
This is the **published version** of the article:

Páez Mérida, María; Simon Jimenez, Eduardo, dir. *Avaluació de la traducció automàtica d'imatges mitjançant sistemes de reconeixement de text en dispositius mòbils : Google Translate Images i Microsoft Translator Images.* 2020. (1350 Màster en Tradumàtica: Tecnologies de la Traducció)

This version is available at <https://ddd.uab.cat/record/249927>

under the terms of the  license

FACULTAD DE TRADUCCIÓN E INTERPRETACIÓN

MÁSTER EN TRADUMÁTICA: TECNOLOGÍAS DE LA TRADUCCIÓN

TRABAJO DE FIN DE MÁSTER

Curso 2019-2020

**Evaluación de la traducción automática de imágenes
mediante sistemas de reconocimiento de texto en
dispositivos móviles: Google Translate Images y
Microsoft Translator Images**

María Páez Mérida

TUTOR

EDUARDO SIMÓN

Barcelona, 19 de junio de 2020



Resumen

Este trabajo presenta un ejercicio de evaluación de resultados de traducción automática mediante las aplicaciones móviles de traducción de imágenes Google Translate Images y Microsoft Translator Images, en la combinación de idiomas inglés-español. Se comprueba su potencial funcionalidad para el público general y los profesionales de la traducción. La investigación introduce un marco teórico a modo de aproximación al concepto de calidad en la industria de la traducción, más concretamente en la traducción automática, así como a las distintas aplicaciones de la inteligencia artificial en la traducción de imágenes; en la sección práctica se evalúan y comparan los resultados de TA de Google y Microsoft a partir de un muestrario de imágenes seleccionadas aleatoriamente.

Palabras clave

Traducción automática, evaluación de traducción automática, Google, Microsoft, traducción de imágenes

Resum

Aquest treball presenta un exercici d'avaluació de resultats de traducció automàtica mitjançant les aplicacions mòbils de traducció d'imatges Google Translate Images i Microsoft Translator Images, en la combinació d'idiomes anglès-espanyol. Es comprova la seva potencial funcionalitat per al públic general i els professionals de la traducció. La investigació presenta un marc teòric a mode d'aproximació al concepte de qualitat en la indústria de la traducció, més concretament en la traducció automàtica, així com a les diferents aplicacions de la intel·ligència artificial en la traducció d'imatges; a la secció pràctica s'avaluen i comparen els resultats de TA de Google i Microsoft a partir d'una mostra d'imatges seleccionades aleatòriament.

Paraules clau

Traducció automàtica, avaluació de traducció automàtica, Google, Microsoft, traducció d'imatges

Abstract

This paper presents an exercise in evaluating machine translation results using the mobile image translation applications Google Translate Images and Microsoft Translator Images in the English-Spanish language combination. Its potential functionality for the general public and translation professionals is verified. The research presents a theoretical framework as an approach to the concept of quality in the translation industry, more specifically in machine translation, as well as the different applications of artificial intelligence in image translation; in the practical section, Google and Microsoft MT results are evaluated and compared from a sample of randomly selected images.

Key words

Machine translation, machine translation evaluation, Google, Microsoft, image translation

Índice

1.	INTRODUCCIÓN.....	11
2.	TRADUCCIÓN Y CALIDAD	17
2.1.	Traducción automática y calidad	18
2.2.	Métricas de evaluación de TA.....	19
2.2.1.	BLEU.....	20
2.2.2.	COSTA MT Evaluation Tool	22
3.	TRADUCCIÓN AUTOMÁTICA E INTELIGENCIA ARTIFICIAL.....	26
3.1.	Aplicaciones de la IA en la TA	27
3.1.1.	Reconocimiento óptico de caracteres (ROC)	27
3.1.2.	Traducción Automática Neuronal.....	28
4.	EVALUACIÓN DE GOOGLE TRANSLATE IMAGES.....	29
4.1.	Análisis de los textos	31
4.1.1.	Análisis con BLEU.....	36
4.1.2.	Análisis con COSTA MT Evaluation Tool	40
5.	EVALUACIÓN DE MICROSOFT TRANSLATOR IMAGES	43
5.1.	Análisis de los textos	45
5.1.1.	Análisis con BLEU.....	50
5.1.2.	Análisis con COSTA MT Evaluation Tool	54
6.	COMPARACIÓN DE RESULTADOS	57
6.1.	Comparación con BLEU.....	57
6.2.	Comparación con COSTA MT Evaluation Tool	59
6.3.	Cuestiones técnicas de Google y Microsoft.....	61
6.4.	Cuestiones técnicas de BLEU y COSTA	63
7.	CONCLUSIONES.....	66
8.	BIBLIOGRAFÍA	71
9.	ANEXOS	74

Índice de tablas

Tabla 1. Clasificación para categorizar los resultados de BLEU (2020).....	15
Tabla 2. Resultados de TA de Google Translate Images (2020).....	36
Tabla 3. Resultados de TA de Microsoft Translator Images (2020)	50

Índice de figuras

Figura 1. Fórmula de BLEU para calcular las puntuaciones (2020)	20
Figura 2. Interfaz de <i>Interactive BLEU score evaluator</i> de la empresa Tilde (2020)	21
Figura 3. Presentación de resultados de <i>Interactive BLEU score evaluator</i> (2020).....	22
Figura 4. Interfaz de COSTA MT Evaluation Tool (2020).....	23
Figura 5. Interfaz de COSTA MT Evaluation Tool II (2020)	24
Figura 6. Interfaz de COSTA MT Evaluation Tool III (2020).....	24
Figura 7. Ejemplo de arquitectura codificadora-decodificadora (Casacuberta; Peris, 2017).....	28
Figura 8. Logotipo de Google Lens (Google, 2020)	29
Figura 9. Muestra del funcionamiento de Google Lens (Google, 2020).....	30
Figura 10. Puntuación de BLEU para Google (2020)	36
Figura 11. Gráfico con los resultados de BLEU para Google (2020)	37
Figura 12. Ejemplo 1 de la evaluación de BLEU para Google (2020).....	37
Figura 13. Ejemplo 2 de la evaluación de BLEU para Google (2020).....	38
Figura 14. Ejemplo 3 de la evaluación de BLEU para Google (2020).....	38
Figura 15. Ejemplo 4 de la evaluación de BLEU para Google (2020).....	39
Figura 16. Ejemplo 5 de la evaluación de BLEU para Google (2020).....	39
Figura 17. Ejemplo 6 de la evaluación de BLEU para Google (2020).....	39
Figura 18. Puntuación de COSTA para Google (2020)	40
Figura 19. Puntuaciones de fluidez y adecuación de Google según COSTA (2020)	40
Figura 20. Clasificación de errores de Google según COSTA (2020).....	41
Figura 21. Logotipo de Microsoft Translator (Microsoft, 2020).....	43
Figura 22. Interfaz de Microsoft Translator Images (Microsoft, 2020).....	44
Figura 23. Puntuación de BLEU para Microsoft (2020)	50
Figura 24. Gráfico con los resultados de BLEU para Microsoft (2020)	51
Figura 25. Ejemplo 1 de la evaluación de BLEU para Microsoft (2020).....	51
Figura 26. Ejemplo 2 de la evaluación de BLEU para Microsoft (2020).....	52
Figura 27. Ejemplo 3 de la evaluación de BLEU para Microsoft (2020).....	52
Figura 28. Ejemplo 4 de la evaluación de BLEU para Microsoft (2020).....	52
Figura 29. Ejemplo 5 de la evaluación de BLEU para Microsoft (2020).....	53
Figura 30. Ejemplo 6 de la evaluación de BLEU para Microsoft (2020).....	53
Figura 31. Ejemplo 7 de la evaluación de BLEU para Microsoft (2020).....	53

Figura 32. Puntuación de COSTA para Microsoft (2020)	54
Figura 33. Puntuaciones de fluidez y adecuación de Microsoft según COSTA (2020) .	54
Figura 34. Clasificación de errores de Microsoft según COSTA (2020).....	55
Figura 35. Función de añadir comentarios de COSTA (2020)	56
Figura 36. Comparación de Google y Microsoft con BLEU (2020).....	57
Figura 37. Gráfico con los resultados de BLEU para Google y Microsoft (2020).....	58
Figura 38. Ejemplo 1 de la comparación de BLEU sobre Google y Microsoft (2020)..	58
Figura 39. Ejemplo 2 de la comparación de BLEU sobre Google y Microsoft (2020)..	59
Figura 40. Ejemplo 3 de la comparación de BLEU sobre Google y Microsoft (2020)..	59
Figura 41. Puntuación de COSTA para Google y Microsoft (2020)	59
Figura 42. Puntuaciones de fluidez y adecuación de Google y Microsoft según COSTA (2020)	60
Figura 43. Clasificación de errores de Google y Microsoft según COSTA (2020).....	60
Figura 44. Ejemplo 1 de cuestiones técnicas de Google y Microsoft (2020).....	62
Figura 45. Ejemplo 2 de cuestiones técnicas de Google y Microsoft (2020).....	62
Figura 46. Ejemplo 3 de cuestiones técnicas de Google y Microsoft (2020).....	63
Figura 47. Ejemplo 4 de cuestiones técnicas de Google y Microsoft (2020).....	63
Figura 48. Requisitos de COSTA MT Evaluation Tool (2020)	64
Figura 49. Muestra de los archivos de trabajo en Notepad++. TO, TH, TA de Google y TA de Microsoft (2020)	65

Abreviaturas

BLEU	<i>Bilingual Evaluation Understudy</i>
COSTA	COSTA MT Evaluation Tool
IA	inteligencia artificial
LM	lengua meta
LO	lengua origen
PE	posedición
TA	traducción automática
TABR	traducción automática basada en reglas
TAE	traducción automática estadística
TAN	traducción automática neuronal
TH	traducción humana
TM	texto meta
TO	texto original

1. INTRODUCCIÓN

1.1. Motivación

En los últimos años, las nuevas tecnologías y la traducción han evolucionado juntas a gran ritmo. En la actualidad, es difícil imaginar una industria de la traducción sin tecnologías, ya que se encuentran presentes en prácticamente todas las fases de un proyecto de traducción: en la gestión del proyecto, en la traducción propiamente dicha o en las comunicaciones con clientes o lingüistas. Además, las tecnologías, en concreto la inteligencia artificial, han servido para crear nuevas modalidades de traducción, como la traducción automática neuronal (TAN), y nuevas herramientas de traducción al alcance de todo el mundo, tanto para profesionales de la traducción como para el público general.

Una de las formas empleadas actualmente para hacer llegar esas nuevas herramientas de traducción automática (TA) a todo el público de forma rápida y económica son las aplicaciones móviles. Cada día se crean miles de aplicaciones nuevas con diferentes funcionalidades, por lo que también existen las de traducción, propiciadas en gran medida por la alta demanda de información desde cualquier parte del mundo, es decir, por la globalización. Las primeras aplicaciones consistían en traductores automáticos como los que se pueden encontrar en internet, donde el usuario introduce el texto y obtiene la traducción en pocos segundos. No obstante, hoy en día es fácil comprobar que ese tipo de aplicaciones ha evolucionado de tal forma que algunas incluso ya son capaces de traducir imágenes que contienen texto escrito en un soporte físico.

A pesar de la revolución que han generado, la novedad y la falta de estudio sobre este tipo de aplicaciones en ocasiones ponen en duda su credibilidad, tanto por parte de profesionales como de los usuarios que las utilizan. Esa es la principal motivación que ha llevado a la realización de este trabajo: conocer más sobre estas aplicaciones mediante la investigación y el análisis de su funcionamiento y de su utilidad. En concreto, el objetivo planteado ha sido estudiar las funciones de traducción de imágenes de Google Translate Images y Microsoft Translator Images, dos de las aplicaciones móviles de traducción automática más utilizadas. El principal interrogante está relacionado con el sistema de traducción automática empleado, ya que es el elemento principal en el que se basan este tipo de aplicaciones, con el elemento añadido de la traducción de imágenes, el cual añade otras características específicas al funcionamiento

de la aplicación que también deben ser analizadas, principalmente el reconocimiento óptico de caracteres.

1.2. Objetivos y preguntas de investigación

El objetivo principal de este trabajo es comprobar si las aplicaciones de Google y Microsoft para la traducción de imágenes pueden resultar útiles para el público general sin conocimientos lingüísticos de inglés y para el traductor profesional. Para ello, se ha realizado una evaluación de los resultados de traducción automática de Google y Microsoft en sus respectivas aplicaciones y una posterior comparación entre ambos sistemas. De este modo, se intenta concluir cuál ofrece mejores resultados.

Para conjeturar al respecto, en primer lugar se plantean una serie de cuestiones que sirven de guía y ayuda para lograr el objetivo final. Entre los objetivos secundarios se encuentran: conocer las métricas de evaluación de calidad de TA y acotar el concepto de calidad en el sector de la traducción; ahondar en las aplicaciones de la inteligencia artificial en la traducción; analizar en profundidad el funcionamiento de Google Translate Images y Microsoft Translator Images; por último, realizar una comparación final entre ambos sistemas. Las preguntas a las que se pretende dar respuesta son las siguientes:

1. ¿Qué se entiende por «calidad» en la industria de la traducción? ¿Qué son las métricas de evaluación de calidad de TA? ¿Cómo funcionan?
2. ¿Cómo ha afectado la aplicación de la inteligencia artificial en el mundo de la traducción? ¿Cómo funcionan las aplicaciones para traducir imágenes? ¿Qué es el reconocimiento óptico de caracteres y la traducción automática neuronal?
3. ¿Qué es Google Translate Images? ¿Cómo funciona? ¿Qué sistema de TA utiliza? ¿Qué ventajas e inconvenientes tiene?
4. ¿Qué es Microsoft Translator Images? ¿Cómo funciona? ¿Qué sistema de TA utiliza? ¿Qué ventajas e inconvenientes tiene?
5. ¿Qué aplicación funciona mejor? ¿Cuál ofrece mejores resultados de TA y, por lo tanto, puede resultar más útil para nuestro objetivo? ¿Ofrecen traducciones de calidad? ¿Qué diferencias existen entre ambas aplicaciones?
6. ¿Se puede considerar este tipo de aplicaciones como una herramienta práctica para el usuario general? ¿Y para el traductor profesional?

1.3. Metodología

Para poder llevar a cabo esta investigación y responder así a los objetivos y preguntas planteadas en el apartado anterior, el trabajo se divide en dos partes: una sección teórica y un ejercicio práctico. La disposición de cada uno de los capítulos y apartados pretende responder a un objetivo específico.

Previamente a la realización del trabajo, se ha llevado a cabo una recopilación y lectura de todos los materiales necesarios para la base teórica: métricas de evaluación de TA, inteligencia artificial y traducción, calidad y traducción automática, información sobre las herramientas de evaluación BLEU y COSTA e información sobre el funcionamiento de las aplicaciones de Google y Microsoft. Algunas de las principales fuentes de información han sido la *Revista Tradumàtica. Tecnologies de la traducció* y las páginas oficiales de Google y Microsoft, entre otras.

En la parte práctica del trabajo, se lleva a cabo el análisis descriptivo y cualitativo de los resultados de traducción automática en la combinación lingüística inglés-español de ambas aplicaciones. Se divide en tres capítulos: la descripción y el análisis de TA de la aplicación Google Translate Images, la descripción y el análisis de TA de Microsoft Translator Images y la comparación final de ambos resultados. Los dos primeros siguen una misma estructura: descripción de la aplicación y su funcionamiento, análisis de la TA con BLEU y análisis de la TA con COSTA. El último capítulo recoge la comparación final entre los resultados obtenidos con ambas aplicaciones. Todos los informes generados por las herramientas de evaluación y otros materiales empleados en este trabajo se pueden consultar en los ANEXOS o como archivos adjuntos en este mismo documento.

En el último capítulo se exponen y desarrollan las conclusiones a las que se han llegado con este trabajo y se responden las preguntas de investigación presentadas en el apartado anterior. Finalmente, se da respuesta al objetivo principal del trabajo: comprobar si las aplicaciones de Google y Microsoft para la traducción de imágenes pueden resultar útiles para el público general sin conocimientos lingüísticos de la lengua inglesa y para el traductor profesional.

1.3.1. Metodología del estudio

En este apartado se presenta la metodología específica empleada para el análisis y presentación de los resultados de traducción automática de Google y Microsoft.

El objetivo específico de este estudio es ver qué aplicación de traducción de imágenes, Google o Microsoft, ofrece mejores resultados de traducción automática en la combinación lingüística inglés-español.

Google y Microsoft son pioneros en la creación de este tipo de aplicaciones, además de ser empresas reconocidas mundialmente, razón por la que han sido seleccionadas para este trabajo. Asimismo, son dos de las empresas más importantes en la investigación e implementación de las nuevas tecnologías en la industria de la traducción, más concretamente en el sector de la traducción automática. Por ejemplo, en el uso de motores de traducción neuronales.

El corpus utilizado en este trabajo está formado por un total de 16 imágenes de carteles informativos que se pueden encontrar en las instalaciones públicas de Ferrocarrils de la Generalitat de Catalunya (FGC) y en las oficinas de un edificio de coworking. Todos los textos son de tipo informativo e instructivo, ya que intentan informar al usuario sobre algo en específico o dar instrucciones sobre cómo llevar a cabo determinadas acciones, por ejemplo, cómo sacar un billete en las máquinas de FGC. Son textos no especializados, dirigidos a todo tipo de público, y de extensión reducida, puesto que los carteles suelen ser elementos muy visuales que pretenden ayudar al usuario y, por ende, deben ser lo más claros posible. Se ha seleccionado este tipo de texto porque las aplicaciones de traducción de imágenes están pensadas para dudas o problemas lingüísticos puntuales, no para traducir grandes cantidades de texto, como por ejemplo libros o revistas, y las imágenes escogidas contienen una cantidad de texto apropiada para ello.

Las herramientas que se van a utilizar para la evaluación de TA de las imágenes son *Interactive BLEU Score Evaluator*, creada por la empresa Tilde, y COSTA MT Evaluation Tool, creada por Konstantinos Chatzitheodorou y Stamatis Chatzistamatis. La primera realiza el proceso de forma automática y con la segunda el proceso es manual.

Una vez definidos todos los elementos que intervienen en el análisis, en este párrafo se explica con detalle la metodología llevada a cabo en los capítulos 4, 5 y 6. En primer lugar, toda la información con respecto a las traducciones se ha transscrito de

forma manual en una hoja de cálculo (véase documentos adjuntos), donde se incluyen el texto original, la traducción automática y la traducción humana de referencia, ya que las imágenes por si solas no pueden introducirse en las herramientas de evaluación. Esta información se presenta en forma de tablas en los capítulos 4 y 5, donde están recogidos los resultados de Google y Microsoft, respectivamente. Una vez traducidas todas las imágenes, se procede a evaluar la calidad de los resultados de TA con las herramientas de BLEU y COSTA. En el caso de BLEU, la herramienta únicamente proporciona una puntuación global del texto, por lo que, para ofrecer un análisis más detallado de los resultados, se ha diseñado una clasificación específica para ello. Esta categorización ha servido de ayuda para tener una visión global de los resultados en forma de gráfico. La clasificación es la siguiente:

Coincidencia exacta	Traducción automática y traducción humana de referencia son exactamente iguales.
Ambas opciones son correctas	Traducción automática y traducción humana de referencia son opciones igualmente válidas, pero BLEU penaliza el uso de palabras diferentes.
Errores leves	La traducción automática contiene fallos de menor gravedad.
Diferencia dialectal	La traducción automática contiene vocabulario propio del español de América.
Incomprendible	La traducción automática contiene errores graves o sinsentidos.
No traducido	El texto original no se ha traducido.

Tabla 1. Clasificación para categorizar los resultados de BLEU (2020)

En el caso de COSTA, la evaluación se ha llevado a cabo de forma manual. Esta herramienta ya incorpora la función de categorización de errores, por lo que no ha sido necesario buscar una alternativa para presentar los resultados. En el análisis de ambas herramientas, todas las explicaciones se complementan con capturas de pantalla.

Una vez finalizadas las evaluaciones con los dos sistemas, en el capítulo 6 se hace una comparación de los resultados obtenidos en ambos casos para determinar cuál es el motor con mejor puntuación.

1.4. Estructura del trabajo

El trabajo se divide en 9 capítulos: introducción, traducción y calidad, traducción automática e inteligencia artificial, evaluación de Google Translate Images, evaluación de Microsoft Translator Images, comparación de resultados, conclusiones, bibliografía y anexos. Estos capítulos podemos dividirlos en cuatro grandes bloques: introducción, en el capítulo 1; marco teórico, en los capítulos 2 y 3; estudio práctico, en los capítulos 4, 5 y 6; y conclusiones, en el capítulo 7. La bibliografía y los anexos se encuentran en los capítulos 8 y 9. En la introducción se encuentra toda la información con respecto a la motivación, objetivos, metodología y estructura del trabajo.

El marco teórico lo constituyen los capítulos 2 y 3. En el capítulo 2, en primer lugar se intenta hacer una aproximación al concepto de calidad en la industria de la traducción y, más concretamente, en el sector de la traducción automática. A continuación, se explica qué son las métricas de evaluación de traducción automática y se presentan las dos herramientas utilizadas en este trabajo: BLEU y COSTA MT Evaluation Tool. En el capítulo 3, se explica la relación actual entre traducción automática e inteligencia artificial y las aplicaciones de esta última en la industria.

El marco práctico está formado por los capítulos 4, 5 y 6. En el cuarto y quinto capítulos se presentan las aplicaciones de traducción de imágenes y el análisis de los resultados de TA de Google y Microsoft, respectivamente. En el sexto capítulo se lleva a cabo la comparación final de resultados entre ambos.

En el capítulo 7 se desarrollan las conclusiones obtenidas una vez finalizado el trabajo, así como una pequeña reflexión personal de todo lo aprendido. A continuación, se encuentran los dos últimos capítulos dedicados a la bibliografía y los anexos, donde están adjuntos todos los materiales utilizados en el trabajo.

2. TRADUCCIÓN Y CALIDAD

Definir el concepto de calidad en el sector de la traducción es algo complicado. A día de hoy, todavía no existe un único estándar válido para todas las traducciones, ya que cada proyecto de traducción tiene unas características y necesidades diferentes. El concepto de calidad puede aplicarse en los dos planos de la traducción: el proceso de gestión del servicio y la traducción propiamente dicha. La dificultad a la hora de establecer un único criterio o definición de calidad se debe, en buena medida, al alto grado de subjetividad implícito en la traducción, ya que pueden existir miles de traducciones diferentes y que todas sean correctas. No obstante, la traducción es un servicio que conlleva una remuneración y, como tal, es importante definir unos estándares mínimos. A lo largo de los últimos años, son muchas las propuestas de guías de estilo o recomendaciones que se han creado para intentar definir el concepto de calidad en el mundo de la traducción desde un punto de vista objetivo, como por ejemplo la que propone ASATI¹ (2009).

Un servicio de calidad es aquel que logra cumplir el objetivo de satisfacer al cliente. Para ello, se lleva a cabo un proceso organizado y detallado donde se miden los resultados a partir de los cuales se fomentan la mejora del proceso traductor y del producto final. Tanto para grandes empresas como autónomos, el diseño y establecimiento de un buen sistema de gestión de proyectos de traducción es indispensable para poder tener una buena comunicación con el cliente y mejorar la productividad, ya que cuanto más ágil sea la gestión, más tiempo se podrá dedicar a la actividad traductora. Otros aspectos muy importantes son la rapidez y precisión en las respuestas a las preguntas del cliente y los plazos y requisitos de entrega. Como en cualquier actividad comercial, establecer un plazo de entrega razonable y, más importante aún, cumplirlo, es primordial. Su incumplimiento puede suponer la desacreditación como profesionales y conllevar una pérdida de clientes.

Desde el punto de vista del resultado, una traducción de calidad es aquella que es fiel al original y transmite todo su contenido, tanto desde el punto de vista textual como desde el punto de vista del autor, llevando a cabo las adaptaciones necesarias. También es importante que no contenga errores de traducción, ortotipografía u omisión, ni se añadan elementos que no estén presentes en el original. La fase de revisión una vez

¹ Cepero, J. I.; Encuentra, P. et al (2009). «Guía de calidad en la traducción: orientaciones para alcanzar la excelencia en el ejercicio de la profesión de traductor», 2. Recuperado de:
http://www.asati.es/img/web/docs/CALIDAD.TRAD_ASATI.2009.pdf

terminada la traducción resulta así fundamental para detectar y corregir aquellas imprecisiones que se hayan pasado por alto, tanto a nivel textual como intencional, ya que transmitir las sensaciones del autor no es una tarea nada fácil en muchas ocasiones.

Desafortunadamente, en ocasiones la profesión del traductor no cuenta con el merecido reconocimiento. Hay quien asegura que cualquier persona con conocimientos de distintos idiomas es capaz de traducir, pero lo cierto es que ser traductor implica muchos otros factores, además de la parte lingüística. La calidad es un factor diferencial fundamental como traductores. Cuando una traducción alcanza un nivel óptimo de calidad, esta pasa desapercibida para el público: el objetivo de que el lector pueda leer y sentir lo que el autor quería transmitir se habrá cumplido. Si resulta demasiado evidente que se está leyendo una traducción, seguramente se deba al fallo de algún parámetro a lo largo de la gestión y producción del proyecto. La invisibilidad del traductor es, precisamente, lo que hace que muchas veces no se reconozca nuestro trabajo como profesionales e incluso se llegue a subestimar nuestra profesión (Cepero *et al*, 2009).

Una de las propuestas de estandarización de los servicios de traducción humana es el estándar europeo UNE EN-15038:2006, el cual «especifica los requisitos para el proveedor de servicios de traducción (PST) respecto a los recursos humanos y técnicos, la gestión de calidad y de proyectos, el marco contractual y los procedimientos del servicio (AENOR, 2006)».

Desde hace algunos años, la revolución de la TA y la posedición (PE) ha aumentado el número de investigaciones relacionadas con el tema, lo que a su vez ha hecho surgir nuevos estándares y guías de calidad específicas para la TA+PE, los cuales establecen unos criterios mucho más precisos. No obstante, mantienen un punto en común con los estándares de la traducción humana: el componente subjetivo de las traducciones sigue siendo un aspecto difícil de valorar. Aunque es una herramienta que cada vez incorporan más empresas en su flujo de trabajo, todavía queda mucho que descubrir y desarrollar en lo que a TA+PE se refiere.

2.1. Traducción automática y calidad

La traducción automática ha supuesto toda una revolución en la industria de la traducción. Gracias a ella, se pueden traducir grandes cantidades de texto en menos tiempo y a un coste más económico que si la tarea la llevara a cabo un traductor humano. Surgió como solución al aumento del contenido traducible a nivel mundial, especialmente para ser utilizada en los negocios. No obstante, hoy en día la TA está

presente prácticamente en todas nuestras acciones cotidianas. Por ejemplo, al escuchar música, al realizar compras o cuando estamos leyendo prensa extranjera.

A pesar de todas las ventajas que ofrece, la TA sigue mostrando algunas deficiencias. En primer lugar, las máquinas no son capaces de reconocer la intención del texto, por lo que no tienen en cuenta ni al autor ni al público. En segundo lugar, la TA no es capaz de reconocer los elementos culturales presentes en el texto, como las palabras en otros idiomas y las frases hechas, por lo que el sistema las traduce literalmente. Esto, además de cambiar el sentido del original, hace que el traductor cree frases sin sentido y afecte a la fluidez y adecuación del texto (Linguaserve, 2019).

Por todo ello, si decidimos emplear la TA como estrategia de traducción es importante escoger un sistema adecuado para nuestro proyecto, teniendo siempre en cuenta las características del mismo y las especificaciones del cliente. La TA debe ser una herramienta más en nuestro flujo de trabajo que nos facilite el proceso de traducción en determinados escenarios. En caso contrario, la TA no estaría cumpliendo con su objetivo.

A la hora de tomar esa decisión, disponemos de las métricas de evaluación de calidad de TA. Estos sistemas de análisis y medición sirven para comprobar si la TA puede resultarnos útil y decidir qué sistema será el óptimo en nuestro flujo de trabajo. En sus etapas iniciales, la evaluación de TA era un subcampo dentro de la investigación de la TA. Las primeras normas de calidad solo hacían referencia a las posibles mejoras en los sistemas de TA y en el proceso de desarrollo (Babych, 2014). No obstante, hoy en día la idea de calidad en TA está cambiando y se tienen en cuenta muchos otros factores importantes: las especificaciones del cliente, la tipología del texto y el lector al cual va dirigida la traducción, ya que sin uno de estos tres aspectos la traducción no cumpliría con su objetivo de transmitir conocimiento y, por lo tanto, no podría considerarse un servicio de calidad. Para alcanzar un buen nivel de calidad, primero hay que entender que cada proyecto de traducción es diferente y tiene unas características propias, por lo que los estándares también deben adaptarse a cada proyecto (Görög, 2014).

2.2. Métricas de evaluación de TA

Las métricas de evaluación de TA son los estándares empleados por los profesionales de la traducción para evaluar y comparar la calidad que ofrecen distintos sistemas de TA en distintos tipos de texto. Conscientes de la dificultad a la hora de

definir el concepto de «calidad», estas métricas pretenden ayudar a conseguir un buen resultado de TA en cualquier contexto, teniendo siempre en cuenta la adecuación y fluidez del texto. No obstante, la última palabra siempre la tendrá el revisor humano que examine los resultados ofrecidos por estas métricas. En este trabajo, se utilizan las siguientes métricas de evaluación de TA: BLEU y COSTA MT Evaluation Tool.

2.2.1. BLEU

El sistema BLEU (*Bilingual Evaluation Understudy*) fue presentado en 2002 por el grupo de investigación de IBM, compuesto por Kishore Papineni, Salim Roukos, Todd Ward y Wei-Jing Zhu, con el objetivo de hacer que las evaluaciones de TA fueran más fáciles, rápidas y baratas, ya que la evaluación humana podía alargarse meses y suponer un coste demasiado elevado para las empresas. Papineni (2002) resume la idea principal de BLEU en una frase: «The closer a machine translation is to a professional human translation, the better it is».

BLEU es un sistema que consiste en comparar TA en bruto con traducciones humanas de referencia y, a partir de ahí, extraer coincidencias. BLEU otorga a cada una de las propuestas de TA puntuaciones que oscilan entre 0 y 1. La opción que obtenga más coincidencias recibirá una puntuación más cerca de 1 y, por lo tanto, será la catalogada como mejor alternativa (Pan, 2016).

El funcionamiento de BLEU se basa en el análisis de la siguiente información:

- Una o más traducciones humanas de referencia (*input*), idealmente desconocidas por el traductor automático. Se recomienda utilizar un corpus de mil frases o más, ya que una muestra más pequeña podría reducir la calidad de las traducciones al no tener suficiente material de referencia con el que llevar a cabo las comparaciones necesarias.
- El texto original.
- La traducción automática del texto original (*output*).
- La fórmula para realizar la comparación y el cálculo de la puntuación.

$$\text{Precision} = \frac{\text{No. of candidate translation words occurring in any reference translation}}{\text{Total no. of words in the candidate translation}}$$

Figura 1. Fórmula de BLEU para calcular las puntuaciones (2020)

Una vez que BLEU dispone de toda la información, el sistema realiza una comparación entre las traducciones humanas y los resultados de los traductores

automáticos que se están evaluando. El sistema cuya traducción automática contenga más coincidencias con la TH de referencia obtendrá la puntuación más alta y, por lo tanto, será el más fiable según BLEU.

No obstante, y a pesar de estar considerado actualmente como el método estándar de evaluación de TA, el sistema tiene varios puntos débiles:

- Toma como unidad la palabra, por lo que solo busca coincidencias entre las mismas palabras. Esto provoca que otras opciones válidas queden descartadas, simplemente porque emplean paráfrasis o sinónimos.
- Tampoco tiene en cuenta el orden de las palabras. Entre dos oraciones, una ininteligible, aunque con las palabras que aparecen en el original, y una correcta, pero con palabras distintas o sinónimos, BLEU asignará una puntuación más alta a la oración incorrecta, solo porque contiene las palabras del texto de referencia (Vashee, 2019). Es decir, «[BLEU] may only give a small penalty for a change that is superficially small, but completely changes the meaning of a translation (Sennrich, 2016 *citado en* Pan, 2016)».

En el caso de BLEU, se ha utilizado la herramienta en línea *Interactive BLEU score evaluator* creada por la empresa Tilde. Esta implementación ha permitido realizar de forma automática la evaluación de los resultados de TA obtenidos mediante Google y Microsoft, así como comparar ambos sistemas y ver cuál ofrece mejores resultados.

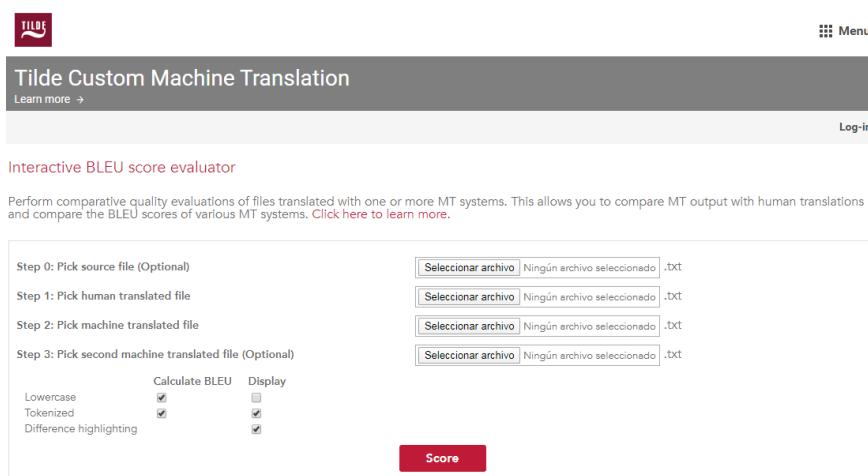


Figura 2. Interfaz de *Interactive BLEU score evaluator* de la empresa Tilde (2020)

Una vez analizados los archivos de traducción, los resultados se exponen en dos apartados: el primero, indica la puntuación global de toda la traducción; el segundo, los resultados y puntuaciones de cada segmento, los cuales aparecen representados en

forma de gráfico y tabla donde pueden observarse el texto original y las diferencias entre la traducción humana de referencia y la TA.



Figura 3. Presentación de resultados de *Interactive BLEU score evaluator* (2020)

2.2.2. COSTA MT Evaluation Tool

COSTA MT Evaluation Tool es una herramienta de evaluación manual asistida de la calidad de TA. Si bien es cierto que la evaluación manual de TA conlleva un elevado coste económico y temporal, es una parte ineludible de cualquier proyecto en el que se trabaje con TA. En la actualidad existen herramientas que pueden hacer este proceso de forma automática pero no ser tan eficaces como la evaluación manual, como es el caso de la propuesta de Tilde sobre la métrica BLEU.

El principal objetivo de esta herramienta es facilitar la labor del evaluador o investigador en un entorno de trabajo sencillo e intuitivo. Permite realizar estadísticas

de evaluación para una combinación lingüística específica y sus datos pueden utilizarse a posteriori en otros proyectos o futuras investigaciones.

Otra de las ventajas de esta herramienta es que es de acceso libre, lo que evita que se tengan que pagar licencias de otras herramientas similares. Además, se trata de un programa Java con un funcionamiento muy simple: cada evaluación se considera un proyecto nuevo. Para que funcione correctamente, solo hay que proporcionar información sobre las lenguas de partida y de llegada y tres archivos en formato UTF-8 segmentados en una oración por línea: el original, el texto traducido con TA y una traducción humana de referencia.



Figura 4. Interfaz de COSTA MT Evaluation Tool (2020)

Una vez creado el proyecto, nos aparece la siguiente interfaz. En ella podemos ver el segmento original y la propuesta de TA. El usuario debe seleccionar primero el grado de fluidez (ordenados de menor a mayor: *Incomprehensible*, *Disfluent language*, *Non-native language*, *Good language* o *Flawless language*) y después el de adecuación (ordenados de menor a mayor: *None*, *Little meaning*, *Much meaning*, *Most meaning* o *All meaning*). Como función opcional, se puede seleccionar qué tipos de error se han cometido y añadir comentarios. Si el segmento de TA es idéntico a la traducción humana, aparecerá un aviso como el que se muestra en la figura 5. Para garantizar una mayor objetividad en la evaluación, primero se evalúa la fluidez, de forma que el

usuario solo ve el TO y la TA. Una vez evaluada la fluidez, aparece la traducción humana para así poder evaluar la adecuación.

The screenshot shows the COSTA MT Evaluation Tool II interface. At the top, there are input fields for 'Source' (Venice) and 'MT' (Venecia). Below these are rating scales for 'Fluency' (ranging from 1. Incomprehensible to 5. Flawless language) and 'Adequacy' (ranging from 1. None to 5. All meaning). A message at the bottom of this section states '100% Fuzzy Matching between Adequacy and Reference translation!'. The next section, 'Translation error classification', contains several checkboxes for categories like Grammar, Words, and Style, each with sub-options such as Verb inflection, Single words, or Acronyms - Abbreviations. A progress bar at the bottom right indicates 'Sentence: 1/6' and a 'Next' button.

Figura 5. Interfaz de COSTA MT Evaluation Tool II (2020)

Al finalizar la evaluación de todos los segmentos, se genera de forma automática un informe en formato UTF-8 con todos los datos recopilados. De esta manera, el archivo puede ser leído por otros programas o aplicaciones como Excel o herramientas de estadística y se pueden aprovechar esos datos para realizar investigaciones complementarias. En el informe, aparecen los datos de adecuación y fluidez representados en una puntuación que oscila entre 0 y 1 y el promedio de tiempo dedicado a cada segmento, lo que también puede resultar útil para conocer nuestra productividad. A continuación, se muestra de forma más descriptiva cuántos segmentos se han clasificado en cada nivel de fluidez y de adecuación, así como la clasificación total del número y tipo de errores registrados.

The screenshot shows the COSTA MT Evaluation Tool III interface. At the top, it displays summary statistics: 'Fluency: 0.8333333 / 1 74 sec average time per sentence' and 'Adequacy: 0.96666664 / 1 16 sec average time per sentence'. Below this are four tables: 'Fluency' (with rows for Incomprehensible, Good language, Non-native language, Disfluent language, and Flawless language), 'Grammar' (with rows for Verb inflection, Noun inflection, Other inflection, Wrong category, Article, Preposition, and Agreement), 'Words' (with rows for Single words, Multi-word units - Idioms, Terminology, Untranslated words, Ambiguous translation, Literal translation, and Conjunctions), and 'Style' (with rows for Acronyms - Abbreviations, Extra words, Country standards, Spelling errors, Accent, Capitalization, and Punctuation).

Figura 6. Interfaz de COSTA MT Evaluation Tool III (2020)

La clasificación de errores de esta herramienta se basa en los datos recogidos sobre los tipos de error más frecuentes en sistemas basados en Moses y en sistemas de TA gratuitos como Google Translate o Bing Translator. Se establecen tres categorías principales de error: *Grammar* (gramática), referente a aspectos lingüísticos; *Words* (palabras), referente al uso de las palabras; y *Style* (estilo), referente al estilo y formato del texto.

3. TRADUCCIÓN AUTOMÁTICA E INTELIGENCIA ARTIFICIAL

La rápida evolución e implantación de la TA no podría ser posible sin entender primero el rápido avance de las tecnologías que acompaña a la globalización actual. Vivimos en un mundo permanentemente conectado, lo que ha acelerado la necesidad de superar el reto de la barrera lingüística. La TA supuso toda una revolución debido a su enorme capacidad para gestionar y traducir grandes volúmenes de texto de una forma increíblemente rápida y económica. Los primeros sistemas de TA fueron la traducción automática basada en reglas (TABR) y la traducción automática estadística (TAE). Estos, a pesar de cumplir su función de forma correcta, ofrecen resultados cuya calidad sigue siendo deficiente en muchos aspectos, especialmente entre lenguas lejanas entre sí, como por ejemplo el chino y el español.

Asimismo, los primeros motores de TA solo podían utilizarse en un ordenador. Actualmente, además de los cientos de traductores automáticos gratuitos o privados en línea, también existen dispositivos y aplicaciones capaces de traducir conversaciones o incluso imágenes. La posibilidad de poder traducir cualquier contenido en cualquier lugar desde un dispositivo móvil es claramente la opción más popular, ya que es la que mejor se adapta a nuestros hábitos cotidianos. Ahora ya no es necesario introducir las palabras de forma manual en un traductor automático; es suficiente con hacer una foto al texto en cuestión y dejar que una aplicación móvil procese y traduzca el texto. Algunas pueden ofrecer una traducción con un formato exacto o muy similar al original.

Esto ha sido posible gracias a la inteligencia artificial (IA). De ahí nació la TAN, actualmente el sistema adoptado cada vez por más empresas. El secreto de la eficacia de la TAN reside en que se basa en el uso de redes neuronales que imitan el funcionamiento del cerebro humano. En algunos contextos, por ejemplo en artículos no especializados, la calidad con respecto a la TAE ha mejorado tanto que puede llegar a equipararse a la traducción humana profesional. La elevada calidad que proporciona incluso ha hecho resurgir la idea de la potencial desaparición futura de la figura del traductor humano. No obstante, muchos expertos afirman que su trabajo no desaparecerá nunca del todo, ya que siempre será necesaria la presencia humana para revisar procesos y resultados de TA. Además, el componente emocional de los escritores es un aspecto muy humano y un obstáculo que las máquinas hoy día todavía no han podido superar (Barbieri, 2019).

Todo esto demuestra que en la investigación en TA todavía queda mucho por descubrir y mejorar. También queda claro, una vez más, que esa investigación no puede llevarse a cabo, hoy por hoy, sin la presencia del traductor humano, la persona que se encarga de la creación, mantenimiento y revisión de todos los sistemas y procesos de TA.

3.1. Aplicaciones de la IA en la TA

Un gran avance en IA+TA podemos verlo reflejado en el desarrollo y comercialización de aplicaciones capaces de producir traducciones automáticas en combinación con el reconocimiento óptico de caracteres y técnicas de TAN, como es el caso de Google Translate Images y Microsoft Translator Images, aplicaciones analizadas en este trabajo.

3.1.1. Reconocimiento óptico de caracteres (ROC)

Una de las aplicaciones de la IA en la traducción ha sido la aparición de aplicaciones móviles capaces de traducir texto en un soporte físico mediante el reconocimiento óptico de caracteres (ROC). Según el Instituto Tecnológico de Informática (ITI), el ROC «consiste en la identificación automatizada de símbolos o caracteres pertenecientes a un determinado alfabeto, a partir de una imagen recogida mediante la lectura óptica de un texto grabado en un apoyo real», es decir, es el proceso digital mediante el cual se identifican y reproducen los caracteres escritos en un soporte físico. El ROC funciona en combinación con las redes neuronales artificiales, las cuales se encargan de detectar e identificar objetos o puntos de referencia. Esta tecnología surgió debido a la gran demanda de digitalización de documentos de los últimos años, ya que era la forma más rápida de llevar a cabo ese proceso (Arlandis, 2010).

Gracias a este sistema, hoy en día es posible procesar documentos antiguos de todo tipo, principalmente históricos y legales. El ROC en combinación con la TA ha agilizado mucho la transcripción y traducción de todos esos documentos. No obstante, el ROC también tiene algunos puntos débiles. Al llevar a cabo la transcripción del texto, a menudo comete errores de reconocimiento de algunos caracteres, lo que provoca la aparición de errores en el texto y, a su vez, menoscaba la eficiencia de la TA. Por esta razón, hace unos años, Haithem Aflì y Andy Way llevaron a cabo un estudio sobre cómo perfeccionar el ROC para mejorar también la eficacia de la TA. Su propuesta consistió en un nuevo sistema de corrección de errores basado en modelos de lenguaje y

métodos de la TAE. En los resultados, observaron una mejora en las traducciones automáticas de textos históricos transcritos mediante ROC, lo que abre la puerta a futuras investigaciones más exhaustivas sobre esta cuestión (Afli y Way, 2016).

3.1.2. Traducción Automática Neuronal

Otra de las aplicaciones de la IA en el sector de la traducción es la creación de motores neuronales de TA. La TAN se está incorporando paulatinamente en motores como los de Google o Microsoft. No obstante, todavía se encuentra en desarrollo debido a su enorme complejidad. El funcionamiento de estos sistemas se basa en la imitación de las redes neuronales del cerebro humano, lo que explica el número de aciertos y el tono natural de sus traducciones.

La arquitectura más utilizada en el uso de redes neuronales consiste en un codificador y un decodificador. El codificador despliega una red que se encarga de analizar la frase de derecha a izquierda (y viceversa) y transformar las palabras en vectores. A continuación, el decodificador genera la traducción en base a la oración original (Casacuberta; Peris, 2017). Otro elemento importante en la TAN es el entrenamiento de los motores, ya que sin la cantidad suficiente de corpus de referencia el motor no podría funcionar correctamente.

A pesar de los esperanzadores resultados obtenidos, la TAN todavía necesita recorrer un largo camino de investigación. Entrenar y mantener un motor neuronal es un proceso muy lento y costoso, además de la falta de precisión en cuestiones de terminología o en lenguas lejanas entre sí (López, 2018).

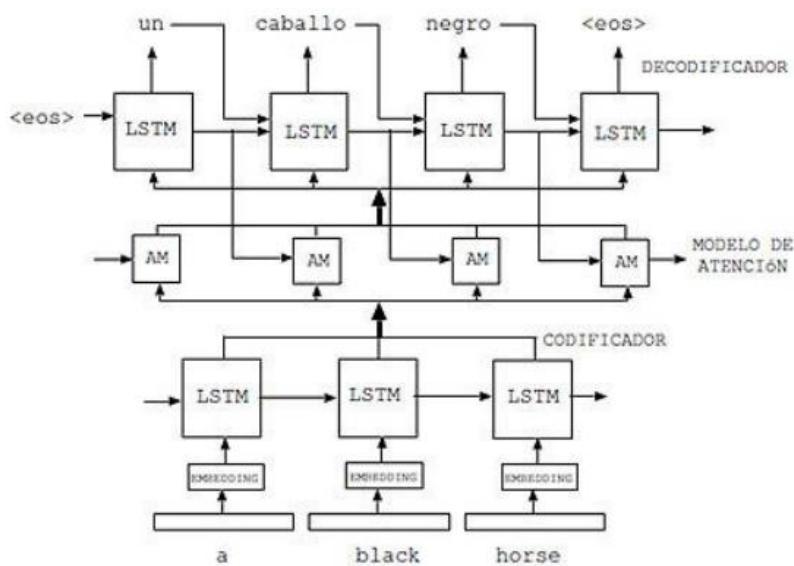


Figura 7. Ejemplo de arquitectura codificadora-decodificadora (Casacuberta; Peris, 2017)

4. EVALUACIÓN DE GOOGLE TRANSLATE IMAGES

Google Translate es el traductor automático de Google. Nació en el año 2006 y originalmente empleaba sistemas de TABR. A partir de 2007, empezó a utilizar la TAE y, desde 2016, emplea también sistemas de TAN, actualmente el sistema de TA con más relevancia a nivel mundial. La TAN todavía no está disponible en muchos idiomas, razón por la cual sigue haciendo uso de la TAE. Si hacemos una prueba en la combinación inglés-español con cualquier texto, la calidad lograda con el uso de redes neuronales es más que evidente.

La función de traducción automática de imágenes nació en 2017 gracias a la creación de Google Lens, una aplicación móvil de reconocimiento de la imagen cuyo lanzamiento tuvo lugar en el Google I/O 2017. El objetivo principal de la aplicación es ofrecer una serie de acciones que llevar a cabo sobre el elemento reconocido. Para ello, Google hace uso de todas sus aplicaciones. Por ejemplo, para conocer el nombre de una flor determinada basta con iniciar la aplicación y enfocar a la flor en cuestión. En apenas unos segundos, Google Lens es capaz de identificarla y mostrarnos la información correspondiente a través de su buscador. Para averiguar el nombre de un restaurante o de una calle, debemos seguir el mismo proceso con la cámara y Google nos facilitará la información gracias a Google Maps. Una vez ha identificado el objeto en cuestión, Google ofrece la posibilidad de llevar a cabo diferentes acciones: comprar (en el caso de productos), cómo llegar (en el caso de lugares), más información (sobre un elemento determinado) o incluso escanear y añadir a tu teléfono direcciones de correo o números de teléfono (Yúbal, 2019 y Google, 2020).



Figura 8. Logotipo de Google Lens (Google, 2020)

La función objeto de este trabajo es la de traducción de imágenes. Para utilizarla, el funcionamiento es el mismo: basta con acercar la cámara del teléfono al texto que queremos traducir y Google Lens nos proporciona su traducción. La propia aplicación es capaz de reconocer el idioma original, de manera que solo debemos indicar a qué idioma queremos traducir. La última actualización de esta aplicación incluye tres formas distintas para hacerlo: la traducción instantánea con solo acercar la cámara al texto en cuestión, hacer fotos al texto y subrayar la parte que queremos traducir y traducir texto de imágenes importadas desde el móvil.



Figura 9. Muestra del funcionamiento de Google Lens (Google, 2020)

4.1. Análisis de los textos

En este apartado se procede al análisis de los resultados de TA proporcionados por Google Translate Images. En primer lugar, se presentan los resultados de TA en forma de tabla y, a continuación, se analizan con detalle las puntuaciones obtenidas con BLEU y COSTA. La metodología del estudio puede consultarse en la INTRODUCCIÓN y los informes generados están en los archivos adjuntos.

	Sg.	Texto original	Traducción automática	Traducción humana
1	1	Only for clients	Solo para clientes	Solo para clientes
	2	Timetable. While this station is open.	Calendario. Mientras esta estación está abierta.	Horario. Durante las horas de servicio de esta estación
	3	Ferrocarrils de la Generalitat de Catalunya assumes no responsibility for bicycles parked on its premises.	Ferrocarrils de la Generalitat de Catalunya no asume ninguna responsabilidad por las bicicletas aparcadas en sus locales.	Ferrocarrils de la Generalitat de Catalunya no se responsabiliza de las bicicletas estacionadas en sus dependencias.
2	4	In case of fire	In case of fire	En caso de incendio
	5	Do not use the lift	No use el ascensor	No use el ascensor
	6	Use the stairs	Use las escaleras	Use las escaleras
3	7	Going down on to the track is prohibited	Ir a la pista está prohibido	No se permite bajar a la vía
4	8	Non-opening of doors	No abrir las puertas	Si las puertas no se abren
	9	If you insert the ticket into the validator and the door does not open, go to the blue vending machine in the platform 1 of this station and follow the instructions on the screen.	Si inserta el boleto en el validador y la puerta no se abre, vaya a la máquina expendedora azul en la plataforma 1 de esta estación y siga las instrucciones en la pantalla.	Si, una vez introducido el billete en la canceladora, la puerta no se abre, diríjase a la máquina expendedora de color azul, situada en el andén de vía 1, y siga las instrucciones de la pantalla.

5	10	Allow way out	Permitir salida	Dejen salir
6	11	Instructions of use:	Instrucciones de uso:	Instrucciones de uso:
	12	To purchase a ticket:	Para comprar un boleto:	Para comprar un billete:
	13	1. Choose the ticket you require by pressing the corresponding button.	1. Elija el boleto que necesita presionando el botón correspondiente.	1. Elegir el billete pulsando el botón correspondiente.
	14	2. Insert coins, notes or credit card.	2. Inserte monedas, billetes o tarjetas de crédito.	2. Insertar monedas, billetes o tarjeta de crédito.
	15	3. Collect your ticket and the change or the receipt, if required.	3. Recoja su boleto y el cambio o el recibo, si es necesario.	3. Recoger el título o billete y el cambio o el recibo, si corresponde.
7	16	Space for preferential use	Espacio de uso preferencial	Espacio de uso preferente
8	17	Attention	Atención	Atención
	18	If the lift is jammed:	Si el elevador está atascado:	En el caso de que el ascensor quede bloqueado:
	19	Do not attempt to open the doors	No intente abrir las puertas	No intente abrir las puertas
	20	Use the intercom to communicate with the Control Centre	Usar el intercomunicador para comunicarse con el centro de control	Use el interfono para comunicarse con el Centro de Control
	21	If the event of a power cut, press the button for outside communication consecutively	Si se produce un corte de energía, presione el botón para comunicación externa consecutivamente	En caso de corte del fluido eléctrico, pulse el botón para comunicarse con el exterior
	22	Press the button. The TV camera will start operating	Presione el botón. La cámara de TV comenzará a funcionar.	Pulse el botón. La cámara de TV se activará.
	23	In a few moments the lift will stop at	En unos momentos el ascensor se detendrá en	En breves momentos el ascensor se situará en el

		the lower floor and the doors will open	el piso inferior y las puertas se abrirán	piso inferior y se abrirán las puertas
9	24	Price	Precio	Precio
	25	Validity period	Periodo de validez	Tiempo de validez
	26	A single person timed ticket for 10 journeys on any means of transport within the integrated system, for the number of zones purchased (maximum charge 6 zones). This ticket is not valid for the Aeroport T1 and Aeroport T2 metro station stops on line L9 Sud.	cronometrado para una sola persona para 10 viajes en cualquier de transporte dentro del sistema integrado, por el número de un boleto zona de medios comprados (carga máxima 6 zonas). Este boleto no es válido para las estaciones de metro Aeroport T1 y Aeroport T2 se detiene en la línea 2	Título unipersonal y horario, de 10 viajes integrados en todos los modos de transporte, según las zonas por las que se transite (pago máximo 6 zonas). No es válido en las estaciones de metro de Aeroport T1 y Aeroport T2 de la línea L9 Sud.
10	27	Station where all trains stop	Estación donde paran todos los trenes	Estación donde paran todos los trenes
	28	Station where only trains from the line shown stop	Estación donde solo paran los trenes de la línea mostrada	Estación donde solo paran los trenes de las líneas señaladas
	29	Metro connection	Conexión de metro	Enlace con Metro
	30	Rodalies de Catalunya connection	Conexión Rodalies de Catalunya	Enlace con Rodalies de Catalunya
	31	Contribution Bus	Contribution Bus	Autobús de aportación
	32	Local bus	Autobús local	Autobús urbano
	33	Funicular	Funicular	Funicular
	34	"Tramvia Blau" (tram)	"Tramvia Blau" (tren)	"Tramvia Blau"
11	35	Press the button and then release it	Press the button and then release it	Pulse el botón y suéltelo
	36	Wait while the light is flashing	Espere mientras la luz parpadea	Espere mientras la luz parpadea

	37	When the light remains steady, you may speak to FGC staff	Cuando la luz permanece estable, puede hablar con el personal de FGC	Con la luz fija, ya puede hablar con el personal de FGC
	38	The light will turn off upon completion of the call	La luz se apagará al finalizar la llamada	La luz se apagará cuando finalice la llamada
	39	You can also contact FGC staff at:	También puede contactar al personal de FGC en:	También pueden comunicarse con el personal de FGC a través de:
12	40	Office	Oficina	Oficina
	41	Offices	Oficinas	Oficinas
13	42	AT UTOPICUS, WE EMPTY THE FRIDGE EVERY FRIDAY TO CARRY OUT A DEEP CLEAN	EN UTOPICUS, VACÍAMOS LA FRIGORÍA CADA VIERNES PARA REALIZAR UNA LIMPIEZA PROFUNDA	EN UTOPICUS, LOS VIERNES VACIAMOS LA NEVERA PARA HACER UNA LIMPIEZA EN PROFUNDIDAD
	43	PLEASE DONT FORGET TO TAKE YOUR TUPPERWARE AND FOOD HOME	NO OLVIDES TOMAR TUPPERWARE Y ALIMENTOS A CASA	NO OLVIDES LLEVARTE TUS TUPPERS Y COMIDA
14	44	KEEPING THE COMMUNAL AREAS CLEAN BY WASHING UP OR USING THE DISHWASHER FORMS PART OF THE UTOPICUS CULTURE	MANTENER LIMPIAS LAS ÁREAS COMUNES LAVANDO O UTILIZANDO LAS FORMAS DE LAVAJILLAS PARTE DE LA CULTURA DE UTOPICUS	FREGAR O USAR EL LAVAJILLAS PARA MANTENER LOS ESPACIOS COMUNES LIMPIOS FORMA PARTE DE LA CULTURA DE UTOPICUS
	45	IF THE DISHWASHER IS IN USE OR FULL, PLEASE WASH THE DISHES THAT YOU HAVE USED. WE	SI EL LAVAJILLAS ESTÁ EN USO O COMPLETO, LAVE LOS PLATOS QUE HA UTILIZADO. PROPORCIONAMOS TODO LO QUE	SI EL LAVAJILLAS ESTÁ EN USO O LLENO, FRIEGA EL MENAJE QUE HAYAS USADO. TENEMOS A TU DISPOSICIÓN TODO

		PROVIDE EVERYTHING YOU NEED TO DO SO	NECESITAS HACER	EL MATERIAL PARA HACERLO
15	46	HERE TO KEEP ALL YOUR STUFF	AQUÍ PARA MANTENER TODAS SUS	AQUÍ PUEDES GUARDAR TUS COSAS
	47	BUT PLEASE,	COSAS PERO POR	PERO POR FAVOR,
	48	TAKE CARE OF MY INNER PEACE	FAVOR, CUIDEN MI PAZ INTERIOR	CUIDA DE MÍ
16	49	FIX COWORKERS AND PRIVATE OFFICES	ARREGLAR A TRABAJADORES Y OFICINAS PRIVADAS	COWORKERS Y OFICINAS PRIVADAS
	50	Fix coworkers and private offices have 24/7 access to the centre, however, from 10pm to 7am the security alarm is automatically activated.	Los compañeros de trabajo de Fix y las oficinas privadas tienen acceso 24/7 al centro, sin embargo, de 10pm a 7am la alarma de seguridad se activa automáticamente.	Coworkers y oficinas privadas tienen acceso al centro 24 horas al día los 7 días de la semana. No obstante, de 10 de la noche a 7 de la mañana la alarma de seguridad se activa automáticamente.
	51	WHAT SHOULD I DO IF I AM STAYING DURING THIS TIME FRAME?	¿QUÉ DEBO HACER SI ME QUEDO DURANTE ESTE MARCO DE TIEMPO?	¿QUÉ DEBO HACER SI ESTOY EN EL EDIFICIO DURANTE ESA FRANJA HORARIA?
	52	Notify the security company over the intercom:	Notifique a la compañía de seguridad por el intercomunicador:	Notificar a la compañía de seguridad a través del interfono:
	53	1) Please identify yourself with your full name and ID.	1) Identifíquese con su nombre completo e identificación.	1) Identificate con tu nombre completo y DNI.
	54	2) Notify until what time you will stay in the building to enable them to re-activate the alarm after your	2) Notificar hasta qué hora permanecerá en el edificio para permitirles reactivar la alarma después de su partida	2) Notifica hasta qué hora permanecerás en el edificio para que puedan reactivar la alarma una vez te hayas ido.

		departure.		
55	Thank you	Gracias	Gracias	
56	*YOU CAN FIND FURTHER INFORMATION IN THE WELCOME TUTORIAL	*PUEDES ENCONTRAR MÁS INFORMACIÓN EN EL TUTORIAL DE BIENVENIDA	*PUEDES ENCONTRAR MÁS INFORMACIÓN EN EL TUTORIAL DE BIENVENIDA	

Tabla 2. Resultados de TA de Google Translate Images (2020)

4.1.1. Análisis con BLEU

En este apartado se presenta la evaluación de TA de Google de acuerdo con el estándar BLEU. Recordamos que el funcionamiento de este sistema se basa en otorgar puntuaciones que oscilan entre 0 y 1, siendo 1 el valor más alto y, por lo tanto, las traducciones que se encuentren más cerca de ese valor serán las mejor valoradas por BLEU (ver apartado [BLEU](#)). No obstante, este análisis se ha llevado a cabo con la herramienta *Interactive BLEU score evaluator*, de Tilde, que ofrece las puntuaciones ponderadas sobre 100.

Una vez introducidos el texto original en inglés, la traducción humana de referencia en español y el resultado de TA, la puntuación que BLEU otorga a Google es de 33.81 puntos sobre 100.

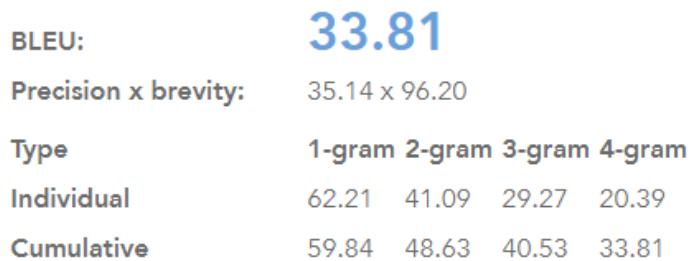


Figura 10. Puntuación de BLEU para Google (2020)

Para poder entender mejor ese resultado, en primer lugar se lleva a cabo un análisis segmento por segmento para ver cómo los ha puntuado BLEU. En este apartado se exponen algunos de los casos más llamativos. Para ver el informe completo de BLEU, véase los documentos adjuntos.

Para poder representar de forma gráfica este análisis y tener una visión global de toda la información extraída, se ha optado por clasificar los segmentos en seis categorías (véase la metodología del estudio en INTRODUCCIÓN): **Coincidencia exacta** (TA y TH son exactamente iguales), **Ambas opciones correctas** (ambas opciones son válidas pero BLEU penaliza la TA por utilizar palabras diferentes), **Incomprensible** (la TA no tiene sentido o contiene errores graves), **No traducido** (TO y TA se mantienen igual), **Diferencia dialectal** (en la TA se encuentra vocabulario propio del español de América) y **Errores leves** (la TA es comprensible, aunque con algunos fallos de menor importancia).

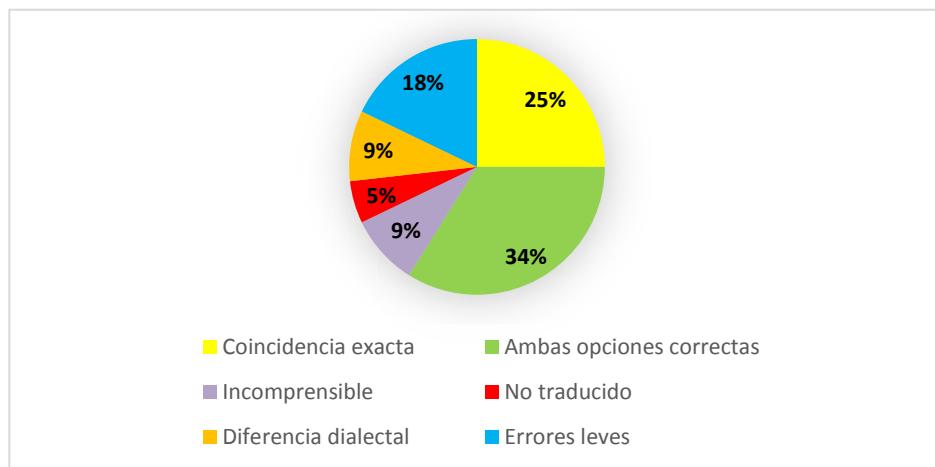


Figura 11. Gráfico con los resultados de BLEU para Google (2020)

Como se puede observar en el gráfico, la mayoría de los segmentos de TA (34%) son opciones igualmente válidas y que podrían funcionar en sus contextos. Sin embargo, BLEU les ha asignado una puntuación menor por diferencias en la elección de las palabras, por usar mayúsculas en vez de minúsculas o por los acentos, entre otras diferencias. Estos son algunos de los ejemplos más llamativos:

En este primer ejemplo, BLEU ha restado mucha puntuación al segmento por utilizar palabras distintas que son correctas. En este contexto, no existe diferencia alguna entre ambas opciones traductoras, por lo que la penalización no estaría justificada desde un punto de vista lingüístico.

Source	-	-	Validity period
Human	100.00	1.00	Tiempo de validez
Machine	63.89	1.00	Periodo de validez

Figura 12. Ejemplo 1 de la evaluación de BLEU para Google (2020)

Lo mismo ocurre en el siguiente ejemplo, donde ambas opciones expresan la misma idea, solo que con otras palabras:

Source	-	-	The light will turn off upon completion of the call
Human	100.00	1.00	La luz se apagará cuando finalice la llamada
Machine	41.11	1.00	La luz se apagará al finalizar la llamada

Figura 13. Ejemplo 2 de la evaluación de BLEU para Google (2020)

Otro dato que llama la atención es el porcentaje de casos de Coincidencia exacta (25%). En total se han encontrado 14 segmentos que se han traducido de forma idéntica a la traducción humana. La gran mayoría de estos casos corresponden a segmentos que contienen entre una y seis palabras, por lo que la probabilidad de que se cometan errores es menor.

El tercer porcentaje más alto es el de Errores leves (18%). Estos errores afectan principalmente a la fluidez del texto y, en ocasiones, a la adecuación, sin que ello implique grandes cambios o pérdidas de significado del TO. En el siguiente ejemplo, al final del texto se produce un error de significado, pero ello no supone la pérdida de información, ya que el resto del texto se puede comprender sin dificultad. No obstante, este tipo de errores podrían confundir al usuario y generar desconfianza, ya que la información con errores o sinsentidos en los paneles informativos podría provocar la sensación de que no es un servicio serio y responsable.

Source	-	-	If the event of a power cut, press the button for outside communication consecutively
Human	100.00	1.00	En caso de corte del fluido eléctrico , pulse el botón para comunicarse con el exterior
Machine	10.88	0.94	Si se produce un corte de energía , presione el botón para comunicación externa consecutivamente

Figura 14. Ejemplo 3 de la evaluación de BLEU para Google (2020)

Además de los errores ya comentados, aunque con menos frecuencia, se han encontrado algunos de mayor gravedad que llevan la pérdida parcial o total de información del original y que incluso podrían tener graves consecuencias. En este ejemplo, la TA da indicaciones diferentes a las del original, lo que, además de confundir al usuario, podría causar graves daños. En el caso de que alguien resulte herido, ambas partes sufrirían consecuencias: el usuario y la compañía. Si el usuario decide denunciar a la empresa, esta podría verse envuelta en graves problemas.

Source	-	-	Going down on to the track is prohibited
Human	100.00	1.00	No se permite bajar a la vía
Machine	13.74	0.86	Ir a la pista está prohibido

Figura 15. Ejemplo 4 de la evaluación de BLEU para Google (2020)

Se produce una situación similar con el siguiente ejemplo. La ininteligibilidad del texto no permite entender con claridad las condiciones de compra y uso del billete, lo que conllevaría una pérdida económica para el usuario si, una vez comprado, no puede usarlo.

Source	-	-	A single person timed ticket for 10 journeys on any means of transport within the integrated system, for the number of zones purchased (maximum charge 6 zones). This ticket is not valid for the Aeroport T1 and Aeroport T2 metro station stops on line L9 Sud.
Human	100.00	1.00	Título unipersonal y horario , de 10 viajes integrados en todos los modos de transporte , según las zonas por las que se transite (pago máximo 6 zonas) . No es válido en las estaciones de metro de Aeroport T1 y Aeroport T2 de la línea L9 Sud .
Machine	20.48	1.08	cronometrado para una sola persona para 10 viajes en cualquier de transporte dentro del sistema integrado , por el número de un boleto zona de medios comprados (carga máxima 6 zonas) . Este boleto no es válido para las estaciones de metro Aeroport T1 y Aeroport T2 se detiene en la línea 2

Figura 16. Ejemplo 5 de la evaluación de BLEU para Google (2020)

El uso del español de América es otro error frecuente en la TA, ya que todavía no existe distinción alguna en los motores entre las variantes del español, todas están recogidas en un único idioma. Esto no supone grandes problemas a la hora de traducir, ya que los hablantes de cada comunidad son capaces de comunicarse entre sí. No obstante, para el usuario puede resultar extraño encontrar palabras, expresiones o estructuras propias de otros países hispanohablantes. A veces se pueden llegar a comprender por contexto, pero en otras podrían provocar confusión. Además, no se estaría cumpliendo con uno de los objetivos de la traducción y la localización: adaptación al público al que va dirigido el texto.

Source	-	-	To purchase a ticket:
Human	100.00	1.00	Para comprar un billete :
Machine	42.73	1.00	Para comprar un boleto :

Figura 17. Ejemplo 6 de la evaluación de BLEU para Google (2020)

4.1.2. Análisis con COSTA MT Evaluation Tool

En este apartado presentamos la evaluación de TA de Google realizada con la herramienta *COSTA MT Evaluation Tool*. Para ello, crearemos el proyecto de evaluación *COSTAMT_Evaluation_Google*.

A la hora de puntuar, COSTA distingue entre la fluidez y la adecuación. Primero se evalúa la fluidez de un segmento y, a continuación, se muestra la traducción humana de referencia para poder evaluar la adecuación (véase apartado [COSTA MT Evaluation Tool](#)). Las puntuaciones obtenidas son las siguientes:

Fluency: 0.7785714 / 1

Adequacy: 0.83214283 / 1

Figura 18. Puntuación de COSTA para Google (2020)

Como podemos observar, la adecuación ha recibido un porcentaje más alto que la fluidez. Esto significa que las traducciones contienen errores de menor gravedad que no impiden que el texto pueda entenderse, ya que el 83% del contenido está presente en la traducción, la cual es una puntuación bastante alta. El porcentaje de fluidez también es elevado, un 77%, pero resulta evidente que Google todavía necesita mejorar la calidad de sus traducciones. En los siguientes párrafos se analiza en profundidad cómo se ha llevado a cabo la evaluación con COSTA y la clasificación de errores y se comentan algunos de los casos más relevantes. Para ver el informe completo de COSTA, véase los documentos adjuntos.

A la hora de puntuar la fluidez y la adecuación de un segmento, se dispone de cinco opciones según el grado de calidad de cada segmento. En el caso de la fluidez, se puede escoger entre las siguientes opciones, ordenadas de menor a mayor puntuación: *Incomprehensible, Disfluent language, Non-native language, Good language* y *Flawless language*. Lo mismo ocurre con la adecuación, ordenadas también de menor a mayor puntuación: *None, Little meaning, Much meaning, Most meaning* y *All meaning*. Los resultados son los siguientes:

Fluency		Adequacy:	
Incomprehensible	5	None	4
Good language	4	Little meaning	6
Non-native language	9	Much meaning	3
Disfluent language	12	Most meaning	7
Flawless language	26	All meaning	36

Figura 19. Puntuaciones de fluidez y adecuación de Google según COSTA (2020)

Las opciones más seleccionadas han sido la de *Flawless language*, en fluidez, y *All meaning*, en adecuación. No obstante, en el caso de la fluidez los segmentos están más repartidos entre todas las categorías. La opción *Disfluent language* es la segunda con más segmentos clasificados y, si se suman las puntuaciones de las tres categorías más bajas (*Incomprehensible*, *Disfluent language* y *Non-native language*), vemos que no existe una gran diferencia con las categorías más altas. En definitiva, las puntuaciones de fluidez evidencian que, aunque la calidad de los resultados de TA es buena, a Google todavía le queda mucho camino por recorrer.

En el caso de la adecuación sucede lo contrario, hay una clara diferencia entre la categoría de *All meaning*, donde están clasificados la mayoría de los segmentos, y las demás. De hecho, si sumamos el resto de las categorías de adecuación, vemos que *All meaning* sigue obteniendo la mayoría. En muchas ocasiones, el significado del texto original está presente en la TA, pero la fluidez no es la adecuada para el público destinatario; por ejemplo, los casos en los que se ha empleado la variante del español de América.

Una de las funciones que incluye COSTA, a diferencia de BLEU, es que el evaluador, de forma opcional, puede clasificar los errores encontrados en la traducción y, si lo considera necesario, añadir comentarios. En este estudio se ha decidido utilizar esta función para ver qué tipos y cantidad de errores se obtienen en cada caso. La clasificación obtenida con Google es la siguiente:

Grammar		Words		Style	
Verb inflection	5	Single words	1	Acronyms - Abbreviations	0
Noun inflection	0	Multi-word units - Idioms	0	Extra words	0
Other inflection	1	Terminology	8	Country standards	8
Wrong category	1	Untranslated words	4	Spelling errors	0
Article	2	Incomprehensible	0	Accent	0
Preposition	3	Literal translation	9	Capitalization	1
Agreement	0	Conjunctions	0	Punctuation	0

Figura 20. Clasificación de errores de Google según COSTA (2020)

De las tres grandes categorías que propone COSTA, la que más errores contiene es la de *Words*, en concreto errores de traducción literal y de terminología. Muchos de los errores encontrados están provocados por traducciones demasiado literales que han derivado en traducciones erróneas o incomprendibles. Los errores de terminología están relacionados con el mal uso del vocabulario propio del sector, en su mayoría del sector ferroviario. Por ejemplo, la traducción de «vía» por «pista» en el segmento 7 o la de «canceladora» por «validador» en el segmento 9.

La segunda clasificación con más errores es la de *Grammar*. Los errores están relacionados principalmente con los verbos y con el uso de preposiciones y artículos. En gran parte del texto, la TA ha cambiado los modos verbales. Si bien esto no convierte la traducción en errónea, no se están siguiendo las directrices del original. Los carteles de FGC interactúan en todo momento con el usuario. Es decir, emplean el imperativo para dar instrucciones u órdenes al público. Por ejemplo, la traducción de referencia del segmento 10 es «Dejen salir», pero se ha traducido como «Permitir salida», por lo que en cierto modo se está perdiendo el matiz de dar una orden y de interacción con el usuario.

En cuanto a *Style*, prácticamente todos los errores se encuentran en la categoría *Country Standards*. Esta categoría está relacionada con la adaptación de direcciones, monedas o números de teléfono al público destinatario. Por ejemplo, en el segmento 50 no se ha adaptado el sistema horario: «Los compañeros de trabajo de Fix y las oficinas privadas tienen acceso 24/7 al centro, sin embargo, de 10pm a 7am la alarma de seguridad se activa automáticamente». No obstante, y ante la falta de una categoría específica para ello, los errores de uso de otras variantes del español se incluyen dentro de esta categoría, ya que, al no adaptarse a la comunidad de hablantes a la que va dirigida la traducción, se considera que no se están cumpliendo los estándares del país, en este caso España. Por ejemplo, el uso de palabras como «boleto (billete)», «elevador (ascensor)» o «identificación (DNI)».

5. EVALUACIÓN DE MICROSOFT TRANSLATOR IMAGES

El primer traductor automático de Microsoft fue desarrollado entre 1999 y el 2000 por Microsoft Research, sección de la compañía dedicada a la investigación, y estaba basado en el corrector gramatical de Microsoft Word. Así como muchos traductores automáticos en la actualidad, su funcionamiento se rige por los datos almacenados en el sistema. Es decir, en lugar de utilizar las normas lingüísticas de cada idioma, los algoritmos son capaces de interpretar textos paralelos traducidos y así aprender a traducir textos nuevos de forma natural. El traductor en línea dirigido al consumidor es conocido como Bing Translator, lanzado en 2007, y ofrece traducciones gratuitas en diversos idiomas. También ofrece la posibilidad de traducir sitios web mediante Bilingual Viewer (Wikipedia, 2020).

En 2011, y con el objetivo de ampliar su oferta, Microsoft creó una API de Microsoft Translator basada en la nube para poder incluirla en muchos de sus productos. Esta API forma parte de la colección de APIs de Microsoft, llamada Microsoft Cognitive Services, y puede utilizarse para crear sitios web, aplicaciones, herramientas o cualquier otro producto que necesite un servicio multilingüe. Las principales características de esta API son: el uso de TAN (los idiomas no disponibles en TAN utilizan TAE), la posibilidad de traducir a varios idiomas a la vez, la detección automática del idioma del TO, la transliteración (pensado para lenguas con alfabeto distinto, como el chino) y un diccionario bilingüe capaz de ofrecer alternativas de traducción (Microsoft, 2020).



Figura 21. Logotipo de Microsoft Translator (Microsoft, 2020)

El servicio de APIs de Microsoft (Cognitive Services) y, por lo tanto, Microsoft Translator, se encuentra almacenado en la nube en una plataforma llamada Azure. Aquí podemos encontrar todos los servicios que Microsoft puede ofrecer, entre ellos los necesarios para el correcto funcionamiento de la aplicación Microsoft Translator Images. Para esta aplicación, se utilizan: Computer Vision, encargado de reconocer y extraer el texto impreso o escrito a mano, y Microsoft Translator, que se encarga de llevar a cabo la traducción (Microsoft Azure, 2020).

Esta aplicación móvil es el equivalente de Google Translate Images, ya que ambas funcionan mediante el reconocimiento óptico de caracteres y TAN. La diferencia entre ambas es que Google Lens incluye muchas más funciones además de la traducción de imágenes, mientras que los servicios de la aplicación de Microsoft Translator se orientan exclusivamente a la traducción. En 2016, Microsoft lanzó su primer servicio de traducción de voz, también incluido en la aplicación. En total, ofrece cuatro posibilidades distintas de traducción: traducción de texto, traducción de imágenes, traducción de audio y traducción de conversaciones.

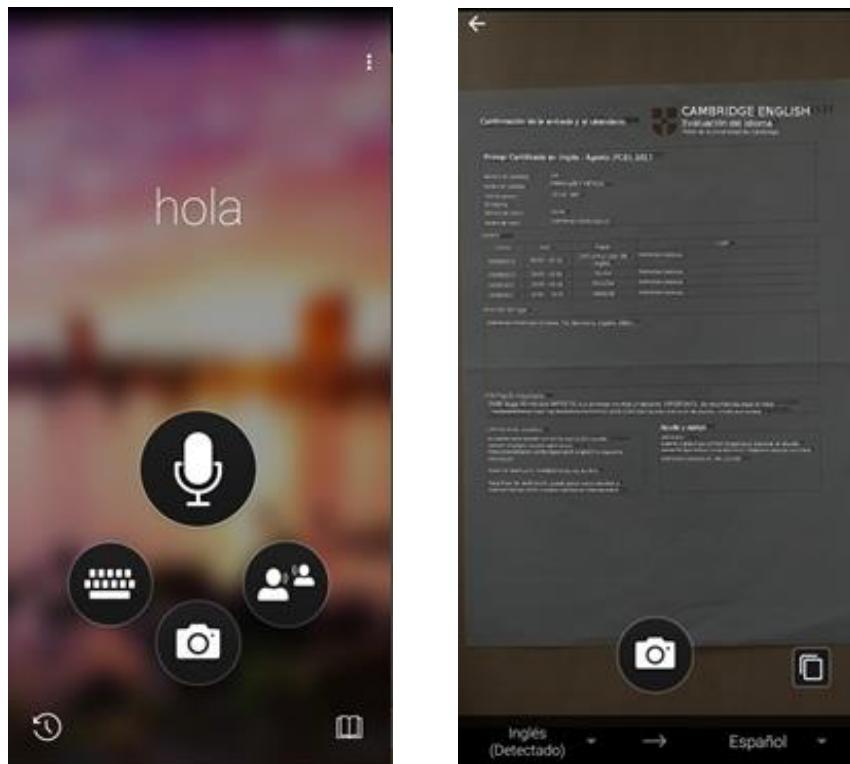


Figura 22. Interfaz de Microsoft Translator Images (Microsoft, 2020)

5.1. Análisis de los textos

En este apartado se procede al análisis de los resultados de TA proporcionados por Microsoft Translator Images. En primer lugar, se presentan los resultados de TA en forma de tabla y, a continuación, se analizan con detalle las puntuaciones obtenidas con BLEU y COSTA. La metodología del estudio puede consultarse en la INTRODUCCIÓN y los informes generados están en los archivos adjuntos.

	Seg.	Texto original	Traducción automática	Traducción humana
1	1	Only for clients	Sólo para clientes	Solo para clientes
	2	Timetable. While this station is open.	Horario. Mientras que esta estación está abierto.	Horario. Durante las horas de servicio de esta estación
	3	Ferrocarrils de la Generalitat de Catalunya assumes no responsibility for bicycles parked on its premises.	Ferrocarrils de la Generalitat de Catalunya asume ninguna responsabilidad por bicicletas estacionadas en sus locales.	Ferrocarrils de la Generalitat de Catalunya no se responsabiliza de las bicicletas estacionadas en sus dependencias.
2	4	In case of fire	En caso de incendio	En caso de incendio
	5	Do not use the lift	No utilice el ascensor	No use el ascensor
	6	Use the stairs	Usa las escaleras	Use las escaleras
3	7	Going down on to the track is prohibited	Bajando a la pista está prohibida	No se permite bajar a la vía
4	8	Non-opening of doors	Sin apertura ofdoors	Si las puertas no se abren
	9	If you insert the ticket into the validator and the door does not open, go to the blue vending machine in the platform 1 of this station and follow the instructions on the screen.	Si inserta el billete en el validador y la puerta no abrir, ir a la azul máquina expendedora en la plataforma 1 de este estación y seguir el instrucciones sobre el / pantalla.	Si, una vez introducido el billete en la canceladora, la puerta no se abre, diríjase a la máquina expendedora de color azul, situada en el andén de vía 1, y siga las instrucciones de la pantalla.

5	10	Allow way out	Al/ow way out	Dejen salir
6	11	Instructions of use:	Instrucciones de uso:	Instrucciones de uso:
	12	To purchase a ticket:	Para comprar un billete:	Para comprar un billete:
	13	1. Choose the ticket you require by pressing the corresponding button.	o Elija el boleto que necesita para pulsando los botones correspondientes	1. Elegir el billete pulsando el botón correspondiente.
	14	2. Insert coins, notes or credit card.	Inserte monedas, billetes o tarjeta de crédito.	2. Insertar monedas, billetes o tarjeta de crédito.
	15	3. Collect your ticket and the change or the receipt, if required.	o Recoge tu billete y el cambio o el recibo, si es necesario.	3. Recoger el título o billete y el cambio o el recibo, si corresponde.
	7	16	Space for preferential use	Espacio para uso preferencial
8	17	Attention	Atención	Atención
	18	If the lift is jammed:	Si el ascensor está atascado	En el caso de que el ascensor quede bloqueado:
	19	Do not attempt to open the doors	No intente abrir las puertas	No intente abrir las puertas
	20	Use the intercom to communicate with the Control Centre	Utilice el intercomunicador para comunicarse con el Centro de Control	Use el interfono para comunicarse con el Centro de Control
	21	If the event of a power cut, press the button for outside communication consecutively	Si el evento de un corte de energía, presione el abeto. Botón fuera de la comunicación consecutjve y	En caso de corte del fluido eléctrico, pulse el botón para comunicarse con el exterior
	22	Press the button. The TV camera will start operating	Press the button. The TV camera will start operating	Pulse el botón. La cámara de TV se activará.
	23	In a few moments the lift will stop at the lower floor and	En unos momentos la llft se detendrá en el suelo más bajo y las	En breves momentos el ascensor se situará en el piso inferior y se

		the doors will open	puertas se abren	abrirán las puertas
9	24	Price	Price	Precio
	25	Validity period	Validity period	Tiempo de validez
	26	A single person timed ticket for 10 journeys on any means of transport within the integrated system, for the number of zones purchased (maximum charge 6 zones). This ticket is not valid for the Aeroport T1 and Aeroport T2 metro station stops on line L9 Sud.	Un billete cronometrado para una sola persona para 10 viajes por cualquier medio transporte dentro del sistema integrado, por el número de zonas compradas (carga máxima 6 zonas). Este billete es no es válido para la estación de metro Airport T1 y Airport T2 paradas en la línea L9 Sud.	Título unipersonal y horario, de 10 viajes integrados en todos los modos de transporte, según las zonas por las que se transite (pago máximo 6 zonas). No es válido en las estaciones de metro de Aeroport T1 y Aeroport T2 de la línea L9 Sud.
10	27	Station where all trains stop	Statjon wtwe todos los trenes se detienen	Estación donde paran todos los trenes
	28	Station where only trains from the line shown stop	Estación donde sólo trenes de la línea se muestra parada	Estación donde solo paran los trenes de las líneas señaladas
	29	Metro connection	Conexión metro	Enlace con Metro
	30	Rodalies de Catalunya connection	Rodalies de Catalunya connection	Enlace con Rodalies de Catalunya
	31	Contribution Bus	Contribución	Autobús de aportación
	32	Local bus	Autobús local	Autobús urbano
	33	Funicular	Funicular	Funicular
	34	"Tramvia Blau" (tram)	"Blau tram" (tranvía)	"Tramvia Blau"
	35	Press the button and then release it	Pulse el botón y, a continuación, liberarlo	Pulse el botón y suéltelo
11	36	Wait while the light is flashing	Espere mientras la luz está Parpadear	Espere mientras la luz parpadea

	37	When the light remains steady, you may speak to FGC staff	Cuando la luz permanece constante, usted puede hablar con Personal de la FGC	Con la luz fija, ya puede hablar con el personal de FGC
	38	The light will turn off upon completion of the call	la luz se apagará finalización de la convocatoria	La luz se apagará cuando finalice la llamada
	39	You can also contact FGC staff at:	También puede ponerse en contacto con el personal de FGC en:	También pueden comunicarse con el personal de FGC a través de:
12	40	Office	Oficina	Oficina
	41	Offices	Oficinas	Oficinas
13	42	AT UTOPICUS, WE EMPTY THE FRIDGE EVERY FRIDAY TO CARRY OUT A DEEP CLEAN	EN UTOPICUS EMPTY EL FRIDGE CADA VIERNES PARA LLEVAR UN DEEP CLEAN	EN UTOPICUS, LOS VIERNES VACIAMOS LA NEVERA PARA HACER UNA LIMPIEZA EN PROFUNDIDAD
	43	PLEASE DONT FORGET TO TAKE YOUR TUPPERWARE AND FOOD HOME	POR FAVOR, NO OLVIDE TOMAR SU TUPPERWARE Y HOGAR DE ALIMENTOS	NO OLVIDES LLEVARTE TUS TUPPERS Y COMIDA
14	44	KEEPING THE COMMUNAL AREAS CLEAN BY WASHING UP OR USING THE DISHWASHER FORMS PART OF THE UTOPICUS CULTURE	MANTENER LIMPIAS LAS ZONAS COMUNES LAVAR O UTILIZAR LOS FORMS DE DISHWASHER PARTE DE LA CULTURA UTOPICUS	FREGAR O USAR EL LAVAJILLAS PARA MANTENER LOS ESPACIOS COMUNES LIMPIOS FORMA PARTE DE LA CULTURA DE UTOPICUS
	45	IF THE DISHWASHER IS IN USE OR FULL, PLEASE WASH THE DISHES THAT YOU HAVE USED. WE PROVIDE EVERYTHING	SI EL DISHWASHER ES EN USO O COMPLETO, POR FAVOR, lava los platos que has usado. PROPORCIONAMOS TODO LO QUE	SI EL LAVAJILLAS ESTÁ EN USO O LLENO, RIEGA EL MENAJE QUE HAYAS USADO. TENEMOS A TU DISPOSICIÓN

		YOU NEED TO DO SO	NECESITA PARA HACERLO.	TODO EL MATERIAL PARA HACERLO
15	46	HERE TO KEEP ALL YOUR STUFF	AQUÍ PARA GUARDAR TODAS SUS COSAS	AQUÍ PUEDES GUARDAR TUS COSAS
	47	BUT PLEASE,	PERO POR FAVOR,	PERO POR FAVOR,
	48	TAKE CARE OF MY INNER PEACE	CUIDA DE MI PAZ INTERIOR	CUIDA DE MÍ
16	49	FIX COWORKERS AND PRIVATE OFFICES	COMPAÑEROS DE TRABAJO Y OFICINAS PRIVADAS	COWORKERS Y OFICINAS PRIVADAS
	50	Fix coworkers and private offices have 24/7 access to the centre, however, from 10pm to 7am the security alarm is automatically activated.	compañeros de trabajo y oficinas privadas tienen acceso 24/7 al centro, sin embargo, de la masion a la lam la alarma de seguridad se activa automáticamente.	Coworkers y oficinas privadas tienen acceso al centro 24 horas al día los 7 días de la semana. No obstante, de 10 de la noche a 7 de la mañana la alarma de seguridad se activa automáticamente.
	51	WHAT SHOULD I DO IF I AM STAYING DURING THIS TIME FRAME?	LO QUE DEBO HACER SI ESTOY QUEDÁNDOME ¿DURANTE ESTE TIEMPO?	¿QUÉ DEBO HACER SI ESTOY EN EL EDIFICIO DURANTE ESA FRANJA HORARIA?
	52	Notify the security company over the intercom:	Notificar a la compañía de seguridad a través del intercomunicador:	Notificar a la compañía de seguridad a través del interfono:
	53	1) Please identify yourself with your full name and ID.	1) Por favor, identifíquese con su nombre completo y su identificación.	1) Identificate con tu nombre completo y DNI.
	54	2) Notify until what time you will stay in the building to enable them to re-	2) Notificar hasta qué hora permanecerá en el edificio para permitirles reactivar la alarma	2) Notifica hasta qué hora permanecerás en el edificio para que puedan reactivar la

		activate the alarm after your departure.	después de su Salida.	alarma una vez te hayas ido.
55	Thank you	Gracias	Gracias	
56	*YOU CAN FIND FURTHER INFORMATION IN THE WELCOME TUTORIAL	*PUEDE		*PUEDES ENCONTRAR MÁS INFORMACIÓN EN EL TUTORIAL DE BIENVENIDA

Tabla 3. Resultados de TA de Microsoft Translator Images (2020)

5.1.1. Análisis con BLEU

Una vez finalizado el análisis y la evaluación de TA de Google, en este apartado se presentan los resultados obtenidos con Microsoft. En primer lugar, se lleva a cabo la evaluación con BLEU y, a continuación, la evaluación con COSTA. En el caso de BLEU, el análisis también se ha llevado a cabo con la herramienta *Interactive BLEU score evaluator*, de Tilde (ver apartado [BLEU](#)). Recordamos que esta herramienta pondera las puntuaciones sobre 100.

En el caso de Microsoft, la puntuación que BLEU ha otorgado al resultado de TA de su traductor automático es de 25.33 puntos sobre 100.

BLEU:	25.33
Precision x brevity:	25.91 x 97.74
Type	1-gram 2-gram 3-gram 4-gram
Individual	54.58 31.20 19.81 13.37
Cumulative	53.34 40.33 31.58 25.33

Figura 23. Puntuación de BLEU para Microsoft (2020)

Para la evaluación de Microsoft se ha seguido el mismo procedimiento. En primer lugar, se analiza segmento por segmento para ver cómo ha asignado BLEU las puntuaciones. A continuación, se comentan algunos casos relevantes. Para ver el informe completo de BLEU para Microsoft, véase los documentos adjuntos.

Una vez realizado el análisis, se ha utilizado la misma clasificación que en el caso de Google para representar de forma gráfica los resultados: **Coincidencia exacta** (TA y TH son exactamente iguales), **Ambas opciones correctas** (ambas opciones son válidas pero BLEU penaliza la TA por utilizar palabras diferentes), **Incomprendible** (la TA no tiene sentido), **No traducido** (TO y TA se mantienen igual), **Diferencia dialectal**

(en la TA se encuentra vocabulario propio del español de América) y **Errores leves** (la TA es comprensible, aunque con algunos fallos de menor importancia).

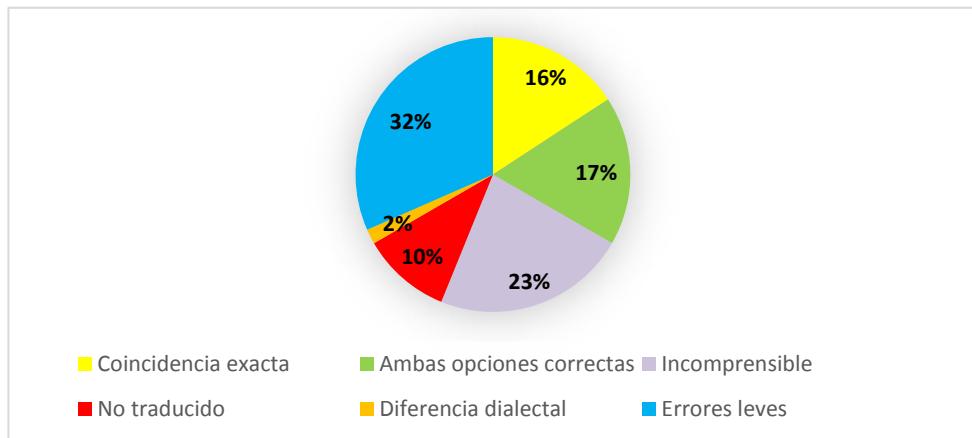


Figura 24. Gráfico con los resultados de BLEU para Microsoft (2020)

En el gráfico se observa que en el caso de Microsoft los errores más frecuentes (de acuerdo con nuestra clasificación) son los errores leves con un 32%, los cuales no implican grandes problemas a la hora de comprender el texto, y los errores incomprensibles con un 23%, que dificultan en gran medida la fluidez y adecuación del texto. Al observar el elevado porcentaje de estos errores y la puntuación global de BLEU, resulta evidente que la calidad de los resultados es bastante deficiente, ya que la gravedad de los errores encontrados es mayor e incluso hace que en ocasiones el texto resulte del todo incomprensible.

Un dato que llama bastante la atención, es que, en algunos casos, a pesar de no haber ninguna coincidencia entre la TA y la TH, BLEU ha asignado una puntuación al segmento en cuestión. Esto podría deberse al uso de bigramas y trigramas en el sistema.

Sentence 8	BLEU	Length ratio	
Source	-	-	Non-opening of doors
Human	100.00	1.00	Si las puertas no se abren
Machine	8.31	0.50	Sin apertura ofdoors

Figura 25. Ejemplo 1 de la evaluación de BLEU para Microsoft (2020)

Los casos leves corresponden a errores gramaticales, incoherencias, cambios en la forma de dirigirse al usuario o puntuación, entre otros. Por lo tanto, el usuario puede llegar a comprender el mensaje o hacerse una idea de lo que quiere expresar el TO. No

obstante, la suma de todos estos pequeños errores hace que todo el texto pierda calidad en cuanto a fluidez y adecuación.

Source	-	-	Timetable. While this station is open.
Human	100.00	1.00	Horario . Durante las horas de servicio de esta estación
Machine	12.07	0.90	Horario . Mientras que esta estación está abierto .

Figura 26. Ejemplo 2 de la evaluación de BLEU para Microsoft (2020)

A lo largo del texto también se encuentran con la misma frecuencia segmentos en los que TA y TH se han traducido exactamente igual, por lo que han recibido la máxima puntuación, y en los que ambas opciones eran correctas, pero BLEU no las ha dado como válidas al usar palabras distintas, lo que reduce en gran medida la puntuación global de todo el texto.

Source	-	-	Do not use the lift
Human	100.00	1.00	No use el ascensor
Machine	35.36	1.00	No utilice el ascensor

Figura 27. Ejemplo 3 de la evaluación de BLEU para Microsoft (2020)

Este caso sirve también de ejemplo para observar algunas incoherencias en el sistema de puntuaciones de BLEU. En el ejemplo 3, ambas oraciones son correctas y representan la misma idea, incluso emplean la misma forma de trato al usuario, pero BLEU penaliza con mucha puntuación el hecho de utilizar una palabra distinta. Por otro lado, en el ejemplo 4 vemos que existe una clara diferencia entre ambos segmentos, ya que la TA ha cambiado la forma de trato al usuario y la importancia de este cambio es superior a la del ejemplo 3. Sin embargo, BLEU otorga una puntuación más alta al ejemplo 4 porque solo cambia una letra, mientras que en el ejemplo 3 cambia toda una palabra, a pesar de que la traducción es más adecuada.

Source	-	-	Use the stairs
Human	100.00	1.00	Use las escaleras
Machine	63.89	1.00	Usa las escaleras

Figura 28. Ejemplo 4 de la evaluación de BLEU para Microsoft (2020)

En Microsoft aparecen con frecuencia casos de segmentos no traducidos o con algunas partes no traducidas. La causa podría ser que, o bien el motor no reconocía las palabras, es decir, una cuestión de falta de entrenamiento del motor, o bien se trata de un fallo técnico de la aplicación relacionado con el sistema de reconocimiento óptico de caracteres. Esta segunda opción parece la más probable, ya que en muchas ocasiones,

además de no traducir la palabra, se han añadido o cambiado algunos caracteres por otros que no tienen ningún sentido en el segmento, como se ve en los siguientes ejemplos:

Source	-	-	AT UTOPICUS, WE EMPTY THE FRIDGE EVERY FRIDAY TO CARRY OUT A DEEP CLEAN
Human	100.00	1.00	EN UTOPICUS , LOS VIERNES VACIAMOS LA NEVERA PARA HACER UNA LIMPIEZA EN PROFUNDIDAD
Machine	6.82	0.86	EN UTOPICUS EMPTY EL FRIDGE CADA VIERNES PARA LLEVAR UN DEEP CLEAN

Figura 29. Ejemplo 5 de la evaluación de BLEU para Microsoft (2020)

Source	-	-	Allow way out
Human	100.00	1.00	Dejen salir
Machine	7.99	2.00	Al/ ow way out

Figura 30. Ejemplo 6 de la evaluación de BLEU para Microsoft (2020)

En los resultados de Microsoft no se han encontrado muchos ejemplos de casos en los que se haga uso de otras variantes del español, tan solo ha ocurrido en una ocasión y no afecta a la adecuación del texto. No obstante, sí puede resultar extraño para el público al que vaya dirigido, en este caso al público de España.

Source	-	-	1. Choose the ticket you require by pressing the corresponding button.
Human	100.00	1.00	1. Elegir el billete pulsando el botón correspondiente .
Machine	4.46	1.22	o Elija el boleto que necesita para pulsando los botones correspondientes

Figura 31. Ejemplo 7 de la evaluación de BLEU para Microsoft (2020)

5.1.2. Análisis con COSTA MT Evaluation Tool

En este apartado se presentan los resultados de la evaluación de TA de Microsoft obtenidos con la herramienta *COSTA MT Evaluation Tool*. En este caso, se ha creado el proyecto de evaluación *COSTAMT_Evaluation_Microsoft*.

El procedimiento ha sido prácticamente el mismo que en el caso de Google. Se han obtenido dos puntuaciones: una para la fluidez del texto y otra para la adecuación. Estos son los resultados:

Fluency: 0.66071427 / 1 Adequacy: 0.675 / 1

Figura 32. Puntuación de COSTA para Microsoft (2020)

Como podemos observar en la imagen, ambas puntuaciones son bastante bajas, lo cual significa que en la gran mayoría del texto no se ha cumplido el objetivo de trasladar el mensaje original a la lengua meta y, por lo tanto, la calidad global del texto se ha visto afectada. Prueba de ello es que solo un 67% del TO está presente en la traducción y la fluidez ha obtenido un porcentaje del 66%. En los siguientes párrafos, se analiza en profundidad el resultado de COSTA sobre Microsoft. El informe completo de la evaluación se puede consultar en los documentos adjuntos.

Recordamos que, a la hora de puntuar la fluidez y la adecuación de un segmento, disponemos de 5 opciones según el grado de calidad de cada segmento. En el caso de la fluidez: *Incomprehensible*, *Disfluent language*, *Non-native language*, *Good language* y *Flawless language*. Y en el caso de la adecuación: *None*, *Little meaning*, *Much meaning*, *Most meaning* y *All meaning*. Los resultados son los siguientes:

Fluency	Adequacy:		
Incomprehensible	14	None	11
Good language	9	Little meaning	11
Non-native language	4	Much meaning	5
Disfluent language	4	Most meaning	4
Flawless language	25	All meaning	25

Figura 33. Puntuaciones de fluidez y adecuación de Microsoft según COSTA (2020)

En la tabla se puede observar que las opciones más seleccionadas en ambos casos son las que representan una buena y una mala calidad del texto: *Flawless language* y *Incomprehensible* en fluidez; *All meaning* y *None* y *Little meaning* en adecuación.

Con respecto a la fluidez, el error más frecuente es el de *Flawless language*. Sin embargo, si se suman las puntuaciones del resto de categorías la cifra es mayor a la de esa categoría en concreto. Esto quiere decir que, aunque la opción más seleccionada es la de «lenguaje perfecto», lo cierto es que desde un punto de vista global la fluidez dificulta la comprensión del texto, algo que se puede comprobar en la puntuación final obtenida. Asimismo, la segunda categoría más frecuente es la de *Incomprehensible*, con 14 casos, la cual es una cifra bastante elevada. Por ejemplo, en el caso del segmento 21 la traducción resultante es del todo ininteligible: «Si el evento de un corte de energía, presione el abeto. Botón fuera de la comunicación consecutve y».

Lo mismo sucede en el caso de la adecuación. Vemos que la opción más seleccionada es la de *All meaning*, pero si sumamos el resto de las categorías, sus puntuaciones la superan. De hecho, los siguientes errores más frecuentes son también los que indican una calidad muy baja del texto: *None* y *Little meaning*.

A continuación, se analiza y comenta la clasificación de errores obtenida a partir de las categorías de error que ofrece COSTA. Esta es una función opcional que incluye el programa. Los resultados son los siguientes:

Grammar		Words		Style	
Verb inflection	6	Single words	4	Acronyms - Abbreviations	2
Noun inflection	0	Multi-word units - Idioms	0	Extra words	0
Other inflection	1	Terminology	5	Country standards	3
Wrong category	0	Untranslated words	14	Spelling errors	0
Article	1	Incomprehensible	1	Accent	0
Preposition	2	Literal translation	14	Capitalization	3
Agreement	6	Conjunctions	0	Punctuation	2

Figura 34. Clasificación de errores de Microsoft según COSTA (2020)

En la categoría de *Grammar* los errores más frecuentes que se han encontrado son los de *Verb inflection* y *Agreement*. La primera categoría está relacionada con un mal uso de las formas verbales, es decir, utilizar un tiempo o modo verbal incorrecto. La segunda, con errores de concordancia entre las palabras del texto, por ejemplo, entre verbo y sujeto o verbo y adjetivo. Encontramos algunos ejemplos en el segmento 2, «Horario. Mientras que esta estación está abierto», y en el segmento 36, «Espere mientras la luz está Parpadear».

La categoría de *Words* es la que más errores contiene. En primer lugar, se encuentran los de *Untranslated words* y *Literal translations*. En el caso de las palabras no traducidas, puede deberse a fallos técnicos de la aplicación relacionados con el

reconocimiento óptico de caracteres, lo que ha impedido que el sistema pudiera reconocer los caracteres y, por ende, traducirlos. Por otra parte, las traducciones demasiado literales de Microsoft han provocado que en muchas ocasiones el texto no se entendiera y esto a su vez ha afectado a la adecuación del texto. Por ejemplo, en el segmento 8 «Sin apertura ofdoors», donde una parte del texto se ha quedado sin traducir, o el segmento 28 «Estación donde sólo trenes de la línea se muestra parada», donde la traducción es tan literal que es prácticamente incomprensible. No obstante, se han registrado pocos errores de terminología.

Por último, la categoría con menos errores ha sido la de *Style*. Principalmente se trata de errores de *Capitalization*, uso incorrecto de mayúsculas y minúsculas, y *Country Standards*, la no adaptación de direcciones, monedas, números de teléfono, entre otros. De la misma manera que con Google, se han incluido dentro de esta categoría los errores relacionados con diferencias dialectales, en este caso con el uso de vocabulario propio del español de América: «o Elija el boleto que necesita para pulsando los botones correspondientes», en el segmento 13.

Como ya se comenta anteriormente, la función de clasificación de errores y comentarios es opcional con la herramienta COSTA. Además de la categorización de errores, también se ha probado la funcionalidad de añadir comentarios, lo cual puede resultar bastante útil por si el revisor quiere matizar o justificar alguna elección. Estos comentarios quedan registrados en el informe Excel que se genera una vez terminada la evaluación (todos los comentarios pueden consultarse en el informe incluido en los documentos adjuntos).

Translation error classification											
Grammar:	<input type="checkbox"/> Verb inflection	<input type="checkbox"/> Noun inflection	<input type="checkbox"/> Other inflection	<input type="checkbox"/> Wrong category	<input type="checkbox"/> Article	<input type="checkbox"/> Preposition	<input checked="" type="checkbox"/> Agreement				
Comments:											
Words:	<input type="checkbox"/> Single words	<input type="checkbox"/> Multi-word units	<input checked="" type="checkbox"/> Terminology	<input type="checkbox"/> Untranslated words	<input type="checkbox"/> Ambiguous translation	<input checked="" type="checkbox"/> Literal translation	<input type="checkbox"/> Conjunctions				
Comments:	Wrong translation of "track". Translation too literal that makes no sense.										
Style:	<input type="checkbox"/> Acronyms - Abbreviations	<input type="checkbox"/> Extra words	<input type="checkbox"/> Country standards	<input type="checkbox"/> Spelling errors	<input type="checkbox"/> Accent	<input type="checkbox"/> Capitalization	<input type="checkbox"/> Punctuation				
Comments:											

Figura 35. Función de añadir comentarios de COSTA (2020)

6. COMPARACIÓN DE RESULTADOS

En el siguiente apartado se presenta la comparación de resultados entre Google y Microsoft. En primer lugar, se comparan los resultados obtenidos con BLEU y, a continuación, los obtenidos con COSTA. Por último, se exponen las diferencias y problemas técnicos encontrados con las herramientas de evaluación y con el funcionamiento de ambas aplicaciones.

6.1. Comparación con BLEU

Recordamos los resultados obtenidos con la herramienta en línea *Interactive BLEU score evaluator*. En la parte izquierda de la imagen, pueden observarse los resultados de Google y, a la derecha, los de Microsoft.

BLEU:	33.81	BLEU:	25.33
Precision x brevity:	35.14 x 96.20	Precision x brevity:	25.91 x 97.74

Figura 36. Comparación de Google y Microsoft con BLEU (2020)

Como se puede apreciar, Google ha obtenido una puntuación notablemente más alta, por lo que, según el sistema BLEU, sería la mejor alternativa a la hora de decidir qué aplicación de traducción de imágenes utilizar. Asimismo, la puntuación en precisión es casi diez puntos más alta que la de Microsoft. No obstante, en cuanto a la brevedad de los segmentos, la puntuación de Microsoft es ligeramente más elevada. Esto significa que las traducciones de Microsoft se han ajustado más al número de caracteres del texto original.

A pesar de la claridad de los resultados, llama la atención que en ambos casos la puntuación está por debajo de 50 puntos (sobre un total de 100), lo que significaría que ambas opciones han proporcionado resultados con un gran margen de mejora. No obstante, esto puede deberse a la cantidad y tamaño del corpus de referencia empleado. Tal y como se explica en el apartado de BLEU, «se recomienda utilizar un corpus de mil frases o más, ya que una muestra más pequeña podría reducir la calidad de las traducciones al no tener suficiente material de referencia con el que llevar a cabo las comparaciones necesarias». En nuestro caso, solo se ha utilizado como material de referencia la traducción humana de los 56 segmentos, por lo que esto podría explicar los bajos resultados obtenidos.

En cuanto a la clasificación creada para poder categorizar los segmentos y así tener una visión global del análisis de BLEU, estas son las principales diferencias entre Google y Microsoft:

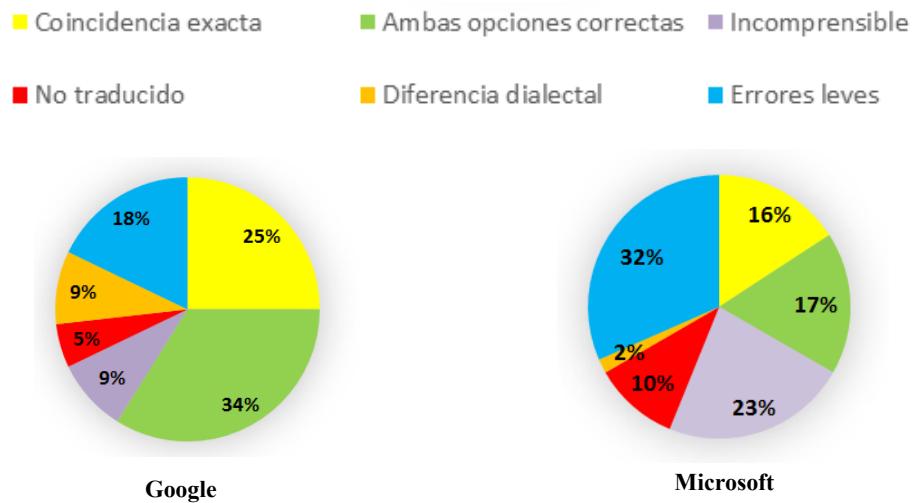


Figura 37. Gráfico con los resultados de BLEU para Google y Microsoft (2020)

Más de la mitad de los resultados de Google se encuentran en las categorías **Coincidencia exacta** y **Ambas opciones correctas**, lo que también quiere decir que más de la mitad de los resultados, un 59%, son buenas opciones de traducción. Sin embargo, en el caso de Microsoft ocurre lo contrario, más de la mitad de los resultados están dentro de las categorías **Errores leves**, **No traducido** e **Incomprensible**, por lo tanto, la gran mayoría de las traducciones obtenidas, un 65%, no se aceptarían como válidas. En ambos casos hay segmentos traducidos correctamente, segmentos que Google ha traducido bien pero Microsoft no, y viceversa, y segmentos del todo incomprensibles. Se toman como ejemplo las siguientes imágenes, donde Google corresponde al color azul y Microsoft al color verde (codificación de colores establecida por la herramienta):

Source	-	-	Press the button. The TV camera will start operating
Human	100.00	1.00	Pulse el botón . La cámara de TV se activará .
Machine	53.32	1.09	Presione el botón . La cámara de TV comenzará a funcionar .
Machine	4.52	0.91	Press the button . The TV camera will start operating

Figura 38. Ejemplo 1 de la comparación de BLEU sobre Google y Microsoft (2020)

Source	-	-	KEEPING THE COMMUNAL AREAS CLEAN BY WASHING UP OR USING THE DISHWASHER FORMS PART OF THE UTOPICUS CULTURE
Human	100.00	1.00	FREGAR O USAR EL LAVAJILLAS PARA MANTENER LOS ESPACIOS COMUNES LIMPIOS FORMA PARTE DE LA CULTURA DE UTOPICUS
Machine	29.28	1.00	MANTENER LIMPIAS LAS ÁREAS COMUNES LAVANDO O UTILIZANDO LAS FORMAS DE LAVAJILLAS PARTE DE LA CULTURA DE UTOPICUS
Machine	16.97	0.94	MANTENER LIMPIAS LAS ZONAS COMUNES LAVAR O UTILIZAR LOS FORMSDE DE DISHWASHER PARTE DE LA CULTURA UTOPICUS

Figura 39. Ejemplo 2 de la comparación de BLEU sobre Google y Microsoft (2020)

Un problema frecuente tanto en Google como Microsoft es que en muchas ocasiones ambas opciones de traducción son correctas, pero BLEU asigna puntuaciones muy bajas solo por emplear sinónimos o palabras distintas. Esto pone en duda una vez más la efectividad del sistema de evaluación, ya que no tiene en cuenta alternativas que son igualmente válidas.

Source	-	-	Use the intercom to communicate with the Control Centre
Human	100.00	1.00	Use el interfono para comunicarse con el Centro de Control
Machine	66.06	1.00	Usar el intercomunicador para comunicarse con el centro de control
Machine	66.06	1.00	Utilice el intercomunicador para comunicarse con el Centro de Control

Figura 40. Ejemplo 3 de la comparación de BLEU sobre Google y Microsoft (2020)

6.2. Comparación con COSTA MT Evaluation Tool

A diferencia de BLEU, la cantidad de corpus utilizada con COSTA en la evaluación no es tan importante. Solo se necesita la traducción humana del segmento en cuestión, ya que la evaluación la lleva a cabo un revisor humano, no una máquina mediante fórmulas y algoritmos. COSTA solo utiliza las fórmulas para el cálculo de la puntuación final en función de las opciones seleccionadas por el evaluador humano.

Recordamos los resultados obtenidos con la herramienta de evaluación manual asistida COSTA MT Evaluation Tool:

Fluency: 0.7785714 / 1	Fluency: 0.66071427 / 1
Adequacy: 0.83214283 / 1	Adequacy: 0.675 / 1

Google

Microsoft

Figura 41. Puntuación de COSTA para Google y Microsoft (2020)

Con los resultados obtenidos en esta herramienta se confirman por segunda vez los ya obtenidos con BLEU: Google sigue siendo la opción con mejores resultados. No obstante, con esta herramienta la diferencia de puntuaciones entre ambos motores es mayor. Google obtiene una media de 8 puntos sobre 10 en fluidez y adecuación, mientras que Microsoft no logra llegar al 7 en ninguno de los dos casos. En cuanto a las diferentes opciones de calidad a nivel de segmento, ambos obtienen buenas

puntuaciones en las categorías más altas, *Flawless language* y *All meaning*, pero Microsoft es el que comete más errores en las categorías de *Incomprehensible*, *Little meaning* y *None*, las cuales imposibilitan la correcta fluidez del texto y, además, indican que falta buena parte de su contenido.

Fluency		Adequacy:		Fluency		Adequacy:	
Incomprehensible	14	None	11	Incomprehensible	5	None	4
Good language	9	Little meaning	11	Good language	4	Little meaning	6
Non-native language	4	Much meaning	5	Non-native language	9	Much meaning	3
Disfluent language	4	Most meaning	4	Disfluent language	12	Most meaning	7
Flawless language	25	All meaning	25	Flawless language	26	All meaning	36

Google
Microsoft

Figura 42. Puntuaciones de fluidez y adecuación de Google y Microsoft según COSTA (2020)

En ambos casos, se ha hecho uso de las funciones de añadir comentarios y clasificación de errores de esta herramienta. Los resultados son los siguientes:

Grammar		Words		Style	
Verb inflection	5	Single words	1	Acronyms - Abbreviations	0
Noun inflection	0	Multi-word units - Idioms	0	Extra words	0
Other inflection	1	Terminology	8	Country standards	8
Wrong category	1	Untranslated words	4	Spelling errors	0
Article	2	Incomprehensible	0	Accent	0
Preposition	3	Literal translation	9	Capitalization	1
Agreement	0	Conjunctions	0	Punctuation	0

Google

Grammar		Words		Style	
Verb inflection	6	Single words	4	Acronyms - Abbreviations	2
Noun inflection	0	Multi-word units - Idioms	0	Extra words	0
Other inflection	1	Terminology	5	Country standards	3
Wrong category	0	Untranslated words	14	Spelling errors	0
Article	1	Incomprehensible	1	Accent	0
Preposition	2	Literal translation	14	Capitalization	3
Agreement	6	Conjunctions	0	Punctuation	2

Microsoft

Figura 43. Clasificación de errores de Google y Microsoft según COSTA (2020)

Los datos que más llaman la atención, y que ya se comentan en capítulos anteriores, son la cantidad de errores en la categoría *Words* en ambos casos, donde la mayoría se concentran en *Terminology* y *Literal Translation*. A lo largo del proceso de análisis de las traducciones aparecen con frecuencia casos en los que el sistema de TA traduce literalmente el TO, lo que provoca la aparición de numerosos sinsentidos tanto

en Google como en Microsoft. Este último, además, obtiene muchos errores en la categoría de *Untranslated Words*.

Otros errores menos frecuentes son los de *Verb inflection, Article, Preposition* y *Agreement*, en la categoría de *Grammar*. Todos consisten en errores de menor gravedad que afectan a la fluidez del texto, por ejemplo errores en el tiempo verbal, la adición u omisión de preposiciones y artículos y la concordancia entre diferentes sintagmas del texto (verbo-sujeto). En estos casos, Microsoft sigue siendo el que más errores comete. No obstante, con respecto a *Style*, en la categoría *Country Standards* (como se explica en el apartado Análisis con COSTA MT Evaluation Tool, esta categoría se ha utilizado para los errores relacionados con diferencias dialectales), queda reflejado que Google utiliza de manera más frecuente variantes del español en sus traducciones y Microsoft comete más errores en *Capitalization* y *Punctuation*.

Todos los datos recopilados con COSTA contrastan los obtenidos con BLEU, los cuales indican que más de la mitad de las traducciones de Google se pueden entender por parte del lector sin dificultad alguna y, por el contrario, las de Microsoft contienen muchas partes sin traducir y sinsentidos que imposibilitan la comprensión del texto traducido.

6.3. Cuestiones técnicas de Google y Microsoft

A la hora de utilizar las aplicaciones de Google y Microsoft han surgido varios problemas técnicos que han dificultado el proceso de análisis y evaluación de la TA. En este apartado se comentan esas y otras cuestiones generales relacionadas con el uso y funcionalidad de ambas aplicaciones.

Con respecto a Google Translate Images, el principal obstáculo es el formato en el que se muestra la TA. Una de las características de esta aplicación es que intenta presentar la traducción con un formato idéntico o lo más parecido posible al original y así ofrecer un resultado de calidad. No obstante, en ocasiones esto dificulta la tarea de leer y encontrar las partes del texto traducidas que nos interesan. Esto supone otra dificultad añadida debido a que, en ocasiones, Google traduce partes del texto que están en un idioma de origen diferente al que hemos seleccionado manualmente en la aplicación, lo que puede provocar confusión en el usuario y poner en duda el funcionamiento de la aplicación con respecto al reconocimiento óptico de caracteres y la detección automática del idioma. En el siguiente ejemplo se puede observar claramente este problema: a pesar de haber seleccionado el inglés como idioma de origen, la

aplicación realiza modificaciones en la versión en catalán, tanto en el formato como en el contenido del texto.



Figura 44. Ejemplo 1 de cuestiones técnicas de Google y Microsoft (2020)

Microsoft también presenta algunas dificultades con respecto al formato de la traducción. Al contrario que Google, Microsoft presenta la TA sobre la imagen original pero en un fondo oscuro y con la letra en color blanco. Esto también es un obstáculo a la hora de identificar la parte del texto traducida que nos interesa, ya que todo aparece en el mismo formato y, en ocasiones, con letra demasiado pequeña y sin espacios entre párrafos. En el siguiente ejemplo, Google Lens (izquierda) ofrece una apariencia mucho más clara y fácil de entender para el usuario. Sin embargo, Microsoft (derecha) transmite la sensación de que todo está desordenado y cuesta encontrar la información.

Bitllets Gratis		Billetes integrados / Tickets integrados					
T-casual 1 zona		2 zones	3 zones	4 zones	5 zones	6 zones	
Precio / Precio	11,35	22,40	30,50	39,20	45,05	47,90	
Tiempo de validez / Periodo de validez	1h 15' 1h 30 1h 45' 2h 00' 2h 15'					2h 30'	
Título unipersonal i horario, de 10 viatges integrats en tots els modes de transport segons les zones per les quals es transit (pago màxim 6 zones). No és vàlid a les estacions de metro d'Aeroport T1 i Aeroport T2 de la L9 Sud.							
Títol unipersonal i horario, de 10 viatges integrats en tots els modes de transport segons les zones per les quals es transit (pago màxim 6 zones). No és vàlid a les estacions de metro d'Aeroport T1 i Aeroport T2 de la línia L9 Sud.							
T-Usual		1 zona	2 zones	3 zones	4 zones	5 zones	6 zones
Precio / Precio	40,00	53,85	75,60	92,55	106,20	113,75	
Títol personal i intranferible, amb un nombre il·limitat de viatges integrats en 30 dies consecutius, a validació en tots els modes de transport segons les zones per les quals es transit.							
Notas "Tegrats rats		mm de Barcelona					
Billetes integrados / Integrated tickets							
T-c zona aual I							
Precio	11,35	22,40	30,50	39,20	45,05	47,90	
Período de validez	1h 15'	1h 30'	1h 45'	2h 00'	2h 15'	2h 30'	
Título y tarjetario Unipersonal, de 10 viajes integrados en todos los medios de transporte según las zonas para (Pago máximo 6 zonas). No es válido en las estaciones de metro d'Aeroport T1 y d'Aeroport T2 de la L9 Sud.							
Título y tarjetario Unipersonal, de 10 viajes integrados en todos los medios de transporte según las zonas para (Pago máximo 6 zonas). No es válido en las estaciones de metro d'Aeroport T1 y d'Aeroport T2 de la L9 Sud.							
T-usual		1 zona	2 zones	3 zones	4 zones	5 zones	6 zones
Precio / Precio	40,00	53,85	75,60	92,55	106,20	113,75	
Título personal e intranferible, con un número ilimitado de viajes integrados en 30 días consecutivos, en suscursales, a validación en todos los medios de transporte según las áreas, para las zonas de metro y tren.							

Figura 45. Ejemplo 2 de cuestiones técnicas de Google y Microsoft (2020)

Otras observaciones al utilizar Google Translate Images son los cambios en la puntuación entre el TO y la TA, el uso en ocasiones de variantes dialectales del español y errores en la segmentación de las oraciones que han causado malas traducciones o sinsentidos. Todo ello está relacionado con el reconocimiento óptico de caracteres y, si este factor falla, todo el funcionamiento de la aplicación puede hacerlo también. En el siguiente ejemplo, Google ha reconocido unos símbolos sin texto como parte traducible y los ha trasladado al texto meta por caracteres chinos y otros símbolos completamente distintos.

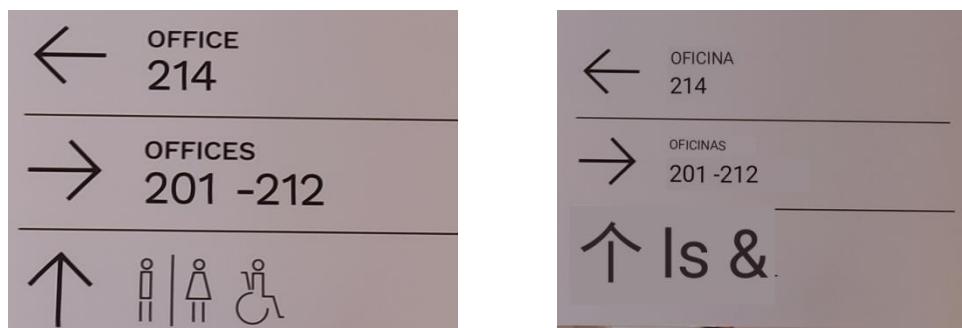


Figura 46. Ejemplo 3 de cuestiones técnicas de Google y Microsoft (2020)

Uno de los errores frecuentes que se ha comentado en el apartado Análisis con BLEU es que, en el caso de Microsoft, hay un elevado porcentaje de segmentos que no se han traducido o que contienen algunas partes sin traducir, lo cual está relacionado con fallos técnicos de la aplicación, más concretamente con el reconocimiento óptico de caracteres. Además, igual que en el caso de Google, en muchas ocasiones Microsoft también traduce todo el texto que reconoce, sin tener en cuenta la combinación lingüística seleccionada. El resultado son traducciones sin sentido y confusas para el usuario.

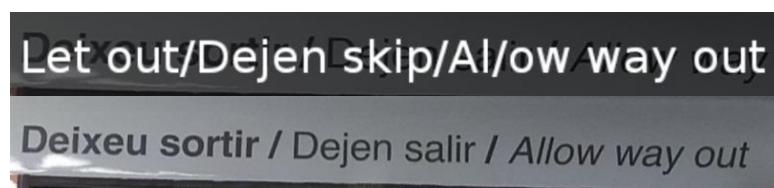


Figura 47. Ejemplo 4 de cuestiones técnicas de Google y Microsoft (2020)

6.4. Cuestiones técnicas de BLEU y COSTA

Con respecto a las herramientas de evaluación de TA utilizadas en este trabajo, el recurso en línea de la empresa Tilde para evaluar la calidad de TA con el sistema BLEU

no comporta complicaciones técnicas importantes, simplemente consiste en cargar los archivos en los apartados indicados y descargar los resultados en un documento CSV o Excel (véase los documentos adjuntos). No obstante, y a pesar de las facilidades técnicas encontradas, se ha podido comprobar que la efectividad del sistema no es del todo completa en textos y corpus pequeños, en este caso el que podemos encontrar en imágenes, ya que las puntuaciones obtenidas han sido bastante bajas. Además, en el caso de BLEU también hubo que transcribir los textos de las imágenes en documentos de texto para que la herramienta pudiera evaluarlos. No obstante, los requisitos técnicos no fueron tan específicos como en el caso de COSTA, como se explica a continuación.

A la hora de utilizar la herramienta COSTA, es muy importante cumplir los requisitos técnicos requeridos por la herramienta con los archivos de trabajo, ya que de no cumplirse el proceso de evaluación podría no realizarse correctamente. Como se explica en el apartado de COSTA MT Evaluation Tool, se necesitan archivos de texto sin formato en codificación UTF-8 y solo puede haber un segmento por línea.

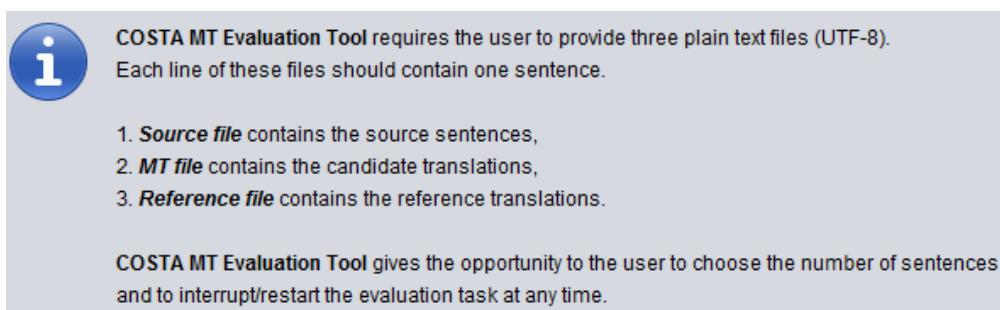


Figura 48. Requisitos de COSTA MT Evaluation Tool (2020)

Para ello, se han creado documentos de texto sin formato *.txt* para cada archivo de trabajo: texto original, traducción humana, traducción automática de Google y traducción automática de Microsoft, sin separación entre segmentos. Todos los archivos deben contener exactamente el mismo número de segmentos y cada segmento debe corresponderse en todos los archivos. Es decir, el segmento 1 tiene que ser el mismo en los tres archivos para que la evaluación funcione. Se trata de una tarea laboriosa que requiere bastante tiempo, ya que todo debe realizarse de forma manual para evitar posibles errores técnicos a la hora de subirlos a COSTA.

```
1 Only for clients
2 Timetable. While this station is open.
1 Solo para clientes
2 Horario. Durante las horas de servicio de esta estación
1 Solo para clientes
2 Calendario. Mientras esta estación está abierta.
1 Sólo para clientes
2 Horario. Mientras que esta estación está abierto.
```

Figura 49. Muestra de los archivos de trabajo en Notepad++. TO, TH, TA de Google y TA de Microsoft (2020)

Otro punto débil de esta herramienta está relacionado con la categorización de errores que proporciona, la cual es algo limitada. Además, si en una oración existen varios errores del mismo tipo, el programa solo puede contabilizar uno, por lo que no queda registro de los demás. No obstante, y a pesar de las dificultades técnicas encontradas, los resultados de la evaluación con COSTA se corresponden más con la realidad que los ofrecidos por BLEU. Esto se debe a que se trata de una herramienta de evaluación manual asistida, es decir, una herramienta de ayuda, ya que todo el trabajo de evaluación es llevado a cabo por un revisor humano, de ahí la fiabilidad de los resultados.

7. CONCLUSIONES

Una vez finalizado el estudio llevado a cabo en este trabajo, en el presente capítulo se describen las conclusiones obtenidas y se da respuesta a las preguntas y objetivos planteados al principio del trabajo, en el apartado Objetivos y preguntas de investigación.

En primer lugar, se ha podido comprobar la dificultad a la hora de definir el concepto de calidad en la industria de la traducción, ya que lo que en algunos casos puede considerarse mala calidad, puede ser buena calidad en otros escenarios. Esto se debe a que no existen dos proyectos de traducción exactamente iguales, todos tienen sus propias características específicas. No obstante, el principal objetivo de cualquier proyecto es lograr la satisfacción del cliente que realiza el encargo. Si el cliente queda satisfecho con el resultado, puede considerarse que se ha realizado un trabajo de calidad.

Esta dificultad con respecto al concepto de calidad también aparece cuando se trabaja con TA. Según las características del proyecto, los resultados de TA se podrán considerar o no como válidos. Por ejemplo, no es lo mismo traducir un texto dirigido al gran público general que documentos para uso interno de una empresa pequeña. Para intentar superar esta dificultad, en los últimos años han surgido diversas propuestas de métricas de evaluación de calidad de TA, como las que se utilizan en este trabajo: BLEU y COSTA. Estas métricas proponen una serie de recomendaciones y clasificaciones de errores que sirven de guía en la evaluación de un determinado motor de TA. Los resultados pretenden ayudar a profesionales de la traducción a escoger qué sistema puede ser el más apropiado para su proyecto de traducción y, de esta manera, lograr un resultado de calidad.

Estas métricas surgieron como respuesta al creciente uso de la TA, provocado a su vez por el aumento de la demanda de información a nivel mundial. Hoy en día, cualquier persona puede estar al corriente de lo que sucede en la otra punta del mundo sin necesidad de moverse de casa. Para un traductor humano resultaría imposible traducir un volumen tan grande de contenido en tan poco tiempo, principal motivo por el que nació la TA. No obstante, la calidad de los primeros motores de TA era muy deficiente. Desde hace algunos años, y gracias al rápido avance de las tecnologías, ha sido posible mejorar los sistemas de TA con la implementación de sistemas neuronales, principal aportación de la inteligencia artificial en el mundo de la traducción. En la

actualidad, son muchas las contribuciones de la IA en el mundo de la traducción que han facilitado en gran medida la labor del traductor y, además, han ayudado a que el uso de la TA esté al alcance de todo el mundo.

Un claro ejemplo de ello son las aplicaciones móviles capaces de traducir imágenes, como las de Google y Microsoft, que han sido objeto de estudio en este trabajo. Estas aplicaciones consisten principalmente en enfocar con la cámara o tomar fotografías del texto que se quiere traducir y dejar que el sistema nos proporcione la traducción en unos pocos segundos. El funcionamiento de estas aplicaciones se basa en la combinación de sistemas de reconocimiento óptico de caracteres y sistemas de traducción neuronales, los cuales no podrían existir sin la intervención de la IA. En primer lugar, el ROC reconoce y extrae los caracteres del texto y, a continuación, se traducen con el sistema de TA que tenga incorporado la aplicación, en este caso sistemas de TAN.

La primera aplicación que se analiza en este trabajo es la de Google Translate Images. En concreto, la traducción de imágenes es una de las funciones incorporadas dentro de la aplicación Google Lens, la cual consiste en el reconocimiento por imagen de elementos con los que interactuar, tales como traducir texto, encontrar información sobre un edificio o una calle o realizar compras por internet. La traducción de imágenes funciona en combinación con el traductor automático de Google, el cual utiliza sistemas de TAN desde 2016. El formato de presentación de las traducciones intenta imitar el original, aspecto que podría considerarse tanto una ventaja, si se trata de un texto reducido, como un inconveniente, ya que en el caso de que la imagen contenga varias líneas de texto los formatos podrían mezclarse y crear una traducción confusa para el usuario. Aun así, visualmente las traducciones resultan mucho más claras que las de Microsoft.

La segunda aplicación analizada es la de Microsoft Translator Images. En este caso, la traducción de imágenes está incluida dentro de la aplicación de Microsoft dedicada exclusivamente a la traducción automática. En esta aplicación se pueden encontrar otras modalidades de traducción, como la traducción por voz o de texto normal. De la misma manera que Google, Microsoft también hace uso de la traducción automática neuronal desde 2016. Con respecto al formato de las traducciones, Microsoft intenta prestar más importancia al texto que a la presentación. Por ese motivo, siempre ofrece los resultados de TA en un fondo oscuro con letra en color blanco. Esta metodología podría considerarse una ventaja, ya que de esta forma el usuario lee

únicamente el texto y evita cualquier confusión con otros elementos presentes en la imagen. No obstante, al mismo tiempo esto también supone una desventaja, en especial en textos largos, ya que el texto puede aparecer desordenado y con diferentes tamaños de letra que pueden confundir al usuario, como bien se ha podido mostrar en capítulos anteriores.

Tal y como se muestra en el apartado de COMPARACIÓN DE RESULTADOS, tanto BLEU como COSTA coinciden en que la mejor opción para la traducción de imágenes es Google Translate Images. Asimismo, la diferencia de puntuaciones entre Google y Microsoft obtenida en ambos casos es otra evidencia que justifica este resultado. Las traducciones de Google son más fáciles de entender para el usuario, ya que más de la mitad de los resultados se consideran opciones válidas de TA. Asimismo, la cantidad de errores encontrada está muy por debajo de la mitad y prácticamente no contiene partes sin traducir. Por el contrario, con Microsoft menos de la mitad de los resultados se pueden considerar como opciones válidas de TA y se han detectado más errores de mayor y menor gravedad que impiden la correcta fluidez y adecuación del texto. También contiene muchos segmentos o partes de segmentos sin traducir. No obstante, en el caso de Google se ha observado un uso más frecuente de vocabulario propio el español de América. Con respecto al funcionamiento de las aplicaciones, en ningún caso se han encontrado fallos o complicaciones técnicas importantes. El uso de ambas aplicaciones es sencillo e intuitivo para el usuario. El único aspecto a destacar es que Google puede ofrecer las traducciones de forma más rápida, simplemente con acercar la cámara es suficiente.

Con respecto a las herramientas de evaluación utilizadas, a nivel técnico BLEU no ha presentado grandes dificultades. La única carencia del estudio con respecto a esta herramienta está relacionada con el tamaño del corpus utilizado. Para su correcto funcionamiento, es necesario aportar un corpus de un volumen considerable. Sin embargo, en este estudio solo se ha tenido en cuenta la traducción de referencia de los 56 segmentos, razón que podría explicar en parte los bajos resultados obtenidos en las puntuaciones. Al tratarse de textos tan pequeños se ha considerado que el corpus de referencia era suficiente y no había necesidad de añadir grandes cantidades de texto. La herramienta COSTA sí ha presentado más dificultades técnicas, tal y como se expone en capítulos anteriores. Para su correcto funcionamiento se han tenido que crear archivos de trabajo con características específicas, lo cual ha resultado una tarea laboriosa. Al contrario que BLEU, los resultados ofrecidos se ajustan más a la realidad, ya que la

evaluación se ha llevado a cabo de forma manual por un revisor humano, de ahí la mayor fiabilidad de los resultados.

Después de ver el análisis del funcionamiento y de las traducciones de estas aplicaciones, consideramos que su uso quedaría restringido al público general, ya que se trata de una herramienta que, aunque los resultados que ofrece no son perfectos, sí pueden ayudar a obtener una idea general de lo que el TO pretende transmitir. Al tratarse de imágenes, el texto es muy reducido, por lo que el margen de error también es menor. Es una aplicación práctica y sencilla para el público general que puede resultar de gran ayuda, por ejemplo, al viajar al extranjero o consultar información en otro idioma. Para el traductor profesional, este tipo de herramientas supone una innovación más en el campo de la TA y abre la puerta a futuras investigaciones en esta dirección. No obstante, no parece una herramienta que pueda utilizarse en el ámbito profesional de la traducción, ya que la calidad de los resultados sigue siendo baja y existen otras herramientas más potentes que ofrecen mejores resultados.

Cabe indicar que la funcionalidad de estas aplicaciones ha quedado demostrada únicamente en la combinación inglés-español. Sería necesario realizar investigaciones con otras combinaciones lingüísticas para determinar si el resultado es el mismo, ya que tanto Google como Microsoft siguen utilizando motores estadísticos para lenguas de comunidades lingüísticas muy reducidas que todavía no han incorporado los motores neuronales y, en tal caso, los resultados podrían ser sustancialmente distintos.

8. BIBLIOGRAFÍA

FUENTES CONSULTADAS Y CITADAS

- AENOR (2006). «Norma española: UNE-EN 15038. Servicios de traducción. Requisitos para la prestación del servicio». Recuperado de: <https://www.une.org/encuentra-tu-norma/busca-tu-norma/norma/?c=N0037193> [Consulta: 18 febrero 2020]
- Afli, H., y Way, A. (2016). «Integrating Optical Character Recognition and Machine Translation of Historical Documents». In *Proceedings of the Workshop on Language Technology Resources and Tools for Digital Humanities (LT4DH)* (pp. 109–116). Recuperado de: <https://www.aclweb.org/anthology/W16-4015.pdf> [Consulta: 26 febrero 2020]
- Arlandis, J. (2010). «Sistemas de reconocimiento óptico de caracteres». *Instituto Tecnológico de Informática*. Recuperado de: <https://web.archive.org/web/20100409143206/http://www.iti.es/media/about/docs/tic/13/articulo2.pdf> [Consulta: 26 febrero 2020]
- Babych, B. (2014). «Mètriques d'avaluació automatitzada de TA i les seves limitacions». *Revista Tradumàtica: Tecnologies de La Traducció*. Recuperado de: <https://revistes.uab.cat/tradumatica/article/view/n12-babych/pdf> [Consulta: 23 febrero 2020]
- Barbieri, A. (2019). «Traductores e intérpretes, ¿las próximas víctimas de la inteligencia artificial?». *La Vanguardia*. Recuperado de: <https://www.lavanguardia.com/tecnologia/20190601/462559134761/traductores-interpretes-victimas-inteligencia-artificial-google-translate.html> [Consulta: 2 febrero 2020]
- Casacuberta, F., y Peris, Á. (2017). «Traducción automática neuronal». *Revista Tradumàtica: Tecnologies de La Traducció*. Recuperado de: https://revistes.uab.cat/tradumatica/article/view/n15-casacuberta-peris/pdf_48 [Consulta: 29 febrero 2020]
- Cepero, J. I.; Encuentra, P. et al (2009). «Guía de calidad en la traducción: orientaciones para alcanzar la excelencia en el ejercicio de la profesión de traductor», 2. Recuperado de: http://www.asati.es/img/web/docs/CALIDAD.TRAD_ASATI.2009.pdf [Consulta: 18 febrero 2020]
- Chatzitheodorou, K. y Chatzistamatis, S. (2013). «COSTA MT Evaluation Tool: An Open Toolkit for Human Machine Translation Evaluation». *The Prague Bulletin of Mathematical Linguistics*. Recuperado de: <https://doi.org/10.2478/pralin-2013-0014> [Consulta: 29 febrero 2020]
- COSTA MT Evaluation Tool (2020). «Costa MT Evaluation Tool – Download». Recuperado de: <https://code.google.com/archive/p/costa-mt-evaluation-tool/> [Consulta: 7 marzo 2020]
- Google (2020). «Google Lens.» Recuperado de: <https://lens.google.com/> [Consulta: 29 febrero 2020]

- Görög, A. (2014). «Quantificació i evaluació comparativa de la qualitat: el Dynamic Quality Framework de TAUS». *Revista Tradumàtica: Tecnologies de La Traducció*. Recuperado de: <https://revistes.uab.cat/tradumatica/article/view/n12-gorog2/pdf> [Consulta: 23 febrero 2020]
- Khandelwal, R. (2020). «BLEU — Bilingual Evaluation Understudy». *Towards data science*. Recuperado de: <https://towardsdatascience.com/bleu-bilingual-evaluation-understudy-2b4eab9bcfd1> [Consulta: 5 febrero 2020]
- Linguaserve (2019). «Traducción automática, inteligencia artificial e I+D+I». Recuperado de: <https://blog.linguaserve.com/traducion-automatica-inteligencia-artificial-i-d-i> [Consulta: 2 febrero 2020]
- López, A. (2018). «Traducción automática neuronal y traducción automática estadística: percepción y productividad». *Trabajo de fin de máster*. Recuperado de: https://ddd.uab.cat/pub/trerecpro/2018/hdl_2072_348878/LopezPereira_Ariana_TFM_Tradumatica_2018.pdf [Consulta: 29 febrero 2020]
- Microsoft (2020). «Machine Translation - Microsoft Translator for Business». Recuperado de: <https://www.microsoft.com/en-us/translator/business/machine-translation/#whatmachine> [Consulta: 1 marzo 2020]
- Microsoft Azure (2020). «Cognitive Services—APIs for AI Developers». Recuperado de: <https://azure.microsoft.com/en-us/services/cognitive-services/> [Consulta: 1 marzo 2020]
- Microsoft Azure. (2019). «Printed, handwritten text recognition - Computer Vision». Recuperado de: <https://docs.microsoft.com/en-us/azure/cognitive-services/computer-vision/concept-recognizing-text> [Consulta: 1 marzo 2020]
- Pan, H. M. (2016). «How BLEU Measures Translation and Why It Matters». Recuperado de: <https://slator.com/technology/how-bleu-measures-translation-and-why-it-matters/> [Consulta: 1 febrero 2020]
- Papineni, K., Roukos, S. et al (2002). «BLEU: a Method for Automatic Evaluation of Machine Translation». *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)* (pp. 311–318). Philadelphia: Association for Computational Linguistics. Recuperado de: <https://www.aclweb.org/anthology/P02-1040.pdf> [Consulta: 1 febrero 2020]
- Tilde (2020). «Interactive BLEU score evaluator». Recuperado de: <https://www.letsmt.eu/Bleu.aspx> [Consulta: 1 febrero 2020]
- Tomás, J., Mas, J. À. et al (2003). «A Quantitative Method for Machine Translation Evaluation». Recuperado de: <https://www.aclweb.org/anthology/W03-2804/> [Consulta: 23 febrero 2020]
- Vashee, K. (2019). «Understanding MT Quality: BLEU Scores». *SDL Blog*. Recuperado de: <https://www.sdl.com/blog/understanding-mt-quality-bleu-scores.html> [Consulta: 1 febrero 2020]
- WIKIPEDIA (2020). *Microsoft Translator*. Recuperado de: https://en.wikipedia.org/wiki/Microsoft_Translator [Consulta: 1 marzo 2020]
- Yúbal FM (2019). «Google Lens: qué es, cómo instalarlo y todo lo que puedes hacer con ella». *Xataka*. Recuperado de: <https://www.xataka.com/basics/google-lens-que-como-instalarlo-todo-que-puedes-hacer-ella> [Consulta: 29 febrero 2020]

FUENTES CONSULTADAS Y NO CITADAS

Google (2017). «Google I/O 2017». Recuperado de: <https://events.google.com/io2017/> [Consulta: 29 febrero 2020]

Microsoft (2016). «Microsoft Translator Adds Image Translation to Android».

Recuperado de: <https://www.microsoft.com/en-us/translator/blog/2016/04/20/microsoft-translator-adds-image-translation-to-android/> [Consulta: 1 marzo 2020]

Penalva, J. (2017). «Google Lens: reconocer lo que tenemos delante será cuestión de apuntar con la cámara de nuestro smarpthon». *Xataka*. Recuperado de: <https://www.xataka.com/moviles/google-lens-reconocer-lo-que-tenemos-delante-sera-cuestion-de-apuntar-con-la-camara-de-nuestro-smarphone> [Consulta: 29 febrero 2020]

WIKIPEDIA (2019). *Google Lens*. Recuperado de:

https://es.wikipedia.org/wiki/Google_Lens [Consulta: 29 febrero 2020]

9. ANEXOS

9.1. Archivos utilizados para la evaluación (adjuntos en el documento)

- EN_TextoOriginal
- ES_TAGoogle
- ES_TAMicrosoft
- ES_TraducciónHumana
- ResultadosTA_Google&Microsoft

9.2. Documentos obtenidos con las herramientas de evaluación (adjuntos en el documento)

- Informe de evaluación de Google con BLEU
- Informe de evaluación de Google con COSTA MT Evaluation Tool
- Informe de evaluación de Microsoft con BLEU
- Informe de evaluación de Microsoft con COSTA MT Evaluation Tool
- Informe de evaluación de Google y Microsoft con BLEU

9.3. Imágenes originales y procesadas con Google Lens y Microsoft

Imagen 1: Original, Google y Microsoft

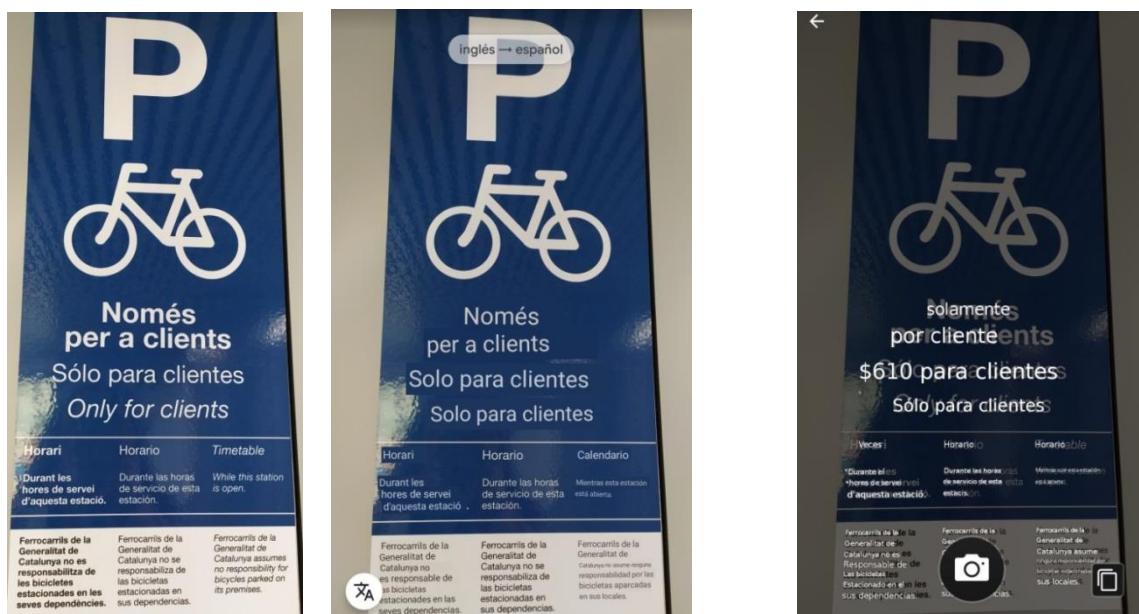


Imagen 2: original, Google y Microsoft



Imagen 3: original, Google y Microsoft



Imagen 4: original, Google y Microsoft

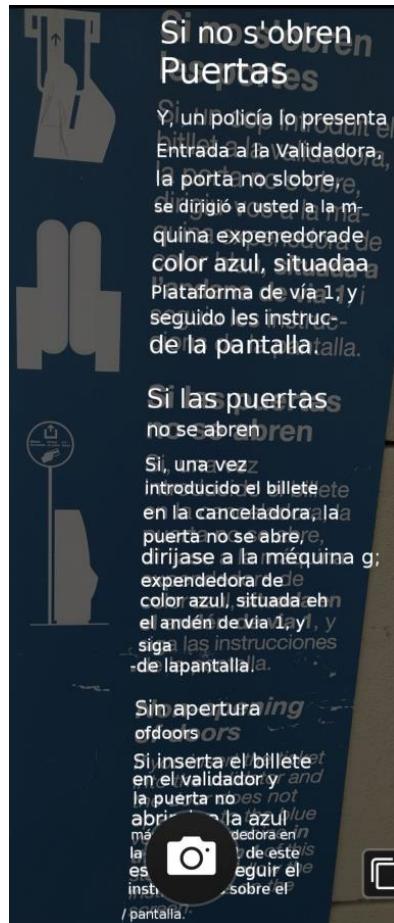
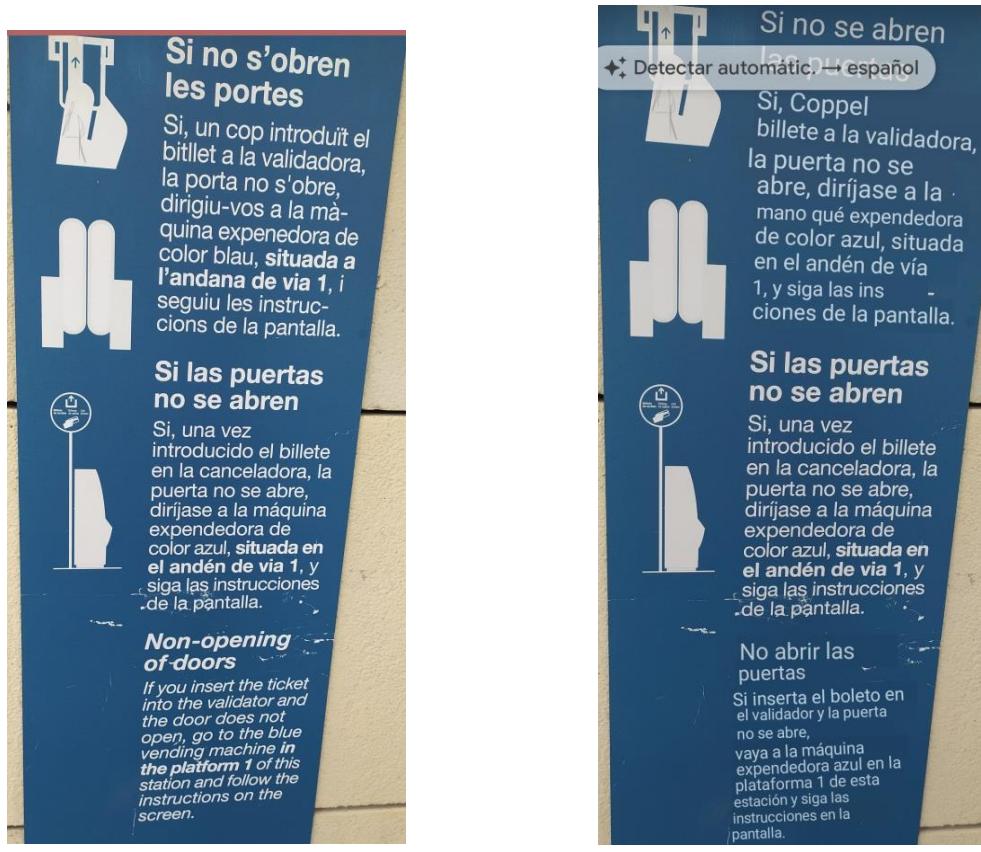


Imagen 5: original, Google y Microsoft

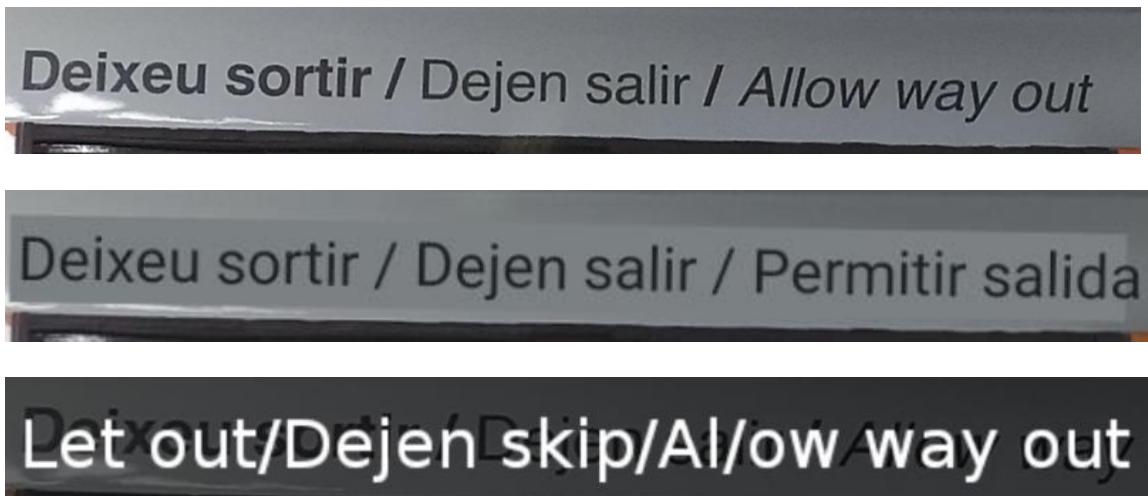


Imagen 6: original, Google y Microsoft

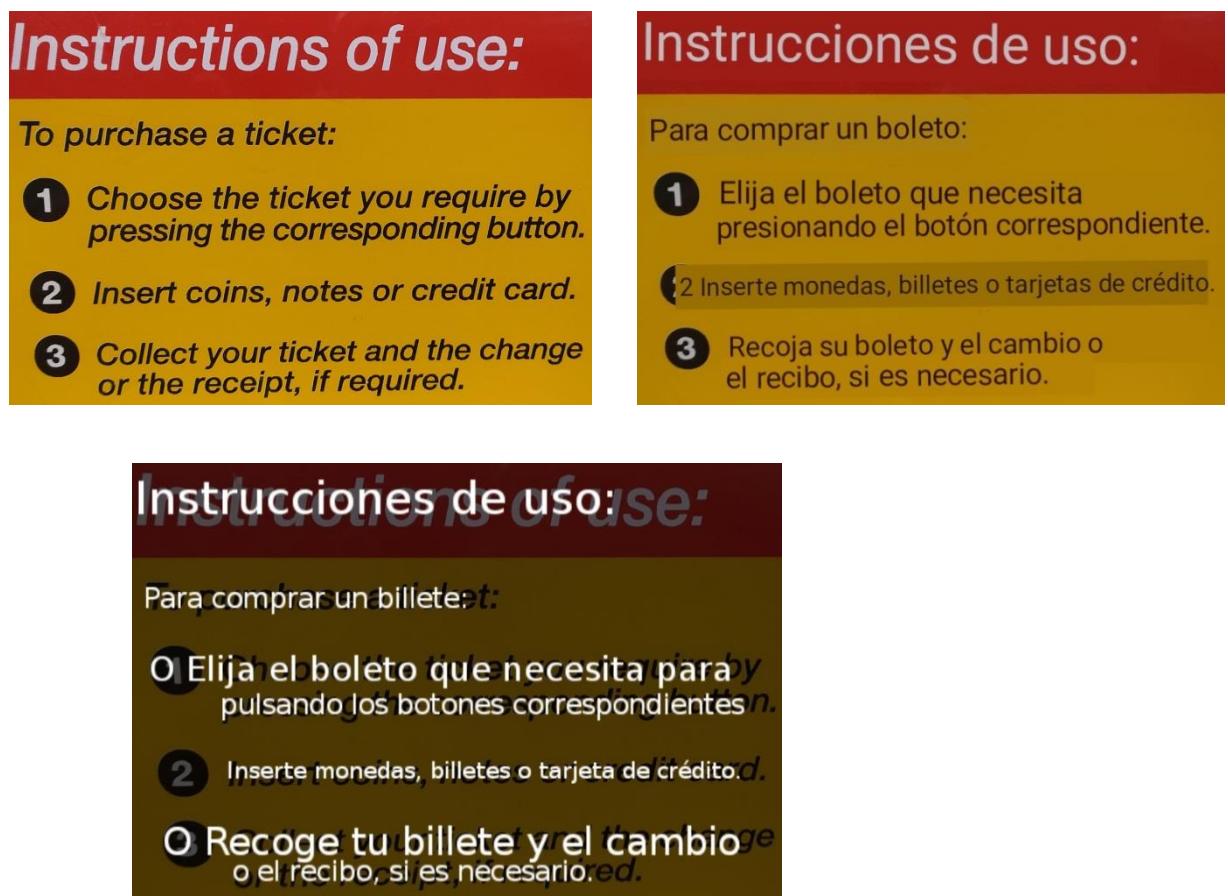


Imagen 7: original, Google y Microsoft



Imagen 8: original, Google y Microsoft

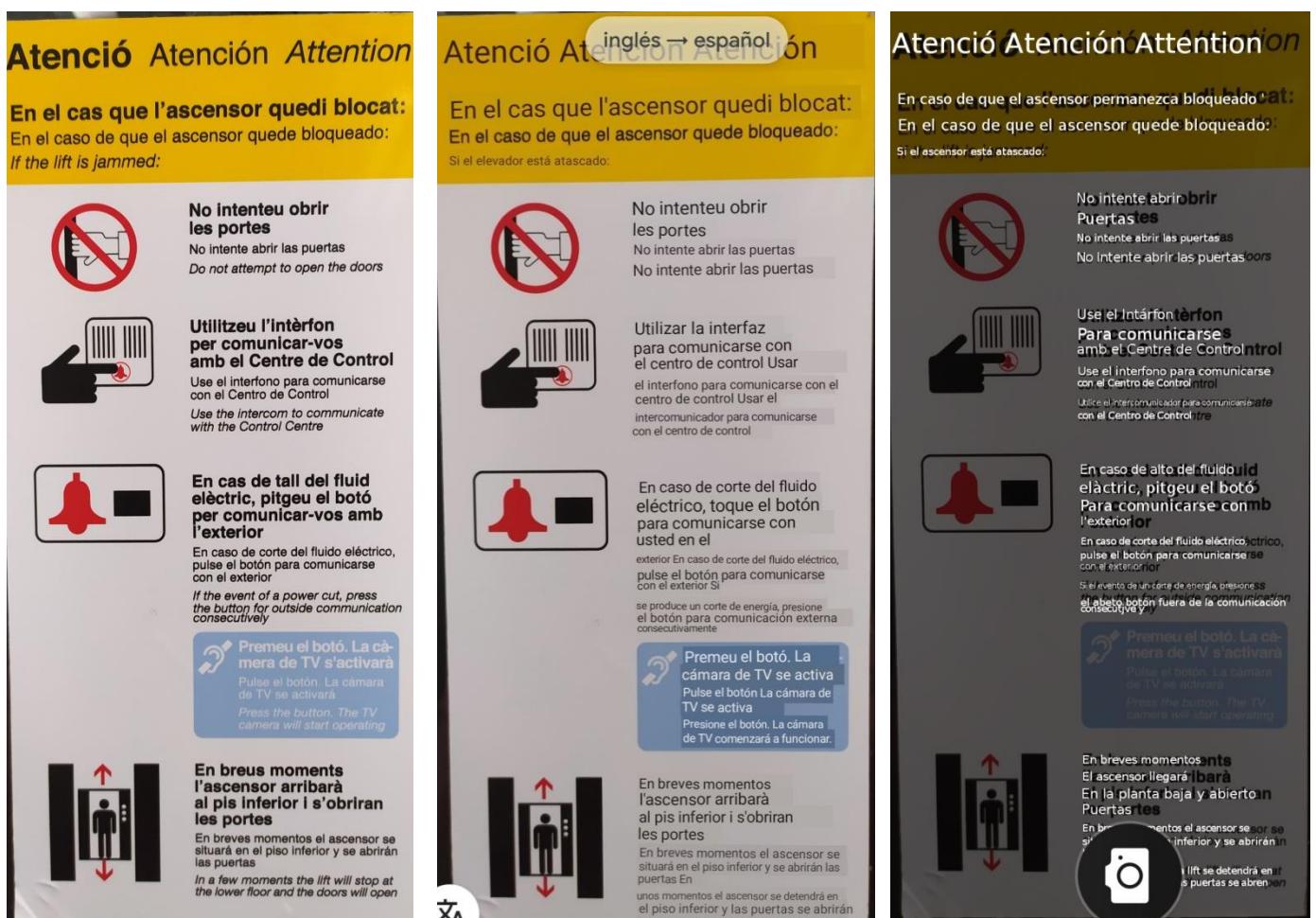


Imagen 9: original, Google y Microsoft

Bitllets integrats ATM Àrea de Barcelona Autoritat del Transport Metropolità TMB Bus FGC Tram Rodalies

Billetes integrados / Integrated tickets

T-casual	1 zona	2 zones	3 zones	4 zones	5 zones	6 zones
Preu / Precio	11,35	22,40	30,50	39,20	45,05	47,90
Temps de validesa / Tiempo de validez / Validity period	1h 15'	1h 30'	1h 45'	2h 00'	2h 15'	2h 30'

Títol unipersonal i horari, de 10 viatges integrats en tots els modes de transport segons les zones per les quals es transiti (pago màxim 6 zones). No és vàlid a les estacions de metro d'Aeroport T1 i Aeroport T2 de la L9 Sud.

Título unipersonal y horario, de 10 viajes integrados en todos los modos de transporte, según las zonas por las que se transite (pago máximo 6 zonas). No es válido en las estaciones de metro de Aeroport T1 y Aeroport T2 de la linea L9 Sud.

A single person timed ticket for 10 journeys on any means of transport within the integrated system, for the number of zones purchased (maximum charge 6 zones). This ticket is not valid for the Aeroport T1 and Aeroport T2 metro station stops on line L9 Sud.

Bitllets Gratis Cajero automático Área de Barcelona Autoritat del Transport Metropolità TMB Bus FGC Tram Rodalies

Billetes integrados / Tickets integrados

T-casual 1 zona	2 zones	3 zones	4 zones	5 zones	6 zones	
Preu / Precio	11,35	22,40	30,50	39,20	45,05	47,90
Temps de validesa / Periodo de validez	1h 15' 1h 30 1h 45' 2h 00' 2h 15'					2h 30'

Títol unipersonal i horario, de 10 viatges integra en tots els modos de transporte segons les zones per les quals es transit (pago màxim 6 zones). No és vàlid a les estacions de metro d'Aeroport T1 i Aeroport T2 de la L9 Sud.

Título unipersonal y horario, de 10 viajes integrados en todos los modos de transporte, según las zonas por las que se transite (pago máximo 6 zonas). No es válido en las estaciones de metro de Aeroport T1 y Aeroport T2 de la linea L9 Sud.

.cronometrado para una sola persona para 10 viajes en cualquier zona de medios comprados (carga máxima 6 zonas). Este boleto no es válido para las estaciones de metro Aeroport T1 y Aeroport T2 se detiene en la linea L9 Sud.

T-usual	1 zona	2 zones	3 zones	4 zones	5 zones	6 zones
Preu / Precio	40,00	53,85	75,60	92,55	106,20	113,75

Títol personal i intransferible, amb un nombre il·limitat de viatges integrats en 30 dies consecutius, a zones delimitades per la primera validació, en tots els modos de transport segons les zones per les

Notas "Tegrats rats mm de Barcelona Autoritat del Transport Metropolità TMB Bus FGC Tram Rodalies

Billetes integrados / Integrated tickets

T-c zona usual	1 zona	2 zones	3 zones	4 zones	5 zones	6 zones
Preu / Precio	11,35	22,40	30,50	39,20	45,05	47,90
Período de valideza / Periodo de validez	1h 15'15'	1h 30'30'	1h 45'45'	2h 00'	2h 15'15'	2h 30'

Título y calendario unipersonal, de 10 viajes integrados en todos los modos de transporte según las zonas (Pago máximo 6 zonas). No es válido en las estaciones de metro d'Aeroport T1 y Aeroport T2 de la L9 Sud.

Título unipersonal y horario, de 10 viajes integrados en todos los modos de transporte, según las zonas por las que se transite (pago máximo 6 zonas). No es válido en las estaciones de metro de Aeroport T1 y Aeroport T2 de la linea L9 Sud.

Un billete cronometrado para una sola persona para 10 viajes por cualquier medio de transporte dentro del sistema integrado, por el número de zonas compradas (carga máxima 6 zonas). Este billete es válido para la estación de metro Aeroport T1 y Aeroport T2 paradas en la linea L9 Sud.

T-usual	1 zona	2 zones	3 zones	4 zones	5 zones	6 zones
Preu / Precio	40,00	53,85	75,60	92,55	106,20	113,75

Título personal e intransferible, con un número ilimitado de viajes integrados en 30 días consecutivos, en las zonas delimitadas por la primera validación, en todos los modos de transporte según las áreas, por las

Imagen 10: original, Google y Microsoft

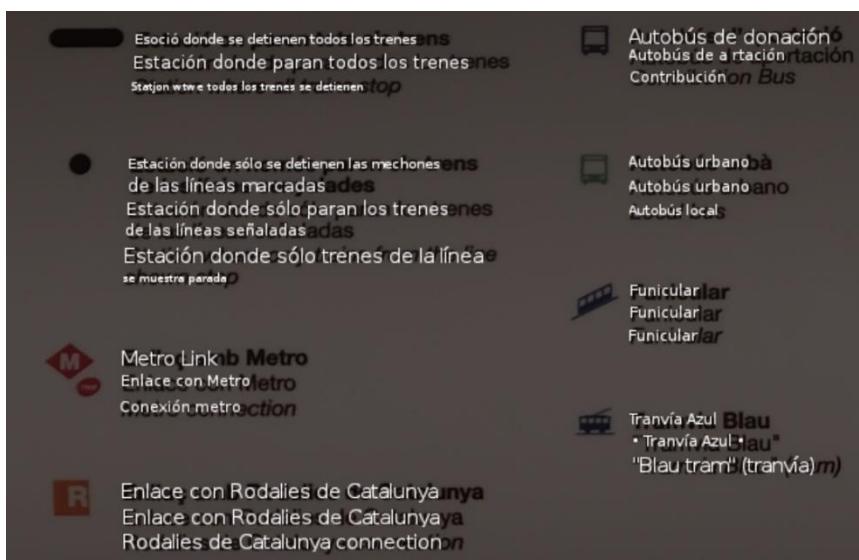
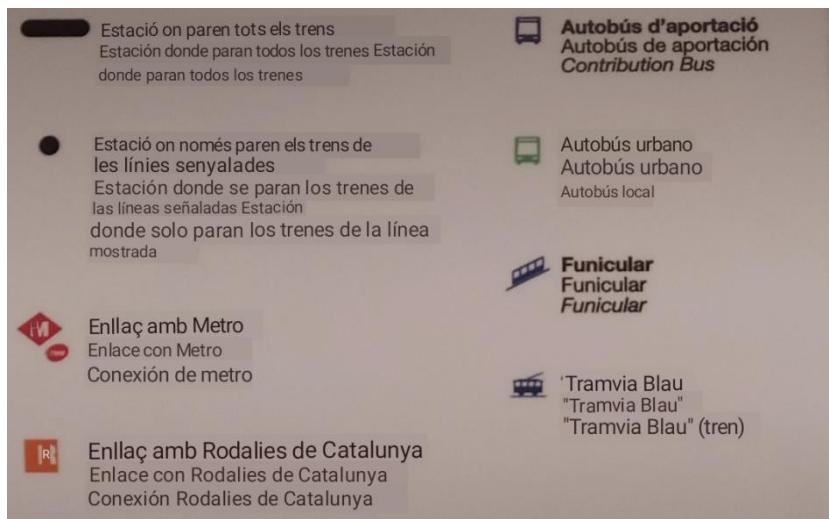
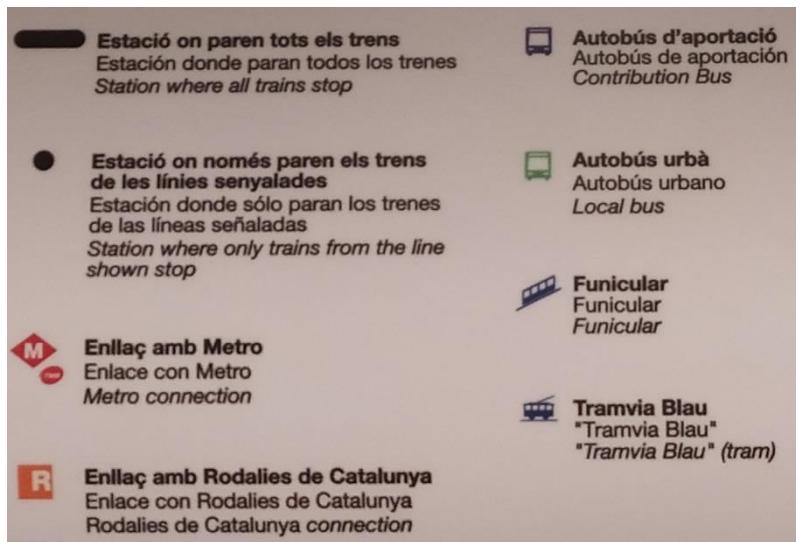


Imagen 11: original, Google y Microsoft

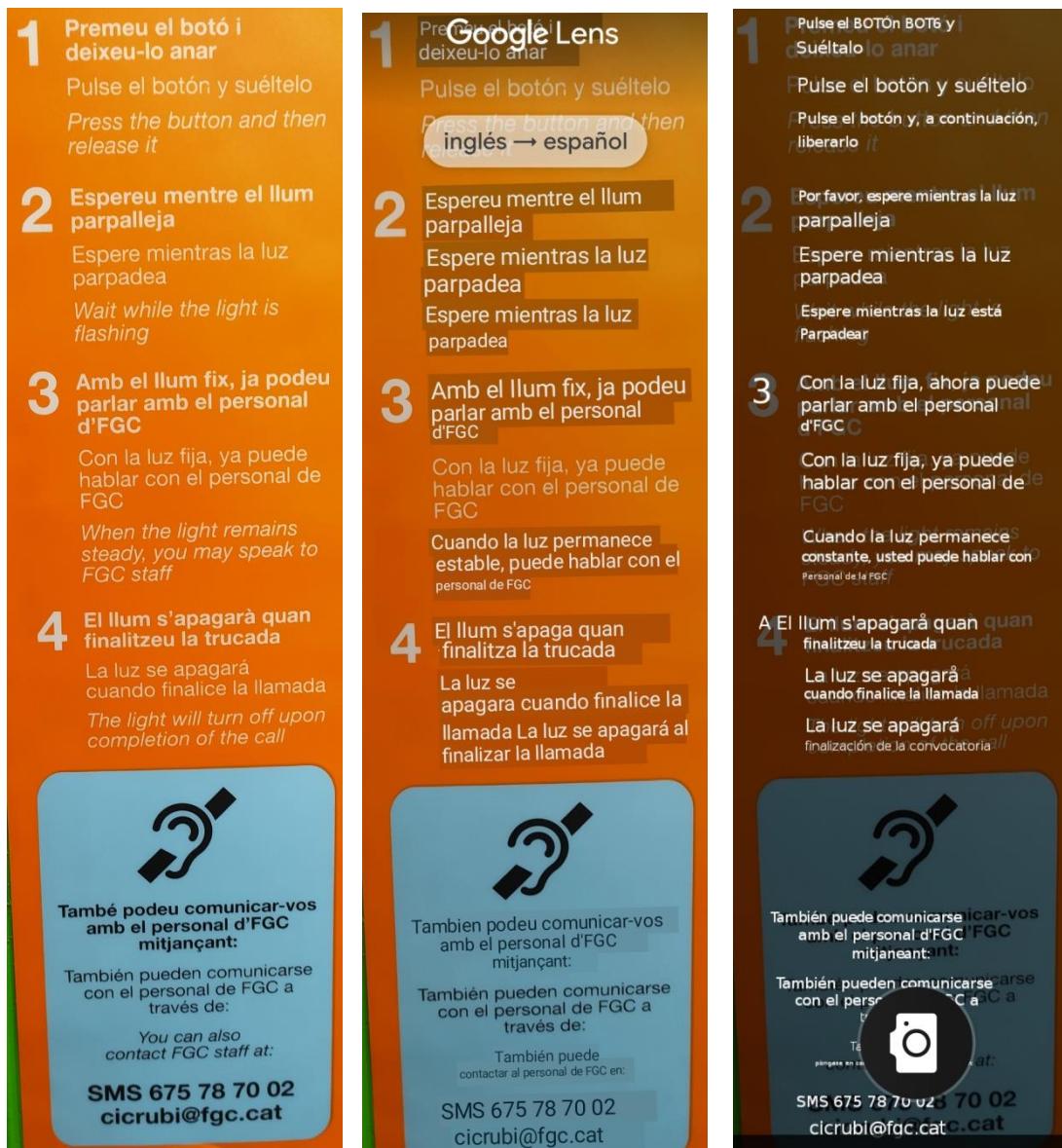


Imagen 12: original, Google y Microsoft

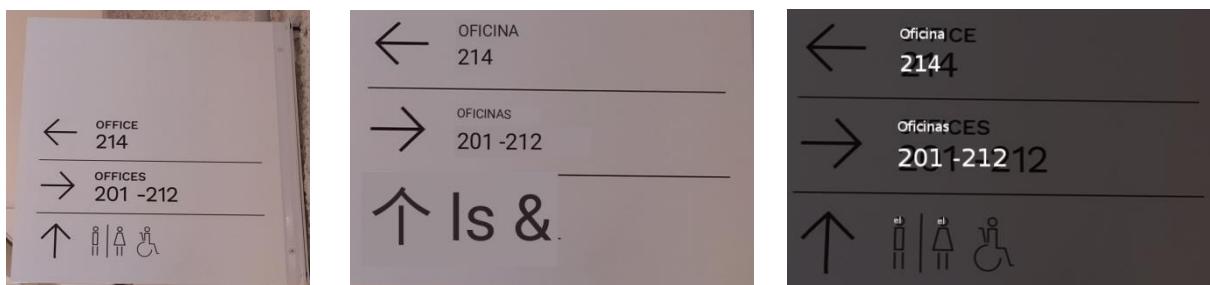


Imagen 13: original, Google y Microsoft

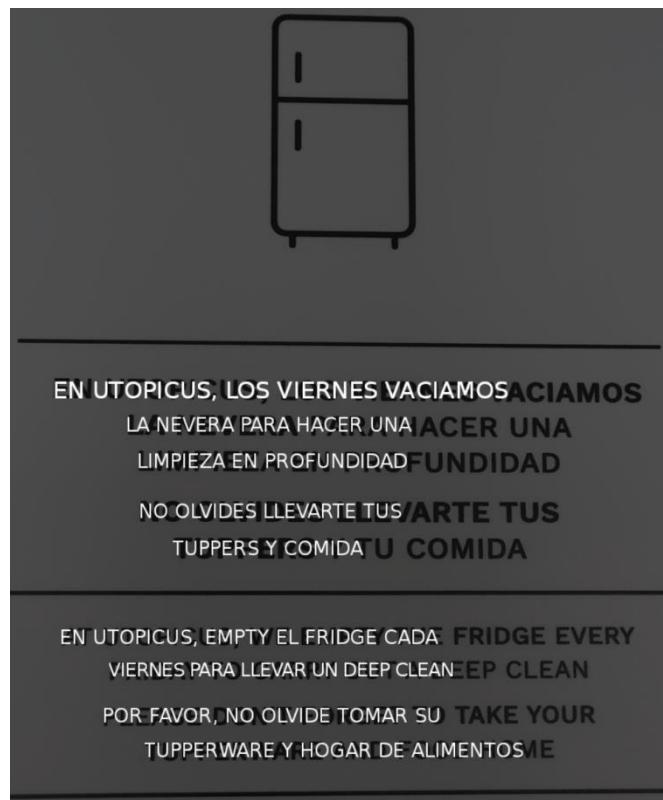
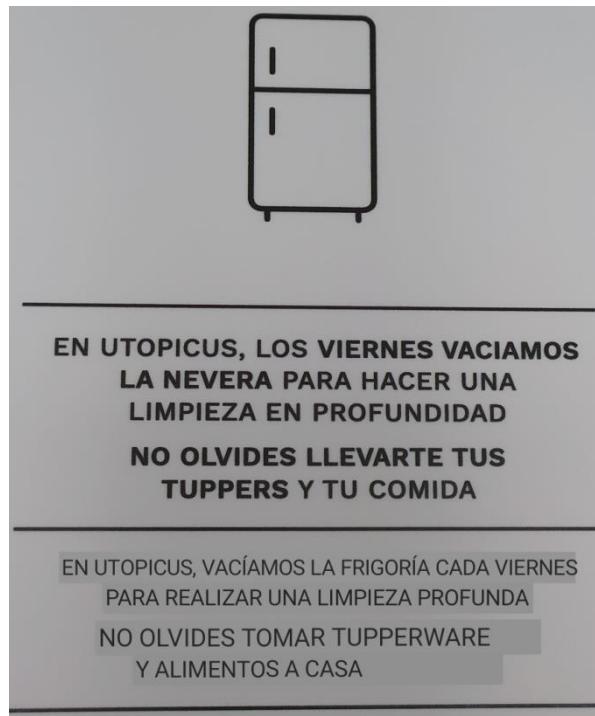
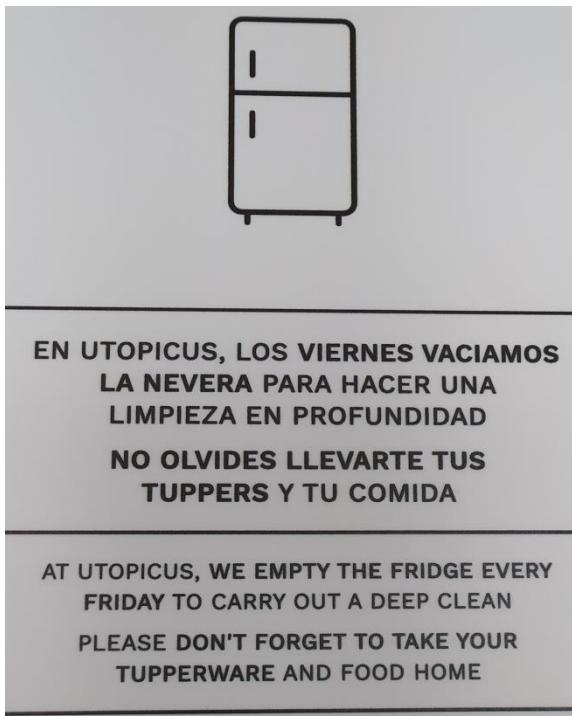


Imagen 14: original, Google y Microsoft



Imagen 15: original, Google y Microsoft

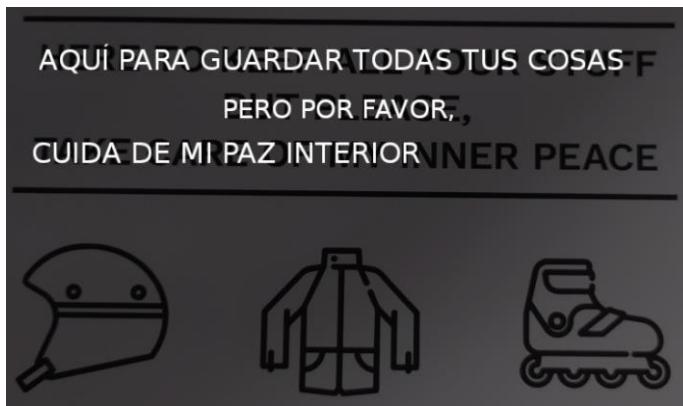


Imagen 16: original, Google y Microsoft

The image displays three versions of a security notice, likely for a company named Fix. Each version features a large exclamation mark icon at the top.

Version 1 (Original):

FIX COWORKERS AND PRIVATE OFFICES

WHAT SHOULD I DO IF I AM STAYING DURING THIS TIME FRAME?

Notify the security company over the intercom:

1) Please identify yourself with your full name and ID.
2) Notify until what time you will stay in the building to enable them to re-activate the alarm after your departure.

Thank you

* YOU CAN FIND FURTHER INFORMATION IN THE WELCOME TUTORIAL

Version 2 (Google Translate - English to Spanish):

ARREGLAR A TRABAJADORES Y OFICINAS PRIVADAS

¿QUÉ DEBO HACER SI ME QUEDO DURANTE ESTE MARCO DE TIEMPO?

Notifique a la compañía de seguridad por el intercomunicador:

1) Identifíquese con su nombre completo e identificación.
2) Notificar hasta qué hora permanecerá en el edificio para permitirles reactivar la alarma después de su partida.

Gracias

* PUEDES ENCONTRAR MÁS INFORMACIÓN EN EL TUTORIAL DE BIENVENIDA

Version 3 (Microsoft Translate - English to Spanish):

FIX COMPAÑEROS DE TRABAJO Y OFICINAS PRIVADAS

LO QUE DEBO HACER SI ESTOY QUEDÁNDOME DURANTE ESTE TIEMPO?

Notificar a la compañía de seguridad a través del intercomunicador:

1) Por favor, identifíquese con su nombre completo y su identificación ID.
2) Notificar hasta qué hora permanecerá en el edificio para permitirles reactivar la alarma después de su Salida.

Gracias

* YOU CAN FIND FURTHER INFORMATION IN THE WELCOME TUTORIAL

