



Ciencia Latina
Internacional

Ciencia Latina Revista Científica Multidisciplinar, Ciudad de México, México.
ISSN 2707-2207 / ISSN 2707-2215 (en línea), marzo-abril 2024,
Volumen 8, Número 2.

https://doi.org/10.37811/cl_rcm.v8i2

SUBTITULADO PARA LA VISUALIZACIÓN DE VIDEOS SIN SONIDO: UNA NUEVA FORMA DE CONSUMIR CONTENIDO

**SUBTITLING FOR WATCHING VIDEOS WITHOUT SOUND:
A NEW WAY OF CONSUMING CONTENT**

Milagro de Guadalupe Chablé Bautista
Universidad Juárez Autónoma de Tabasco, México

Eleazar Morales Vázquez
Universidad Juárez Autónoma de Tabasco, México

Mtro. Julio César Arias Ovando
Universidad Juárez Autónoma de Tabasco, México

Maritza Aguilera Ramírez
Universidad Juárez Autónoma de Tabasco, México

DOI: https://doi.org/10.37811/cl_rcm.v8i2.10576

Subtitulado para la Visualización de Videos sin Sonido: Una Nueva Forma de Consumir Contenido

Milagro de Guadalupe Chablé Bautista¹

milagrodeguadalupechablebautis@gmail.com

<https://orcid.org/0009-0000-7795-3692>

Universidad Juárez Autónoma de Tabasco
Villahermosa – México

Eleazar Morales Vázquez

eleazarmove@gmail.com

<https://orcid.org/0000-0003-1596-5043>

Universidad Juárez Autónoma de Tabasco
Villahermosa – México

Mtro. Julio César Arias Ovando

jcarias30@hotmail.com

<https://orcid.org/0000-0001-5101-9391>

Universidad Juárez Autónoma de Tabasco
Villahermosa – México

Maritza Aguilera Ramírez

maritzaguilera75@gmail.com

<https://orcid.org/0000-0002-8161-667X>

Universidad Juárez Autónoma de Tabasco
Villahermosa – México

RESUMEN

¿Por qué los usuarios de esta nueva realidad para el mundo eligen la visualización de videos sin sonido? Este cuestionamiento es el que da pauta para analizar las funciones de la Traducción Audiovisual (TAV) y redescubrir su objetivo como modalidad de traducción. Actualmente, la reproducción automática de subtítulos en plataformas como YouTube se ha vuelto un recurso de suma relevancia para los usuarios, especialmente para aquellos nacidos durante la era digital. Esta nueva manera de consumir productos audiovisuales ha revolucionado la industria de la comunicación, del entretenimiento e incluso de la mercadotecnia, por lo cual resulta interesante reflexionar sobre los sistemas de subtitulación automática, considerando los estándares de calidad que busca la TAV y las percepciones de los usuarios. Por ello, en este artículo se presenta una revisión de literatura en la que se buscó conocer resultados acerca de la efectividad y la calidad de la generación automática de subtítulos en la plataforma YouTube, con el fin de determinar posibles campos de estudio.

Palabras clave: plataformas digitales, productos audiovisuales, traducción audiovisual

¹ Autor principal

Correspondencia: milagrodeguadalupechablebautis@gmail.com

Subtitling for Watching Videos Without Sound: A New Way of Consuming Content

ABSTRACT

Why do users of this new reality for the world choose to watch videos without sound? This questioning is what sets the tone for analyzing the functions of Audiovisual Translation (AVT) and rediscovering its objective as a translation modality. Currently, the automatic playback of subtitles on platforms such as YouTube has become an extremely relevant resource for users, especially for those born during the digital age. This new way of consuming audiovisual products has revolutionized the communication, entertainment and even marketing industries, which is why it is interesting to reflect on automatic subtitling systems, considering the quality standards sought by TAV and the perceptions of users. Therefore, this article presents a literature review in which we sought to know results about the effectiveness and quality of the automatic generation of subtitles on the YouTube platform, in order to determine possible fields of study.

Keywords: digital platforms, audiovisual products, audiovisual translation

Artículo recibido 20 febrero 2024

Aceptado para publicación: 25 marzo 2024



INTRODUCCIÓN

La Traducción Audiovisual (TAV) tiene sus orígenes en el cine mudo, en donde el objetivo era representar textualmente lo que sucedía en la pantalla, la cual mostraba un contenido visual, pero sin sonido. Sin embargo, no siempre se le denominó de esta manera, ya que ha sido llamada de diferentes formas, como Screen translation (Mason, 1989) o inclusive Traducción cinematográfica (Hurtado, 1994).

Posteriormente, el término TAV pasó a denominar la traducción que de forma intra o interlingüística traduce textos audiovisuales. Esto es, textos que se transmiten a través del canal acústico y del canal visual, en donde entran en juego códigos de significación como imágenes, voz, efectos especiales, carteles, entre otros (Chaume, 2004).

Actualmente, la función del subtítulo ha vuelto a sus orígenes, siendo empleado para la visualización de videos sin activar el sonido. Asimismo, esto ha revolucionado el estudio del subtítulo profesional y de la definición misma de la TAV.

El subtítulo como opción para ver videos sin sonido ha emergido como una solución fundamental para atender diversas necesidades en el mundo contemporáneo debido a la creciente influencia de la tecnología y la digitalización en la vida diaria. En la era actual, donde el acceso a contenido audiovisual y la diversidad de plataformas de consumo de medios es vasta, el subtítulo se ha vuelto esencial para adaptarse a las diversas preferencias y circunstancias del espectador.

A lo largo de los años, la tendencia hacia la preferencia de videos sin sonido ha influido notablemente en las plataformas digitales. Tanto las principales redes sociales como los navegadores web han adaptado sus funciones para no reproducir automáticamente el audio de los videos.

En este sentido, navegadores populares como Google Chrome o Firefox han implementado cambios significativos. Ahora, estos navegadores no activan el sonido de ningún video de manera automática, permitiendo al usuario tener el control total sobre su experiencia de navegación (Nuteco, s. f.).

En las redes sociales es en donde se ve más este fenómeno, puesto que los videos, ya sean publicitarios, informativos o de entretenimiento, destacan más que los textos en cualquier forma (Gómez, 2017). Al respecto, Patel (17 de mayo de 2016) indicó que Facebook registró un total de 8 000 millones de reproducciones diarias, y que de estas el 85 % eran sin sonido.

Pero ¿por qué los usuarios eligen la visualización sin sonido? Báez (23 de abril de 2018) menciona que una de las razones principales es que no interrumpe las actividades que el usuario está haciendo, y que el hecho de que tenga subtítulos ayuda a que no se pierda la finalidad del video.

Este comportamiento revela una preferencia creciente por el contenido visual que se puede consumir de manera discreta y conveniente, lo que no solo ayuda al espectador, sino a quien hace el video a lograr un mayor alcance y accesibilidad. Así, pueden alcanzar a personas que hablan el mismo idioma del video, pero que pertenecen a diferentes regiones o países, a personas que hablan otros idiomas o a las personas con discapacidad auditiva.

Por otro lado, los subtítulos son ideales en situaciones donde el usuario no puede escuchar el audio, ya sea porque se encuentra en un lugar público en donde el ruido no es tolerado o porque no tiene auriculares, dado que su uso también se ha vuelto indispensable para muchos.

En estos casos, los subtítulos proporcionan una alternativa efectiva para comprender el contenido del video sin necesidad de audio, mejorando así la accesibilidad y la experiencia del espectador (Hotmart, 21 de octubre de 2022). De esta manera, el video se ha adaptado al estilo de vida contemporáneo, donde el ritmo de trabajo es acelerado y el entretenimiento debe ser, de preferencia, rápido y en cualquier sitio. Otra ventaja de este tipo de subtítulo es que brinda información sobre el contenido de un video con el fin de motivar al espectador para que lo reproduzca plenamente con sonido. Esto también forma parte de la nueva forma de atraer a los usuarios, ya que en muchos casos los primeros instantes de un video contienen un breve resumen, por lo que el espectador puede interesarse a profundidad o simplemente descartar el producto audiovisual.

Según Gómez (2017), la creación del *autoplay*, que básicamente es la reproducción automática de un video, pero sin activar el sonido (para dejar esa decisión al usuario), es una estrategia de marketing que ofrece un contenido que es mucho menos intrusivo, teniendo en mente que la saturación de contenido en internet es evidente y que actualmente hay demasiadas opciones en la red que sobrecargan de información al espectador.

Por lo anterior, Gómez (2017) argumenta que la transformación de los contenidos y productos, de los procesos de creación, de los canales de distribución, de los formatos y de las formas de explotar lo audiovisual han sido a partir de la innovación tecnológica que ha generado un escenario en el que las

empresas se han visto orilladas a buscar nuevas formas de relación con el público, lo que también ha llevado a la segmentación de audiencias y a la individualización del consumo.

En otras palabras, el contenido actual debe pensarse para satisfacer las necesidades de determinados sectores cuyas preferencias son muy específicas. Y estas preferencias se encuentran a su vez condicionadas por todo lo que acontece en el mundo, lo que genera gratificación, entretenimiento o simplemente utilidad.

En este sentido, de acuerdo con lo señalado por Navimedia (s. f.), la población mundial prefiere acceder a la información de manera audiovisual dado que esto les permite retener el 95% de un mensaje, además de que es un medio de comunicación generalmente accesible y rápido.

En añadidura, no debe olvidarse que las generaciones que mayor uso hacen de las plataformas y redes sociales han crecido durante la evolución de la digitalización, lo cual implica que han sido predispuestos para procesar la información transmitida por medios audiovisuales por encima del texto.

Para complementar, Ribera (2016) comentó que, para estas generaciones, lo gráfico fue desplazado por lo audiovisual, ya que crecieron aprendiendo a decodificar el lenguaje audiovisual de las películas, videojuegos, anuncios o series de televisión, lo que hace que sea más sencillo para ellos ver un video que leer un artículo o un post, a su vez que generan una conexión emocional. De este modo, puede decirse que la generación de nativos digitales ha dictado la pauta para la transformación de los medios de comunicación.

Para enfatizar lo anterior, basta con analizar algunas estimaciones. En el estudio de Hernández (2016), se consideró que, de acuerdo con el crecimiento desde el 2016, tan solo en España el tráfico IP sería dominado por el video en el 2021, representando el 82 %. Así, se dedujo para dicho año un aproximado de 1900 millones de usuarios de video, siendo así el consumo de 3 billones de minutos de video por mes.

Todo esto genera una necesidad de analizar el papel del subtítulo en los productos de video, especialmente ahora que forman parte de una nueva forma de consumir contenido en internet, puesto que no solamente es una manera de alcanzar espectadores que no hablan el idioma en el que se produce el audio de un producto (haciendo uso del subtítulo interlingüístico); pero, tampoco es solo una vía para brindar accesibilidad a personas con discapacidad auditiva (haciendo uso del subtítulo

intralingüístico). Ahora, el subtítulado, en cualquiera de ambas formas, es una opción para cambiar la experiencia de consumo audiovisual del usuario que vive en una era donde el silencio también es un producto.

Así, en este artículo se hace una revisión de diversos estudios que abordan el subtítulado en una de las plataformas de consumo de video más importantes: YouTube. Se hace énfasis en el subtítulado automático que permite el consumo de contenido sin sonido, subrayando las ventajas y desventajas que se han identificado para este recurso.

METODOLOGÍA

El enfoque de esta revisión de literatura se determinó sobre la plataforma YouTube, principalmente, porque es una de las que emplean el subtítulado automático sin sonido; pero, también porque está considerada como el segundo sitio web con mayor tráfico a nivel mundial (Vizcaíno-Verdú et al., 2019), lo cual permite sugerir que su influencia en la transformación de los medios de comunicación es relevante.

Según la información del sitio Platzi (s. f.), YouTube combina la tecnología de reconocimiento automático de voz de Google con su sistema de subtitulación automática. Este sistema se basa en un algoritmo que reconoce las palabras en el video y puede convertirlas a texto, empleando redes neuronales profundas y mezclas gaussianas. Además, su sistema de inteligencia artificial se enriquece constantemente con las transcripciones manuales que son subidas a las plataformas y con las correcciones que se hacen sobre los subtítulos, por lo que puede considerarse que nunca deja de aprender.

De acuerdo con Google Support (s. f.), las redes neuronales profundas que se emplean mejoran con el tiempo al tener mayores cantidades de datos, lo cual se denomina entrenamiento de modelos. Dicho entrenamiento se divide en tres clases:

- 1) Aprendizaje convencional: Google recoge y almacena fragmentos reales de audio en sus servidores para que revisores humanos anoten parte de estos y un algoritmo de entrenamiento aprenda a partir de muestras de datos de audio anotadas.
- 2) Aprendizaje federado: es un entrenamiento de modelos de inteligencia artificial que puede aprender directamente desde el dispositivo móvil sin recoger esta información para sus servidores, por lo que

se protege la privacidad de los usuarios. Esto permite crear un nuevo modelo de voz combinando los diferentes aprendizajes adquiridos en distintos dispositivos.

- 3) Aprendizaje efímero: esta técnica, al igual que la anterior, protege la privacidad de los usuarios cuando el modelo de voz se ejecuta en servidores de Google mediante su almacenamiento en una memoria a corto plazo, es decir, una memoria RAM, con la que el algoritmo puede aprender de tales muestras de audio en tiempo real y posteriormente se eliminan sin ser almacenadas ni compartidas.

De esta manera, en esta revisión de literatura se analizaron estudios acerca del subtítulo automático en YouTube, en donde se buscó conocer los hallazgos sobre las características del subtítulo automático que ofrece esta plataforma y la calidad con la que cuenta, teniendo en mente la tecnología con la que trabaja y así también la funcionalidad desde la perspectiva de los usuarios.

RESULTADOS Y DISCUSIÓN

Tatman y Kasten (2017), en su trabajo de investigación *Effects of Talker Dialect, Gender & Race on Accuracy of Bing Speech and Youtube Automatic Captions*, analizaron la tasa de error de palabras, conocida como *Word Error Rate* (WER), una medida que se emplea para evaluar los sistemas de reconocimiento automático de voz, tomando en cuenta las variables dialecto, raza y género dentro del inglés americano. Los autores incluyeron en su estudio dialectos acústicamente distintos para realizar una evaluación tanto del sistema Bing Speech (antes Proyecto Oxford) como del subtítulo automático de YouTube. De acuerdo con los autores, las variaciones en el habla han representado un desafío para los sistemas de reconocimiento automático de voz, especialmente las variaciones de dialecto y género que, pese a los avances tecnológicos, siguen ocasionando dificultades, sobre todo para el subtítulo automático de YouTube. Así, en el estudio se utilizaron grabaciones de un archivo de dialectos del inglés, siendo un total de 39 hablantes con cuatro dialectos distintos de inglés americano. Tatman y Kasten (2017) identificaron que el sistema Bing Speech tuvo diferencias en la WER entre dialectos y etnias, pero no eran estadísticamente confiables; mientras que pudieron determinar que el subtítulo automático de YouTube sí tuvo WERs estadísticamente diferentes entre dialectos y razas. Según estos resultados, las tasas menores de errores fueron con hablantes blancos de inglés estándar y ningún

sistema mostró tasas confiables entre género. Lo más preocupante, según los autores, es que la tasa de errores es mucho más alta para hablantes de color, lo cual puede reducir la utilidad de este recurso para ellos y ocasiona una brecha tecnológica.

Para el 2018, Rodríguez y Brosa plantearon en su trabajo *El subtitulado para personas sordas en YouTube: una aproximación a su estudio* que, si bien el subtitulado automático de esta plataforma puede ser de utilidad para las personas con discapacidad auditiva, no siempre prestan las funciones de accesibilidad adecuadas. Después de un análisis de 10 programas diferentes en dicho sitio de videos, los autores concluyeron que, en función del subtitulado para sordos (SPS), el sistema de YouTube es deficiente. Según los autores, el reconocimiento de voz empleado por la plataforma resultó poco fiable y solamente podía usarse como una guía aproximada para comprender el contenido del video. Sin embargo, se destacó que los resultados pueden ser relativos si se comparan con los obtenidos tras analizar el subtitulado automático de un programa televisivo en directo, ya que en estos se puede llegar a perder más de la mitad de la información total, en tanto con el subtitulado de YouTube el espectador (en términos de palabras) no pierde casi nada de información. Si bien, la pérdida de contenido recae en aspectos específicos de este subtitulado automático, como las transcripciones ininteligibles o que no tienen un significado en el idioma meta y que hacen difícil encontrar un sentido al mensaje. Estos errores en la transcripción, aunados a las restricciones espacio-temporales del subtitulado, hacen muy complicada la extracción del significado del contenido con tan poco tiempo de visualización. En este sentido, teniendo en mente que el SPS debe ofrecer un mensaje tan claro como sea posible (dada la privación del contenido acústico) podría sugerirse que, por sí solos, los subtítulos no logran transmitir adecuadamente el mensaje.

Jeong-Hwa y Kyung-Whan (2020), en su trabajo *An Analysis of the Errors in the Auto-Generated Captions of University Commencement Speeches on YouTube*, buscaron identificar y analizar los errores en el subtitulado automático de la plataforma YouTube en discursos iniciales de diferentes oradores que fueron presentados a lo largo de 12 años en distintos momentos de sus vidas. Los autores plantearon que, aunque el subtitulado automático de YouTube es útil, este no siempre es exacto y puede ocasionar confusión en los usuarios.

Los resultados de su investigación apuntaron a los sustantivos como la categoría gramatical en la que más ocurrieron errores, siendo más del 30 % de los casos, seguido de los verbos con más del 20 % y las preposiciones con el 8.1 %. Además, se pudo determinar que la omisión, la adición, la sustitución y el orden de palabras fueron los aspectos negativos con mayor incidencia. Con estos resultados, Jeong-Hwa y Kyung-Whan (2020) pudieron concluir que el reconocimiento de voz automático que emplea la plataforma YouTube necesita estar en continuo desarrollo para generar subtítulos con más eficiencia. Recientemente, Corrochano (2023) realizó el trabajo *Estudio de caso de la subtitulación automática en YouTube*, en el que buscó determinar si la subtitulación automática de YouTube es aceptable, tomando en cuenta las convenciones de subtitulación dentro de la traducción profesional. Según la autora, esta plataforma presenta errores en los subtítulos generados automáticamente, por lo que implementó un análisis basado en dos baremos de corrección de subtítulos. Los resultados de su análisis indicaron que la mayoría de los errores en los subtítulos generados automáticamente en la plataforma YouTube fueron de tipo ortográfico, seguidos de errores de formato y errores de supresión. Además, la autora destaca que los subtítulos no siempre aparecen en el momento exacto de intervención, no se respetan pausas, cambios de escena, texto en pantalla ni cambios de plano y no se siguen las normas ortotipográficas. El aspecto negativo más resaltado por la autora, dada la relevancia para la transmisión del mensaje, se relaciona con la segmentación de los subtítulos, ya que se detectó una gran cantidad de errores en los que las líneas perdieron sentido debido a una segmentación incorrecta, por ejemplo, al terminar las oraciones en preposiciones, conjunciones, artículos, determinantes, verbos auxiliares o adverbios. Si bien, algo que subraya la autora es que en momentos específicos los errores podrían considerarse de transcripción, ya que la falta de vocalización de quienes participan en el video puede llevar a no detectar adecuadamente palabras o frases, a su vez que YouTube detecta elementos de más cuando se escucha un doble audio (como al reproducir una canción de fondo). Por otro lado, en este estudio también se señalan ciertos aspectos positivos o ventajas del subtulado automático que ofrece YouTube. Algunos de estos son: posibilidad de mover la posición de los subtítulos en pantalla; elección libre de la fuente, así como del color, fondo, opacidad, ventana, borde y tamaño; posibilidad de modificar la velocidad de reproducción del video (y por lo tanto de los subtítulos).

Como conclusión, Corrochano (2023) comenta que, aunque la subtítulos automática de YouTube tiene ventajas, en el caso estudiado más de la mitad del producto audiovisual pudo considerarse inadecuadamente subtítulo, ya que los errores hallados afectaron considerablemente el sentido del mensaje.

En este mismo año, Ríos (2023) desarrolló el trabajo *Análisis de la calidad. El subtítulo en vivo interlingüístico de YouTube en un programa de noticias estadounidense accesible a usuarios peruanos*.

En este, tal y como el título indica, la autora analizó la calidad de los subtítulos generados por YouTube mediante el reconocimiento de voz y la traducción automática, haciendo partícipe a usuarios de dicha plataforma. La autora plantea en su estudio que el subtítulo en vivo permite acceder, por ejemplo, a programas informativos que se encuentran en otros idiomas y que en países como Perú (donde los medios de comunicación han perdido objetividad y seriedad, según sus percepciones), la población pueda estar adecuadamente informada sobre lo que acontece en el mundo, siendo el ejemplo mencionado la pandemia por Covid-19 que en años recientes sacudió la realidad. En este sentido, desde su punto de vista, la plataforma YouTube es una opción como medio de comunicación que merece la pena analizar. Así, en su trabajo Ríos (2023) evaluó la tasa de exactitud del subtítulo en vivo generado automáticamente, tomando en cuenta la comprensión y percepción por parte de los usuarios y una serie de factores establecidos por distintos organismos e instituciones reguladores. Cabe resaltar que se tuvo en cuenta que YouTube genera los subtítulos mediante algoritmos de aprendizaje automático, así como que Google provee la tecnología de reconocimiento de voz a YouTube. Para desarrollar la investigación, se seleccionó un programa sobre la salud mental y su relación con el confinamiento por el Covid-19, en donde se utilizó un lenguaje sencillo, con pronunciación clara de inglés nativo a una velocidad de 171 palabras por minuto sin interrupción, así como libre de ruido de fondo ni fallas de sonido. En cuanto a los participantes, se optó por usuarios sin conocimiento de inglés, pero con educación completa hasta el grado superior. Para determinar la comprensión lectora, se contemplaron los niveles de lenguaje literal, inferencial y crítico, en tanto para determinar la percepción de la calidad de los subtítulos se consideraron los siguientes aspectos: tiempo de aparición del subtítulo; tiempo para su lectura; sincronización con los hablantes y; coherencia entre imagen, sonido y subtítulo.

De este modo, los resultados revelaron que los errores con mayor incidencia fueron errores menores de traducción, siendo los principales aquellos de forma en los que no se incluyó puntuación apropiada, mayúsculas y hubo errores morfológicos. Así también, aunque en menor medida, se hallaron errores mayores de traducción, en especial aquellos de contenido en donde hubo una sustitución, omisión o adición que, conjuntamente con el desorden de palabras, ocasionaron sinsentidos en el mensaje. Finalmente, se pudieron identificar errores de reconocimiento, por ejemplo, cuando la pronunciación original no fue clara. Otro de los aspectos que se resaltaron en los hallazgos es que el subtítulo, pese a no presentar un retraso considerable en su aparición (lo que se calificó positivamente por los usuarios), sí causó problemas para leer al desaparecer muy rápidamente. Además, si bien no se identifica de ninguna manera a la persona que está hablando en los subtítulos, los participantes comentaron que sí hubo coherencia entre imagen y subtítulos. Finalmente, los usuarios calificaron la calidad del subtítulo en vivo de YouTube con un 50 %. La autora concluye que el subtítulo en vivo de esta plataforma no tiene suficiente calidad de traducción, teniendo en consideración que la mayor parte de errores fueron precisamente de traducción y no de reconocimiento, lo cual para personas que no tienen conocimientos del idioma original (en este caso inglés) obstaculiza su recepción del mensaje completo. Por último, Shintemirova (2023) realizó la investigación *Translation of metaphors in official and automatic subtitling and MT evaluation*, en la que comparó y analizó la traducción de metáforas en subtítulos generados automáticamente y subtítulos creados por humanos en la plataforma YouTube de inglés a ruso. Su objetivo era identificar las similitudes y diferencias entre ambas versiones de subtítulo, para después evaluar el algoritmo de traducción automática y determinar sus errores al traducir metáforas. La autora tomó en consideración por qué las metáforas se emplearon en cada momento específico, cómo fueron identificadas en la traducción y cómo fueron traducidas. Los resultados de su estudio demostraron que la traducción automática presenta dificultades para reconocer metáforas originales creadas por el autor del producto audiovisual y aplicarlas en el contexto adecuado, pero sí es capaz de traducir metáforas comunes o usadas con frecuencia. Estos hallazgos pueden relacionarse con el origen mismo del sistema de traducción automática, puesto que el uso de grandes bases de datos en las que se encuentra el vocabulario más empleado de una lengua y las traducciones comunes es el principal motor.

Ahora bien, el primero de los cuestionamientos surgidos después de realizar la revisión de literatura es si el sistema de reconocimiento de voz que emplea el subtítulo automático de YouTube puede hacer frente de forma satisfactoria a la diversidad de lenguas y de habla del mundo.

Por ejemplo, los resultados del estudio de Tatman y Kasten (2017) permiten plantear, por un lado, que el sistema de reconocimiento de voz automático puede presentar fallas al detectar variaciones en el habla humana que no son muy comunes o que, en este caso en específico, no son suficientemente comunes para que exista un registro real vasto que permita a Google tomar en consideración sus características y con ello enriquecer el trabajo de sus redes neuronales profundas. Desde este punto de vista, se podría suponer que, en tanto no existan suficientes registros reales de la inmensa cantidad de variantes del habla humana en cada una de las lenguas que el reconocimiento de voz abarca, las variaciones poco comunes serán un conflicto que entorpecerá la exactitud de estos sistemas, lo cual podría representar una desventaja para que determinados contenidos sean subtítulos, siendo entonces desigual el recurso ofrecido por dicha plataforma de videos. Esto, a su vez, supondría que los usuarios que visualicen contenido en YouTube subtítulo automáticamente sin activar el sonido no podrán tener la certeza de que cualquier video estará correctamente subtítulo, independientemente del idioma original empleado en el audio y sus posibles variantes dialectales, de género, raza o cualquier aspecto que haga una diferencia acústica del habla estándar o, en su defecto, de lo más registrado y asimilado por la inteligencia artificial de Google.

Por otro lado, los hallazgos de Shintemirova (2023) abren el debate sobre las variaciones de la lengua desde un punto de vista individual: la creación. Entonces, si se considera que el ser humano posee la capacidad de modificar, moldear, alterar el lenguaje, así como de crear nuevas formas de este a partir de las reinterpretaciones pragmáticas que puede impregnar con su experiencia de vida y conocimiento íntimo del mundo, puede sugerirse que el reconocimiento automático de voz de Google, cuya funcionalidad recae en el registro del habla humana real y su aprendizaje, estaría siempre sujeto al riesgo de no reconocer aquello que no está ya creado o, en este caso, que no está dicho y registrado por Google para ser estudiado, lo cual origina cabos sueltos para el trabajo exitoso de la inteligencia artificial y a su vez puede ocasionar fallas en el subtítulo automático de YouTube, sobre todo si el usuario depende únicamente de la información brindada por este sin poder corroborar el audio.

Por otra parte, al revisar varios de los trabajos contemplados en este estudio se pudo notar que en muchos casos el subtítulo automático de YouTube es impreciso o erróneo debido a las fallas en la transcripción que hacen difícil e incluso imposible obtener una información clara y comprensible por el usuario. Un ejemplo de ello es que en los resultados de Rodríguez y Brosa (2018) se detectaron muchas imprecisiones en la transcripción, desde palabras ininteligibles hasta palabras que no son propias de la lengua meta. Otro ejemplo es el trabajo de Jeong-Hwa y Kyung-Whan (2020), donde se identificaron muchos errores en categorías gramaticales importantes como los sustantivos y los verbos, lo cual compromete el traslado del mensaje.

Finalmente, al hablar específicamente de calidad del subtítulo, tomando en consideración las normativas para esta modalidad de TAV, así como las perspectivas de los usuarios sobre ello, se pudo llegar a la conclusión de que el subtítulo automático no cumple con los criterios de calidad que deberían tener los subtítulos de un producto audiovisual, ya que no se respetan aspectos de suma importancia como la puntuación, la segmentación o el tiempo de aparición en pantalla. Prueba de ello son los resultados de los trabajos de Corrochano (2023) y Ríos (2023), en donde se detectaron muchas inconsistencias en la calidad de los subtítulos generados automáticamente por YouTube.

CONCLUSIONES

Después de llevar a cabo esta revisión de literatura, se pudo concluir que el subtítulo automático de YouTube no responde a los estándares de calidad que la TAV determina. Sin embargo, independientemente de las convenciones de la traducción profesional para un subtítulo de calidad, es necesario recalcar que lo más importante es lo que el usuario percibe y si para este el subtítulo automático es funcional, si cumple con lo que espera recibir. En este sentido, sería conveniente poder estudiar si el subtítulo automático que se reproduce con el audio de los videos desactivado es suficientemente efectivo para los espectadores, dependiendo de sus necesidades y de sus motivos para visualizar el contenido de esta manera en particular. Por supuesto, esto abre una nueva puerta para redescubrir las funciones de la TAV y con ello replantearse las normativas para esta modalidad de traducción profesional.

REFERENCIAS BIBLIOGRÁFICAS

Chaume, F. (2004). *Cine y traducción*. Madrid: Cátedra.

Corrochano Muñoz, L. (2023). *Estudio de caso de la subtítulos automática en YouTube* (Trabajo de grado). Universidad Jaime I.

Google Support. (s. f.). Cómo mejora Google los modelos de voz.

<https://support.google.com/assistant/answer/11140942?hl=es#zippy=%2Caprendizaje-convencional%2Caprendizaje-federado%2Caprendizaje-ef%C3%ADmero>

Gómez-Aguilar, A. (2017). El video online: la eficacia del silencio. En Á. Martínez-García (Ed.), *La imagen en la era digital* (pp. 41-51). Editorial Egregius. <https://hdl.handle.net/11441/91565>

Hurtado, A. (1994). Modalidades y tipos de traducción. *Vasos Comunicantes*, (4), 19-27.

Jeong-Hwa, L. y Kyung-Whan, C. (2020). An Analysis of the Errors in the Auto-Generated Captions of University Commencement Speeches on YouTube. *The Journal of Asia TEFL*, 17(1), 143-159. <http://dx.doi.org/10.18823/asiatefl.2020.17.1.9.143>

Patel, S. (17 de mayo de 2016). 85 percent of Facebook video is watched without sound. *Digiday* <https://digiday.com/media/silent-world-facebook-video/>

Platzi. (s. f.). Cómo funciona el algoritmo de subtítulos automática de YouTube.

<https://platzi.com/blog/como-funciona-el-algoritmo-de-subtitulacion-automatica-de-youtube/>

Baez, J. (23 de abril de 2018). La importancia de los subtítulos en los videos sin sonido de Facebook. *Sandia Films*.

<https://www.sandiafilms.com/la-importancia-de-los-subtitulos-en-los-videos-sin-sonido-de-facebook/>

Hernández, A. (2016). El tráfico IP global se multiplicará por tres en los próximos cinco años, superando los 3 Zettabytes en 2021. Cisco Global Newsroom,

<http://globalnewsroom.cisco.com/es/es/re-lease/El-tr%C3%A1fico-IP-global-semultiplicar%C3%A1-por-tres-en-los-pr%C3%B3ximos-cinco-a%C3%93os>

Hotmart. (21 de octubre de 2022). ¡Descubre por qué necesitas ponerle subtítulos a un video enseguida!

Hotmart Blog. <https://hotmart.com/es/blog/poner-subtitulos-a-un-video>



- Mason, I. (1989). Speaker meaning and reader meaning: preserving coherence in screentranslating. En R. Kölmel y J. Payne (Eds.), *Babel. The Cultural and Linguistic Barriers between Nations* (p. 13-24), Aberdeen: Aberdeen University Press.
- Navimedia. (s. f.). ¿Por qué el cerebro prefiere vídeo vs. texto? Navimedia.
<https://navimedia.es/por-que-el-cerebro-prefiere-el-video-vs-texto/>
- Nuteco. (s. f.). ¿Tus vídeos son igual de efectivos sin sonido? <https://www.nutecoweb.com/videos-igual-efectivos-sin-sonido/>
- Ribera de Gracia, F. (2016). “Las marcas ante el reto del vídeo”.
- Ríos Valero, L. (2023). Análisis de la calidad. El subtítulo en vivo interlingüístico de YouTube en un programa de noticias estadounidense accesible a usuarios peruanos. *Entreculturas*, 13, 143-168.
- Rodríguez Campillo, M. J. y Brosa Rodríguez, A. (2018). El subtítulo para personas sordas en YouTube: una aproximación a su estudio. *Universitat Rovira i Virgili*.
- Shintemirova, M. (2023). Translation of metaphors in official and automatic subtitling and MT evaluation. *JCAL*, 1, 77-93. <https://doi.org/10.33919/JCAL.23.1.4>
- Tatman, R. y Kasten, C. (2017). Effects of Talker Dialect, Gender & Race on Accuracy of Bing Speech and YouTube Automatic Captions. *Interspeech*, 934-938 .
<http://dx.doi.org/10.21437/Interspeech.2017-1746>
- Vizcaíno-Verdú, A., Contreras-Pulido, P., y Guzmán-Franco, M.-D. (2019). Lectura y aprendizaje informal en YouTube: El booktuber. *Comunicar*, 27(59), 95–104.
<https://doi.org/10.3916/C59-2019-09>