pdf)>. [Accessed: 20220711].
- Castilho, S.; Doherty, S.; Gaspari, F.; Moorkens, J. (2018). Approaches to human and machine translation quality assessment. In: Moorkens, J.; Castilho, S.; Gaspari, F.; Doherty, S. (eds.). Translation quality assessment: From Principle to Practice. Charm: Springer, pp. 9-38. <https://doi.org/10.1007/978-3-319-91241-7_2>. [Accessed: 20220712].
- Castilho, S.; Moorkens, J.; Gaspari, F.; Calixto, I.; Tinsley, J.; Way, A. (2017). Is neural machine translation the new state of the art? The Prague Bulletin of Mathematical Linguistics, n. 108 (June), pp. 109-120. <<http://dx.doi.org/10.1515/pralin-2017-0013>>. [Accessed: 20220717].

- Chatzikoumi, E. (2020). How to evaluate machine translation: A review of automated and human metrics. *Natural Language Engineering*, v. 26, n. 2, pp. 137-161.
<<http://dx.doi.org/10.1017/S1351324919000469>>. [Accessed: 20220715].
- Doherty S.; O'Brien S. (2014) Assessing the usability of raw machine translated output: a usercentred study using eye tracking. *International Journal of Human-Computer Interaction*, v. 30, n. 1, pp. 40-51.
<<http://dx.doi.org/10.1080/10447318.2013.802199>>. [Accessed: 20220720].
- España. Ministerio del Interior (2022), 26 de abril). España ha concedido 75.000 protecciones temporales a desplazados por la guerra en Ucrania. La Moncloa (26 abril).
<<https://www.lamoncloa.gob.es/serviciosdeprensa/notasprensa/interior/Paginas/2022/260422-desplazados-ucranianos.aspx>>. [Accessed: 20220505].
- Federici, F. M.; O'Brien, S. (eds.). (2020). *Translation in cascading crises*. 1st ed. New York: Routledge.
- Fiederer, R.; O'Brien, S. (2009). Quality and machine translation: A realistic objective. *The journal of Specialised translation*, n. 11, pp. 52-74.
<https://www.jostrans.org/issue11/art_fiederer_obrien.pdf>. [Accessed: 20220715].
- Freitag, M.; Foster, G.; Grangier, D.; Ratnakar, V.; Tan, Q.; Macherey, W. (2021). Experts, errors, and context: A large-scale study of human evaluation for machine translation. *Transactions of the Association for Computational Linguistics*, n. 9, pp. 1460-1474. <https://doi.org/10.1162/tacl_a_00437>. [Accessed: 20220715].
- Graham, Y.; Baldwin, T.; Moffat, A.; Zobel, J. (2013). Continuous measurement scales in human evaluation of machine translation. In: *Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse*: Sofia, Bulgaria, August 8-9 2013, pp. 33-41. <<https://aclanthology.org/W13-2305.pdf>>. [Accessed: 20220711].
- Hamon, O.; Hartley, A.; Popescu-Belis, A.; Choukri, K. (2007). Assessing human and automated quality judgments in the French MT evaluation campaign CESTA. In: *Proceedings of Machine Translation Summit XI*, pp. 231-238.
<<https://aclanthology.org/2007.mtsummit-papers.31.pdf>>. [Accessed: 20220711].
- Han, L.; Jones, Gareth, J. F.; Smeaton, A. F. (2021). Translation Quality Assessment: A Brief Survey on Manual and Automatic Methods. In: *Proceedings for the First Workshop on Modelling Translation: Translatology in the Digital Age*, pp. 15-33.
<<https://arxiv.org/pdf/2105.03311.pdf>>. [Accessed: 20221012].
- Helguera, D. (2022). Integración laboral para refugiados ucranianos. *Radio Nervión* (7 octubre). <<https://www.radionervion.com/2022/10/07/integracion-laboral-para-refugiados-ucranianos/>>. [Accessed: 20221012].
- House, J. (2014). Translation quality assessment: Past and present. In: *Translation: A multidisciplinary approach*. London: Palgrave Macmillan, pp. 241-264.
<http://dx.doi.org/10.1057/9781137025487_13>. [Accessed: 20220715].

- Hutchins, J. (2001). Machine translation and human translation: in competition or in complementation. *International Journal of Translation*, n. 13 (1-2 Jan-Dec.), pp. 5-20. <https://www.translationdirectory.com/article404.htm>. [Accessed: 20220715].
- Hutchins, W. J.; Somers, H. L. (1992). *An Introduction to Machine Translation*. London [etc.]: Academic Press.
- Instituto Cervantes (2022). El Instituto Cervantes facilitará el aprendizaje del español a los refugiados ucranianos del programa de Cruz Roja Española. Instituto Cervantes. (19 mayo). https://www.cervantes.es/sobre_instituto_cervantes/prensa/2022/noticias/aprendizaje-refugiados-ucranianos.htm#:~:text=El%20Instituto%20Cervantes%20ha%20firmado,a%20sus%20cursos%20en%20el%20C3%ADnea. [Accessed: 20220718].
- International Organization for Standardisation (2002). ISO/TR 16982:2002 ergonomics of human-system interaction—usability methods supporting human centred design. Geneva: International Organization for Standardisation. Available at: http://www.iso.org/iso/catalogue_detail?csnumber=31176. [Accessed: 20220713].
- Kim, A.; Colominas Ventura, C. (2020). Human evaluation of NMT& annual progress report: A case study on Spanish to Korean. *Revista Tradumàtica: technologies de la traducció*, n. 18 (diciembre), pp. 33-48. <http://dx.doi.org/10.5565/rev/tradumatica.241>. [Accessed: 20220715].
- Koby, G. S.; Fields, P.; Hague, D. R.; Lommel, A.; Melby, A. (2014). Defining translation quality. *Revista Tradumàtica: technologies de la traducció*, n. 12 (diciembre), pp. 413-420. <http://dx.doi.org/10.5565/rev/tradumatica.76>. [Accessed: 20220715].
- Koehn, P. (2009). *Statistical machine translation*. Cambridge: Cambridge University.
- Lewis, W.; Munro, R.; Vogel, S.; (2011). Crisis MT: Developing a cookbook for MT in crisis situations. In: *Proceedings of the Sixth Workshop on Statistical Machine Translation* (July), pp. 501-511. <https://nlp.stanford.edu/pubs/Munro2011wmt.pdf>. [Accessed: 20220715].
- Matas, A. (2018). Diseño del formato de escalas tipo Likert: un estado de la cuestión. *Revista electrónica de investigación educativa*, v. 20, n. 1, pp. 38-47. <http://dx.doi.org/10.24320/redie.2018.20.1.1347>. [Accessed: 20220721].
- Mendoza, M. H. (2017). La posesición de traducciones de textos técnicos del alemán al castellano. Tesis doctoral. Departament de Traducció i d'Interpretació i d'Estudis de l'Àsia Oriental, Universitat Autònoma de Barcelona. <https://www.tdx.cat/handle/10803/457362?locale-attribute=es#page=1>. [Accessed: 20220721].
- Naciones Unidas (2022). Operational Data Portal: Ukraine Refugee Situation. <https://data.unhcr.org/en/situations/ukraine>. [Accessed: 20220622].

- O'Brien, S. (2019). Translation technology and disaster management. In: The Routledge handbook of translation and technology. Routledge: Abingdon, Oxon [etc.]: Routledge, pp. 304-318.
- Oliver, A. (2020). Human translation and machine translation: Specificities, uses, advantages and disadvantages. In: Arenas, C. (coord.). Translating Diversity = Traduir la diversitat. [S.l.]: Linguapax, pp. 111-130.
- Przybocki, M.; Peterson, K.; Bronsart, S.; Sanders, G. (2009). The NIST 2008 Metrics for machine translation challenge-overview, methodology, metrics, and results. Machine Translation, v. 23, n. 2, pp. 71-103. <<https://doi.org/10.1007/s10590-009-9065-6>>. [Accessed: 20220728].
- Ramlow, M. (2008). Maschinelle Übersetzungssysteme im Vergleich. Maschinelle Übersetzung und XML im Übersetzungsprozess, 19.
- Reiss, K.; Vermeer, H. J. (2014). Towards a general theory of translational action: Skopos theory explained. Abingdon, Oxon [etc.]: Routledge.
- Rico, C. (2017). La formación de traductores en Traducción Automática. Revista Tradumàtica: tecnologies de la traducció, n. 15 (deseembre), pp. 75-96. <<http://dx.doi.org/0000-0002-5056-8513>>. [Accessed: 20220715].
- Roturier, J. (2006). An investigation into the impact of controlled English rules on the comprehensibility, usefulness and acceptability of machine-translated technical documentation for French and German users. PhD Tesis. Dublin City University. <<https://doras.dcu.ie/18190>>. [Accessed: 20220715].
- Specia, L.; Raj, D.; Turchi, M. (2010). Machine translation evaluation versus quality estimation. Machine translation, v. 24, n. 1, pp. 39-50. <<http://dx.doi.org/10.1007/s10590-010-9077-2>>. [Accessed: 20220715].
- TAUS. (2013). TAUS: the language data network. <<https://www.taus.net/>>. [Accessed: 20220715].
- TAUS. (2013). Adequacy/Fluency Guidelines. <<https://www.taus.net/academy/best-practices/evaluate-best-practices/adequacy-fluencyguidelines>>. [Accessed: 20220715].
- Translators without Borders (2016). Translators without Borders Develops the World's First Crisis-Specific Machine Translation System for Kurdish Languages. Accessed June 25, 2019. <https://translatorswithoutborders.org/translators-without-borders-develops-worlds-first-crisisspecific-machine-translation-system-kurdish-languages/>
- Van Slype, G. (1979). Evaluation of the 1978 version of the SYSTRAN English-French automatic system of the commission of the European communities. The Incorporated Linguist, v. 18, n. 3 (Summer). <<https://aclanthology.org/www.mt-archive.info/70/InLinguist-1979-VanSlype.pdf>>. [Accessed: 20220715].
- White, J. (2003). How to Evaluate Machine Translation. In: Somers, H. (ed.). Computers and Translation: A Translator's Guide. Amsterdam [etc.]: John Benjamins, pp. 211-244. <<http://dx.doi.org/10.1075/btl.35.16whi>>. [Accessed: 20220715].